**Formal Languages and Automata (CS322)**

Mid-term exam (20 October 2025)

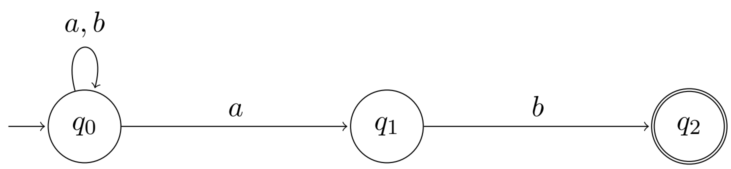
Lecturer: Eunjung KIM

♠ The points, over 5 questions in total, sum up to 113 including 13 bonus points. Your final score will be the minimum of 100 and the total points you earned out of 113.

♠ Whenever you present an automaton, make the tuple explicit, except for the transition function, which you can present as a transition diagram. You don't need to prove the correctness of your automaton. If an automaton is obtained from other automata, present clearly how the modification is done (without proof).

♠ Except for Q2 (c) - if you want bonus points - and Q3, you don't need to provide a formal proof.

**[16 points]** Q1. Consider the NFA $N$ given by the state transition diagram below.



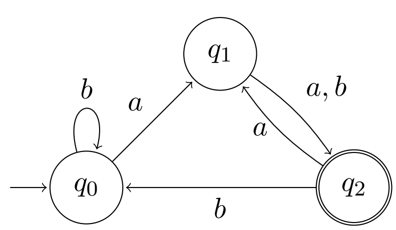

(a) (4 points) Define the language that $N$ recognizes in plain English.

(b) (6 points) Present a deterministic finite automaton $D$ which recognizes the same language as $N$.

(c) (6 points) Present an MSO-sentence (Monadic Second Order sentence) on $\{a, b\}$-strings which define $L(N)$. Specify the vocabulary of MSO logic on $\{a, b\}$-strings and explain the meaning of each relation name in the vocabulary.

♠ An MSO-sentence can use the logical connectives $\neg, \wedge, \vee, \rightarrow$, the quantifiers $\forall, \exists$, brackets $($, $)$. The order of precedence goes (from higher to lower): the quantifiers, then $\neg, \wedge, \vee, \rightarrow$, which are all overridden by the bracket. E.g. $\forall x \exists z \varphi(x) \wedge \exists y \phi(y) \vee \beta \rightarrow \neg\gamma$ is $\big((\forall x \exists z \varphi(x, z) \wedge \exists y \phi(y)\big) \vee \beta) \rightarrow (\neg\gamma)$, and $\exists x \forall z \varphi \wedge \phi$ equals $(\exists x \forall z \varphi) \wedge \phi$, which is different from $\exists x(\forall z \varphi \wedge \phi)$. When confused, use the brackets!

**[20 points + 10 bonus points]** Q2. Consider the DFA $D$ given by the state transition diagram below.



(a) (6 points) Present a regular expression $R$ such that $L(R) = L(D)$ using generalized NFA. Present each step of constructing generalized NFAs ending in $R$.

(b) (6 points) A grammar $G = (V, \Sigma, R, S)$ is said to be *right-linear* if *every rule in* $R$ is in the form $X \to xY$, where $x \in \Sigma \cup \{\epsilon\}$ and $Y \in V \cup \{\epsilon\}$. Write a right-linear context-free grammar $G$ which generates the language $L(D)$ and explain how you obtained the grammar.

(c) (8 points + 10 bonus points) Present how to construct a right-linear context-free grammar $G$ such that $L(G) = L$ for an arbitrary regular language $L$. Bonus points apply if you give a formal proof for the correctness of your grammar construction.

**[32 points]** Q3. Prove or disprove that the following language over $\{0, 1\}$ is regular.

(a) (8 points) $A = \{w \mid w$ contains the same number of occurrences of the substrings 01 and 10$\}$.

(b) (8 points) $B = \{0^n 1^m 0^n \mid n, m \geq 0\}$.

(c) (8 points) The language $L_1$-avoid-$L_2$ defined as

$$\{w \mid w \in L_1 \text{ and } w \text{ does not contain any string in } L_2 \text{ as a substring}\}$$

for two regular languages $L_1$ and $L_2$.

(d) (8 points) $C = \{0^m \mid m = 2^n$ for some integer $n \geq 0\}$.

♠ You can "prove or disprove" in the following way.

- When you claim that a language $L$ is not regular, provide a full proof using pumping lemma or Myhill-Nerode Theorem.
- When you claim that a language $L$ is regular, your proof can use any properties about regular languages covered in the classroom / homework assignments.

**[12 points + 3 bonus points]** Q4. Present a context-free grammar generating the following language; don't forget to specify the 4-tuple!

(a) (6 points) $A = \{0^n 1^m \mid n \leq m + 3\}$.

(b) (6 points + 3 bonus points) $\{w \in \{a, b\} \mid$ the number of $a$'s and the number of $b$'s in $w$ is the same$\}$. Bonus points apply if you present an *unambiguous* CFG.

**[20 points]** Q5. Present a pushdown automaton recognizing the following language.

(a) (6 points) $L = \{0^n 1^n \mid n$ is a positive odd integer$\}$.

(b) (6 points) The language of the grammar $G = (\{S, A, B\}, \{a, b\}, R, S)$, where the production rules $R$ are given as

$$S \to aSB \mid bSA \mid \epsilon \qquad\qquad A \to a \qquad\qquad B \to b$$

(c) (8 points) The language $A \star B = \{xy \mid x \in A, y \in B, |x| = 2|y|\}$, where $A$ is the language recognized by the NFA of Question 1 and $B$ is the language recognized by the DFA of Question 2.