Middle East Technical University Department of Computer Engineering

Formal Languages and Abstract Machines
Take Home Exam 2

Ertuğrul Aypek
2171270

1 Context-Free Grammars (10 pts)

a) Give the rules of the Context-Free Grammars to recognize strings in the given languages where $\Sigma = \{a, b\}$ and S is the start symbol.

$L(G) = \{w \mid \ w \in \Sigma^*;\ |w| \geq 3;$ (2/10 pts)
the first and the second from the last symbols of w are the same$\}$

G=(V, Σ, R, S) where $V = \{a, b, S, A\}$ is the alphabet, Σ and S is given in the question,
R={
$S \rightarrow aAaa \mid aAab \mid bAba \mid bAbb$,
$A \rightarrow aA \mid bA \mid e\}$ are the rules

$L(G) = \{w \mid \ w \in \Sigma^*;$ the length of w is odd$\}$ (2/10 pts)

G=(V, Σ, R, S) where $V = \{a, b, S, A\}$ is the alphabet, Σ and S is given in the question,
R={
$S \rightarrow aA \mid bA$,
$A \rightarrow aaA \mid abA \mid baA \mid bbA \mid e\}$ are the rules.

$L(G) = \{w \mid \; w \in \Sigma^*;\; n(w,a) = 2 \cdot n(w,b)\}$ where $n(w,x)$ is the number of x symbols in w (3/10 pts)

G=(V, Σ, R, S) where $V = \{a, b, S, X\}$ is the alphabet, Σ and S is given in the question,
R={
$S \rightarrow X,$
$X \rightarrow e \mid aab \mid aba \mid baa,$
$X \rightarrow Xaab \mid Xaba \mid Xbaa,$
$X \rightarrow aXab \mid aXba \mid bXaa,$
$X \rightarrow aaXb \mid abXa \mid baXa,$
$X \rightarrow aabX \mid abaX \mid baaX\}$ are the rules.

b) Find the set of strings recognized by the CFG rules given below: (3/10 pts)

$S \rightarrow X \mid Y$
$X \rightarrow aXb \mid A \mid B$
$A \rightarrow aA \mid a$
$B \rightarrow Bb \mid b$
$Y \rightarrow CbaC$
$C \rightarrow CC \mid a \mid b \mid \varepsilon$

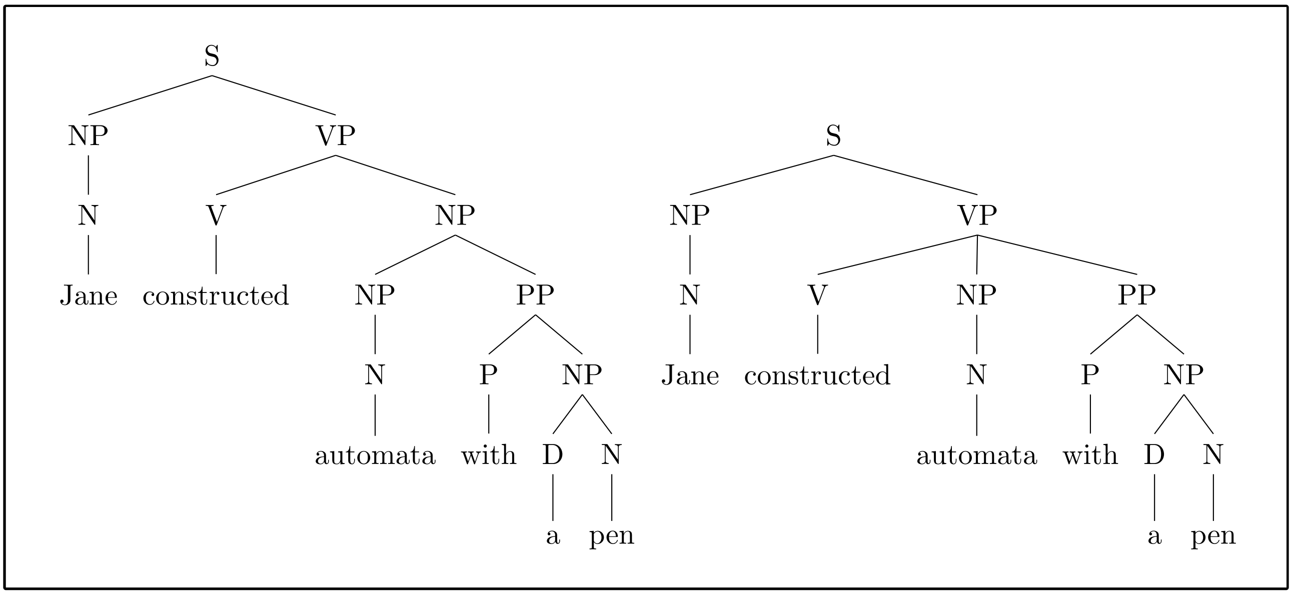
Let the name of given CFG be G.
G=(V, Σ, R, S) where V=$\{a, b, S, X, Y, A, B, C\}$, $\Sigma = \{a, b\}$, R is given in the question and S is the start symbol.
L(G)=$\{aa^* \cup bb^* \cup a^{n+1}a^*b^n \cup a^nb^*b^{n+1} \cup (a \cup b)^*ba(a \cup b)^* \mid n \geq 1; n \in \mathbb{N}^+\}$ is the set of strings recognized by G.

