

This assignment is **due in Week 4** and should be submitted to Gradescope.

All work must be **done individually** without consulting anyone else's solutions in accordance with the University's "[Academic Dishonesty and Plagiarism](#)" policies.

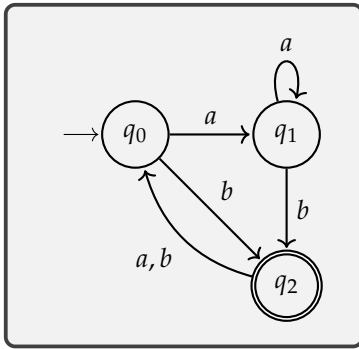
Go to the last page of this document and read the **Submission Instructions**. For clarifications and updates, monitor "[Assignment FAQ](#)".

Problem 1. (5 marks)

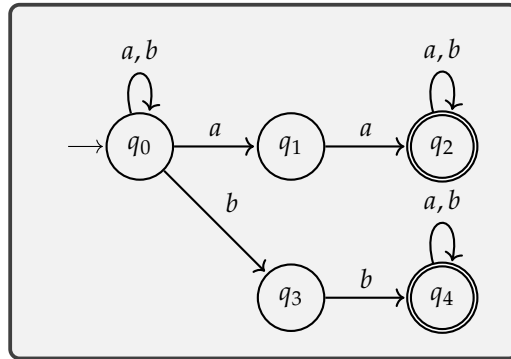
For the regular expression $(a|(ab))^*|b^*$ over $\Sigma = \{a, b\}$, state 5 strings that are in its language, and 5 strings that are not. You score $\max(0, 5 - \text{num_errors})$ points.

Problem 2. (15 marks)

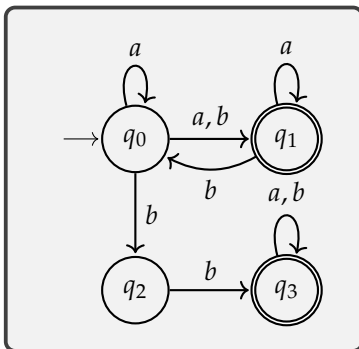
For each of the following three automata over $\Sigma = \{a, b\}$, state 5 strings that are in its language, and 5 strings that are not. You score $\max(0, 5 - \text{num_errors})$ points per part.



Part 1



Part 2



Part 3

Problem 3. (40 marks)

For each of the following languages over $\Sigma = \{a, b\}$, provide a regular expression and an NFA with at most 10 states for the language (note: sometimes a DFA will be sufficient, othertimes an NFA will be needed). Partial marks may be awarded for submissions recognising a sufficiently close language or with more than 10 states.

1. The set of strings that contain at least two as .
2. The set of strings that start with baa and have even length.
3. The set of strings where the final letter appears at least 4 times in the string (including the final letter).
4. The set of strings where the length is even and aa does not occur as a substring.

Problem 4. (10 Marks)

You're a secret agent, receiving coded messages over the alphabet $\Sigma = \{a, b\}$ behind enemy lines telling you whether to keep going or retreat. You've been issued a DFA to interpret the messages for you, based on whether or not the DFA accepts the string. Unfortunately, recently enemy forces have started jamming the airwaves, inserting random garbage into the messages you receive.

Fortunately, you have a solution. You've determined that the noise occurs periodically: there will always be 2 bits of noise, followed by the 1 bit of signal sent by your country, and so on. Based on this, you'll need to use your expert skills in DFA construction to modify your automaton, so that it works on the strings you receive, and you can understand the messages.

That is: Let $f : \Sigma^* \rightarrow \Sigma^*$ be the function that keeps every 3rd letter of a string, and discards the rest. For example, $f(abbbababab) = bba$, $f(babababab) = ba$, $f(ab) = f(b) = f(\epsilon) = \epsilon$. If the clean string was ba , then the noisy string you receive might be $bbbaaa$, $babababa$, $aabbbab$, etc. as these are all strings x where $f(x) = ba$. (Underlines are for clarity only and not part of the actual strings.)

Given a DFA $A = (Q, \Sigma, q_0, \delta, F)$, show how to construct a DFA $B = (Q', \Sigma, q'_0, \delta', F')$ for the language $\{x \in \Sigma^* : f(x) \in L(A)\}$ in terms of A . For example, if ba is accepted by A , then $babababa$ should be accepted by B , as $f(babababa) = ba$.

For full marks you should:

1. Give a short English description of the idea of your solution (including, e.g., what "information" is stored in the states of Q' , see Lecture slides).
2. Define all the components of B in terms of A .

Submission Instructions

You will submit answers to all the problems to Gradescope.

Problems 1, 2 and 3 are autograded. It is essential that you ensure that your submission is formatted so that the autograder can understand it. Upon submitting your responses, you should **wait** for the autograder to provide feedback on whether your submission format was correct. **An incorrectly formatted submission for a question will receive zero marks for that question.** A scaffold will be provided on Ed with the file names the autograder expects.

Problem 1 and 2 format:

The first line of each answer should contain a comma separated sequence of five strings that are in the language, and the second line should contain a comma separated sequence of five strings that are not in the language. For example, if the regular expression was b^* , an example of a correct text file would be:

```
epsilon, b, bb, bbb, bbbb
a, aa, aaa, aaaa, aaaaa
```

Problem 3 format, regular expressions:

Write your regular expressions using the following notation: a b epsilon emptyset Sigma | * ()

E.g., $((a|\epsilon)^*baba)|\Sigma^*|(\text{emptyset}((baa)|b))$

You may use the abbreviations in this course (outermost parentheses may be dropped, and parentheses may be dropped for associative operations of 3+ regexes). Otherwise it is recommended that you follow the syntax strictly. Leaving out parentheses may have unexpected results and is done at your own risk. Check the feedback from the autograder to see how your regex was interpreted and ensure it matches what you intended.

Problem 3 format, NFAs:

```
Sigma = a b
Q = q0 q1 q2 q3
start = q0
F = q1 q2
q0 a q1
q0 b q0
q1 a q1
q1 b q2
q2 a q2
q2 b q3
q3 a q3
q3 b q3
```

1. The first line is the set of alphabet symbols, as a space-separated list
2. The second line is the set of states (Q), as a space-separated list
3. The third line indicates the start state of the automata
4. The fourth line is set of final states (F), as a space-separated list
5. The fifth line is a transition from q_0 to q_1 on input character 'a' (i.e. $\delta(q_0, a) = q_1$)

6. The remaining lines are all transitions. For an NFA, multiple transitions on the same (state, character) pair can occur, on different lines.

Problem 4 format:

Problem 4 is hand graded. You will submit a single **typed pdf (no pdf containing text as images, no handwriting)**. Start by typing your student ID at the top left of the first page of your submission. Do **not** type your name. Do not include a cover page. Submit only your answers to the questions. Do **not** copy the questions. Your pdf must be readable by Turnitin.