Question Paper for Midsem Examination
Course CS310

21 September, 2023

- The question paper carries 100 marks in total and consists of 7 questions.
- Partial answers carry partial marks. Hence it is advantageous to show working of your answers.
- This is a paper and pen examination. Answers must be written in an answersheet which must be submitted with role number clearly marked.
- Additionally, Answer to each question must ALSO be uploaded on SAFE as an image answer.
- The examination including uploading on SAFE must be completed in 120 minutes. Do not spend too much time on a single questions.
- Students may keep 3 printed or handwritten A4 size sheets with them for reference. Use of books, notebooks, laptops etc. is not allowed.
- Good Luck!

Q1 (16 marks) For each of the following statements please state whether it is true or false. Please provide a short justification in no more than 3 lines for your answer.

A) A language L is co-finite if $\Sigma^* - L$ is finite. If L is co-finite then L is regular. (True)

$\sim L$ is finite $\Rightarrow$ $\sim L$ is regular
$\Rightarrow$ $\sim\sim L$ is regular
(as regular languages are closed under complement)
$\Rightarrow$ L is regular

B) If complement $\sim L$ is context-free then L must be context-free. (False)

Context free languages are NOT closed under complement
$\therefore$ $\sim L$ is context-free $\not\Rightarrow$ $\sim\sim L$ is Context free
$\therefore$ $\sim L$ is context-free $\not\Rightarrow$ L is context-free.

C) If $L_1 \cap L_2$ is regular then L_1 and L_2 must be regular. (False)

Consider $L_1 = \{a^n b^n\} \cup \{a^*\}$ Not regular

$L_2 = \{b^n a^n\} \cup \{a^*\}$ Not regular

Then $L_1 \cap L_2 = \{a^*\}$ regular

D) If L_1 and L_2 are context-free then $(L_1 - L_2)$ is necessarily context-free. (False)

Let $L_1 = \Sigma^*$. It is regular and hence context-free.

Then $L_1 - L_2 = \sim L_2$ complement of L_2.

$\therefore$ CFL closed under $L_1 - L_2 \Rightarrow$ CFL closed under $\cap$

But CFL not closed under $\cap$

E) If L is context-free then L^* is necessarily context-free. (True)

Given grammer (N_1, Σ, P_1, S_1) for L we can construct grammer for L^*. Add rules

$S \rightarrow S_1 S \mid S_1$

F) $\{a^i b^j \mid i \neq j\}$ is context-free. (True)

$S \rightarrow A \mid B$ $A \rightarrow aA \mid aE$ $B \rightarrow Bb \mid Eb$

$E \rightarrow aEb \mid \epsilon$

G) $\{a^i b^j a^k b^l \mid i = j \wedge k = l\}$ is context-free. (True)

$L_1 = \{a^i b^j \mid i = j\}$ and $L_2 = \{a^k a^l \mid k = l\}$ are CFL

$\therefore L_1 \cdot L_2$ is CFL

H) $\{a^i b^j a^k b^l \mid i = k \wedge j = l\}$ is context-free. (False)

Give Full marks ignoring justification for correct answer

Rubric:
* 1 mark each for correct option (true/false)
* 1 mark for correct justification.

Q2 (9 marks) In showing that $A = \{\ ww \mid w \in \{a, b\}^*\}$ is not regular using the pumping lemma, the Demon chooses $k = 4$. Which of the following choice of x,y,z s.t. $xyz \in L$ is adequate for showing that A is not regular using the pumping lemma? Mark all correct answers. Please briefly justify your answers.

A) $x = \epsilon$, $y = abab$, $z = \epsilon$?

B) $x = aaaab$, $y = aaaa$, $z = b$ **(correct)**

C) $x = aaaab$, $y = aaa$, $z = ab$

D) $x = \epsilon$, $y = abba$, $z = abba$ **(correct)**

Rubric

1) 1 Mark each for correct true/False (4 marks)

2) 1 Mark each for correct justification (4 mk)

3) 1 Mark for correct statement of pumping lemma requirements. (1 mk)

Solution:

By Pumping Lemma, x, y, z must satisfy $x \cdot y \cdot z \in A$ & $|y| \geq k$ and y is not pumpable, i.e.

