Limbaje Formale și Automate
Tutoriat 2

Gabriel Majeri

Lema de pompare

Să se demonstreze că limbajul

$$\mathcal{L} = \{ \ a^n b^n c^2 \ \big| \ n \in \mathbb{N}^* \ \}$$

nu este regulat.

Pentru a demonstra că acesta nu este un limbaj regulat, trebuie să demonstrăm că nu există niciun DFA/NFA/λ-NFA care să îl accepte.

Când vrei să demonstrezi că ceva nu este posibil în matematică, de obicei folosești *reducerea la absurd*. Presupui că o propoziție este adevărată, și ajungi la o contradicție.

O demonstrație că un limbaj nu este regulat/nu este independent de context decurge în felul următor:

1. Presupun că limbajul este regulat/independent de context.
2. **Opțional:** Dacă mă ajută, pot să aplic orice operații la care este închis (deoarece am presupus că limbajul inițial este regulat/independent de context, în urma acestor operații obțin tot un limbaj regulat/independent de context):
   - pentru limbaje regulate: cam toate operațiile
   - pentru limbaje independente de context: intersecție cu limbaje regulate, morfisme și morfisme inverse
3. Pe rezultat aplic lema de pompare. Caut un contra-exemplu de cuvânt și i pentru care lema este falsă, și astfel am o contradicție
4. În concluzie, limbajul nu este regulat/independent de context.

Observație. Lemele de pompare pot fi folosite doar pentru a demonstra că un limbaj **nu** este regulat/independent de context. Pentru a demonstra afirmativa trebuie să construim automatul/gramatica corespunzătoare.

Pentru limbaje regulate

Fie $\mathcal{L}$ un limbaj regulat. Există un $n_0 \in \mathbb{N}^*$ care depinde de limbaj, cu proprietatea că orice cuvânt $w \in \mathcal{L}$, cu $|w| \geq n_0$, se poate descompune în $w = xyz$ cu proprietățile:

- $|y| \geq 1$
- $|xy| \leq n_0$
- $\forall i \in \mathbb{N}$, $xy^i z \in \mathcal{L}$

Pentru limbaje independente de context

Fie $\mathcal{L}$ un limbaj independent de context. Există un $n_0 \in \mathbb{N}^*$ care depinde de limbaj, cu proprietatea că orice cuvânt $w \in \mathcal{L}$, cu $|w| \geq n_0$, se poate descompune în $w = xuyvz$ cu proprietățile:

- $|uv| \geq 1$
- $|uyv| \leq n_0$
- $\forall i \in \mathbb{N}$, $xu^i yv^i z \in \mathcal{L}$

Exerciții

Exercițiu 1 (exercițiul 8 din examen iunie 2011)**.** Demonstrați folosind lema de pompare că limbajul

$$\mathcal{L} = \{\, a^{2k} b^l a^k \mid k \geq 0, l \geq 1 \,\}$$

nu este regulat.

Rezolvare. Partea din limbaj care „ne deranjează” este cea cu k, unde avem o corelare a numărului de a-uri de la început cu numărul de a-uri de la sfârșit.

Ca să simplificăm problema, vom intersecta limbajul cu un limbaj regulat.

Presupunem că $\mathcal{L}$ este regulat. Atunci limbajul

$$\mathcal{L}' = \mathcal{L} \cap a^* b a^* = \{\, a^{2k} b a^k \mid k \geq 0 \,\}$$

este la rândul lui regulat (am folosit o expresie regulată ca să descriu limbajul cu care intersectez).

Fie $n_0 \in \mathbb{N}^*$ constanta din lema de pompare pentru limbaje regulate.

Să luăm cuvântul

$$w = a^{2n_0}\, b\, a^{n_0}$$

Acesta are lungimea $|w| = 2n_0 + 1 + n_0 = 3n_0 + 1 \geq n_0$, deci putem aplica lema pe el.