2 Parse Trees and Derivations (20 pts)

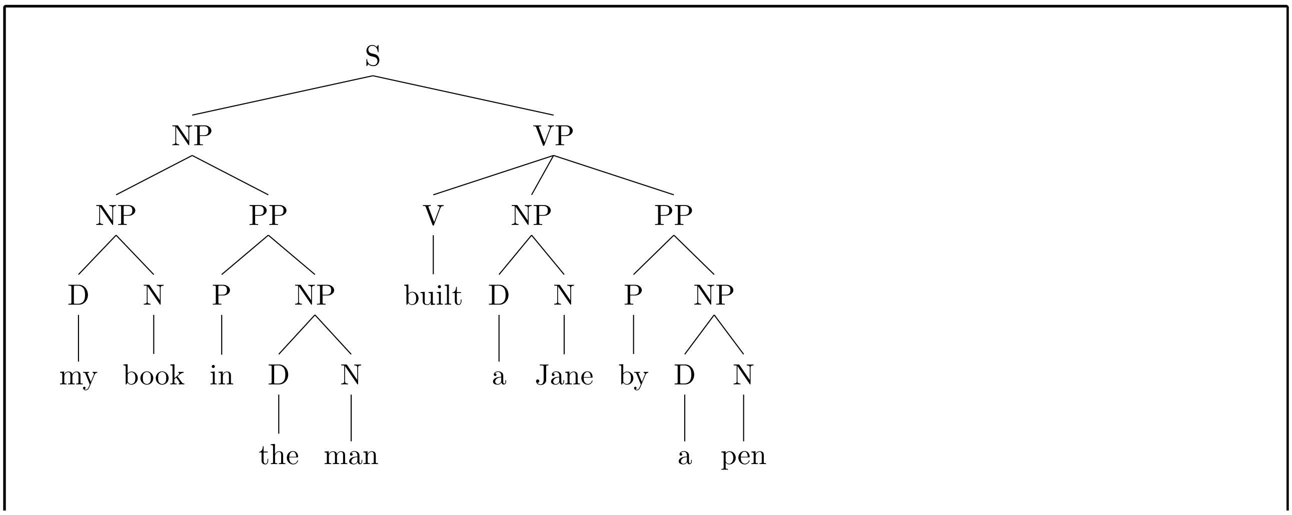
Given the CFG below, provide parse trees for given sentences in **a** and **b**.

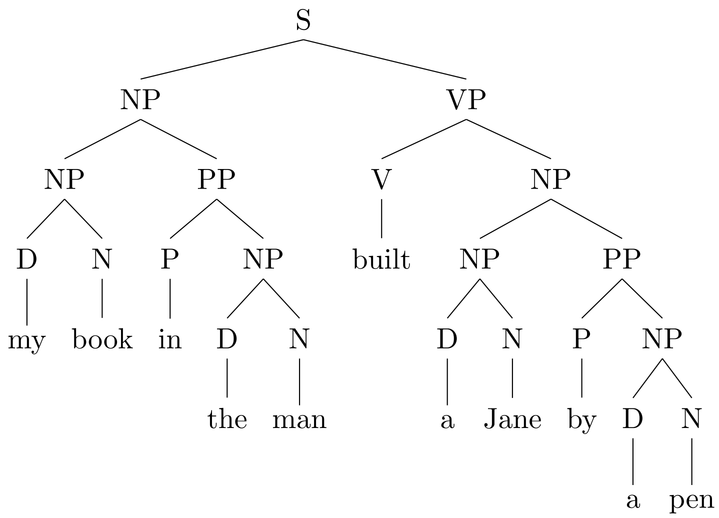
```
S   → NP VP
VP  → V NP | V NP PP
PP  → P NP
NP  → N | D N | NP PP
V   → wrote | built | constructed
D   → a | an | the | my
N   → John | Mary | Jane | man | book | automata | pen | class
P   → in | on | by | with
```

a) Jane constructed automata with a pen (4/20 pts)



b) my book in the man built a Jane by a pen (4/20 pts)




