11. Closure Properties of Context-free Languages

11.1 Positive Results

Goal: Show closure under $\cup$, $\cdot$, *, h, h^{-1}, substitution, and regular intersection.

Theorem:
The context-free languages are closed under union, concatenation, and Kleene star.

Proof:
Assume the languages are given by context-free grammars. □
The constructions are easy.

Definition:

- A substitution is a function
$$\sigma : \Sigma \longrightarrow \mathcal{P}(\Delta^*)$$
that assigns to each letter a a language $\sigma(a)$ over Δ.
A substitution is • regular, if $\sigma(a)$ is a regular language for each $a \in \Sigma$
• context-free, if $\sigma(a)$ is a context-free language for each $a \in \Sigma$.
- The application of σ to $w \in \Sigma^*$
yields the language
$$\sigma(w) := \{\varepsilon\} \quad , \text{ if } w = \varepsilon$$
$$\sigma(w) := \sigma(a_1) \dots \sigma(a_n), \text{ if } w = a_1 \dots a_n.$$
The application of σ to $L \subseteq \Sigma^*$ yields
$$\sigma(L) := \bigcup_{w \in L} \sigma(w).$$

Note: Homomorphisms are substitutions where $\sigma(a)$ consists of a single word for each $a \in \Sigma$.

<u>Theorem:</u>

- The regular languages are closed under regular substitutions.
- The context-free languages are closed under context-free substitutions:

  If L is context-free and $\sigma(a) \subseteq \Delta^*$ is context-free for all $a \in \Sigma$,

  then $\sigma(L) \subseteq \Delta^*$ is context-free.

<u>Proof:</u>

We show the statement for context-free languages.

Let $G = (N, \Sigma, P, S)$ be the grammar for L.

Let $G_a = (N_a, \Delta, P_a, S_a)$ be the grammar for $\sigma(a)$, for each $a \in \Sigma$.

Wlog. we assume the non-terminals in G and in the G_a are pairwise disjoint.

We replace every occurrence of a in a right-hand side of a rule in P by S_a, the start symbol of G_a. $\square$

<u>Corollary:</u> The context-free languages are closed under homomorphisms.

<u>Theorem:</u>

The context-free languages are closed under inverse homomorphisms:

If $L \subseteq \Delta^*$ is context-free and $h: \Sigma^* \to \Delta^*$ is a homomorphism,

then $h^{-1}(L) \subseteq \Sigma^*$ is context-free.

<u>Proof:</u>

We use an automata-theoretic construction.

Let $L = L(M)$ with $M = (Q, \Delta, \Gamma, q_0, \delta, Q_F)$ a PDA.

We construct a PDA M' accepting $h^{-1}(L)$.

<u>Idea:</u> • We let M', on input a, generate the word $h(a)$ and simulate M on $h(a)$.

· Since the length of the words is bounded (by $\max\{|h(a)| \mid a \in \Sigma\}$),
we can store them in the control state of M'.
(There is a general rule-of-thumb in automata theory:
If some information is bounded, put it into the control state.)

· Every time M wants to read an input symbol,
M' removes a letter from the currently stored word
(which is a suffix of some $h(a)$).
If the word becomes empty (ε), M' accepts the next input symbol.

· A state of M' is final, if the auxiliary word is empty (ε)
and the control state of M is final.

<u>Construction</u>:

We define
$$M' := (Q', \Sigma, \Gamma, (q_0, \varepsilon), \delta', Q_F \times \varepsilon)$$
where
$$Q' := \{(q, x) \mid q \in Q,\ x \text{ a suffix of } h(a) \text{ for some } a \in \Sigma\}.$$

The transition relation δ' is defined as follows:

↳ $(q, \varepsilon) \xrightarrow[\varepsilon/\varepsilon]{a} (q, h(a))$ for all $a \in \Sigma$ // Get new input.

↳ $(q, ax) \xrightarrow[v_1/v_2]{\varepsilon} (p, x)$, if $q \xrightarrow[v_1/v_2]{a} p \in \delta$ // M reads an input from the auxiliary word.

