Algebraic Approaches to the Study of the Polynomial Closure of Classes of Regular Languages

Jana Bartoňová

Department of Mathematics and Statistics, Faculty of Science, Masaryk University

Brno, Czech Republic

Regular languages

- A ... fixed finite set (*alphabet*, elements of A are called *letters*)
- *monoid* $(M, \cdot, 1)$ ... a set M with an associative binary operation $\cdot$ and a neutral element 1
- A^* ... free monoid over A, with the operation of *concatenation*; elements of A^* are called *words*
- $L \subseteq A^*$ ... *language* over A
- *regular* language ... language which can be created from languages of the form $\emptyset$ and $\{a\}$, where $a \in A$, by means of finite number of applications of the following operations:
  - concatenation: $K \cdot L = \{x \cdot y \mid x \in K, y \in L\}$
  - iteration: $L^* = \bigcup_{n=0}^{\infty} L^n$ (submonoid of A^* generated by L)
  - union.
- The set of all regular languages is closed also under complementation (and hence also under intersection).

Quotienting lattices of regular languages

- *lattice* of regular languages - a set of regular languages containing $\emptyset$, A^* and closed under supremum $\cup$ (union) and infimum $\cap$ (intersection)
- *left/ right quotient* of a language $L \subseteq A^*$ by a word $u \in A^*$:

$$u^{-1}L = \{x \in A^* \mid ux \in L\},\ Lu^{-1} = \{x \in A^* \mid xu \in L\}$$

Example

$L = A^*aA^*bA^* \quad$ where $a, b \in A$

- $a^{-1}L = L \cup A^*bA^* = A^*bA^*$
- $b^{-1}L = L$

- *Quotienting lattice* of regular languages is a lattice of regular languages closed under left and right quotients (by every word).

Polynomial closure

Polynomial closure $Pol(\mathcal{C})$ of a set of regular languages $\mathcal{C}$ is a set of regular languages which are finite unions of languages of the form

$$L_0 a_1 L_1 \dots a_n L_n \qquad \text{where } a_i \in A, L_i \in \mathcal{C}.$$

Concatenation hierarchies of regular languages:

$\mathcal{C}_0$ ... a given quotienting lattice of regular languages

$\mathcal{C}_{n+1/2} = Pol(\mathcal{C}_n) = Pol\big(\{L^{\mathrm{C}} \mid L \in \mathcal{C}_{n-1/2}\}\big)$

$\mathcal{C}_{n+1} = B(\mathcal{C}_{n+1/2})$ ... *Boolean* closure of the level $n + 1/2$ (closure under union and complementation)

Example

- $\mathcal{C}_0 := \{A^*, \emptyset\}$
- $\mathcal{C}_{1/2} = Pol(\mathcal{C}_0)$ ... finite unions of languages of the form

$$A^* a_1 A^* \dots a_n A^* \quad \text{where } a_i \in A$$

- How to decide whether a given regular language belongs to $Pol(\mathcal{C})$ for a given $\mathcal{C}$?

Let's learn about some tools for an algebraic description of the polynomial closure.

Metric $d\colon A^* \times A^* \to \mathbb{R}_0$

- Two words $u, v \in A^*$ are "close" w.r.t. d iff for distinguishing them by a homomorphism $\varphi\colon A^* \to M$ there's needed a "big" (finite) monoid M.
- We denote by $\widehat{A^*}$ a completion of a metric monoid (A^*, d).
- $\widehat{A^*}$ still forms a monoid, with an operation of concatenation continuously extended from A^*.
- Two *pseudowords* $u, v \in \widehat{A^*}$ are "close" iff for distinguishing them by a **continuous** homomorphism $\varphi\colon \widehat{A^*} \to M$ there's needed a "big" (finite) monoid M.

Example

For an arbitrary pseudoword $x \in \widehat{A^*}$ we define $x^\omega := lim_{n\to\infty} x^{n!}$.

