1

Exercises

Exercise n°1

Prove the following version of the pumping lemma: If L is regular, then there is an m such that every $w \in L$ of length greater than m can be decomposed as

$$w = xyz,$$

with $|yz| \leq m$ and $|y| \geq 1$, such that xy^iz is in L for all i.

Exercise n°2

Show that the language

$$L = \{w \mid n_a(w) = n_b(w)\}$$

is not regular. Is L^* regular?

Exercise n°3

Prove that the following language is not regular:

$$L = \{w \mid n_a(w) \neq n_b(w)\}$$

Exercise n°4

Prove that the following language is not regular:

$$L = \{ww \mid w \in \{a, b\}^*\}$$

Exercise n°5

Prove that the following language is not regular:

$$L = \{a^p | \text{ p is a prime number}\}$$

Exercise n°6

Find a context-free grammar for the following language, with $n, m0$:

$$L = \{a^n b^m \mid n \neq m - 1\}$$

Exercise n°7

Find a context-free grammar for the following language, with $n, m0$:

$$L = \{w \in \{a, b\}^* \mid n_a(w) \neq n_b(w)\}$$

Exercise n°8

Find a context-free grammar for the following language, with $n, m, k \geq 0$:

$$L = \{a^n b^m c^k \mid n = m \text{ or or } m \leq k\}$$

Exercise n°9

Find a context-free grammar for the following language, with $n, m \geq 0$:

$$L = \{a^n b^m c^k \mid k \geq 3\}$$

Exercise n°10

Consider the derivation tree below:

Find a grammar G for which this is the derivation tree of the string aab. Then find two more sentences of $L(G)$. Find a sentence in $L(G)$ that has a derivation tree of height five or larger.

Exercise n°11

Show that the grammar

$$S \to aSb$$
$$S \to SS$$
$$S \to \lambda$$

is ambiguous, but that the language denoted by it is not.

Exercise n°12

Let $G = (V, T, S, P)$ be any context-free grammar with- out any λ-productions or unit-productions. Let k be the maximum number of symbols on the right of any productions in P. Show that there is an equivalent grammar in Chomsky normal form with no more than $(k-1)|P| + |T|$production rules.

Exercise n°13

Draw the dependency graph for the following grammar:

$$S \to abAB$$
$$A \to bAB \mid \lambda$$
$$B \to BAa \mid \lambda A \mid \lambda$$

Exercise n°14

Show that for every context-free grammar $G = (V, T, S, P)$ there is an equivalent one in which all productions have the form

$$A \to aBC$$

or

$$A \to \lambda$$

where $a \in \Sigma \cup \{\lambda\}, A, B, C \in V$

Solution n°1

We want to prove: If L is regular, then there is an m such that every $w \in L$ of length greater than m can be decomposed as

$$w = xyz,$$

with $|yz| \leq m$ and $|y| \geq 1$, such that xy^iz is in L for all i.

So we need to find an integer m which verifies the statement

Let L be a regular language.

$$L \text{ is regular} \iff \exists \text{ DFA } M = (Q, \Sigma, \delta, q_0, F) \text{ with } L(M) = L$$

Take $m := |Q|$ so m corresponds to the number of state $q_1, \ldots, q_m$ of M. We are also given input string $w \in L$ with $w = w_1 w_2 \cdots w_n$ ($n = |w| \geq m$). Since $w \in L$, w is accepted by M:

$$q_1 \xrightarrow{w_1} q_2 \xrightarrow{w_2} q_3 \xrightarrow{w_3} \cdots \xrightarrow{w_m} q_{m+1} \xrightarrow{w_{m+1}} \cdots \xrightarrow{w_n} q_{n+1}$$

Note that $w_{m+1} = \cdots = w_n = \lambda$ and $q_{m+1} = \cdots = q_{n+1} \in F$ if $n = m$.

M went through at least m + 1 states, but has only m distinct states. By Dirichlet's drawer principle, some state repeats: there exists a cycle / a loop.

$$\begin{tikzcd}
q_1 \arrow[r, "w_1"] & \cdots \arrow[r, "w_{j-1}"] & q_j = q_k \arrow[r, "w_k"] \arrow[dl, "w_j"] & q_{k+1} \arrow[r, "w_{k+1}"] & \cdots \arrow[r, "w_n"] & q_{n+1} \\
& q_{j+1} \arrow[rr, dashed] & & q_{k-1} \arrow[ul, "w_{k-1}"] & &
\end{tikzcd}$$