$\forall u, v, w.\ y = uvw$ & $|v| > 0$

$\exists i.\ xuv^i wz \notin A$.

A) False.

Demon chooses $u = \epsilon$, $v = abab$, $w = \epsilon$.

Then $\forall i.\ \epsilon \cdot \epsilon \cdot (abab)^i \cdot \epsilon \cdot \epsilon \in A$.

B) True.

$y = aaaa$. We have $|y| = 4 \geq k = 4$ and y is not pumpable. See below.

If Demon chooses $u = a^p$, $v = a^q$, $w = a^r$ with $u + v + w = k$ and $v > 0$ then $xuv^i wz$ is $aaaab \cdot a^p a^{q \times i} a^r \cdot b \notin A$ for $i = 0, 2, 3, 4, \ldots$

C) False

$|y| = 3$ does not satisfy $|y| \geq k$

D) True

We have $|y| \geq k$ but y is not pumpable.

Q3 (15 marks) Let $A = \{xy \mid x, y \in \{a,b\}^* \text{ and } \#a(x) = \#b(y)\}$ and $B = \{x\$y \mid x, y \in \{a,b\}^* \text{ and } \#a(x) = \#b(y)\}$ be two languages. Note that the alphabet of A is $\{a.b\}$ whereas the alphabet of B is $\{a.b,\$\}$ with letter \$ in it. One of A and B is regular and the other is not. Identify which is regular and which is not regular. *Justify your answer with formal proofs* (for both A and B).

Rubric

- 2 marks for identifying A regular, B not regular
- 2 marks for stating that $A = \{a,b\}^*$
- 6 marks for proving that $A = \{a,b\}^*$
- 5 marks for proving using pumping lemma or Myhill Nerode Theorem that B is not regular.

Solution

$A = \{a,b\}^*$ which is regular. E.g. $(a+b)^*$

B is not regular.

(1) Claim: $A = \{a,b\}^*$

Proof: $A \subseteq \{a,b\}^*$ ∵ obvious from Defn A.

Conversely, we show $\forall z \in \{a,b\}^*$ that $z \in A$.

(By Induction on structure of z)

Basis ($z = \epsilon$) Let $x = y = \epsilon$. Then $z = x \cdot y$ & $\#a(x) = \#b(y) = 0$

$\therefore z \in A$.

Induction Step 1. ($z = w \cdot a$)

By Ind. Hyp. $\exists x, y.\ w = x \cdot y$ & $\#a(x) = \#b(y)$

Let $x' = x$, $y' = y \cdot a$

Then $z = x' \cdot y'$ & $\#a(x') = \underbrace{\#a(x) = \#b(y)}_{\text{Ind. Hyp.}} = \#b(y')$

$\therefore z \in A$

Induction Step 2 ($z = w \cdot b$)

By Ind. Hyp. $\exists x, y.\ w = x \cdot y$ & $\#a(x) = \#b(y)$

Case 1 ($y = \epsilon$). Let $x' = x \cdot b$, $y' = \epsilon$.

Then $z = x' \cdot y'$ and $\#a(x') = \#a(x) = \#b(y) = 0 = \#_b(y')$

$\therefore z \in A$

Case 2 ($y = aV$) Let $x' = xa$, $y' = Vb$

Then $z = x'y'$ and $\#a(x') = \#a(x) + 1 = \#b(V) + 1$
$= \#b(y')$

$\therefore z \in A$

Case 3 ($y = bV$) Let $x' = xb$, $y' = Vb$

Then $z = x'y'$ and $\#a(x') = \#a(x) = \#b(y)$
$= \#b(V) + 1 = \#b(y')$

$\therefore z \in A$

(2) Claim B is not regular

Proof Using pumping lemma

1). Demon chooses k

2) Let $x = \epsilon$, $y = a^k$, $z = \$ \cdot b^k$.

Hence $|y| \geq k$ and $x \cdot y \cdot z \in B$.

