Characterization of regular languages

Theorem 1.6 (Pumping Lemma) Let Σ be an alphabet and A be a language over Σ. If A is a regular language, then there is a number p, the **pumping length**, where, if s is any string in A of length at least p, then s may be divided into three pieces, $s = xyz$, satisfying the following conditions:

1. for each $i \geq 0$, $xy^iz \in A$,
2. $|y| > 0$, and
3. $|xy| \leq p$.

A regular language $\Rightarrow$ "pumping" possible for strings $s \in A$, $|s| \geq p$

depends on A!

not known as value!

pumping:

all (!) $s \in A$, $|s| \geq p$

$s = xyz$

decomposition such that below conditions are fulfilled

1. $\forall i \geq 0$, $xy^iz \in A$ 2. $|y| > 0$ 3. $|xy| \leq p$

the actual pumping   y is non-trivial   size of xy bound

$\Rightarrow xz \in A, xyz \in A, xyyz \in A, xyyyz \in A, \ldots$

Proving that a language is not (!) regular

Example : $L = \{0^n 1^n \mid n \geq 0\}$

Proof (by contradiction)

standard opening

Assumption : L regular $\overset{P.L.}{\Rightarrow}$ $\exists p$ pumping length

again: unknown! $\Rightarrow$ never (!) select a concrete value!!!

Choose $s = 0^p 1^p \in L$ $\Rightarrow |s| = 2p \geq p$

always has to be checked!

"creative" part

default part

$\Rightarrow$ P.L. guarantees: $\exists$ decomposition $s = xyz$ s.th.
1) $\forall i \geq 0$ $xy^iz \in L$
2) $|y| > 0$
3) $|xy| \leq p$

Showing that for **all** possible splittings of s into xzy, there is always **at least** one condition of 1), 2, 3) not fulfilled :

(!)

main "work"

Case 1: y consists only of 0s $\Rightarrow$ xyyz has more 0s than 1s $\Rightarrow xyyz \notin L$ $\lightning$ cond. 1)

Case 2: y consists only of 1s $\Rightarrow$ xyyz has more 1s than 0s $\lightning$ cond 1)

Case 3: y consists both of 0s and 1s $\Rightarrow$ xyyz contains pattern $\ldots 0 \ldots 1 \ldots 0 \ldots 1 \ldots$ $\Rightarrow xyyz \notin L$ $\lightning$ cond 1)

abbreviated

Attention: Don't give "examples" but argue for arbitrary p.

default closing argument

Since no decomposition $s = xyz$ fulfilling all conditions 1), 2), 3) exists, we have a contradiction to the statement of the (known to be correct) P.L. $\Rightarrow$ $\lightning$ $\Rightarrow$ L must be irregular

$\square$

1 Pumping Lemma

Let the language

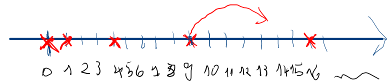
$$D = \{1^{n^2} | n \geq 0\}$$

be given. Show that D is not regular.

$0^2 \quad 1^2 \quad 2^2 \quad 3^2 \quad 4^2 \quad \dots$

$0 \quad 1 \quad 4 \quad 9 \quad 16 \quad \dots$

$D = \{ \varepsilon, 1, 1111, 111111111, \dots \}$



<u>Proof</u>

Assume: D is regular $\overset{P.L.}{\Rightarrow} \exists p$ pumping length

Choose $s = 1^{p^2} \in D \Rightarrow |s| = p^2 \geq p$

Pumping Lemma $\Rightarrow \exists$ decomposition $s = xyz$ s.th.
1) $\forall i \geq 0 \; xy^iz \in D$
2) $|y| > 0$
3) $|xy| \leq p$

Due to 3) we have $|xy| \leq p \Rightarrow |y| \leq p \quad (*)$

By construction $|xyz| = p^2$

appl. P.L. (1) $i=2$): $|xy^2z| \leq p^2 + p \quad (\mathrm{I})$

We further have: $p^2 + p < p^2 + 2p + 1 = (p+1)^2 \quad (\mathrm{II})$

Due to 2) $|y| > 0 \Rightarrow y \neq \varepsilon \Rightarrow |xy^2z| > p^2 \quad (\mathrm{III})$

$\Rightarrow p^2 \overset{\mathrm{III}}{<} |xy^2z| \overset{\mathrm{I}+\mathrm{II}}{<} (p+1)^2 \Rightarrow xy^2z \notin D$

$\Rightarrow D$ not regular

Context-free grammars

Definition 2.1 A **context-free grammar** (CFG) is a 4-tuple (V, Σ, R, S), where

1. V is a finite set called the **variables**,
2. Σ is a finite set, disjoint from V, called the **terminals**,
3. R is a finite set of (substitution) **rules** / **productions** of the form

$$A \to w\,,$$

   with $A \in V$ and $w \in (V \cup \Sigma)^*$ (i.e. $w = \varepsilon$ is possible), and
4. $S \in V$ is the **start variable**.