Trebuie să ne gândim cum ar arăta o descompunere a cuvântului care să respecte proprietățile din lemă:

$$\begin{aligned} w &= x\, y\, z \\ |y| &\geq 1 \\ |xy| &\leq n_0 \end{aligned}$$

Datorită modului în care am ales cuvântul, o descompunere care convine o să fie de forma

$$w = \underbrace{aa \ldots aa}_{x}\, \underbrace{aa \ldots aa}_{y}\, \underbrace{aa \ldots aabaa \ldots aa}_{z}$$

Scris pe scurt:

$$\begin{aligned} x &= a^p \\ y &= a^q \\ z &= a^{2n_0 - q - p}\, b\, a^{n_0} \end{aligned}$$

unde

$$
\begin{aligned}
q &\geq 1 \\
p + q &\leq n_0
\end{aligned}
$$

Din lemă avem că

$$\forall i \in \mathbb{N}, xy^i z \in \mathcal{L}'$$

Pentru $i = 0$ obținem

$$
\begin{aligned}
xy^0 z &= a^p \, a^{2n_0 - q - p} \, b \, a^{n_0} = \\
&= a^{2n_0 - q} \, b \, a^{n_0}
\end{aligned}
$$

Dacă aparține limbajului, ar rezulta că

$$2n_0 - q = 2n_0$$

Dar știm deja că $q \geq 1$, deci contradicție.

Presupunerea că $\mathcal{L}$ este regulat este falsă. $\square$

Exercițiu 2 (exercițiul 8 din examen iunie 2014)**.** Demonstrați că limbajul

$$\mathcal{L} = \{ \, a^k b^{3l} a^l \mid k \geq 1, l \geq 0 \, \}$$

nu este regulat.

Rezolvare. Presupunem că $\mathcal{L}$ este regulat.

Putem intersecta limbajul cu ab^*a^*, și obținem

$$\mathcal{L}' = \left\{ \, ab^{3l} a^l \;\middle|\; l \geq 0 \, \right\}$$

Fie $n_0 \in \mathbb{N}$ din lema de pompare pentru limbaje regulate.

Putem lua cuvântul

$$w = a \, b^{3n_0} \, a^{n_0}$$

care are lungimea $|w| = 1 + 3n_0 + n_0 = 4n_0 + 1 \geq n_0$.

Trebuie să luăm în considerare toate descompunerile posibile pentru acest cuvânt.

Să ne gândim cum arată cuvântul w:

$$w = \underbrace{a}_{\text{un } a} \; \underbrace{bbbbbbb \ldots bbbb}_{3n_0 \text{ de } b\text{-uri}} \; \underbrace{aaa \ldots aaa}_{n_0 \text{ de } a\text{-uri}}$$

Știm că xy trebuie să se afle la începutul cuvântului, și din lema de pompare avem $|xy| \leq n_0$.

1. Cazul în care x îl conține pe a:

$$w = \underbrace{abb \dots bbb}_{x}\, \underbrace{b \dots bb}_{y}\, \underbrace{bbb \dots bbbaa \dots aa}_{z}$$

$$\begin{aligned} x &= a\, b^p \\ y &= b^q \\ z &= b^{3n_0-q-p} a^{n_0} \end{aligned}$$

unde

$$\begin{aligned} |y| \geq 1 &\implies q \geq 1 \\ |xy| \leq n_0 & \end{aligned}$$

Din lema de pompare avem că

$$\forall i \in \mathbb{N}, x\, y^i\, z \in \mathcal{L}'$$

Luăm $i = 0$. Obținem cuvântul

$$\begin{aligned} xy^0 z &= a\, b^p\, b^{3n_0-q-p}\, a^{n_0} \\ &= a\, b^{3n_0-q}\, a^{n_0} \end{aligned}$$

care nu aparține limbajului deoarece $3n_0 - q < 3n_0$.

Contradicție cu lema de pompare, deci $\mathcal{L}'$ nu este regulat.

2. Cazul în care x nu-l conține pe a (este λ):

$$w = \underbrace{}_{x} \underbrace{abb \dots bbb}_{y}\, \underbrace{bbb \dots bbbaa \dots aa}_{z}$$

$$\begin{aligned} x &= \lambda \\ y &= a\, b^p \\ z &= b^{3n_0-p}\, a^{n_0} \end{aligned}$$

unde

$$1 \leq |y| \leq n_0$$

După un raționament analog ajungem la aceeași concluzie. În momentul în care ridicăm y la o putere $i \geq 2$, o să avem mai mult de un a la început.

$\square$

Exercițiu 3 (exercițiul 9 din examen iunie 2016)**.** Demonstrați că limbajul

$$\mathcal{L} = \{\, wa^k w \mid w \in \{\, a, b, c \,\}^* , k \geq 0 \,\}$$

nu este regulat.

