

# Formal languages and automatas

Ruben Triwari



Seminar: Algebra & Computer Science, LMU Munich

July 6, 2024

- 1 Formal Languages Definitions, Examples
- 2 Prove not all Languages are rational (regular)
- 3 Automatas Definitions, Examples
- 4 Prove: Finite automaton only accept rational languages

An Alphabet  $\Sigma$  is a set of characters.

Examples:

①  $\Sigma = \{a, b, c\}$

②  $\Sigma = \{x\}$

③  $\Sigma = \emptyset$

A Language  $L$  is a set of words with characters out of an Alphabet  $\Sigma$ .

Examples:

①  $\Sigma = \{a, b, c\} \rightsquigarrow L = \{aaa, bbb, ccc, abc\}$

②  $\Sigma = \{x\} \rightsquigarrow L = \{x, xx, xxx, xxxx\}$

③  $\Sigma = \emptyset \rightsquigarrow L = \{\epsilon\}$

$\rightsquigarrow L \subset \Sigma^*$

We can also write a Language with a formal series, with a fixed  $\Sigma$ :

$$L = \sum (L, w) w$$

$$\Leftrightarrow L = \bigcup_{w \in \Sigma^*} \underbrace{(L, w)}_{0 \text{ or } 1} \{w\}$$

Examples:

- ①  $\Sigma = \{a, b, c\} \rightsquigarrow L = \{aaa, abc\}$   
 $\rightsquigarrow L = 1aaa + 1abc + 0a + 0b + 0c + \dots$
- ②  $\Sigma = \{x\} \rightsquigarrow L = \{x, xx\}$   
 $\rightsquigarrow L = 1x + 1xx + 0xxx + 0xxxx + 0xxxxx + \dots$
- ③  $\Sigma = \emptyset \rightsquigarrow L = \{\epsilon\}$   
 $\rightsquigarrow L = 1\epsilon$

Now we can define Addition on formal series:

$$U + V = \sum \underbrace{((U, w) + (V, w))}_{\text{Boolean addition}} w$$

$$\Leftrightarrow U + V = U \cup V$$

Example:

Let  $U = \{x, xx\}$  and  $V = \{aaa, abc\}$  languages:

$$\begin{aligned} U + V &= (1 + 0)x + (1 + 0)xx + (0 + 1)aaa + (0 + 1)abc + (0 + 0)a + \dots \\ &= 1x + 1xx + 1aaa + 1abc + 0a + \dots \end{aligned}$$

Next we can define multiplication on formal series:

$$U \cdot V = \sum ( \underbrace{(U, s) \cdot (V, t)}_{\text{Boolean multiplication}} )_w, \text{ such that } st = w$$

$$\iff U \cdot V = \{ st \mid s \in U \wedge t \in V \}$$

Example:

Let  $U = \{x, xx\}$  and  $V = \{aaa, abc\}$  languages:

$$\begin{aligned} U \cdot V &= (1 \cdot 1)xa aa + (1 \cdot 1)xxaa a + (1 \cdot 1)xabc + (1 \cdot 1)xxabc \\ &\quad + (0 \cdot 0)axxx + \dots \\ &= 1xa aa + 1xxaa a + 1xabc + 1xxabc + 0axxx + \dots \end{aligned}$$

With these definitions, all languages with a fixed alphabet  $\Sigma$  form an algebra  $\mathbb{B}\langle\Sigma\rangle$  over  $\mathbb{B}$ .

## Kleene star:

Let  $U$  be a Language.

$$U^* = \epsilon + U + U^2 + U^3 + \dots$$

Example:

Let  $U = x$ .

$$U^* = \epsilon + x + x^2 + x^3 + x^4 + \dots$$

- 1 Formal Languages Definitions, Examples
- 2 Prove not all Languages are rational (regular)**
- 3 Automatas Definitions, Examples
- 4 Prove: Finite automaton only accept rational languages



All languages generated by a finite number of additions, multiplications, and kleene star is a rational (regular) language.

