

# HW1CSE105F24: Homework assignment 1

CSE105F24

Due: October 8th at 5pm, via Gradescope

## In this assignment,

You will practice reading and applying the definitions of alphabets, strings, languages, Kleene star, and regular expressions. You will use regular expressions and relate them to languages and finite automata. You will use precise notation to formally define the state diagram of finite automata, and you will use clear English to describe computations of finite automata informally.

**Resources:** To review the topics for this assignment, see the class material from Weeks 0 and 1. We will post frequently asked questions and our answers to them in a pinned Piazza post.

**Reading and extra practice problems:** Sipser Section 0, 1.3, 1.1. Chapter 1 exercises 1.1, 1.2, 1.3, 1.18, 1.23.

**For all HW assignments:** Weekly homework may be done individually or in groups of up to 3 students. You may switch HW partners for different HW assignments. Please ensure your name(s) and PID(s) are clearly visible on the first page of your homework 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(s) to the Gradescope submission by selecting their name(s) in the “Add Group Members” dialog box. You will need to re-add your group member(s) every time you resubmit a new version of your assignment. Each homework question will be graded either for correctness (including clear and precise explanations and justifications of all answers) or fair effort completeness. For “graded for correctness” questions: collaboration is allowed only with CSE 105 students in your group; if your group has questions about a problem, you may ask in drop-in help hours or post a private post (visible only to the Instructors) on Piazza. For “graded for completeness” questions: collaboration is allowed with any CSE 105 students this quarter; if your group has questions about a problem, you may ask in drop-in help hours or post a public post on Piazza.

All submitted homework for this class must be typed. You can use a word processing editor if you like (Microsoft Word, Open Office, Notepad, Vim, Google Docs, etc.) but you might find it useful to take this opportunity to learn LaTeX. LaTeX is a markup language used widely in

computer science and mathematics. The homework assignments are typed using LaTeX and you can use the source files as templates for typesetting your solutions. To generate state diagrams of machines, you can (1) use the LaTeX tikzpicture environment (see templates in the class notes), or (2)) use the software tools Flap.js or JFLAP described in the class syllabus (and include a screenshot in your PDF), or (3) you can carefully and clearly hand-draw the diagram and take a picture and include it in your PDF. We recommend that you submit early drafts to Gradescope so that in case of any technical difficulties, at least some of your work is present. You may update your submission as many times as you'd like up to the deadline.

## Integrity reminders

- Problems should be solved together, not divided up between the partners. The homework is designed to give you practice with the main concepts and techniques of the course, while getting to know and learn from your classmates.
- You may not collaborate on homework questions graded for correctness with anyone other than your group members. You may ask questions about the homework in office hours (of the instructor, TAs, and/or tutors) and on Piazza (as private notes viewable only to the Instructors). You *cannot* use any online resources about the course content other than the class material from this quarter – this is primarily to ensure that we all use consistent notation and definitions (aligned with the textbook) and also to protect the learning experience you will have when the ‘aha’ moments of solving the problem authentically happen.
- Do not share written solutions or partial solutions for homework with other students in the class who are not in your group. Doing so would dilute their learning experience and detract from their success in the class.

You will submit this assignment via Gradescope (<https://www.gradescope.com>) in the assignment called “hw1CSE105F24”.

## Assigned questions

### 1. Finding examples and edge cases (12 points):

With  $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$  and  $\Gamma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F\}$

- (a) (*Graded for completeness*)<sup>1</sup> Give an example of a string over  $\Sigma$  that is meaningful to you in some way and whose length is between 5 and 20, and explain why this string is meaningful to you.

---

<sup>1</sup>This means you will get full credit so long as your submission demonstrates honest effort to answer the question. You will not be penalized for incorrect answers. To demonstrate your honest effort in answering the question, we expect you to include your attempt to answer *each* part of the question. If you get stuck with your attempt, you can still demonstrate your effort by explaining where you got stuck and what you did to try to get unstuck.