Rezolvare. Presupunem că $\mathcal{L}$ ar fi regulat.

Fie $n_0 \in \mathbb{N}$ din lema de pompare.

Putem alege $w = b^{n_0}\, a\, b^{n_0}$, cu $|w| = n_0 + 1 + n_0 = 2n_0 + 1 \geq n_0$.

Avem

$$\begin{aligned} x &= b^p \\ y &= b^q \\ z &= b^{n_0 - q - p}\, a\, b^{n_0} \end{aligned}$$

cu

$$\begin{aligned} |y| &\geq 1 \\ |xy| &\leq n_0 \end{aligned}$$

Aplicăm lema de pompare pentru $i = 0$ și obținem

$$\begin{aligned} xy^0 z &= b^p\, b^{n_0 - q - p}\, a\, b^{n_0} \\ &= b^{n_0 - q}\, a\, b^{n_0} \end{aligned}$$

Deoarece $n_0 - q \neq n_0$, cuvântul nu aparține limbajului.

Prin urmare, presupunerea noastră este falsă, limbajul nu este regulat.

$\square$

Exercițiu 4 (exercițiul 10 din examen iunie 2017)**.** Demonstrați că limbajul

$$\mathcal{L} = \{\, a^n b^m \mid n \text{ pătrat perfect}, m \geq 5 \,\}$$

nu este independent de context.

Rezolvare. Presupunem că $\mathcal{L}$ este independent de context.

Îl intersectăm cu $a^* b^5$ și obținem

$$\mathcal{L}' = \{\, a^n b^5 \mid n \text{ pătrat perfect } \,\}$$

Fie $n_0 \in \mathbb{N}$ din lema de pompare.

Alegem cuvântul

$$w = a^{(n_0)^2} b^5$$

care este din limbaj și are $|w| = (n_0)^2 + 5 \geq n_0$

Dacă luăm o descompunere în care uyv se află în zona de b-uri, atunci când o să aplicăm lema de pompare pentru $i \neq 1$ nu mai avem b^5.

Dacă luăm o descompunere în care uyv conține și a-uri, de exemplu:

$$\begin{aligned} x &= a^p \\ u &= a^q \\ y &= a^r \\ v &= a^s \\ z &= a^{(n_0)^2 - s - r - q - p}\, b^5 \end{aligned}$$

unde

$$\begin{aligned} |uv| \geq 1 &\iff q + s \geq 1 \\ |uyv| \leq n_0 &\iff q + r + s \leq n_0 \end{aligned}$$

Aplicăm lema de pompare pentru un i oarecare și vedem ce obținem:

$$\begin{aligned} xu^iyv^iz &= a^p\, a^{iq}\, a^r\, a^{is}\, a^{(n_0)^2 - s - r - q - p}\, b^5 \\ &= a^{(n_0)^2 + q(i-1) + s(i-1)}\, b^5 \\ &= a^{(n_0)^2 + (q+s)(i-1)}\, b^5 \end{aligned}$$

Între două pătrate perfecte nu poate exista un alt pătrat perfect. După $(n_0)^2$, următorul pătrat perfect este $(n_0 + 1)^2$, adică $(n_0)^2 + 2n_0 + 1$.

Dacă luăm $i = 2$, obținem

$$a^{(n_0)^2 + (q+s)}\, b^5$$

Cu siguranță $q + s \leq 2n_0 + 1$, deci nu avem un pătrat perfect. Cuvântul nu aparține limbajului.

În concluzie, $\mathcal{L}$ nu este independent de context. $\square$

Exercițiu 5 (exercițiul bonus din examen iunie 2013)**.** Demonstrați că limbajul

$$\mathcal{L} = \{\, 0^n 1^m \mid n > 0, m \text{ prim} \,\}$$

nu este independent de context.

Rezolvare. Presupunem că $\mathcal{L}$ este independent de context.

Îl putem intersecta cu 01^* și obținem

$$\mathcal{L} = \{\, 01^m \mid m \text{ prim } \}$$

Fie $n_0 \in \mathbb{N}$ din lema de pompare.

Notăm cu $P =$ următorul număr prim mai mare decât n_0. Cuvântul $w = 01^P$ aparține limbajului și are lungime $|w| = 1 + P \geq n_0$.