Formally, there exists some j, k with $j \neq k$ such that $q_j = q_k$. Of course this repetition occurs somewhere between q_1 and q_{m+1} thus $j, k \leq m + 1$. We're done by setting:

- $x = w_1 \cdots w_{j-1}$
- $y = w_j \cdots w_{k-1}$ (Without lose of generality $j < k$)
- $z = w_k \cdots w_n$

We have shown that

- For all $i \geq 0, xy^iz \in L$ (because we may exploit the loop)

- $|y| \geq 1$ (because j, k are distinct)
- $|xy| \leq m$ (because $|xy| = k - 1$ and $k \leq m + 1$.) (pumping lemma)
- $|yz| \leq m$ (because $|yz| = n - k$ and $k \leq m + 1$.) (this version)

Solution n°2

Going by contradiction, assume $\{w \mid n_a(w) = n_b(w)\}$ is regular.

Then by the pumping lemma, $\exists\ m \in \mathbb{N}_{>0}$ such that every $w \in L$ with $|w| \geq m$ can be written $w = xyz$ with $|xy| \leq m$, $|y| \geq 1$ and $\forall i\ xy^iz \in L$.

In particular if we take $w = a^m b^m$ then:

Since $|xy| \leq m$, $x = a^j$ and $y = a^k$ with $j + k \leq m$ and $k \geq 1$

Take $i = 0$:

$$xy^0z = xz = a^j b^m = a^{m-k} b^m$$

Note $m - k \neq m$ since $k \geq 1$ so $a^{m-k} b^m \notin L$
Another rationale: $j \leq m - k < m$ since $k \geq 1$ so $a^j b^m \notin L$

This contradicts the pumping lemma. Therefore, our assumption that L is regular cannot be true. Thus L is not regular.

Solution n°3

Going by contradiction, assume $L = \{w \mid n_a(w) \neq n_b(w)\}$ is regular.

Then $\overline{L} = \{w \mid n_a(w) = n_b(w)\}$ is regular.

This contradicts the pumping lemma (seen in the previous exercise). Therefore, our assumption that L is regular cannot be true. Thus L is not regular.

Solution n°4

Going by contradiction, assume $L = \{ww \mid w \in \{a,b\}^*\}$ is regular.

Then by the pumping lemma, $\exists\ m \in \mathbb{N}_{>0}$ such that every $w \in L$ with $|w| \geq m$ can be written $w = xyz$ with $|xy| \leq m$, $|y| \geq 1$ and $\forall i\ xy^iz \in L$.

In particular if we take $w = a^m b^m a^m b^m$ then:

Since $|xy| \leq m$, $x = a^j$ and $y = a^k$ with $j + k \leq m$ and $k \geq 1$

Take $i = 0$:

$$xy^0z = xz = a^j b^m a^m b^m$$

$j \leq m - k < m$ since $k \geq 1$ so $a^j b^m a^m b^m \notin L$

This contradicts the pumping lemma. Therefore, our assumption that L is regular cannot be true. Thus L is not regular.

Solution n°5

Going by contradiction, assume $L = \{a^p \mid$ p is a prime number$\}$ is regular.

Then by the pumping lemma, $\exists\ m \in \mathbb{N}_{>0}$ such that every $w \in L$ with $|w| \geq m$ can be written $w = xyz$ with $|xy| \leq m$, $|y| \geq 1$ and $\forall i\ xy^iz \in L$.

In particular if we take $w = a^p$ where n and $p > m + 1$ is prime then:

Since $|xy| \leq m$, $x = a^j$ and $y = a^k$ with $j + k \leq m$ and $k \geq 1$

Take $i = p + 1$:

$$xy^{p+1}z = xyzy^p = a^{p+kp} = a^{p(1+k)}$$

the number $p(1 + k)$ is divisible by $1 + k \neq 1$ since $k \geq 1$ and $1 + k \neq p$ because $1 + k \leq 1 + m - j < p - j < p$ so $p(1 + k)$ is not prime hence $w \notin L$. This contradicts the pumping lemma. Therefore, our assumption that L is regular cannot be true. Thus L is not regular.