↳ $(q, x) \xrightarrow[v_1/v_2]{\varepsilon} (p, x)$, if $q \xrightarrow[v_1/v_2]{\varepsilon} p \in \delta$ // M does an ε-transition.

□

<u>Theorem (Closure under regular intersection)</u>:
If $L \subseteq \Sigma^*$ is context-free and $R \subseteq \Sigma^*$ is regular,
then $L \cap R$ is context-free.

It is easy to establish the theorem using pushdown automata.

We give a grammar-based construction instead that once more practices the idea of <u>summarizing a computation</u> that we saw when converting a PDA to a CFG.

Proof:
Let $L = L(G)$ with $G = (N, \Sigma, P, S)$ in Chomsky normal form, potentially with $S \to \varepsilon$ and then S not in right-hand sides.
Let $R = L(A)$ with $A = (Q, \Sigma, q_0, \to, Q_F)$ an NFA without ε-transitions.
We construct the grammar $G' = (N', \Sigma, P', S')$ for $L \cap R$ as follows.
The non-terminals are

$$N' := (Q \times N \times Q) \cup \{S'\}.$$

For the productions, we have

$$S' \to \varepsilon \quad \text{iff} \quad \varepsilon \in L \cap R.$$

Moreover, there are productions

$$S' \to (q_0, S, q_f) \quad \text{for each } q_f \in Q_F.$$

For every non-terminal (q_1, A, q_2) we have

$$(q_1, A, q_2) \to a, \text{ if } q_1 \xrightarrow{a} q_2 \text{ in the NFA and } A \to a \in P.$$

$$(q_1, A, q_2) \to (q_1, B, q).(q, C, q_2), \text{ if } A \to BC \in P.$$

There are variants of the latter production for all $q \in Q$.

Intuitively, the grammar guesses the state change in A that is induced by the word derived from B (and from C). $\square$

We again employ the closure properties to disprove context freeness.

Example (Application of closure properties):

We show that

$L = \{ww \mid w \in \{a,b\}^*\}$ is not context-free.

- Assume it was. Then by closure under regular intersection also

$$L_1 = L \cap a^+b^+a^+b^+ = \{a^ib^ja^ib^j \mid i,j \geq 1\}$$

is context-free.

But L_1 is not context-free by an application of the pumping lemma.

↯ L cannot be context-free.

- If we do not want to apply the pumping lemma to L_1, we can further reduce the language to

$$L_2 = \{a^ib^jc^id^j \mid i,j \geq 1\}.$$

Consider $h: \{a,b,c,d\}^* \to \{a,b\}^*$ defined by

$$h(a) := a =: h(c) \quad \text{and} \quad h(b) := b =: h(d).$$

Then

$$h^{-1}(L) = \{x_1x_2x_3x_4 \mid |x_1| = |x_3| \text{ and } x_1, x_3 \in \{a,c\}^* \text{ and } |x_2| = |x_4| \text{ and } x_2, x_4 \in \{b,d\}^*\}.$$

Then $h^{-1}(L) \cap a^*b^*c^*d^* = L_2$.

Moreover, if L was context-free, then so was L_2.

But L_2 is not context-free by the pumping lemma.

↯ L cannot be context-free. $\square$

11.2 Negative Results

Goal: Show that the context-free languages are not closed under intersection and under complement.

Theorem:

(1) There are context-free languages L_1 and L_2 so that $L_1 \cap L_2$ is not context-free.

(2) There is a context-free language L where $\overline{L}$ is not context-free.

Proof:

(1) The languages are

$$L_1 = \{ a^n b^m c^m \mid n, m \in \mathbb{N} \} \quad \text{and}$$
$$L_2 = \{ a^n b^n c^m \mid n, m \in \mathbb{N} \}$$

with $L_1 \cap L_2 = \{ a^n b^n c^n \mid n \in \mathbb{N} \}$, known to be not context-free.

(2) If the context-free languages were closed under complement (which means for each context-free language L, $\overline{L}$ is context-free), then, due to closure under union, we would obtain

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}} \quad \text{being context-free,}$$

for all L_1 and L_2 context-free languages. ↯ Contradiction to (1).

For a constructive proof, one shows that

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

is context-free.

The complement is

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

which we just proved to be not context-free. □