- (b) (*Graded for completeness*) Calculate the number of distinct strings of length 3 over  $\Sigma$  and then explain your calculation.
- (c) (*Graded for completeness*) With the ordering  $0 < 1 < 2 < 3 < 4 < 5 < 6 < 7 < 8 < 9 < A < B < C < D < E < F$ , list the first 50 strings over  $\Gamma$  in string order. Explain how you constructed this list. *Note: you can write a program to generate this list if you'd like, and you may use any external tools to help you write this program. If you do use a program to generate the list, include it (and documentation for how it works) as part of your submission.*
- (d) (*Graded for correctness*)<sup>2</sup> Give an example of a finite set that is a language over  $\Sigma$  and over  $\Gamma$ , or explain why there is no such set. A complete and correct answer will use clear and precise notation (consistent with the textbook and class notes) and will include a description of why the given example is a language over  $\Sigma$  and over  $\Gamma$  and is finite, or an explanation why there is no such example.
- (e) (*Graded for correctness*) Give an example of an infinite set that is a language over  $\Sigma$  and not over  $\Gamma$ , or explain why there is no such set. A complete and correct answer will use clear and precise notation (consistent with the textbook and class notes) and will include a description of why the given example is a language over  $\Sigma$  and not over  $\Gamma$  and is infinite, or an explanation why there is no such example.

## 2. Regular expressions (10 points):

- (a) (*Graded for completeness*) Give three regular expressions that all describe the set of all strings over  $\{a, b\}$  that have odd length. Ungraded bonus challenge: Make the expressions as different as possible!
- (b) (*Graded for completeness*) A friend tells you that each regular expression that has a Kleene star ( $*$ ) describes an infinite language. Are they right? Either help them justify their claim or give a counterexample to disprove it and explain your counterexample.

## 3. Functions over languages (15 points):

For each language  $L$  over the alphabet  $\Sigma_1 = \{0, 1\}$ , we have the associated sets of strings

$$SUBSTRING(L) = \{w \in \Sigma_1^* \mid \text{there exist } x, y \in \Sigma_1^* \text{ such that } xwy \in L\}$$

and

$$EXTEND(L) = \{w \in \Sigma_1^* \mid w = uv \text{ for some strings } u \in L \text{ and } v \in \Sigma_1^*\}$$

- (a) (*Graded for completeness*) Specify an example language  $A$  over  $\Sigma_1$  such that  $SUBSTRING(A) = EXTEND(A)$ , or explain why there is no such example. A complete solution will include

---

<sup>2</sup>This means your solution will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should explain how you arrived at your conclusions, using mathematically sound reasoning. Whether you use formal proof techniques or write a more informal argument for why something is true, your answers should always be well-supported. Your goal should be to convince the reader that your results and methods are sound.

either (1) a precise and clear description of your example language  $A$  and a precise and clear description of the result of computing  $SUBSTRING(A)$ ,  $EXTEND(A)$  (using the given definitions) to justify this description and to justify the set equality, or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.

- (b) (*Graded for correctness*) Specify an example language  $B$  over  $\Sigma_1$  such that

$$SUBSTRING(B) = \{\varepsilon\}$$

and

$$EXTEND(B) = \Sigma_1^*$$

or explain why there is no such example. A complete solution will include either (1) a precise and clear description of your example language  $B$  and a precise and clear description of the result of computing  $SUBSTRING(B)$ ,  $EXTEND(B)$  (using the given definitions) to justify this description and to justify the set equality with  $\{\varepsilon\}$  and  $\Sigma_1^*$  (respectively), or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.

- (c) (*Graded for correctness*) Specify an example **infinite** language  $C$  over  $\Sigma_1$  such that

$$SUBSTRING(C) \neq \Sigma_1^*$$

and

$$EXTEND(C) \neq \Sigma_1^*$$

, or explain why there is no such example. A complete solution will include either (1) a precise and clear description of your example language  $C$  and a precise and clear description of the result of computing  $SUBSTRING(B)$ ,  $EXTEND(B)$  (using the given definitions) to justify this description and to justify the set nonequality claims, or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.

#### 4. Finite automata (13 points):

Consider the finite automaton  $(Q, \Sigma, \delta, q_0, F)$  whose state diagram is depicted below



where  $Q = \{q_0, q_1, q_2\}$ ,  $\Sigma = \{0, 1\}$ , and  $F = \{q_0\}$ , and  $\delta : Q \times \Sigma \rightarrow Q$  is specified by the look-up table

|       | 0     | 1     |
|-------|-------|-------|
| $q_0$ | $q_0$ | $q_1$ |
| $q_1$ | $q_2$ | $q_0$ |
| $q_2$ | $q_2$ | $q_2$ |

- (a) (*Graded for completeness*) A friend tries to summarize the transition function with the formula

$$\delta(q_i, x) = \begin{cases} q_0 & \text{when } i = 0 \text{ and } x = 0 \\ q_2 & \text{when } x < i \\ q_j & \text{when } j = (i + 1) \bmod 2 \text{ and } x = 1 \end{cases}$$

for  $x \in \{0, 1\}$  and  $i \in \{0, 1, 2\}$ . Are they right? Either help them justify their claim or give a counterexample to disprove it and then fix their formula.