We show that y is not pumpable

3) Demon chooses $u = a^p$, $v = a^q$, $w = a^r$, $p+q+r = k$, $q > 0$

4) For any $i \neq 1$ we have $p + iq + r \neq k$

Hence $x \cdot u \cdot v^i \cdot w \cdot z = \epsilon \cdot a^p v^{iq} a^r \cdot \$ b^k \notin B$

for $i = 0, 2, 3, \cdots$

Hence B is not regular.

Q4 (15 marks) Consider the regular expressions $a^*b^* \; + \; b^*a^*$ defining a language R.

1. Determine the equivalence class of the language induced Nerode congruence $\equiv_R$. Give a regular expressions for each of these classes.
2. Draw the DFA $A_{(\equiv_R)}$ for R derived from these equivalence classes.
3. Check using the Hopcroft DFA minimization algorithm that the DFA you have drawn is minimal. Use the table data-structure proposed in class. Clearly indicate which phase separates a pair of states starting with the initialization phase 0.

Rubric

1. 4 marks for correctly giving 6 regular expressions. (5 marks)
   (• If REST is improper (eg. as two classes) deduct 2 marks)
   1 Mark for stating that words in different classes can be separated.

2. 5 Marks for correctly drawing the automaton (5 marks)

3. 5 Marks for correctly drawing minization table
   (1 mark for correct table format)
   (Deduct 2 marks if phases are not correctly labelled.) (5 marks)

Solution. Let r^+ abbreviate $r \cdot r^*$.

The $\equiv_R$ relation partitions Σ^* into following 6 equivalence classes

1). ϵ $\quad a^+ \quad b^+ \quad a^+b^+ \quad b^+a^+$, REST

where REST $= (a^+b^+a\Sigma^* + b^+a^+b\Sigma^*)$

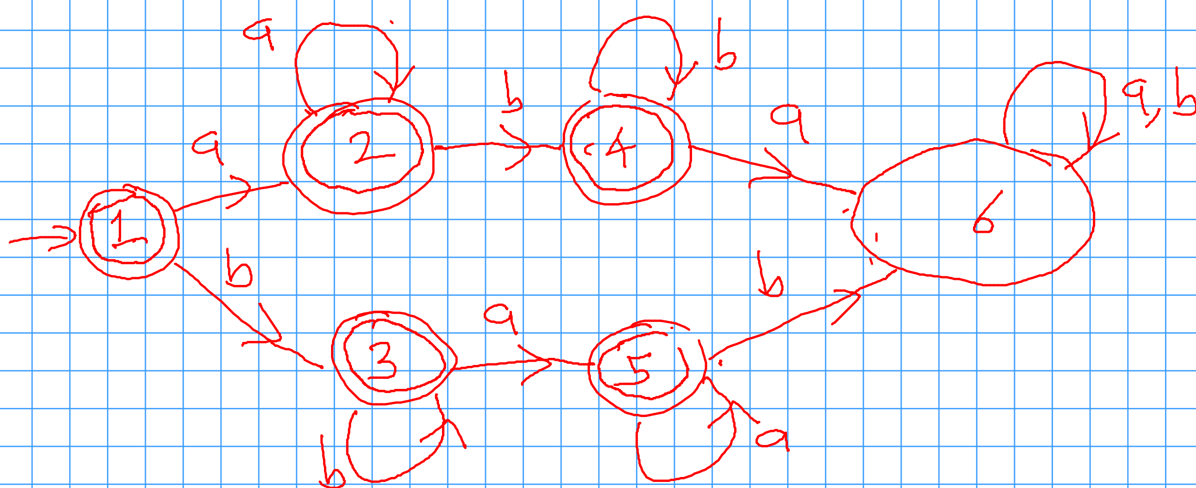
Note that any two classes can be separated.

E.g. $x \in a^+b^+$ & $y = b^+a^+$ then $x \not\equiv_R y$

as $x \cdot b \in R$ but $y \cdot b \notin R$.

No two words in the same class can be separated.

