Nico Bachmann
Ruben Hutter

Exercise 5.1

Step 1: ϵ Rules
$S \to \epsilon$ is the only occurrence of ϵ and S never occurs on the right-hand side. We don't have to do anything here.

Step 2: Chain Rules
Since the CNF can only have a terminal or **two** Variables on the right-hand side, we have to eliminate every Rule that has only a single Variable on the right.

Chains in G:

$$S \to Z \to bb \qquad S \to Z \to Za \qquad X \to Y \to bY \qquad Y \to X \to aZb$$

We eliminate the chains by linking the end result of the chain directly with the first Variable. Since $X \to Y$ and $Y \to X$, we can just write them together as X that points to the results of X and Y.

$$S \to bb \mid Za \mid XX \qquad Z \to bb \mid Za \qquad X \to aZb \mid bX$$

Step 3: Order
Now we have to eliminate all the rules that have terminals **and** Variables on the right-hand side. We do so by adding a new Variable for each terminal and having each new Variable point to exactly one terminal.
So we create the new rules:

$$A \to a \quad \text{and} \quad B \to b$$

and can now rewrite the Rules above to:

$$S \to BB \mid ZA \mid XX \qquad Z \to BB \mid ZA \qquad X \to AZB \mid BX$$

Step 4: Shorten
To get our CNF we have to shorten the right-hand side to exactly two variables per rule. Since $X \to AZB$ is too long, we add $V \to ZB$ and get our final Grammar $G' = \langle \{ S, V, X, Z \}, \{ a, b \}, P, S \rangle$ with the following rules in P:

$$S \to BB \mid ZA \mid XX \qquad Z \to BB \mid ZA \qquad X \to AV \mid BX \qquad V \to ZB$$

$$A \to a \qquad B \to b$$

Nico Bachmann
Ruben Hutter

Exercise 5.2

Proof. Length of Derivations in Chomsky Normal Form:

In the case $|w| = 1$ we can only have one terminal Rule $S \to a$ because $S \to AB$ would be replaced by at least two terminal rules and therefore be $|w| \geq 2$.

With $S \to AB$ we can generate any word w with $|w| \geq 2$ in the language G by replacing the A and B.

$$S \to AB \xrightarrow{(B \to CD)} ACD \xrightarrow{(D \to EF)} ACEF \to \cdots \to X_1 \dots X_n$$

By the definition of the Chomsky Normal Form, the Grammar has either a terminal or two Variables on the right-hand side. Therefore with every Step, exactly one Variable can be added.
So to generate $X_1 \dots X_n$ exactly $n - 1$ steps are needed.

Now we have to replace every Variable X with a terminal symbol. By definition we can only replace one X at a time and therefore need n more steps to generate our final word w.

In total we need exactly $2n - 1$ steps and since $n = |w|$ the statement is true.

□

Exercise 5.3

(a)

Σ read input	step	Stack status
ϵ	$q_0 \to q_1$	$\#$
a	$q_1 \to q_1$	$\#X$
ϵ	$q_1 \to q_2$	$\#X$
a	$q_2 \to q_2$	$\#XY$
a	$q_2 \to q_2$	$\#XYY$
d	$q_2 \to q_3$	$\#XYY$
b	$q_3 \to q_4$	$\#XYY$
a	$q_4 \to q_3$	$\#XY$
b	$q_3 \to q_4$	$\#XY$
a	$q_4 \to q_3$	$\#X$
ϵ	$q_3 \to q_5$	$\#X$
c	$q_5 \to q_6$	$\#X$
c	$q_6 \to q_5$	$\#$
ϵ	$q_5 \to q_7$	ϵ

(b) This PDA accepts the following language: $w = \{ a^x \ d \ (ba)^y \ (cc)^z \mid x = y + z, \quad x, y, z \in \mathbb{N}_0 \}$
The self-loops q_1 and q_2 stack a's. Then follows a single d and the loop $q_3 \to q_4 \to q_3$ (ab) which removes anything stacked by q_1 followed by the loop $q_5 \to q_6 \to q_5$ (cc) which removes anything stacked by q_2.

Nico Bachmann
Ruben Hutter

Exercise 5.4

(a) The statement is false, as there exist context-free languages whose complements are also context-free. Context-free languages are not closed under complement.

(b) No, the language $L = \{ a^n b^n b^m a^m \mid n, m \in \mathbb{N}_0 \}$ is not context-free. This can be shown using the pumping lemma for context-free languages.

(c) Given two context-free grammars G_1 and G_2, we can construct a pushdown automaton P that recognizes the intersection of the languages generated by G_1 and G_2. This can be done by combining PDAs for G_1 and G_2 into a single PDA that accepts a word if and only if both G_1 and G_2 accept the same word. Therefore, the question "Is $w \in L(G_1) \cap L(G_2)$?" is decidable for context-free grammars using the construction of the PDA P.