VII

Seventh Lecture of
AUTOMATA & FORMAL LANGUAGES

29 October 2024

RECAP The main result of Lecture VI:

THEOREM TFAE:

(i) L is regular
(ii) $L = \mathcal{L}(D)$ for D det. automaton
(iii) $L = \mathcal{L}(N)$ for N nondet. aut.

TRANSLATIONS

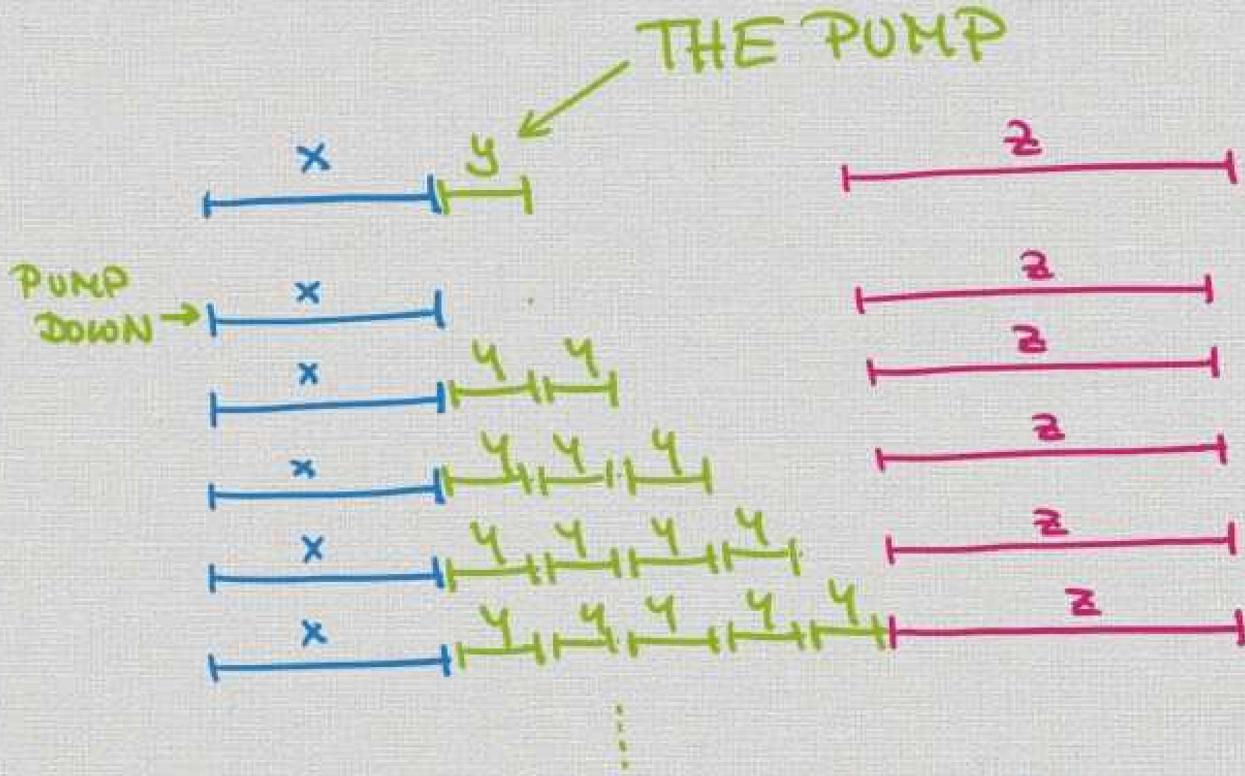
(i) $\rightsquigarrow$ (iii) Grammar (Σ, V, S, P)
$\rightsquigarrow$ Nondeterministic Automaton $(\Sigma, Q, \delta, q_0, F)$
with $|Q| = |V| + 1$

(iii) $\rightsquigarrow$ (ii) SUBSET CONSTRUCTION
Nondeterministic automaton
$\rightsquigarrow$ deterministic automaton
with $|Q'| = 2^{|Q|}$

Goal: Show that there is a context-free language that is not regular!

§ 2.4 The Pumping Lemma

Definition 2.10. Let $L \subseteq \mathbb{W}$ be a language. We say that L *satisfies the (regular) pumping lemma with pumping number n* if for every word $w \in L$ such that $|w| \geq n$ there are words x, y, z such that $w = xyz$, $|y| > 0$, $|xy| \leq n$ and for all $k \in \mathbb{N}$, we have that $xy^kz \in L$. We say that L *satisfies the (regular) pumping lemma* if there is some n such that it satisfies the (regular) pumping lemma with pumping number n.