2)



Number the states as

$\epsilon \to 1$, $a^+ \to 2$, $b^+ \to 3$, $a^+b^+ \to 4$, $b^+a^+ \to 5$,

REST $\to 6$

3) Minimization

	1	2	3	4	5	6
1	•					
2	$\checkmark_2$	•				
3	$\checkmark_2$	$\checkmark_2$	•			
4	$\checkmark_1$	$\checkmark_1$	$\checkmark_1$	•		
5	$\checkmark_1$	$\checkmark_1$	$\checkmark_1$	$\checkmark_1$	•	
6	$\checkmark_0$	$\checkmark_0$	$\checkmark_0$	$\checkmark_0$	$\checkmark_0$	•

We give phasewise marking separating pair of states (p, q).
A $\checkmark_i$ shows separation in phase i.

Since all pairs of states can be separated there are no equivalent states and the automaton is minimal.

Q5 **(15 marks)** Let a grammar G be given by the productions

$S \to aSb \mid bSa \mid SS \mid \epsilon.$

Let language $M = \{x \in \{a,b\}^* \mid PROP(x)\}$. Define mathematically $PROP(x)$ of word x such that $L(G) = M$. Formally, *by giving a complete proof*, show that $L(G) = M$.

Rubric

1) 3 marks for correctly defining PROP(x)

2) 6 Marks for Proving $L(G) \subseteq M$

Give partial marks for mistaken proof.

- 1 mark for induction on length of derivation
- 1 mark for correct Ind. Hypothesis
- 4 marks for proof

3) 6 Marks for Proving $M \subseteq L(G)$

Give Partial marks for mistaken Proof.

- 1 mark for correct induction Hyp
- 2 marks for correct case split in induction step
- 3 marks for correct proof of induction steps.

Solution

(1) $Prop(x) \stackrel{def}{=} \#a(x) = \#b$

$\therefore M = \{x \in \{a,b\}^* \mid \#a(x) = \#b(x)\}$.

(2) Claim $L(G) \subseteq M$.

Proof. We show by Ind. on length n of derivation that:

If $S \xrightarrow{n} \alpha$ then $\#a(\alpha) = \#b(\alpha)$, for $\alpha \in (N \cup \Sigma)^*$

Basic step $S \xrightarrow{0} S$. Also $\#a(S) = \#b(S) = 0$.

Induction Step: Let $S \xrightarrow{n} \beta S \gamma \xrightarrow{1} \alpha$.

By Induction hypothesis: $\#a(\beta S \gamma) = \#b(\beta S \gamma)$.

Now α is obtained by applying one of S productions in the last step.

Case 1 ($S \to \epsilon \mid SS$) $\therefore \alpha = \beta\gamma$ or $\alpha = \beta SS\gamma$.

$\therefore \#a(\alpha) = \#a(\beta S \gamma) = \#b(\beta S \gamma) = \#b(\alpha)$

Case 2 ($S \to aSb \mid bSa$) $\therefore \alpha = \beta aSb\gamma$ or $\beta bSa\gamma$

$\therefore \#a(\alpha) = \#a(\beta S \gamma) + 1 = \#b(\beta S \gamma) + 1 = \#b(\alpha)$

$\square$

(3) Claim: $x \in M$ then $S \xrightarrow{*} x$

Proof: By complete induction on $|x|$.

Basic Step ($x = \epsilon$). From G we have $S \xrightarrow{1} \epsilon$

Induction Step ($|x| > 0$). Then $x = ayb$ OR bya or there exist u, v s.t. $x = u \cdot v$ and $0 < |u| < |x|$ and $u, v \in M$.

Case 1. ($x = ayb$) Since $\#a(y) = \underbrace{\#a(x) - 1 = \#b(x) - 1}_{\text{since } x \in M}$

$= \#b(y)$

$\therefore y \in M$ & $|y| < |x|$

$\therefore S \xrightarrow{*} y$ (by Ind Hyp).

$\therefore S \to aSb \xrightarrow{*} ayb = x$

Case ($x = bSa$) Similar.

Case 3 ($x = u \cdot v$ & $0 < |u| < |x|$ & $u, v \in M$)

$\therefore S \xrightarrow{*} u$ and $S \xrightarrow{*} v$ (by Ind Hyp).