**Examples:**

$$\Sigma = \{x, y, z\}$$

$$L_1 = x + y \quad (1)$$

$$L_2 = (x + y + z)^* \quad (2)$$

$$L_3 = (x + y^*)^* z^* \quad (3)$$

$$L_4 = (xyz)^* (y + x^* zxyx^*)^* \quad (4)$$

**We show that not all languages generated by a single letter are rational.**

Let  $L_x = x$ .

A gap in a language generated by  $L_x$  is the amount of consecutive missing powers.

**Examples:**

$$L_1 = x \rightsquigarrow \text{gaps}(L_1) = \{0\}$$

$$L_2 = x + x^2 + x^3 \rightsquigarrow \text{gaps}(L_2) = \{0\}$$

$$L_3 = x + x^5 + x^7 \rightsquigarrow \text{gaps}(L_3) = \{0, 3, 1\}$$

$$L_4 = x + x^5 + x^7 \rightsquigarrow \text{maxgap}(L_4) = 3$$

**Raional languages generated by  $x$  have alway a finite maximum gap.  
Examples of not rational languages:**

$$L_1 = x + x^2 + x^4 + x^7 + x^{11} \dots \rightsquigarrow \text{gaps}(L_1) = \{0, 1, 2, 3, \dots\}$$

$$L_2 = x^3 + x^{31} + x^{314} + x^{3141} \dots \rightsquigarrow \text{gaps}(L_2) = \{27, 282, 2826, \dots\}$$

- $\rightsquigarrow$  Only way to get an endless sequence is using the star operator.
- $\rightsquigarrow$  The star operator just repeats the language an abritary amount of times
- $\rightsquigarrow$  Informally we can say rational languages are finite or have an repeating pattern.

Claim: **All rational languages generated by  $L_x$  have a maximum gap**

Let  $L$  be language generated by  $L_x$

$$\text{maxgap}(L) < \infty.$$

**Structural induction over  $+, \cdot, *$ :**

Base case:

$$L = L_x = x$$

$$\implies \text{maxgap}(L) = \text{maxgap}(L_x) = \text{maxgap}(x) = 0$$

## Addition case:

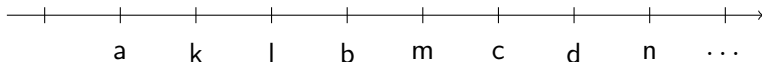
Let  $L_1, L_2$  languages generated by  $L_x$  with a maximum gap.

We need to show that:

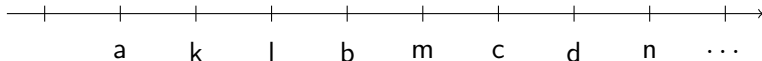
$$\maxgap(L_1 + L_2) < \infty$$

$$L_1 = x^a + x^b + x^c + x^d + \dots$$

$$L_2 = x^k + x^l + x^m + x^n + \dots$$



## case 1: The series of their powers intersect



$\rightsquigarrow$  Gaps are only getting filled

$$\rightsquigarrow \maxgap(L_1 + L_2) \leq \max(\maxgap(L_1), \maxgap(L_2))$$

## case 2: The series of their powers don't intersect



$$\rightsquigarrow \maxgap(L_1 + L_2) \leq \max(\max(\maxgap(L_1), \maxgap(L_2)), k - z + 1)$$

## Multiplication case:

Let  $L_1, L_2$  languages generated by  $L_x$  with a maximum gap.