- (b) (*Graded for correctness*) Give a regular expression  $R$  so that  $L(R)$  is the language recognized by this finite automaton. Justify your answer by referring to the definition of the semantics of regular expressions and computations of finite automata. Include an explanation for why each string in  $L(R)$  is accepted by the finite automaton *and* for why each string not in  $L(R)$  is rejected by the finite automaton.
- (c) (*Graded for correctness*) Keeping the same set of states  $Q = \{q_0, q_1, q_2\}$ , alphabet  $\Sigma = \{0, 1\}$ , same start state  $q_0$ , and same transition function  $\delta$ , choose a new set of accepting states  $F_{new}$  so that the new finite automaton that results accepts at least one string that the original one rejected **and** rejects at least one string that the original one accepted, or explain why there is no such choice of  $F_{new}$ . A complete solution will include either (1) a precise and clear description of your choice of  $F_{new}$  and a precise and clear the two example strings using relevant definitions to justify them, or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.

HW2CSE105F24: Homework assignment 2 Due: October 15th at 5pm, via Gradescope

### **In this assignment,**

You will practice designing multiple representations of regular languages and working with general constructions of automata to demonstrate the richness of the class of regular languages.

**Resources:** To review the topics for this assignment, see the class material from Week 2. We will post frequently asked questions and our answers to them in a pinned Piazza post.

**Reading and extra practice problems:** Sipser Section 1.1, 1.2, 1.3. Chapter 1 exercises 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.10, 1.11, 1.12, 1.14, 1.15, 1.16, 1.17, 1.19, 1.20, 1.21, 1.22. Chapter 1 problem 1.51.

**For all HW assignments:** Weekly homework may be done individually or in groups of up to 3 students. You may switch HW partners for different HW assignments. Please ensure your name(s) and PID(s) are clearly visible on the first page of your homework 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(s) to the Gradescope submission by selecting their name(s) in the “Add Group Members” dialog box. You will need to re-add your group member(s) every time you resubmit a new version of your assignment. Each homework question will be graded either for correctness (including clear and precise explanations and justifications of all answers) or fair effort completeness. For “graded for correctness” questions: collaboration is allowed only with CSE 105 students in your group; if your group has questions about a problem, you may ask in drop-in help hours or post a private post (visible only to the Instructors) on Piazza. For “graded for completeness” questions: collaboration is allowed with any CSE 105 students this quarter; if your group has questions about a problem, you may ask in drop-in help hours or post a public post on Piazza.

All submitted homework for this class must be typed. You can use a word processing editor if you like (Microsoft Word, Open Office, Notepad, Vim, Google Docs, etc.) but you might find it useful to take this opportunity to learn LaTeX. LaTeX is a markup language used widely in computer science and mathematics. The homework assignments are typed using LaTeX and you can use the source files as templates for typesetting your solutions. To generate state diagrams of machines, you can (1) use the LaTeX tikzpicture environment (see templates in the class notes), or (2)) use the software tools Flap.js or JFLAP described in the class syllabus (and include a screenshot in your PDF), or (3) you can carefully and clearly hand-draw the diagram and take a picture and include it in your PDF. We recommend that you submit early drafts to Gradescope so that in case of any technical difficulties, at least some of your work is present. You may update your submission as many times as you’d like up to the deadline.

### **Integrity reminders**

- Problems should be solved together, not divided up between the partners. The homework

is designed to give you practice with the main concepts and techniques of the course, while getting to know and learn from your classmates.

- You may not collaborate on homework questions graded for correctness with anyone other than your group members. You may ask questions about the homework in office hours (of the instructor, TAs, and/or tutors) and on Piazza (as private notes viewable only to the Instructors). You *cannot* use any online resources about the course content other than the class material from this quarter – this is primarily to ensure that we all use consistent notation and definitions (aligned with the textbook) and also to protect the learning experience you will have when the ‘aha’ moments of solving the problem authentically happen.
- Do not share written solutions or partial solutions for homework with other students in the class who are not in your group. Doing so would dilute their learning experience and detract from their success in the class.

You will submit this assignment via Gradescope (<https://www.gradescope.com>) in the assignment called “hw2CSE105F24”.