Avem mai multe cazuri, în funcție de cum este descompunerea:

1.
$$\begin{aligned} x &= 0\,1^p \\ u &= 1^q \\ y &= 1^r \\ v &= 1^s \\ z &= 1^{P-s-r-q-p} \end{aligned}$$

   unde

$$\begin{aligned} |uv| \geq 1 &\iff q + s \geq 1 \\ |uyv| \leq n_0 &\iff q + r + s \leq n_0 \end{aligned}$$

   Aplicăm lema de pompare pentru un $i \in \mathbb{N}$ oarecare:

$$\begin{aligned} xu^iyv^iz &= 0\,1^p\,1^{iq}\,1^r\,1^{is}\,1^{P-s-r-q-p} \\ &= 0\,1^{P+q(i-1)+s(i-1)} \\ &= 0\,1^{P+(q+s)(i-1)} \end{aligned}$$

   Pentru $i = P + 1$, obținem cuvântul

$$0\,1^{P+(q+s)P} = 0\,1^{(q+s)(P+1)}$$

   care nu aparține limbajului, deoarece exponentul lui 1 este număr compus. Contradicție cu lema de pompare.

2.
$$\begin{aligned} x &= \lambda \\ u &= 0\,1^p \\ y &= 1^q \\ v &= 1^r \\ z &= 1^{P-r-q-p} \end{aligned}$$

   Aplicând lema de pompare obținem un rezultat analog cu primul caz.

3.

$$x = u = \lambda$$
$$y = 0\,1^p$$
$$v = 1^q$$
$$z = 1^{P-q-p}$$

Asemănător cu cazurile de mai sus.

4.

$$x = u = y = \lambda$$
$$v = 0\,1^p$$
$$z = 1^{P-p}$$

Asemănător cu cazurile de mai sus.

În concluzie, $\mathcal{L}$ nu este independent de context. □

Exercițiu 6 (exercițiul 11 din examen iunie 2011)**.** Demonstrați că limbajul

$$\mathcal{L} = \{\, a^i b^j c^k \mid i < j \text{ și } i + 2j + 3 < k \,\}$$

nu este independent de context.

Rezolvare. Presupunem că *lang* este independent de context.

Fie n_0 din lema de pompare.

Vrem să alegem un cuvânt în care inegalitățile să fie satisfăcute cât mai la limită. Alegem $i = n_0$, din $i < j$ putem alege $j = n_0 + 1$. Din $i + 2j + 3 < k$, alegem $k = 3n_0 + 6$.

Cuvântul ales este $w = a^{n_0}\, b^{n_0+1}\, c^{3n_0+6}$, unde $|w| = 5n_0 + 7 \geq n_0$.

Obținem foarte multe descompuneri posibile:

Dacă

$$x = a^p$$
$$u = a^q$$
$$y = a^r$$
$$v = a^s$$
$$z = a^{n_0-s-r-q-p}\, b^{n_0+1}\, c^{3n_0+6}$$

putem pompa cu $i > 1$ și vom obține un număr de a-uri care depășește numărul de b-uri, ieșind din limbaj.

Dacă

$$
\begin{aligned}
x &= a^{n_0} b^p \\
u &= b^q \\
y &= b^r \\
v &= b^s \\
z &= b^{n_0+1-s-r-q-p} c^{3n_0+6}
\end{aligned}
$$

putem pompa cu $i > 1$ și vom depăși a doua condiție.
etc. □

Exercițiu 7 (exercițiul 10 din examen iunie 2016)**.** Demonstrați că limbajul

$$\mathcal{L} = \{\, a^n b^m c^r \mid n \geq m \geq r \geq 150 \,\}$$

nu este independent de context.

Rezolvare. Limbajul se poate rescrie ca

$$\mathcal{L} = \{\, a^n b^m c^r \mid n \geq m, m \geq r, r \geq 150 \,\}$$

Alegem un cuvânt cum ar fi $w = a^{n_0} b^{n_0} c^{n_0}$ și aplicăm lema de pompare.

- În cazurile în care pompăm doar a-uri putem lua $i = 0$ și o să avem mai puține a-uri decât b-uri.
- Când pompăm și a-uri și b-uri sau doar b-uri putem lua $i = 0$ și avem mai puține b-uri decât c-uri.
- Când pompăm c-uri putem lua un $i > 1$ și o să avem mai multe c-uri decât b-uri.

□