- For every continuous homomomorphism $\varphi\colon \widehat{A^*} \to M$ to a finite monoid M: $\varphi(x^\omega) = \varphi(\lim_{n\to\infty} x^{n!}) = \lim_{n\to\infty} \varphi(x)^{n!} \ldots$ the unique idempotent power of $\varphi(x)$

Preorder $\leq_L$ on $\widehat{A^*}$

- For a regular language L and pseudowords $u, v \in \widehat{A^*}$ the relation

$$u \leq_L v$$

means that the following property is satisfied:

$$\forall p, q \in A^* : \quad u \in \overline{p^{-1}Lq^{-1}} \;\Rightarrow\; v \in \overline{p^{-1}Lq^{-1}}.$$

- The relation $\sim_L \;\subseteq\; \widehat{A^*} \times \widehat{A^*}$ defined by

$$u \sim_L v \quad \Leftrightarrow \quad u \leq_L v \text{ and } v \leq_L u$$

is a congruence of finite index.

- $\mathbf{M_L} = A^*/\sim_L \ldots$ *syntactic monoid* of L, equipped with a partial order induced by $\leq_L$

  … **computable algorithmically**

An equational description of the polynomial closure

Theorem (Branco, Pin, 2009)

Let $\mathcal{C}$ be a quotienting lattice of regular languages, K a regular language. Then K belongs to $Pol(\mathcal{C})$ if and only if $x^\omega \leq_K x^\omega y x^\omega$ for all $x, y \in \widehat{A^*}$ such that $x =_\mathcal{C} x^2 \leq_\mathcal{C} y$.

- For a quotienting lattice of regular languages $\mathcal{C}$:

$$
\begin{aligned}
u \leq_\mathcal{C} v \quad &\Leftrightarrow \quad \forall L \in \mathcal{C} : \; u \leq_L v \\
&\Leftrightarrow \quad \forall L \in \mathcal{C}\, \forall p, q \in A^* : \; u \in \overline{p^{-1} L q^{-1}} \; \Rightarrow \; v \in \overline{p^{-1} L q^{-1}} \\
&\Leftrightarrow \quad \forall L \in \mathcal{C} : \quad u \in \overline{L} \; \Rightarrow \; v \in \overline{L}.
\end{aligned}
$$

- $u =_\mathcal{C} v \quad \Leftrightarrow \quad u \leq_\mathcal{C} v$ and $v \leq_\mathcal{C} u$
- For the decidability of $Pol(\mathcal{C})$ it suffices to be able to compute the set of pairs

$$\mathcal{C}[K] = \{(S, T) \in M_K \times M_K \mid S = S^2, \exists x \in \overline{S}, \exists y \in \overline{T} : \; x \leq_\mathcal{C} y\}.$$

Connection to the separability of regular languages

- For the decidability of $Pol(\mathcal{C})$ it suffices to be able to compute the set of pairs

$$\mathcal{C}[K] = \{(S,T) \in M_K \times M_K \mid S = S^2, \exists x \in \overline{S}, \exists y \in \overline{T} : \ x \leq_{\mathcal{C}} y\}.$$

For $u \in \widehat{A^*}$ denote $[u]_{\sim_K} = \{v \in A^* \mid v \sim_K u\}$.

Proposition (connection of Theorems by Branco, Pin, 2009 and Place, Zeitoun, 2018)

Let $\mathcal{C}$ be a quotienting lattice of regular languages, K a regular language, $x, y \in \widehat{A^*}$ pseudowords. Then the two following conditions are equivalent:

1. $x \leq_{\mathcal{C}} y$
2. $\forall L \in \mathcal{C} : \quad [x]_{\sim_K} \subseteq L \ \Rightarrow \ [y]_{\sim_K} \cap L \neq \emptyset$

- The condition 2 says that the language $[x]_{\sim_K}$ **is not** *$\mathcal{C}$-separable* from the language $[y]_{\sim_K}$. (Place, Zeitoun)