S
NP VP
NP PP V NP
D N P NP built NP PP
my book in D N D N P NP
the man a Jane by D N
a pen

Given the CFG below, answer **c**, **d** and **e**

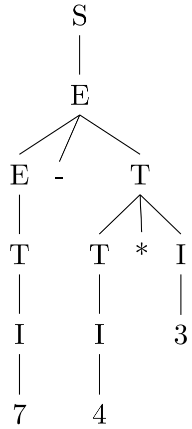
```
S  → E
E  → E + T | E - T | T
T  → T * I | T / I | I
I  → 0 | 1 | 2 | 3 | 4 | 6 | 7 | 8 | 9
```

c) Provide the left-most derivation of 7 - 4 * 3 step-by-step and plot the final parse tree matching that derivation (4/20 pts)

Here is the leftmost derivation:

$$S \overset{L}{\Rightarrow} E \overset{L}{\Rightarrow} E\text{-}T \overset{L}{\Rightarrow} T\text{-}T \overset{L}{\Rightarrow} I\text{-}T \overset{L}{\Rightarrow} 7\text{-}T \overset{L}{\Rightarrow} 7\text{-}T*I \overset{L}{\Rightarrow} 7\text{-}I*I \overset{L}{\Rightarrow} 7\text{-}4*I \overset{L}{\Rightarrow} 7\text{-}4*3$$

Here is the parse tree:

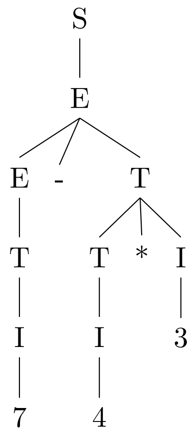


d) Provide the right-most derivation of 7 - 4 * 3 step-by-step and plot the final parse tree matching that derivation (4/20 pts)

Here is the rightmost derivation:

$$S \overset{R}{\Rightarrow} E \overset{R}{\Rightarrow} E\text{-}T \overset{R}{\Rightarrow} E\text{-}T*I \overset{R}{\Rightarrow} E\text{-}T*3 \overset{R}{\Rightarrow} E\text{-}I*3 \overset{R}{\Rightarrow} E\text{-}4*3 \overset{R}{\Rightarrow} T\text{-}4*3 \overset{R}{\Rightarrow} I\text{-}4*3 \overset{R}{\Rightarrow} 7\text{-}4*3$$

Here is the parse tree:

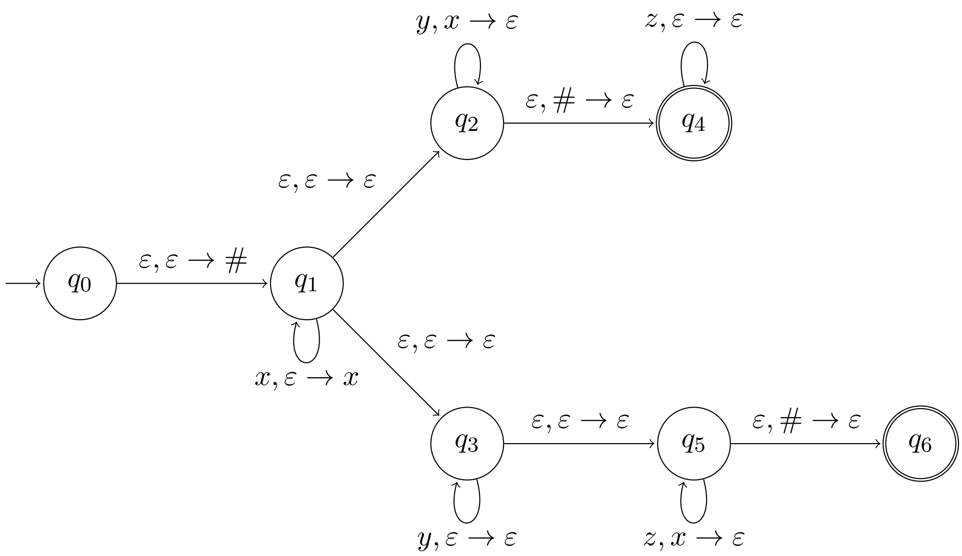


e) Are the derivations in **c** and **d** in the same similarity class? (4/20 pts)

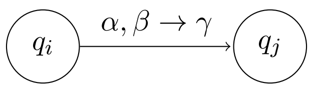
Yes, these two derivations are in the same similarty class because their parse trees are the same as shown.