## Assigned questions

1. **Automata design** (12 points): As background to this question, recall that integers can be represented using base  $b$  expansions, for any convenient choice of base  $b$ . The precise definition is: for  $b$  an integer greater than 1 and  $n$  a positive integer, the **base  $b$  expansion of  $n$**  is defined to be

$$(a_{k-1} \cdots a_1 a_0)_b$$

where  $k$  is a positive integer,  $a_0, a_1, \dots, a_{k-1}$  are nonnegative integers less than  $b$ ,  $a_{k-1} \neq 0$ , and

$$n = \sum_{i=0}^{k-1} a_i b^i$$

Notice: *The base  $b$  expansion of a positive integer  $n$  is a string over the alphabet  $\{x \in \mathbb{Z} \mid 0 \leq x < b\}$  whose leftmost character is nonzero.*

An important property of base  $b$  expansions of integers is that, for each integer  $b$  greater than 1, each positive integer  $n = (a_{k-1} \cdots a_1 a_0)_b$ , and each nonnegative integer  $a$  less than  $b$ ,

$$bn + a = (a_{k-1} \cdots a_1 a_0 a)_b$$

In other words, shifting the base  $b$  expansion to the left results in multiplying the integer value by the base. In this question we’ll explore building deterministic finite automata that recognize languages that correspond to useful sets of integers.

- (a) (*Graded for completeness*) <sup>3</sup> Design a DFA that recognizes the set of binary (base 2) expansions of positive integers that are powers of 2. A complete solution will include the state diagram of your DFA and a brief justification of your construction by explaining the role each state plays in the machine, as well as a brief justification about how the strings accepted and rejected by the machine connect to the specified language.

*Hints:* (1) A power of 2 is an integer  $x$  that can be written as  $2^y$  for some nonnegative integer  $y$ , (2) the DFA should accept the strings 100, 10 and 100000 and should reject the strings 010, 1101, and  $\varepsilon$  (can you see why?).

- (b) (*Graded for completeness*) Consider arbitrary positive integer  $m$ . Design a DFA that recognizes the set of binary (base 2) expansions of positive integers that are multiples of  $m$ . A complete solution will include the formal definition of your DFA (parameterized by  $m$ ) and a brief justification of your construction by explaining the role each state plays in the machine, as well as a brief justification about how the strings accepted and rejected by the machine connect to the specified language.

*Hints:* (1) Consider having a state for each possible remainder upon division by  $m$ . (2) To determine transitions, notice that reading a new character will shift what we already read over by one slot.

- (c) (*Graded for correctness*) <sup>4</sup> Choose a positive integer  $m_0$  between 5 and 8 (inclusive) and draw the state diagram of a DFA recognizing the following language over  $\{0, 1, 2, 3\}$

$$\{w \in \{0, 1, 2, 3\}^* \mid w \text{ is a base 4 expansion of a positive integer that is a multiple of } m_0\}$$

A complete solution will include the state diagram of your DFA and a brief justification of your construction by explaining the role each state plays in the machine, as well as a brief justification about how the strings accepted and rejected by the machine connect to the specified language.

*Bonus extension to think about (ungraded):* Which other languages related to sets of integers can be proved to be regular using a similar strategy?

2. **Nondeterminism** (15 points): For this question, the alphabet is  $\{a, b\}$ .

- (a) (*Graded for completeness*) Design a DFA that recognizes the language

$$\{w \in \{a, b\}^* \mid w \text{ contains at most one } a \text{ and at least two } bs\}$$

---

<sup>3</sup>This means you will get full credit so long as your submission demonstrates honest effort to answer the question. You will not be penalized for incorrect answers. To demonstrate your honest effort in answering the question, we expect you to include your attempt to answer *each* part of the question. If you get stuck with your attempt, you can still demonstrate your effort by explaining where you got stuck and what you did to try to get unstuck.

<sup>4</sup>This means your solution will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should explain how you arrived at your conclusions, using mathematically sound reasoning. Whether you use formal proof techniques or write a more informal argument for why something is true, your answers should always be well-supported. Your goal should be to convince the reader that your results and methods are sound.



You can design this DFA directly or use the constructions from class (and the footnote to Theorem 1.25 in the book) to build this DFA from DFA for the simpler languages that are intersected to give this language.

A complete solution will include the state diagram of your DFA and a brief justification of your construction either by explaining the role each state plays in the machine, as well as a brief justification about how the strings accepted and rejected by the machine connect to the specified language, or by justifying the design of the DFA for the simpler languages and then describing how the Theorem was used.

- (b) (*Graded for correctness*) Design a NFA with at most 6 states that recognizes the language

$$\{w \in \{a, b\}^* \mid w \text{ contains at most one } a \textbf{ and at least two } bs\}$$

A complete solution will include the state diagram of your NFA and a brief justification of your construction by explaining the role each state plays in the machine, as well as a brief justification about how the strings accepted and rejected by the machine connect to the specified language. Give one example string in the language and explain the computation of the NFA that witnesses that the machine accepts this string. Also, give one example string not in the language and explain why the NFA rejects this string.