<u>Remark</u> If L satisfies RPL w PN n and $w \in L$ with $|w| \geq n$, then L is infinite.

Thm 2.11 For every regular language L there is n s.t. L satisfies RPL w PN n.

Proof. By Lecture $\underline{\text{VI}}$, we know that $L = \mathcal{L}(D)$ for D det aut. $= (\Sigma, Q, \delta, q_0, F)$

I claim that $n := |Q|$ is the PN.

Suppose $w \in \mathcal{L}(D)$ with $|w| \geq n$.

$w = a_0 \dots a_{n-1} v$ with $a_i \in \Sigma$, $v \in W$

State sequence of w:

$$\underbrace{q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \longrightarrow \dots \xrightarrow{a_{i-1}} q_i}_{n+1 \text{ many states}} \xrightarrow{a_i} \dots \xrightarrow{a_{j-1}} q_j \xrightarrow{a_j} \dots \xrightarrow{a_{n-1}} q_n$$

$$q_i = q_j$$

Thus by pigeon hole principle, there are $i < j \leq n$ s.t. $q_i = q_j$

Define
$$x := a_0 \dots a_{i-1}$$
$$y := a_i \dots a_{j-1}$$
$$z := \begin{cases} a_j \dots a_{n-1} v & \text{if } j < n \\ v & \text{if } j = n \end{cases}$$

__Properties__ $w = xyz$

$y \neq \varepsilon \implies |y| \neq 0$

$|xy| = j \leq n.$

__Check__ (a) $\hat{\delta}(q_0, x) = q_i$

(b) $\hat{\delta}(q_i, y) = q_j = q_i = \hat{\delta}(q_j, y)$

(c) $\hat{\delta}(q_i, z) = \hat{\delta}(q_j, z) \in F.$

So the split x, y, z satisfies the requirements of the RPL

__CLAIM__ $xy^k z \in \mathcal{L}(D).$

[First observe: for all k $\hat{\delta}(q_0, xy^k) = q_i$

By induction using (a) & (b).

Thus, by (c)

$\hat{\delta}(q_0, xy^k z) = \hat{\delta}(q_i, z) \in F.$]

q.e.d.

First example (E 2.12)

$L = \{ 0^n 1^n ; n > 0 \}$ is context-free, not regular.

Context-free. $S \to 0S1$ produces L.
$S \to 01$

Not regular: Suppose it is regular. So by T 2.11 it satisfies RPL w PN, say N.
Consider $w = 0^N 1^N$. $|w| = 2N \geq N$.
By RPL find $xyz = w$ s.t. $|xy| \leq N$ & $|y| > 0$.
So by choice of w $x = 0^k$ $k \in \mathbb{N}$
$y = 0^\ell$ $\ell > 0$
Pumping y down gives: $0^{N-\ell} 1^N \notin L$ since $\ell > 0$.
Contradiction!

2nd example

Example 2.13. Fix some positive number $n \in \mathbb{N}$. Then the language $L := \{\mathbf{0}^n w\,;\, w \in \mathbb{W}\}$ is regular and there cannot be an automaton D with n or fewer states such that $\mathcal{L}(D) = L$.

[Towards a contradiction, let's assume that there is such an automaton. By the proof of Theorem 2.11, we get that L satisfies the pumping lemma with pumping number n. Consider the word $w = \mathbf{0}^n \in L$. Clearly, $|w| = n$, so the word can be pumped, in particular, pumped down. Since it consists entirely of zeros, we know that for $w = xyz$, the words x, y, and z also consist entirely of zeros and $xy^0z = xz$ is a sequence of $n - |y| < n$ zeros. Hence it's not in L: contradiction!]

NATURAL Q

Does the RPL characterise the regular languages ?

Since the pumping lemma is a very useful tool to prove that languages are not regular, it is quite natural to wonder whether the statement of the pumping lemma is equivalent to regularity, i.e., whether a language L is regular if and only if it satisfies the regular pumping lemma. The answer is "No" as we shall show now.

NO!