3 Pushdown Automata (30 pts)

a) Find the language recognized by the PDA given below (5/30 pts)



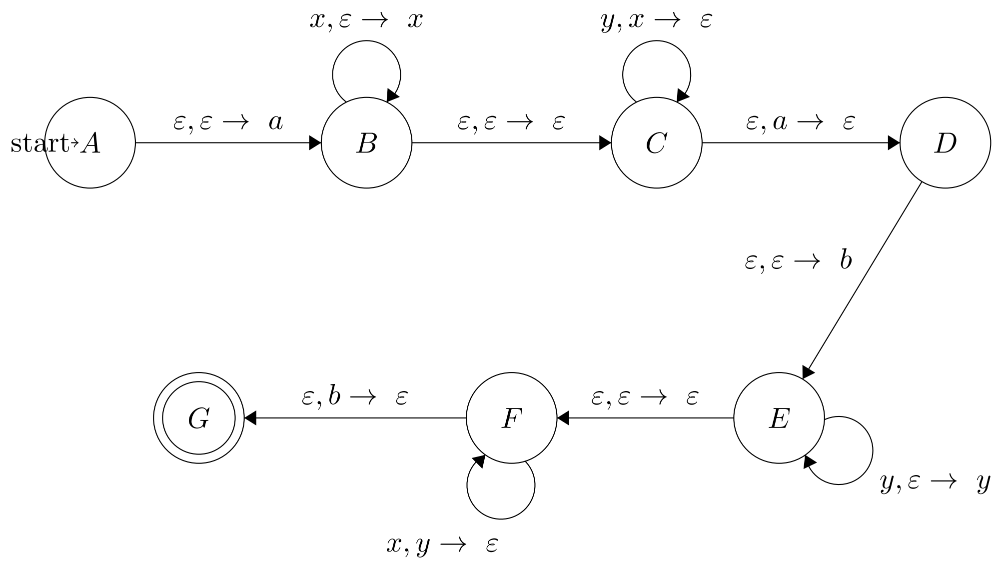
where the transition $((q_i, \alpha, \beta), (q_j, \gamma))$ is represented as:



L=$\{x^n y^n z^* \cup x^n y^* z^n \mid n \geq 0; n \in \mathbb{N}\}$ is the language recognized by given PDA.

b) Design a PDA to recognize language $L = \{x^n y^{m+n} x^m \mid \; n, m \geq 0; \; n, m \in \mathbb{N}\}$ (5/30 pts)