- (c) (*Graded for correctness*) Design a NFA with at most 6 states that recognizes the language

$$\{w \in \{a, b\}^* \mid w \text{ contains at most one } a \textbf{ or at least two } bs\}$$

A complete solution will include the state diagram of your NFA and a brief justification of your construction by explaining the role each state plays in the machine, as well as a brief justification about how the strings accepted and rejected by the machine connect to the specified language. Give one example string in the language and explain the computation of the NFA that witnesses that the machine accepts this string. Also, give one example string not in the language and explain why the NFA rejects this string.

*Bonus extension to think about (ungraded):* Did you need all 6 states? Could you design DFA with 6 states that recognize each of these languages?

**3. General constructions** (15 points): In this question, you'll practice working with formal general constructions for NFAs and translating between state diagrams and formal definitions.

- (a) (*Graded for correctness*) Consider the following general construction: Let  $N_1 = (Q, \Sigma, \delta_1, q_1, F_1)$  be a NFA and assume that  $q_0 \notin Q$ . Define the new NFA  $N_2 = (Q \cup \{q_0\}, \Sigma, \delta_2, q_0, \{q_1\})$  where

$$\delta_2 : (Q \cup \{q_0\}) \times \Sigma_\epsilon \rightarrow \mathcal{P}(Q \cup \{q_0\})$$

is defined by

$$\delta_2(q, a) = \begin{cases} \{q' \in Q \mid q \in \delta_1(q', a)\} & \text{if } q \in Q, a \in \Sigma_\epsilon \\ F_1 & \text{if } q = q_0, a = \epsilon \\ \emptyset & \text{if } q = q_0, a \in \Sigma \end{cases}$$

Illustrate this construction by defining a specific example NFA  $N_1$  and applying the construction above to create the new NFA  $N_2$ . Your example NFA should

- Have exactly four states (all reachable from the start state),
- Accept at least one string and reject at least one string, and
- Not have any states labelled  $q_0$ .

Apply the construction above to create the new NFA. A complete submission will include the state diagram of your example NFA  $N_1$  and the state diagram of the NFA  $N_2$  resulting from this construction and a precise and clear description of  $L(N_1)$  and  $L(N_2)$ , justified by explaining the role each state plays in the machine, as well as a brief justification about how the strings accepted and rejected by the machine connect to the language.

- (b) In Week 2's review quiz, we saw the definition that a set  $X$  is said to be **closed under an operation** if, for any elements in  $X$ , applying to them gives an element in  $X$ . For example, the set of integers is closed under multiplication because if we take any two integers, their product is also an integer .

Recall the definitions we have: For each language  $L$  over the alphabet  $\Sigma_1 = \{0, 1\}$ , we have the associated set of strings

$$EXTEND(L) = \{w \in \Sigma_1^* \mid w = uv \text{ for some strings } u \in L \text{ and } v \in \Sigma_1^*\}$$

We will prove that the collection of languages over  $\{0, 1\}$  that are each recognizable by some NFA is closed under the *EXTEND* operation.

- (*Graded for completeness*) As a helpful tool in our construction<sup>5</sup>, prove that every NFA can be converted to an equivalent one that has a single accept state. Note: this is exercise 1.11 in the textbook.
- (*Graded for correctness*) Prove that the collection of languages over  $\{0, 1\}$  that are each recognizable by some NFA is closed under the *EXTEND* operation. You can assume that you are given a NFA with a single accept state  $N = (Q, \{0, 1\}, \delta, q_0, \{q_{acc}\})$  and you need to define a new NFA,  $N_{new} = (Q_{new}, \{0, 1\}, \delta_{new}, q_{new}, F_{new})$ , so that  $L(N_{new}) = EXTEND(L(N))$ .

A complete solution will include precise definitions for  $Q_{new}$ ,  $\delta_{new}$ ,  $q_{new}$ , and  $F_{new}$ , as well as a brief justification of your construction by explaining why these definitions work, referring specifically to the definition of *EXTEND* and to acceptance of NFA.

**4. Multiple representations** (8 points): For any language  $L \subseteq \Sigma^*$ , recall that we define its *complement* as

$$\overline{L} := \Sigma^* - L = \{w \in \Sigma^* \mid w \notin L\}$$

That is, the complement of  $L$  contains all and only those strings which are not in  $L$ . Our notation for regular expressions does not include the complement symbol. However, it turns out that the complement of a language described by a regular expression is guaranteed to also be describable by a (different) regular expression.<sup>6</sup>

<sup>5</sup>A result that is proved in order to work towards a larger theorem is called a Lemma.