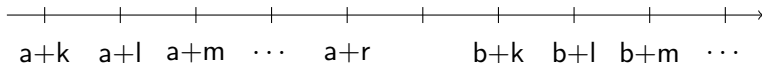
We need to show that:

$$\maxgap(L_1 \cdot L_2) < \infty$$

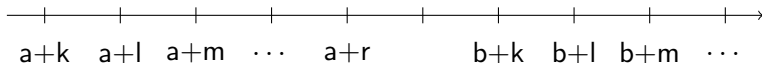
$$L_1 = x^a + x^b + x^c + x^d + \dots$$

$$L_2 = x^k + x^l + x^m + x^n + \dots$$

$$L_1 \cdot L_2 = x^{a+k} + x^{a+l} + \dots + x^{b+k} + x^{b+l} + \dots$$



## case 1: The groups don't intersect



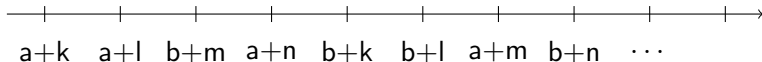
$\rightsquigarrow$  Gaps in one group stay the same

$\rightsquigarrow$  For gaps between groups hold:

$$b + k - (a + r) = b - a + k - r \leq \maxgap(L_1) + 0 = \maxgap(L_1)$$

$$\rightsquigarrow \maxgap(L_1 \cdot L_2) \leq \max(\maxgap(L_1), \maxgap(L_2))$$

## case 2: The groups intersect



$$\rightsquigarrow \maxgap(L_1 \cdot L_2) \leq \max(\maxgap(L_1), \maxgap(L_2))$$



## Kleene star case:

Let  $L$  be a language generated by  $L_x$  with a maximum gap.

We need to show that:

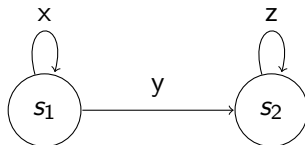
$$\text{maxgap}(L^*) < \infty$$

It's easy to see with induction:

$$\forall n \in \mathbb{N}_0 : \text{maxgap}(L^n) \leq \text{maxgap}(L)$$

$$\rightsquigarrow L^* = \underbrace{\underbrace{\epsilon}_{\leq \text{maxgap}(L)} + \underbrace{L}_{\leq \text{maxgap}(L)} + \underbrace{L^2}_{\leq \text{maxgap}(L)} + \underbrace{L^3}_{\leq \text{maxgap}(L)} + \underbrace{L^4}_{\leq \text{maxgap}(L)} + \dots}_{\leq \text{maxgap}(L)}$$

- 1 Formal Languages Definitions, Examples
- 2 Prove not all Languages are rational (regular)
- 3 Automatas Definitions, Examples**
- 4 Prove: Finite automaton only accept rational languages



$$\begin{array}{c} s_1 \\ s_2 \end{array} \left[ \begin{array}{cc} x & y \\ 0 & z \end{array} \right]$$

$s_1 \quad s_2$

**Formally, we can write:**

Let  $S = \{s_1, s_2, \dots, s_n\}$ ,  $A \in (\Sigma^*)^{n \times n}$ .

Then a transition from state  $s_1$  to  $s_2$ :

$$s_i A_{ij} = s_j$$

Exmample:

Let  $S = \{s_1, s_2\}$  and  $A = \begin{pmatrix} x & y \\ 0 & z \end{pmatrix}$

$$s_1 A_{11} = s_1 x = s_1$$

$$s_1 A_{12} = s_1 y = s_2$$

$$s_2 A_{22} = s_2 z = s_2$$

**What happens if we apply matrix multiplication on  $A \in (\Sigma^*)^{n \times n}$ ?**

$C = A^1$  :  $c_{ij} = a_{ij} \rightsquigarrow$  from state  $i$  to state  $j$  in 1 steps

$C = A^2$  :  $c_{ij} = \sum_{k=1}^n a_{ik} a_{kj} \rightsquigarrow$  from state  $i$  to state  $j$  in 2 steps

$C = A^3$  :  $c_{ij} = \sum_{k_1=1}^n \sum_{k_2=1}^n a_{ik_1} a_{k_1 k_2} a_{k_2 j} \rightsquigarrow$  from state  $i$  to state  $j$  in 3 steps