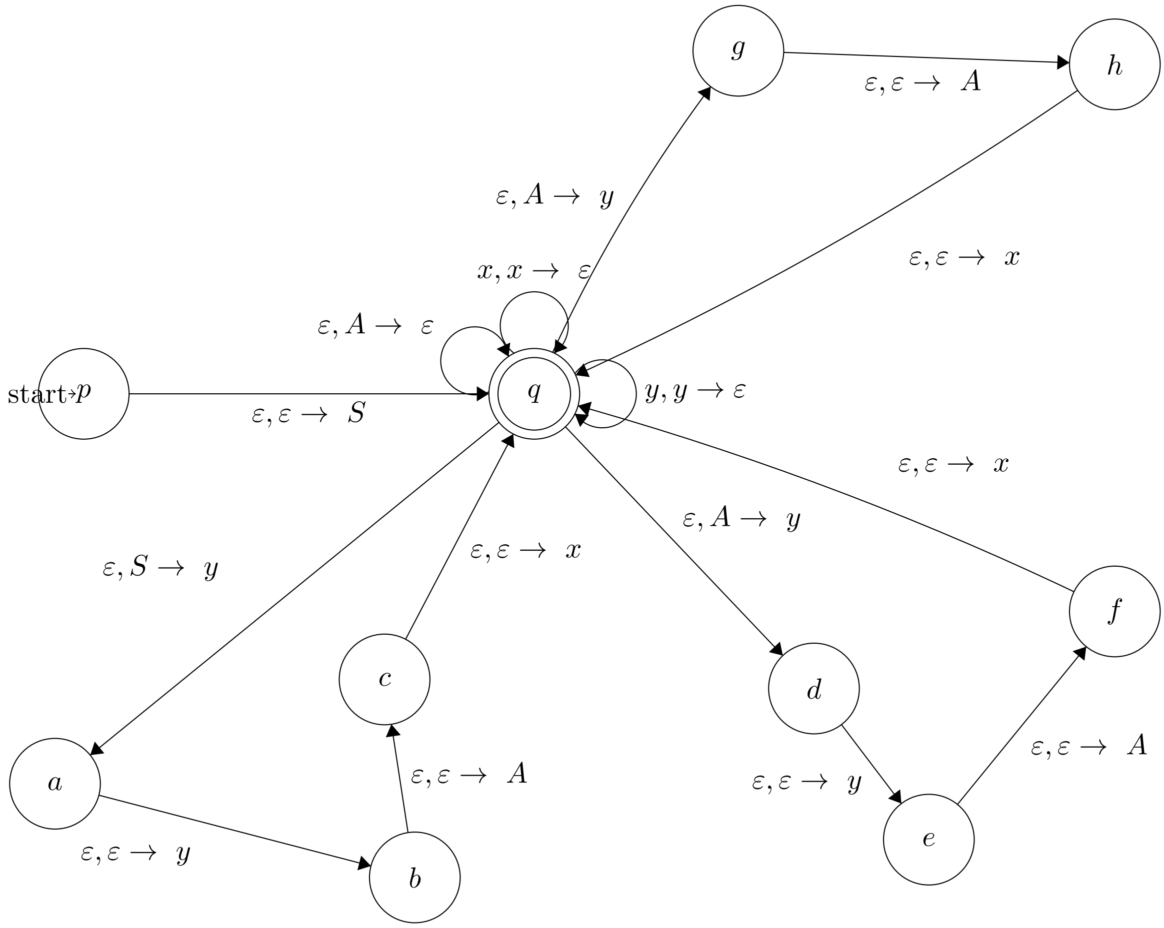
Let $M = (K, \Sigma, \Gamma, \Delta, A, F)$ be a pushdown automata.Where;
K=$\{A, B, C, D, E, F, G\}$,
$\Sigma = \{x, y\}$,
$\Gamma = \{a, b, x, y\}$,
A is the initial state,
$\Delta = \{((A, \varepsilon, \varepsilon), (B, a)),$
$((B, x, \varepsilon), (B, x)),$
$((B, \varepsilon, \varepsilon), (C, \varepsilon)),$
$((C, y, x), (C, \varepsilon)),$
$((C, \varepsilon, a)(D, \varepsilon)),$
$((D, \varepsilon, \varepsilon), (E, b)),$
$((E, y, \varepsilon), (E, y)),$
$((E, \varepsilon, \varepsilon), (F, \varepsilon)),$
$((F, x, y), (F, \varepsilon)),$
$((F, \varepsilon, b), (G, \varepsilon))\}$
F=$\{G\}$



start
A
B
C
D
E
F
G
$\varepsilon, \varepsilon \rightarrow a$
$x, \varepsilon \rightarrow x$
$\varepsilon, \varepsilon \rightarrow \varepsilon$
$y, x \rightarrow \varepsilon$
$\varepsilon, a \rightarrow \varepsilon$
$\varepsilon, \varepsilon \rightarrow b$
$y, \varepsilon \rightarrow y$
$\varepsilon, \varepsilon \rightarrow \varepsilon$
$x, y \rightarrow \varepsilon$
$\varepsilon, b \rightarrow \varepsilon$

c) Design a PDA to recognize language $L = \{x^n y^m \mid n < m \leq 2n;\ n, m \in \mathbb{N}^+\}$ (10/30 pts)
Do not use multi-symbol push/pop operations in your transitions.
Simulate the PDA on strings *xxy* (with only one rejecting derivation) and *xxyyyy* (accepting derivation) with transition tables.