<sup>6</sup>We'll see that this is connected to the result we proved in class that the complement of each language recognizable by a DFA is recognizable by a(nother) DFA.

For example, over the alphabet  $\Sigma = \{a, b\}$ , the complement of the language described by the regular expression  $\Sigma^*b$  is described by the regular expression  $\varepsilon \cup \Sigma^*a$  because any string that does not end in  $b$  must either be the empty string or end in  $a$ .

For each of the regular expressions  $R$  over the alphabet  $\Sigma = \{a, b\}$  below, write the regular expression for  $\overline{L(R)}$ . Your regular expressions may use the symbols  $\emptyset$ ,  $\varepsilon$ ,  $a$ ,  $b$ , and the following operations to combine them: union, concatenation, and Kleene star.

Briefly justify why your solution for each part works by giving plain English descriptions of the language described by the regular expression and of its complement and connecting them to the regular expression via relevant definitions. An English description that is more detailed than simply negating the description in the original language will likely be helpful in the justification.

Alternatively, you can justify your solution by first designing a DFA that recognizes  $L(R)$ , using the construction from class and the book to modify this DFA to get a new DFA that recognizes  $\overline{L(R)}$ , and then applying the constructions from class and the book to convert this new DFA to a regular expression.

For each part of the question, clearly state which approach you're taking and include enough intermediate steps to illustrate your work.

- (a) (*Graded for correctness*)  $(a \cup b)^*a(a \cup b)^*$
- (b) (*Graded for correctness*)  $(a \cup b)(a \cup b)(a \cup b)$

HW3CSE105F24: Homework assignment 3 Due: October 22nd at 5pm, via Gradescope

### **In this assignment,**

You will demonstrate the richness of the class of regular languages, as well as its boundaries.

**Resources:** To review the topics for this assignment, see the class material from Week 3. We will post frequently asked questions and our answers to them in a pinned Piazza post.

**Reading and extra practice problems:** Sipser Chapter 1. Chapter 1 exercises 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.10, 1.11, 1.12, 1.14, 1.15, 1.16, 1.17, 1.19, 1.20, 1.21, 1.22. Chapter 1 problem 1.51.

**For all HW assignments:** Weekly homework may be done individually or in groups of up to 3 students. You may switch HW partners for different HW assignments. Please ensure your name(s) and PID(s) are clearly visible on the first page of your homework 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(s) to the Gradescope submission by selecting their name(s) in the “Add Group Members” dialog box. You will need to re-add your group member(s) every time you resubmit a new version of your assignment. Each homework question will be graded either for correctness (including clear and precise explanations and justifications of all answers) or fair effort completeness. For “graded for correctness” questions: collaboration is allowed only with CSE 105 students in your group; if your group has questions about a problem, you may ask in drop-in help hours or post a private post (visible only to the Instructors) on Piazza. For “graded for completeness” questions: collaboration is allowed with any CSE 105 students this quarter; if your group has questions about a problem, you may ask in drop-in help hours or post a public post on Piazza.

All submitted homework for this class must be typed. You can use a word processing editor if you like (Microsoft Word, Open Office, Notepad, Vim, Google Docs, etc.) but you might find it useful to take this opportunity to learn LaTeX. LaTeX is a markup language used widely in computer science and mathematics. The homework assignments are typed using LaTeX and you can use the source files as templates for typesetting your solutions. To generate state diagrams of machines, you can (1) use the LaTeX tikzpicture environment (see templates in the class notes), or (2)) use the software tools Flap.js or JFLAP described in the class syllabus (and include a screenshot in your PDF), or (3) you can carefully and clearly hand-draw the diagram and take a picture and include it in your PDF. We recommend that you submit early drafts to Gradescope so that in case of any technical difficulties, at least some of your work is present. You may update your submission as many times as you’d like up to the deadline.

### **Integrity reminders**

- Problems should be solved together, not divided up between the partners. The homework is designed to give you practice with the main concepts and techniques of the course, while

getting to know and learn from your classmates.

- You may not collaborate on homework questions graded for correctness with anyone other than your group members. You may ask questions about the homework in office hours (of the instructor, TAs, and/or tutors) and on Piazza (as private notes viewable only to the Instructors). You *cannot* use any online resources about the course content other than the class material from this quarter – this is primarily to ensure that we all use consistent notation and definitions (aligned with the textbook) and also to protect the learning experience you will have when the ‘aha’ moments of solving the problem authentically happen.
- Do not share written solutions or partial solutions for homework with other students in the class who are not in your group. Doing so would dilute their learning experience and detract from their success in the class.