Definition 2.2 Let $G = (V, \Sigma, R, S)$ be a CFG, $u, v, w \in (V \cup \Sigma)^*$ and $A \in V$.

- If $A \to w$ is a rule of the grammar, we say that uAv **yields** uwv, written

$$uAv \Rightarrow uwv\,.$$

- We write

$$u \overset{*}{\Rightarrow} v\,,$$

  if $u = v$ or if a sequence $u_1, u_2, \ldots, u_k \in (V \cup \Sigma)^*$, $k \geq 0$ exists, such that

$$u \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \ldots \Rightarrow u_k \Rightarrow v\,.$$

- The **language of the grammar** G, $L(G)$, is

$$L(G) = \{w \in \Sigma^* | S \overset{*}{\Rightarrow} w\}\,.$$

- A language that is generated by a CFG is called **context-free language** (CFL).

Lemma 2.1 For every FA M there exists a CFG G generating the same language, i.e. $L(G) = L(M)$. Hence, every regular language can be generated by a CFG.

$$G(\{S, S_1, S_2\}, \{0,1\}, R, S)$$

S

$$\begin{aligned} S &\to S_1 | S_2 \\ S_1 &\to 0S_11 | \varepsilon \\ S_2 &\to 1S_20 | \varepsilon \end{aligned}$$

$$L(G) = \{0^n1^n | n \geq 0\} \cup \{1^n0^n | n \geq 0\}$$

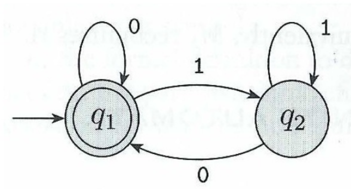
$\{L \text{ language} \mid L \text{ regular}\}$

$\not\subset$

$\{L \text{ lang.} \mid L \text{ generated by CFG}\}$

2 From FA to CFG

Let the FA *M* be given by its STD:



$M = (\{q_1, q_2\}, \{0,1\}, \delta, q_1, \{q_1\})$

Follow the lines of the proof for Lemma 2.1 to build a CFG *G* generating *L*(*M*).

$\delta(q, a) = q'$

$q \to a q'$

$G = (\{q_1, q_2\}, \{0, 1\}, R, q_1)$

Rules R:

$q_1 \to 0 q_1$

$q_1 \to 1 q_2$

$q_2 \to 0 q_1$

$q_2 \to 1 q_2$

$q_1 \to \varepsilon$

Chomsky normal form

Definition 2.3 A context-free grammar $G = (V, \Sigma, R, S)$ is in **Chomsky normal form** if every rule is of the form

$$A \rightarrow BC \tag{2.1}$$
$$A \rightarrow a \tag{2.2}$$

where $a \in \Sigma$, $A \in V$, $B, C \in V \setminus S$. In addition, we permit the rule

$$S \rightarrow \varepsilon\,.$$

Obvious: Every CFG in CNF is also a general CFG

Theorem 2.1 Any context-free language is generated by a context-free grammar in Chomsky normal form.

$\Rightarrow$ CNF and "general" form are equivalent

$S \rightarrow S_1 \mid S_2$

$S_1 \rightarrow 0 S_1 1 \mid \varepsilon$

$S_2 \rightarrow 1 S_2 0 \mid \varepsilon$

aa aa

aaaa

$S_1 \rightarrow 0 S_1 1 \mid 01$

$S' \rightarrow S$

$S \rightarrow AX \mid AB$

$X \rightarrow S_1 B$

$S_1 \rightarrow AY \mid AB$

$Y \rightarrow S_1 B$

$S \rightarrow BX' \mid BA$

$X' \rightarrow S_2 A$

$S_2 \rightarrow BY' \mid BA$

$Y' \rightarrow S_2 A$

$A \rightarrow 0$

$B \rightarrow 1$