Computation of $\mathcal{C}[K]$

$$\mathcal{C}[K] = \{(S,T) \in M_K \times M_K \mid S = S^2, \exists x \in \overline{S}, \exists y \in \overline{T} :\ x \leq_{\mathcal{C}} y\}$$

Proposition

Let $\mathcal{C}$ be a quotienting lattice of regular languages, K a regular language, $x, y \in \widehat{A^*}$ pseudowords. Then the two following conditions are equivalent:

1. $x \leq_{\mathcal{C}} y$
2. $\forall L \in \mathcal{C} : \quad [x]_{\sim_K} \subseteq L \ \Rightarrow\ [y]_{\sim_K} \cap L \neq \emptyset$

- For given $S, T \in M_L$ and a given regular language L we can check if the conditions $S \subseteq L$ and $T \cap L \neq \emptyset$ are satisfied.
- So for a **finite** quotienting lattice of regular languages $\mathcal{C}$ we can compute $\mathcal{C}[K]$.

Example - Computation of $\mathcal{C}_{1/2}[K]$

$\mathcal{C}_0 = \{\emptyset, A^*\}$
$\mathcal{C}_{1/2} = Pol(\mathcal{C}_0) \ldots$ finite unions of languages of the form

$$A^* a_1 A^* a_2 A^* \ldots a_n A^* \quad \text{where } a_1, \ldots, a_n \in A$$

$\ldots$ quotienting lattice of regular languages
Stratification: $\mathcal{C}_{1/2} = \bigcup_{k=1}^{\infty} \mathcal{C}_{1/2}^k$
$\mathcal{C}_{1/2}^k \ldots$ a (quotienting) lattice of regular languages generated by languages of the form

$$A^* a_1 A^* a_2 A^* \ldots a_n A^* \quad \text{where } \mathbf{n} \leq \mathbf{k}, a_1, \ldots, a_n \in A$$

The lattice $\mathcal{C}_{1/2}^k$ is finite for every $k \in \mathbb{N}$.

Proposition

Let K be a regular language. Then $\mathcal{C}_{1/2}[K] = \mathcal{C}_{1/2}^{|M_K|}[K]$.

Work in progress - Computation of $\mathcal{C}_{3/2}[K]$

$\mathcal{C}_{3/2} = Pol\big(\{L^{\mathrm{C}} \mid L \in \mathcal{C}_{1/2}\}\big) \ldots$ finite unions of languages of the form

$$B_0^* a_1 B_1^* a_2 B_2^* \ldots a_n B_n^* \quad \text{where } a_i \in A, B_i \subseteq A \quad \text{(Arfi, 1991)}$$

$\mathcal{C}_{3/2} = \bigcup_{k=1}^{\infty} \overline{\mathcal{C}_{3/2}^k}$ where $\overline{\mathcal{C}_{3/2}^k}$ are certain finite quotienting lattices defined by terms of logic FO[<] on words (Place, Zeitoun)

Theorem (Place, Zeitoun, 2017)

Let K be a regular language. Then $\mathcal{C}_{3/2}[K] = \overline{\mathcal{C}_{3/2}^k}[K]$ for $k = 18 \cdot |M_K|^2 \cdot 2^{|M_K|}$.

$\mathcal{C}_{3/2} = \bigcup_{k=1}^{\infty} \mathcal{C}_{3/2}^k$ where $\mathcal{C}_{3/2}^k$ is a lattice of regular languages generated by languages of the form

$$B_0^* a_1 B_1^* a_2 B_2^* \ldots a_n B_n^* \quad \text{where } \mathbf{n} \leq \mathbf{k}, a_i \in A, B_i \subseteq A$$

$\mathcal{C}_{3/2}^k$ is finite for every $k \in \mathbb{N}$ (even $\mathcal{C}_{3/2}^k \subseteq \overline{\mathcal{C}_{3/2}^k}$).
GOAL: Find k (as small as possible) such that $\mathcal{C}_{3/2}[K] = \mathcal{C}_{3/2}^k[K]$.

Thank you for your attention.