You will submit this assignment via Gradescope (<https://www.gradescope.com>) in the assignment called “hw3CSE105F24”.

### Assigned questions

1. **Using general constructions** (16 points): Consider the NFA  $N$  over  $\{0, 1, 2\}$  with state diagram



- (*Graded for completeness*)<sup>7</sup> Give two examples of strings of length greater than 2 that are accepted by  $N$  and two examples of strings of length greater than 2 that are rejected by  $N$ . For each example string, list at least one of the computations of  $N$  on this string and label whether this computation witnesses that the string is accepted by  $N$ .
- (*Graded for correctness*)<sup>8</sup> Use the “macro-state” construction from Theorem 1.39 and class to create the DFA  $M$  recognizing the same language as  $N$ . You only need to include states

<sup>7</sup>This means you will get full credit so long as your submission demonstrates honest effort to answer the question. You will not be penalized for incorrect answers. To demonstrate your honest effort in answering the question, we expect you to include your attempt to answer *each* part of the question. If you get stuck with your attempt, you can still demonstrate your effort by explaining where you got stuck and what you did to try to get unstuck.

<sup>8</sup>This means your solution will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should explain how you arrived at your conclusions, using mathematically sound reasoning. Whether you use formal proof techniques or write a more informal argument for why something is true, your answers should always be well-supported. Your goal should be to convince the reader that your results and methods are sound.

that are reachable from the start state. For full credit, submit (1) a state diagram that is deterministic (there should be arrows labelled 0, 1, and 2 coming out of each state) and where each state is labelled by a subset of the states in  $N$ ; and (2) for one of your example strings that is accepted by  $N$ , give the computation of  $M$  on this string as a sequence of states visited; and (3) for one of your example strings that is rejected by  $N$ , give the computation of  $M$  on this string as a sequence of states visited.

- (c) (*Graded for completeness*) Give a mathematical description either using set builder notation or a regular expression for  $L(N)$  and for  $L(M)$ .

## 2. Multiple representations (12 points):

- (a) Consider the language  $A_1 = \{uw \mid u \text{ and } w \text{ are strings over } \{0, 1\} \text{ and have the same length}\}$  and the following argument.

“Proof” that  $A_1$  is not regular using the Pumping Lemma: Let  $p$  be an arbitrary positive integer. We will show that  $p$  is not a pumping length for  $A_1$ .

Choose  $s$  to be the string  $1^p 0^p$ , which is in  $A_1$  because we can choose  $u = 1^p$  and  $w = 0^p$  which each have length  $p$ . Since  $s$  is in  $A_1$  and has length greater than or equal to  $p$ , if  $p$  were to be a pumping length for  $A_1$ ,  $s$  ought to be pumpable. That is, there should be a way of dividing  $s$  into parts  $x, y, z$  where  $s = xyz$ ,  $|y| > 0$ ,  $|xy| \leq p$ , and for each  $i \geq 0$ ,  $xy^i z \in A_1$ . Suppose  $x, y, z$  are such that  $s = xyz$ ,  $|y| > 0$  and  $|xy| \leq p$ . Since the first  $p$  letters of  $s$  are all 1 and  $|xy| \leq p$ , we know that  $x$  and  $y$  are made up of all 1s. If we let  $i = 2$ , we get a string  $xy^2 z$  that is not in  $A_1$  because repeating  $y$  twice adds 1s to  $u$  but not to  $w$ , and strings in  $A_1$  are required to have  $u$  and  $w$  be the same length. Thus,  $s$  is not pumpable (even though it should have been if  $p$  were to be a pumping length) and so  $p$  is not a pumping length for  $A_1$ . Since  $p$  was arbitrary, we have demonstrated that  $A_1$  has no pumping length. By the Pumping Lemma, this implies that  $A_1$  is nonregular.

- i. (*Graded for completeness*) Find the (first and/or most significant) logical error in the “proof” above and describe why it’s wrong.
  - ii. (*Graded for completeness*) Prove that the set  $A_1$  is actually regular (by finding a regular expression that describes it or a DFA/NFA that recognizes it, and justifying why) **or** fix the proof so that it is logically sound.
- (b) Consider the language  $A_2 = \{u1w \mid u \text{ and } w \text{ are strings over } \{0, 1\} \text{ and have the same length}\}$  and the following argument.

