Problem 1

1. The set of strings over the alphabet Σ ={a,b}with more a's than b's

Solution➔

S → TaT

T → T T | aTb | bTa | a | ε

T generates all strings with at least as many a's as b's, and S forces an extra a.

2. The complement of the language{a^n b^n|n≥0}

Idea: we can break this language into the union of several simpler languages: $L = \{a^i b^j | i > j\} \cup \{a^i b^j | i < j\} \cup (a \cup b)^* b (a \cup b)^* a (a \cup b)^*$. That is, all strings of a's followed by b's in which the number of a's and b's differ, unioned with all strings *not* of the form $a^i b^j$.

First, we can achieve the union of the CFGs for the three languages:

$$S \to S_1 | S_2 | S_3$$

Now, the set of strings $\{a^i b^j | i > j\}$ is generated by a simple CFG:

$$S_1 \to a S_1 b | a S_1 | a$$

Similarly for $\{a^i b^j | i < j\}$:

$$S_2 \to a S_2 b | S_2 b | b$$

Finally, $(a \cup b)^* b (a \cup b)^* a (a \cup b)^*$ is easily generated as follows:

$$\begin{aligned} S_3 &\to XbXaX \\ X &\to aX | bX | \epsilon \end{aligned}$$

Problem 2

Give context-free grammars (CFGs) generating the following language:

A={a^i b^j c^k | i =j or j = k where i, j, k ≥ 0}

Is your grammar ambiguous? Why or why not? If yes, please provide an example of two different left-most derivations that generate the same string.

Idea: this language is simply the union of $A_1 = \{a^i b^j c^k | i, j, k \geq 0, i = j\}$ and $A_2 = \{a^i b^j c^k | i, j, k \geq 0, j = k\}$. We can create simple grammars for the separate languages and union them:

$$S \rightarrow S_1 | S_2$$

For A_1, we simply ensure that the number of a's equals the number of b's:

$$\begin{aligned} S_1 &\rightarrow S_1 c | A | \epsilon \\ A &\rightarrow aAb | \epsilon \end{aligned}$$

Similarly for ensuring that the number of b's equals the number of c's:

$$\begin{aligned} S_2 &\rightarrow aS_2 | B | \epsilon \\ B &\rightarrow bBc | \epsilon \end{aligned}$$

This grammar is ambiguous. For $x = a^n b^n c^n$, we may use either S_1 or S_2 to generate x.

Problem 3

Exercise 2.14. Convert the following CFG into an equivalent CFG in Chomsky normal form, using the procedure given in Theorem 2.9.Please provide all intermediate steps with comments on how you transform from the grammar from one version to another (these steps are critical for your work to be graded).

$$A \rightarrow BAB \mid B \mid \epsilon$$
$$B \rightarrow 00 \mid \epsilon$$

Given CFG (Context-free Grammar) is

$$A \rightarrow BAB \mid B \mid \varepsilon$$
$$B \rightarrow 00 \mid \varepsilon$$

Now, construct an equivalent CFG (Context-free Grammar) in Chomsky normal form.

Chomsky normal form:

A context i– free grammar s in **Chomsky normal form** if every rule is of the form

$$A \rightarrow BC$$
$$A \rightarrow a$$

Here, a is terminal,

A, B and C are variables,

In addition, it permits the rule $S \rightarrow \varepsilon$, here S is the start variable.

convert the given CFG into an equivalent CFG in Chomsky normal form.

Let's add a new start variable S_0 and the rule $S_0 \rightarrow A$.

Thus the obtained grammar is

$$S_0 \rightarrow A$$
$$A \rightarrow BAB \mid B \mid \varepsilon$$
$$B \rightarrow 00 \mid \varepsilon$$

In the addition of new start variable guarantees that the start variable doesn't occur on the right-hand side of a rule.

Removing all rules that containing ε.

Removing $A \to \varepsilon$ and $B \to \varepsilon$ gives

$$S_0 \to A \mid \varepsilon$$
$$A \to BAB \mid BA \mid AB \mid A \mid B \mid BB$$
$$B \to 00$$

The rule $S_0 \to \varepsilon$ is accepted since S_0 is the start variable and that is allowed in Chomsky normal form.

Now remove the unit rules.

Removing $A \to A$ gives

$$S_0 \to A \mid \varepsilon$$
$$A \to BAB \mid BA \mid AB \mid B \mid BB$$
$$B \to 00$$

Removing $S \to B$ gives

$$S_0 \to A \mid \varepsilon$$
$$A \to BAB \mid BA \mid AB \mid 00 \mid BB$$
$$B \to 00$$

Removing $S_0 \to S$ gives

$$S_0 \to BAB \mid BA \mid AB \mid 00 \mid BB \mid \varepsilon$$
$$A \to BAB \mid BA \mid AB \mid 00 \mid BB$$
$$B \to 00$$

Now replace ill placed terminals 0 by variable U with new

$$S_0 \to BAB \mid BA \mid AB \mid UU \mid BB \mid \varepsilon$$
$$A \to BAB \mid BA \mid AB \mid UU \mid BB$$
$$B \to UU$$
$$U \to 0$$

Shorten the right-hand side of rules with only 2 variables each.

To shorten the rules, replace $S_0 \to BAB$ with two rules $S_0 \to BA_1$ and $A_1 \to AB$.

The rule $A \to BAB$ is replaced by the two rules $A \to BA_2$ and $A_2 \to AB$.

After replacing these rules, the final Context-free grammar in Chomsky normal form is $G = (V, \Sigma, R, S_0)$,

Here the set of variables is $V = \{S_0, S, B, U, A_1, A_2\}$,

the start variable is S_0.

The set of terminals is $\Sigma = \{0\}$, and the rules R are given by

$$
\begin{aligned}
S_0 &\to BA_1 \mid BA \mid SB \mid UU \mid BB \mid \varepsilon \\
A &\to BA_1 \mid BA \mid SB \mid UU \mid BB \\
B &\to UU \\
U &\to 0 \\
A_1 &\to AB
\end{aligned}
$$

This is the final CFG in Chomsky normal form equivalent to the given CFG.