If $w \in \mathbb{B}$ is a binary word that contains at least one zero, we write $\mathrm{tail}(w)$ for the number of ones in w that follow the last occurring zero. E.g., $\mathrm{tail}(\mathbf{0101111}) = 4$. Let $X \subseteq \mathbb{N}$ be an arbitrary set of natural numbers (by Proposition 1.3, there are uncountably many of those). We define a language $L_X \subseteq \mathbb{B}$ by $w \in L_X$ if w consists entirely of ones or if w has some zero, then $\mathrm{tail}(w) \in X$. Let us show that if $X \neq Y$, then $L_X \neq L_Y$: w.l.o.g., we can assume that there is some $n \in X \backslash Y$. Then $\mathbf{01}^n \in L_X \backslash L_Y$. This shows that $X \mapsto L_X$ is an injection from the power set of $\mathbb{N}$ into the collection of languages of the form L_X, so there are uncountably many such languages.

Proposition 2.15. Every language L_X satisfies the (regular) pumping lemma.

Proof. We shall prove that it satisfies the pumping lemma with pumping number 2. Let w be an arbitrary binary word with $|w| \geq 2$.

Case 1. It starts with $\mathbf{0}$. Let $x = \varepsilon$, $y = \mathbf{0}$, and z such that $w = xyz = \mathbf{0}z$. Pumping up produces $\mathbf{0}^k z$ (for $k > 1$), but clearly $\mathrm{tail}(\mathbf{0}^k z) = \mathrm{tail}(\mathbf{0}z) \in X$, so $\mathbf{0}^k z \in L_X$. Pumping down produces z: if z still contains a $\mathbf{0}$, then $\mathrm{tail}(z) = \mathrm{tail}(\mathbf{0}z) \in X$, so $z \in L_X$; if z contains no $\mathbf{0}$s, then $z \in L_X$ anyway.

Case 2. It starts with $\mathbf{1}$. Let $x = \varepsilon$, $y = \mathbf{1}$, and z such that $w = xyz = \mathbf{1}z$. If z does not contain any $\mathbf{0}$s, then all results of pumping y will result in a word without $\mathbf{0}$s, so they are all in L_X. If z contains a $\mathbf{0}$, then all results of pumping y will result in a word that has the same tail as $\mathbf{1}z$, and hence they are all in L_X. Q.E.D.

Corollary 2.16. There are languages satisfying the (regular) pumping lemma that are not regular.

Proof. There are only countably many regular languages (by Proposition 1.16), but uncountably many languages satisfying the regular pumping lemma by Proposition 2.15. Q.E.D.

Second example $L = \{ 0^n w , w \in W \}$ This is regular, but no D with
fixed n $|Q| \leq n$ can produce $\mathcal{L}(D) = L$.

RPL does not characterise regular languages

For $w \in W$ say $\text{tail}(w) = n$ if n is the maximal number s.t. $w = v\, 1^n$. $w = 1\,0\,1\,1\,1\,0\,\boxed{1\,1\,1}$

$\text{tail}(w) = 3$.

For $X \subseteq \mathbb{N}$, define $L_X := 1^* \cup \{ w ; \text{tail}(w) \in X \}$

Clearly, there are uncountably many languages of the type L_X.

Thus, almost all of them are not regular.

Every L_X satisfies RPL w PN 2.

[If w starts with 0, set $x = \varepsilon$, $y = 0$

If w start with 1, set $x = \varepsilon$, $y = 1$.]

Details: Prop 2.15.

DECISION PROBLEMS

We already know that the **WORD PROBLEM** for regular grammars is solvable [since they are noncontracting].

THE EMPTINESS PROBLEM FOR AUTOMATA

Given a det. automaton D, is $\mathcal{L}(D) = \emptyset$?

By the proof of RPL for $\mathcal{L}(D)$, know that $\mathcal{L}(D)$ satisfies RPL w PN $|Q| = N$. So, suppose $w \in \mathcal{L}(D)$, then if $|w| \geq N$, it can be pumped down, so it's not the shortest word. So there is w s.t. $|w| < N$ s.t. $w \in \mathcal{L}(D)$.
So: list all words w with $|w| < N$ and check them ("use the WORD PROBLEM"). If none of them is accepted, $\mathcal{L}(D) = \emptyset$.

THE EMPTINESS PROBLEM FOR REGULAR GRAMMARS

Given a regular grammar G, is $\mathcal{L}(G) = \emptyset$?

By the proof of the characterisation theorem, find ALGORITHMICALLY an automaton D s.t. $|Q| = 2^{|V|+1}$ and $\mathcal{L}(G) = \mathcal{L}(D)$. Run the automaton argument above.