“Proof” that  $A_2$  is not regular using the Pumping Lemma: Let  $p$  be an arbitrary positive integer. We will show that  $p$  is not a pumping length for  $A_2$ .

Choose  $s$  to be the string  $1^{p+1} 0^p$ , which is in  $A_2$  because we can choose  $u = 1^p$  and  $w = 0^p$  which each have length  $p$ . Since  $s$  is in  $A_2$  and has length greater than or equal to  $p$ , if  $p$  were to be a pumping length for  $A_2$ ,  $s$  ought to be pumpable. That is, there should be a way of dividing  $s$  into parts  $x, y, z$  where  $s = xyz$ ,  $|y| > 0$ ,  $|xy| \leq p$ , and for each  $i \geq 0$ ,  $xy^i z \in A_2$ . When  $x = \varepsilon$  and  $y = 1^{p+1}$  and  $z = 0^p$ , we have satisfied that  $s = xyz$ ,  $|y| > 0$  (because  $p$  is positive) and  $|xy| \leq p$ . If we let

$i = 0$ , we get the string  $xy^iz = 0^p$  that is not in  $A_2$  because its middle symbol is a 0, not a 1. Thus,  $s$  is not pumpable (even though it should have been if  $p$  were to be a pumping length) and so  $p$  is not a pumping length for  $A_2$ . Since  $p$  was arbitrary, we have demonstrated that  $A_2$  has no pumping length. By the Pumping Lemma, this implies that  $A_2$  is nonregular.

- i. (*Graded for completeness*) Find the (first and/or most significant) logical error in the “proof” above and describe why it’s wrong.
- ii. (*Graded for completeness*) Prove that the set  $A_2$  is actually regular (by finding a regular expression that describes it or a DFA/NFA that recognizes it, and justifying why) **or** fix the proof so that it is logically sound.

### 3. Pumping (10 points):

- (a) (*Graded for correctness*) Give an example of a language over the alphabet  $\{a, b\}$  that has cardinality 5 and for which 4 is a pumping length and 3 is not a pumping length. Is this language regular? A complete solution will give (1) a clear and precise description of the language, (2) a justification for why 4 is a pumping length, (3) a justification for why 3 is not a pumping length, (4) a correct and justified answer to whether the language is regular.
- (b) (*Graded for completeness*) In class and in the reading so far, we’ve seen the following examples of nonregular sets:

$$\begin{array}{lll} \{0^n 1^n \mid n \geq 0\} & \{0^n 1^m \mid 0 \leq m \leq n\} & \{0^n 1^m 0^n \mid n, m \geq 0\} \\ \{0^n 1^n \mid n \geq 2\} & \{0^i 1^{2i} \mid 0 \leq i\} & \{w \in \{0, 1\}^* \mid w = w^R\} \\ \{0^n 1^m \mid 0 \leq n \leq m\} & \{0^i 1^{i+1} \mid 0 \leq i\} & \{ww^R \mid w \in \{0, 1\}^*\} \end{array}$$

Modify one of these sets in some way and use the Pumping Lemma to prove that the resulting set is still nonregular.

**4. Regular and nonregular languages (12 points):** In Week 2’s review quiz, we saw the definition that a set  $X$  is said to be **closed under an operation** if, for any elements in  $X$ , applying to them gives an element in  $X$ . For example, the set of integers is closed under multiplication because if we take any two integers, their product is also an integer .

Prove or disprove each closure claim statement below about the class of regular languages and the class of nonregular languages. Your arguments may refer to theorems proved in the textbook and class, and if they do, should include specific page numbers and references (i.e. write out the claim that was proved in the book and/or class).

Recall the definitions we have:

For language  $L$  over the alphabet  $\Sigma_1 = \{0, 1\}$ , we have the associated sets of strings

$$SUBSTRING(L) = \{w \in \Sigma_1^* \mid \text{there exist } a, b \in \Sigma_1^* \text{ such that } awb \in L\}$$

and

$$EXTEND(L) = \{w \in \Sigma_1^* \mid w = uv \text{ for some strings } u \in L \text{ and } v \in \Sigma_1^*\}$$

- (a) (*Graded for completeness*) The set of regular languages over  $\{0,1\}$  is closed under the *SUBSTRING* operation.
- (b) (*Graded for completeness*) The set of nonregular languages over  $\{0,1\}$  is closed under the *SUBSTRING* operation.
- (c) (*Graded for correctness*) The set of regular languages over  $\{0,1\}$  is closed under the *EXTEND* operation.
- (d) (*Graded for correctness*) The set of nonregular languages over  $\{0,1\}$  is closed under the *EXTEND* operation.