The grammmar rules for given language (where V=$\{x, y, S, A\}$, $\Sigma = \{x, y\}$,S is the start symbol) is:
S→xAyy
A→xAy|xAyy| ε
Following the algorithm to convert a CFG to PDA in Lemma 3.4.1 in the book we have a PDA call it M:
$M = (K, \Sigma, \Gamma, \Delta, p, F)$
where;
K=$\{p, q, a, b, c, d, e, f, g, h\}$,
$\Sigma = \{x, y\}$
$\Gamma = \{x, y, S, A\}$
$\Delta = \{((p, \varepsilon, \varepsilon)(q, S)),$
$((q, \varepsilon, S), (a, y)),$
$((a, \varepsilon, \varepsilon), (b, y)),$
$((b, \varepsilon, \varepsilon), (c, A)),$
$((c, \varepsilon, \varepsilon), (q, x)),$
$((q, \varepsilon, A), (d, y)),$
$((d, \varepsilon, \varepsilon), (e, y)),$
$((e, \varepsilon, \varepsilon), (f, A)),$
$((f, \varepsilon, \varepsilon), (q, x)),$
$((q, \varepsilon, A), (g, y)),$
$((g, \varepsilon, \varepsilon), (h, A)),$
$((h, \varepsilon, \varepsilon), (q, x)),$
$((q, \varepsilon, A), (q, \varepsilon)),$
$((q, x, x), (q, \varepsilon)),$
$((q, y, y), (q, \varepsilon))\},$
p is the start symbol,
F=$\{q\}$
Here is the PDA diagram:



Simulation of xxy

State	Unread Input	Stack	Transition
p	xxy	ε	-
q	xxy	S	$((p, \varepsilon, \varepsilon)(q, S))$
a	xxy	y	$((q, \varepsilon, S), (a, y))$
b	xxy	yy	$((a, \varepsilon, \varepsilon), (b, y))$
c	xxy	Ayy	$((b, \varepsilon, \varepsilon), (c, A))$
q	xxy	$xAyy$	$((c, \varepsilon, \varepsilon), (q, x))$
q	xy	Ayy	$((q, x, x), (q, \varepsilon))$
q	xy	yy	$((q, \varepsilon, A), (q, \varepsilon))$

Simulation of $xxyyyy$

State	Unread Input	Stack	Transition
p	$xxyyyy$	ε	-
q	$xxyyyy$	S	$((p,\varepsilon,\varepsilon)(q,S))$
a	$xxyyyy$	y	$((q,\varepsilon,S),(a,y))$
b	$xxyyyy$	yy	$((a,\varepsilon,\varepsilon),(b,y))$
c	$xxyyyy$	Ayy	$((b,\varepsilon,\varepsilon),(c,A))$
q	$xxyyyy$	$xAyy$	$((c,\varepsilon,\varepsilon),(q,x))$
q	$xyyyy$	Ayy	$((q,x,x),(q,\varepsilon))$
d	$xyyyy$	yyy	$((q,\varepsilon,A),(d,y))$
e	$xyyyy$	$yyyy$	$((d,\varepsilon,\varepsilon),(e,y))$
f	$xyyyy$	$Ayyyy$	$((e,\varepsilon,\varepsilon),(f,A))$
q	$xyyyy$	$xAyyyy$	$((f,\varepsilon,\varepsilon),(q,x))$
q	$yyyy$	$Ayyyy$	$((q,x,x),(q,\varepsilon))$
q	$yyyy$	$yyyy$	$((q,\varepsilon,A),(q,\varepsilon))$
q	yyy	yyy	$((q,\varepsilon,A),(q,\varepsilon))$
q	yy	yy	$((q,\varepsilon,A),(q,\varepsilon))$
q	y	y	$((q,\varepsilon,A),(q,\varepsilon))$
q	e	e	$((q,\varepsilon,A),(q,\varepsilon))$

d) Given two languages L' and L as $L' = \{w \mid\ w \in L;\ |w| = 4n + 2\ for\ n \in \mathbb{N}\}$ (10/30 pts)
If L is a CFL, show that L' is also a CFL by constructing an automaton for L' in terms of another automaton that recognizes L.