Solution n°6

$$L = \{a^n b^m \mid n \neq m - 1\}$$
$$L = \{a^n b^m \mid n \geq m \;\; \wedge \;\; n < m - 1\}$$

We can create the grammar for the language $\{a^n b^m \mid n \geq m\}$

$$S \rightarrow aS \mid aSb \mid \lambda$$

We can create the grammar for the language $\{a^n b^m \mid n < m - 1\}$

$$S \rightarrow aSb \mid Sb \mid bb$$

Based one these two cases one can then create the following context-free grammar:

$$S \rightarrow aS \mid aSb \mid Sb \mid bb \mid \lambda$$

Solution n°7

Find a context-free grammar for the following language, with $n, m \geq 0$:

$$L = \{w \in \{a, b\}^* \mid n_a(w) \neq n_b(w)\}$$

First we construct $L = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w)\}$

$$S \rightarrow E$$
$$E \rightarrow EaEbE \mid EbEaE \mid \lambda$$

Then use this grammar to solve two simpler versions of the original problem. Namely, we find grammars that generate the language:

$$L = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w) + 1\}$$

$$S \rightarrow A$$
$$A \rightarrow EaE$$
$$E \rightarrow EaEbE \mid EbEaE \mid \lambda$$

$$L = \{w \in \{a, b\}^* \mid n_a(w) + 1 = n_b(w)\}$$

$$S \to A$$
$$A \to EbE$$
$$E \to EaEbE \mid EbEaE \mid \lambda$$

We observe that $L = L_a^+ \cup L_b^+$, and with this knowledge define the grammar that generates L:

$$S \to A$$
$$A \to EaE \mid EbE$$
$$E \to EaEbE \mid EbEaE \mid \lambda$$

Solution n°8

$$L = \{a^n b^m c^k \ \mid\ n = m\}$$

$$S \to XSc \mid \lambda \mid Sc$$
$$X \to aXb \mid \lambda$$

Better version:

$$S_1 \to WX$$
$$W \to aWb \mid \lambda$$
$$X \to cX \mid \lambda$$

$$L = \{a^n b^m c^k \ \mid\ m \leq k\}$$

$$S \to AS \mid \lambda \mid bSc \mid Sc$$
$$A \to aA \mid \lambda$$

Better version:

$$S_2 \to YZ$$
$$Y \to aY \mid \lambda$$
$$Z \to bZc \mid Zc \mid \lambda$$

$$L = \{a^n b^m c^k \ \mid\ n = m \text{ or or } m \leq k\}$$
$$S \to S_1 \mid S_2$$
$$W \to aWb \mid \lambda$$
$$X \to cX \mid \lambda$$
$$Y \to aY \mid \lambda$$
$$Z \to bZc \mid Zc \mid \lambda$$

Solution n°9

$$L = \{a^n b^m c^k \ \mid\ k \geq 3\}$$

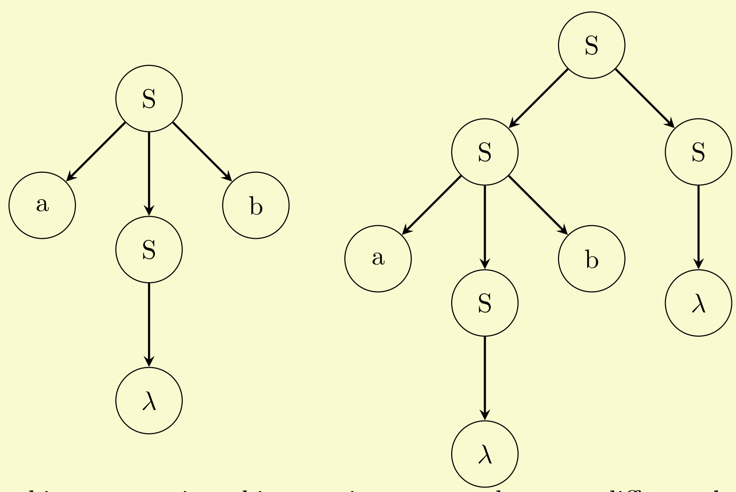
$$S \to XYZccc$$
$$X \to aX \mid \lambda$$
$$Y \to bY \mid \lambda$$
$$Z \to cZ \mid \lambda$$

Solution n°10

$$S \to aSSb \mid \lambda \mid a$$

Solution n°11

$$S \to aSb$$
$$S \to SS$$
$$S \to \lambda$$



Hence this grammar is ambiguous since we can draw two different derivation trees for the word ab

A language is ambiguous if it is context-free and all the grammars that denote the language are ambiguous. Thus we need to find a grammar which is not ambiguous:

$$S \to aSbS$$
$$S \to \lambda$$

Solution n°12

Let $G = (V, T, S, P)$ be any context-free grammar without any λ-productions or unit-productions. Let k be the maximum number of symbols on the right of any productions in P. Show that there is an equivalent grammar in Chomsky normal form with no more than $(k-1)|P| + |T|$production rules.