$\vdots$

$C = A^n$  :  $c_{ij} = \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_{n-1}=1}^n a_{ik_1} a_{k_1 k_2} \cdots a_{k_{n-2} k_{n-1}} a_{k_{n-1} j}$

$\rightsquigarrow$  from state  $i$  to state  $j$  in  $n$  steps

**Combining these will give us all possible inputs from one state in another:**

$$A^* = E_n + A + A^2 + A^3 + \dots$$

$$A^* = \sum_{k \in \mathbb{N}_0} A^k$$

$\leadsto$  In  $a_{ij} \in A$  are all possible inputs carrying state  $i$  to state  $j$  in any arbitrary amount of steps.

- 1 Formal Languages Definitions, Examples
- 2 Prove not all Languages are rational (regular)
- 3 Automatas Definitions, Examples
- 4 Prove: Finite automaton only accept rational languages**

Let  $M \in (\Sigma^*)^{n \times n}$  an arbitrary transition matrix with rational entries.  
We need to show that every entry of  $M^*$  is rational.

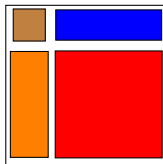
**Devide the matrix in the following block matrices.**

$$M = \begin{pmatrix} M(1,1) & M(1,2) \\ M(2,1) & M(2,2) \end{pmatrix} \in (\Sigma^*)^{n \times n}$$

Where:

$$M(1,1) \in (\Sigma^*)^{1 \times 1}, M(1,2) \in (\Sigma^*)^{1 \times n-1}$$

$$M(2,1) \in (\Sigma^*)^{n-1 \times 1}, M(2,2) \in (\Sigma^*)^{n-1 \times n-1}$$

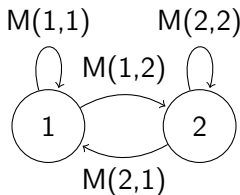




We can divide the all States  $S = \{s_1, s_2, \dots, s_n\}$  in two:

$$1 = \{s_1\} \text{ and } 2 = \{s_2, s_3, \dots, s_n\}$$

**Automata with divided transition matrix and states:**



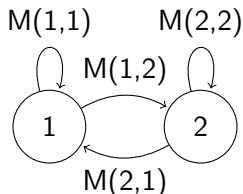
We can write  $M^*$  in terms of our block matrices:

$$M(1,1)^* = M(1,1)^* + (M(1,2)M(2,2)^*M(2,1))^*$$

$$M(1,2)^* = M(1,1)^*M(1,2)(M(2,2)^* + M(2,1)M(1,1)^*M(1,2))^*$$

$$M(2,1)^* = M(2,2)^*M(2,1)(M(1,1)^* + M(1,2)M(2,2)^*M(2,1))^*$$

$$M(2,2)^* = M(2,2)^* + (M(2,1)M(1,1)^*M(1,2))^*$$



**We have reduced the problem to  $n - 1$ :**

$$M(1, 1)^* = M(1, 1)^* + (M(1, 2)M(2, 2)^*M(2, 1))^*$$

$$M(1, 2)^* = M(1, 1)^*M(1, 2)(M(2, 2)^* + M(2, 1)M(1, 1)^*M(1, 2))^*$$

$$M(2, 1)^* = M(2, 2)^*M(2, 1)(M(1, 1)^* + M(1, 2)M(2, 2)^*M(2, 1))^*$$

$$M(2, 2)^* = M(2, 2)^* + (M(2, 1)M(1, 1)^*M(1, 2))^*$$

$\leadsto M(1, 1)^*$  is rational, because its 1-dimensional

$\leadsto M(2, 1), M(1, 2)$  is rational, because  $M$  is rational

$\leadsto M(2, 2)^* \in (\Sigma^*)^{(n-1) \times (n-1)}$  needs to be proven, to be rational

$\leadsto$  With this reduction we can do an induction over the number of states

Thanks for your attention!