If L is a context free language, then $L = L(M_1)$ for some PDA $M_1 = (K_1, \Sigma, \Gamma_1, \Delta_1, s_1, F_1)$. Let $L'' = \{w \mid w \in \Sigma^*; |w| = 4n + 2; n \in \mathbb{N}\}$ be a regular language of length 4n+2 over the same alphabet with L. We said that it is regular since it is very clear to see. For example if $\Sigma = \{a\}$, $L'' = \{aa(aaaa)^*\}$ can be written as a regular expression. And for all finite Σ's, there will be a regular expression or a finite automaton that recognizes L''.
Since L'' is a regular language, there is a deterministic finite automaton $M_2 = (K_2, \Sigma, \delta, s_2, F_2)$. Then, notice that L'=$L \cap L''$. So, for L', we have an intersection of a regular language with a context free language. By Theorem 3.5.2 in the book, we know that it is a context free language. Then, $L' = L(M)$ for some PDA $M = (K, \Sigma, \Gamma, \Delta, s, F)$, where:
$K = K_1$ x K_2, the cartesian product of the state sets of M_1 and M_2,
$\Gamma = \Gamma_1$,
$s = (s_1, s_2)$,
$F = F_1$ x F_2,
Δ is defined as follows:
for each transition of the PDA $((q_1, a, \beta), (p_1, \gamma)) \in \Delta_1$, and for each state $q_2 \in K_2$, $(((q_1, q_2), a, \beta), ((p_1, \delta(q_2, a)), \gamma)) \in \Delta$,
for each transition of the form $((q_1, e, \beta), (p_1, \gamma)) \in \Delta_1$, and for each state $q_2 \in K_2$, $(((q_1, q_2), e, \beta), ((p_1, q_2), \gamma)) \in \Delta$.

So, as in the proof of Theorem 3.5.2, $L' = L(M) = L(M_1) \cap L(M_2)$ is a context free language.

4 Closure Properties (20 pts)

Let L_1 and L_2 be context-free languages which are not regular, and let L_3 be a regular language. Determine whether the following languages are necessarily CFLs or not. If they need to be context-free, explain your reasoning. If not, give one example where the language is a CFL and a counter example where the language is not a CFL.

a) $L_4 = L_1 \cap (L_2 \setminus L_3)$ (10/20 pts)

$L_4 = L_1 \cap (L_2 \setminus L_3) = L_1 \cap (L_2 \cap \overline{L_3})$ by set operations.
$(L_2 \cap \overline{L_3})$ is necessarily a context free language since intersection of a context free language with a regular language is necessarily a context free language ($\overline{L_3}$ is a regular language since regular languages are closed under complementation and L_3 is given a regular language.)
However, intersection of two context free languages is not necessarily a context free language. To see it:

Let L_1=$\{a^n b^n : n \in N\}$, L_2=$\{a^m b^m : m \in N\}$ and L_3=$\emptyset$ be languages over the same alphabet $\Sigma = \{a, b\}$ where L_1 and L_2 are context free but not regular languages, and L_3 is a regular language.
$\overline{L_3}$=$\{a, b\}^*$
L_4=$L_1 \cap (L_2 \cap \overline{L_3})$=$\{a^x b^x : x \in N\}$ is a context free language as shown in the Example 3.1.1 in the book.

Then, let L_1=$\{a^n b^n c^m : n, m \in N\}$, L_2=$\{a^m b^n c^n : n, m \in N\}$ and L_3=$\emptyset$ be languages over the same alphabet $\Sigma = \{a, b, c\}$ where L_1 and L_2 are context free but not regular languages, and L_3 is a regular language.
$\overline{L_3}$=$\{a, b, c\}^*$
L_4=$L_1 \cap (L_2 \cap \overline{L_3})$=$\{a^x b^x c^x : x \in N\}$ is not a context free language as shown in the Example 3.5.4 in the book.

We showed that $L_4 = L_1 \cap (L_2 \setminus L_3)$ may or may not be a context free language.Hence, L_4 is not necessarily a context free language.

b) $L_5 = (L_1 \cap L_3)^*$ (10/20 pts)

First looking at $(L_1 \cap L_3)$:
According to theorem 3.5.2 in the book, intersection of a context free language with a regular language is a context free language. So, since L_1 is context free and L_3 is regular, $(L_1 \cap L_3)$ is context free.
Then looking at $L_5 = (L_1 \cap L_3)^*$:
According to theorem 3.5.1 in the book context free languages are closed under Kleene star. So, $(L_1 \cap L_3)^*$ is necessarily a context free language since we showed that $(L_1 \cap L_3)$ is necessarily a context free language.

Hence, $L_5 = (L_1 \cap L_3)^*$ is necessarily a context free language.

5 Pumping Theorem (20 pts)

a) Show that $L = \{a^n m^n t^i \mid n \leq i \leq 2n\}$ is not a Context Free Language (10/20 pts) using Pumping Theorem for CFLs.