$\therefore S \xrightarrow{1} SS \xrightarrow{*} uS \xrightarrow{*} uv = x$. $\square$

Q6 (15 marks) Given a CFG grammar G, design an algorithm to determine if its language is empty, i.e. $L(G) = \emptyset$.

Rubric:

- 10 marks for correct algorithm
- 5 marks for intuitive explanation of its working and correctness

Solution: Given a CFG G, a nonterminal A is called generative if $A \xrightarrow{*} x$ for some $x \in \Sigma^*$

Then $L(G)$ is non-empty iff S is generative

We give algorithm to find all generative non-term.

① $Old = \{\}$; \\ empty set

② $New = \{A \in N \mid A \to \alpha \in P \text{ and } \alpha \in \Sigma^*\}$

③ while $Old \neq New$

{

④ $Old = New$;

⑤ $New = New \cup \{A \in N \mid A \to \alpha \in P \text{ and } \alpha \in (\Sigma \cup Old)^*\}$

}

⑥ return New

We iteratively expand the set of generative nonterminals. In step 2, we include all nonterms which by 1 application of its productions give terminal words.

Having previous estimate (stored in Old) we expand it by Noterms A which in 1-step give rise to α consisting of terminals and Old non-term. Since all nonterms in α are known to be generative this makes A also generative.

Claim: New returned is exactly the set of generative nonterms.

Claim 1: $A \in New \Rightarrow A$ is generative.

Proof outline: We maintain the claim as loop invariant.

Initialization step (2) as well as progress step (5) maintain this invariant.

Claim 2. If $A \xRightarrow{*} x$ for some $x \in \Sigma^*$ then $A \in New$.

proof outline: Consider the shortest derivation

$A \xRightarrow{n} x$ from A to a terminal string x.

Let New_n denote set New after n iterations.

We can argue by induction on n that

$A \in New_n$.

Hence on termination we have the result.

Q7 (15 marks) Convert the following Chomsky NormalForm (CNF) grammar into language equivalent Griebach Normal Form (GNF) grammar using the procedure discussed in Class/Hopcroft-Ullman book. Give intermediate grammars obtained as you follow each step of the procedure. Assume that the nonterminals are ordered as $S < A < B < X$.

$$\begin{aligned}
&S \to XB \mid AA \\
&A \to a \mid SA \\
&B \to b \\
&X \to a
\end{aligned}$$

Rubric:

- Step (0) → (1) Eliminate backward recursion
  (3 marks)
- Step (1) → (2) Eliminate left recursion
  (5 marks)
- Step (2) → (3) Substitute for leftmost symbol
  (4 marks)
- Step (3) → (4) Substitute leftmost symbol in Z productions
  (3 marks)

Solution: Order $S < A < B < X$

0) $S \to XB \mid AA \quad A \to a \mid SA \quad B \to b \quad X \to a$

$A \to SA$ is backward recursive. Eliminate

1) $S \to XB \mid AA \quad A \to a \mid XBA \mid AAA \quad B \to b \quad X \to a.$

$A \to AAA$ is left recursive. Eliminate.

2) $S \to XB \mid AA \quad A \to aZ \mid XBAZ \mid a \mid XBA \quad B \to b \quad X \to a.$

$Z \to AAZ \mid AA$

Z is new nonterminal. Only Forward rules.

Substitute where leftmost symbol on right is nonterminal.
Ignore Z rules

3) $Z \to AAZ \mid AA \quad X \to a \quad B \to b$

$A \to aZ \mid aBAZ \mid a \mid aBA$

$S \to aB \mid aZA \mid aBAZA \mid aA \mid aBAA$

Substitute for leftmost symbol on right of Z productions

4) $X \to a \quad B \to b$

$A \to aZ \mid aBAZ \mid a \mid aBA$

$S \to aB \mid aZA \mid aBAZA \mid aA \mid aBAA$

unchanged

$Z \to aZAZ \mid aBAZAZ \mid aAZ \mid aBAAZ \mid$

$aZA \mid aBAZA \mid aA \mid aBAA$

Now the grammer is in GNF.