Assume that L is a Context Free Language. Choose w=$a^n m^n t^{2n}$ where n is the pumping length.
Then, by Pumping Theorem, $w = uvxyz$ can be rewritten as shown where $|vxy| \leq$n,
1$\leq |vy| \leq n$ and w can be pumped as $uv^i xy^i z$ for 0$\leq$i.
So, vy can not have all three symbols of a,m,t since for it $|vy|$ must be at least n+2.
If we assume that vy has only a's, m's or t's; then, for some i's, $uv^i xy^i z$ will have unequal numbers of symbols for a's and m's; will not be twice as much of the numbers of a's and m's for t's. So, we reached a contradiction here.
If we assume that vy contains a's and m's or m's and t's; then, for some i's, the order will be broken. For example, for i=2 of vy contains a's and m's, we will see some string like a..am..ma..am..mt..t which is not in the language. So, we reached a contradiction here.
Hence, as we have covered all the possibilities and reached a contradiction after all of them, our assumption seems to be wrong.
Hence, L is not a Context Free Language.

b) Show that $L = \{a^n b^{2n} a^n \mid n \in \mathbb{N}+\}$ is not a Context Free Language (10/20 pts) using Pumping Theorem for CFLs.

Assume that L is a Context Free Language. Choose w=$a^n b^{2n} c^n$ where n is the pumping length.
Then, by Pumping Theorem, $w = uvxyz$ can be rewritten as shown where $|vxy| \leq$n,
1$\leq |vy| \leq n$ and w can be pumped as $uv^i xy^i z$ for 0$\leq$i.
For this language, let us group the substrings. Let all of first seen a's until b's be first group, all of b's be second group and all of a's after that be third group.
So, vy can not have all three groups since for it $|vy|$ must be at least 2n+2.
If we assume that vy contains substrings from only first, second or third group; then, for some i's, $uv^i xy^i z$ will have unequal numbers of a's for first and third groups; will not be twice as much of the length of first and third groups for second group. So, we reached a contradiction here.
If we assume that vy contains symbols from first and second groups or second and third groups; then, for some i's, the order will be broken. For example, for i=2 of vy contains some symbols from first and second groups, we will see some string like a..ab..ba..ab..ba..a which is not in the language. So, we reached a contradiction here.
Hence, as we have covered all the possibilities and reached a contradiction after all of them, our assumption seems to be wrong.
Hence, L is not a Context Free Language.

6 CNF and CYK (not graded)

a) Convert the given context-free grammar to Chomsky Normal Form.

$S \to XSX \mid xY$
$X \to Y \mid S$
$Y \to z \mid \varepsilon$

answer here ...

b) Use the grammar below to parse the given sentence using Cocke–Younger–Kasami algorithm. Plot the parse trees.

S → NP VP
S → X1 VP
X1 → Aux NP
S → book | include | prefer
S → Verb NP
S → X2 PP
S → Verb PP
S → VP PP
NP → I | she | me | Houston
NP → Det Nom
Nom → book | flight | meal | money
Nom → Nom Noun
Nom → Nom PP

VP → book | include | prefer
VP → Verb NP
VP → X2 PP
X2 → Verb NP
VP → Verb PP
VP → VP PP
PP → Prep NP
Det → that | this | the | a
Noun → book | flight | meal | money
Verb → book | include | prefer
Aux → does
Prep → from | to | on | near | through

book the flight through Houston

Empty parse table:

1:5 → 1:1 2:5 1:5 → 1:2 3:5 1:5 → 1:3 4:5 1:5 → 1:4 5:5				
1:4 → 1:1 2:4 1:4 → 1:2 3:4 1:4 → 1:3 4:4	2:5 → 2:2 3:5 2:5 → 2:3 4:5 2:5 → 2:4 5:5			
1:3 → 1:1 2:3 1:3 → 1:2 3:3	2:4 → 2:2 3:4 2:4 → 2:3 4:4	3:5 → 3:3 4:5 3:5 → 3:4 5:5		
1:2 → 1:1 2:2	2:3 → 2:2 3:3	3:4 → 3:3 4:4	4:5 → 4:4 5:5	
1:1	2:2	3:3	4:4	5:5
book	the	flight	through	Houston

rest of the answer here ...

7 Deterministic Pushdown Automata (not graded)

Provide a DPDA to recognize the given languages, the DPDA must read its entire input and finish with an empty stack.

a) $a^*bc \cup a^nb^nc$

answer here ...

b) $(aa)^*c \cup a^n b^n c$

answer here ...