Classes of languages generated by the Kleene star of a word

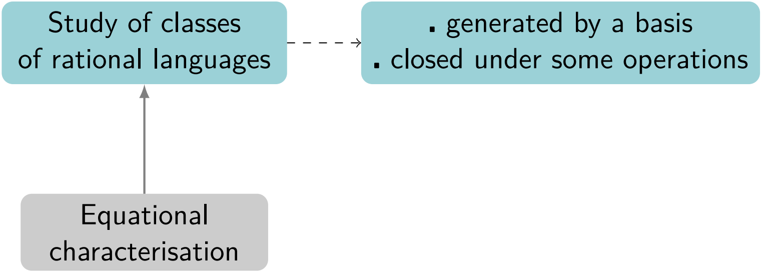
Laure Daviaud
LIF, Aix-Marseille Université

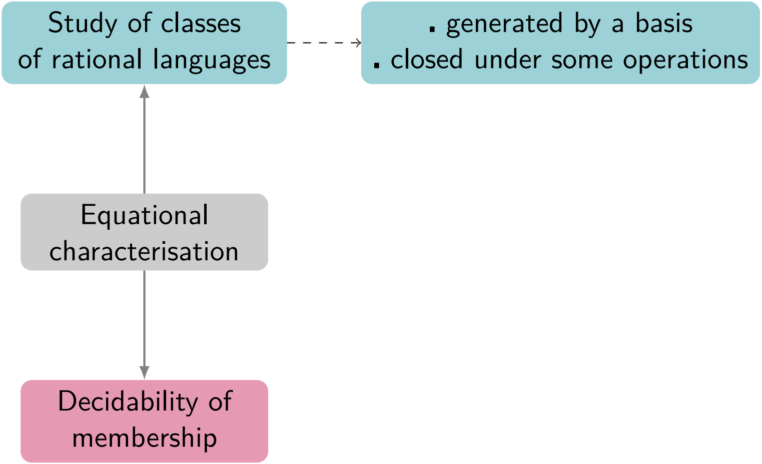
Charles Paperman
Warsaw University

MFCS 2015

Study of classes of rational languages

- generated by a basis
- closed under some operations


Study of classes of rational languages
. generated by a basis
. closed under some operations
Equational characterisation


Study of classes of rational languages
. generated by a basis
. closed under some operations
Equational characterisation
Decidability of membership

Study of classes of rational languages

- generated by a basis
- closed under some operations

Equational characterisation

Decidability of membership

Rational languages
= Recognisable languages

A finite monoid M recognises a language L if there is a morphism $\varphi : A^* \to M$ and $P \subseteq M$ s.t. $L = \varphi^{-1}(P)$

Study of classes of rational languages

- generated by a basis
- closed under some operations

Equational characterisation

Decidability of membership

Rational languages
= Recognisable languages

A finite monoid M recognises a language L if there is a morphism $\varphi : A^* \to M$ and $P \subseteq M$ s.t. $L = \varphi^{-1}(P)$

Classes of languages generated by $\{u^* \mid u \in A^*\}$

Profinite distance on words

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Profinite distance on words

A monoid M separates u and v if:
there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

$$d(u, v) = 2^{-n}$$

where n is the minimal size of a monoid that separates u and v.

Profinite distance on words

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

$$d(u, v) = 2^{-n}$$

where n is the minimal size of a monoid that separates u and v.

Example 1: $a \in A$ - separate a and a^{100}?

$M = \{1, a, a^2 = a^3 = \ldots = 0\}$

Profinite distance on words

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

$$d(u, v) = 2^{-n}$$

where n is the minimal size of a monoid that separates u and v.

Example 1: $a \in A$ - separate a and a^{100}?
$M = \{1, a, a^2 = a^3 = \ldots = 0\}$

Example 2: $u \in A^*$, $n \in \mathbb{N}$ - separate $u^{n!}$ and $u^{(n+1)!}$?
$x \in M$ then $x^{|M|!} = x^{(|M|+1)!} =$ the idempotent power of x in M
$\implies \varphi(u)^{|M|!} = \varphi(u)^{(|M|+1)!}$

$u^{n!}$ and $u^{(n+1)!}$ cannot be separated by a monoid of size less than n

Profinite monoid

Definition

Profinite monoid $\widehat{A^*}$:
completion of A^* with respect to the distance d.

- Compact
- $\varphi : A^* \to M$ can be uniquely extended to a continuous morphism $\widehat{\varphi} : \widehat{A^*} \to M$

Profinite monoid

Definition

Profinite monoid $\widehat{A^*}$:
completion of A^* with respect to the distance d.

- Compact
- $\varphi : A^* \to M$ can be uniquely extended to a continuous morphism $\widehat{\varphi} : \widehat{A^*} \to M$

V.I.P. word (very important profinite word): the Idempotent Power

$$u^\omega = \lim_{n \to \infty} u^{n!}$$

- For all morphisms $\varphi : A^* \to M$ (finite monoid):
  $\widehat{\varphi}(u^\omega)$ is the idempotent power of $\widehat{\varphi}(u)$ in M.

Equations

Definition

Given two profinite words u, v, a rational language L satisfies

$$u \to v$$

if $u \in \bar{L}$ implies $v \in \bar{L}$

$a, b \in A$

Equation $ab \to aba$

$\{L \subseteq A^* \mid ab \notin L\} \cup \{L \subseteq A^* \mid ab, aba \in L\}$

Equations

Definition

Given two profinite words u, v, a rational language L satisfies

$$u \leftrightarrow v$$

if $u \in \overline{L}$ if and only if $v \in \overline{L}$

$a, b \in A$

Equation $ab \leftrightarrow aba$

$\{L \subseteq A^* \mid ab, aba \notin L\} \cup \{L \subseteq A^* \mid ab, aba \in L\}$

Equations

Definition

Given two profinite words u, v, a rational language L satisfies

$$u \leqslant v$$

if for all $w, w' \in \widehat{A^*}, wuw' \in \overline{L}$ implies $wvw' \in \overline{L}$

$a, b \in A$

Equation $ab \leqslant aba$

$\{L \subseteq A^* \mid \text{for all } w, w' \in A^*, \text{ if } wabw' \in L \text{ then } wabaw' \in L\}$

Equations

Definition

Given two profinite words u, v, a rational language L satisfies

$$u = v$$

if for all $w, w' \in \widehat{A^*}, wuw' \in \overline{L}$ if and only if $wvw' \in \overline{L}$

$a, b \in A$

Equation $ab = aba$

$\{L \subseteq A^* \mid \text{for all } w, w' \in A^*,\ wabw' \in L \text{ iff } wabaw' \in L\}$

Equations

Theorem [Gehrke, Grigorieff, Pin 2008]

Classes of rational languages

- Lattice (union, intersection): $\rightarrow$
- Boolean algebra (lattice, complement): $\leftrightarrow$
- Lattice closed under quotient: $\leqslant$
- Boolean algebra closed under quotient: $=$

quotient : $u^{-1}Lv^{-1} = \{w \mid uwv \in L\}$

Why u^*?

- Non trivial applications of the equational theory of rational languages.

- Characterise the variety of languages generated by the languages u^*. [Restivo]

- (long-term perspective) Boolean algebra generated by the languages F^* for F a finite set of words. [key result for the generalised star-height problem]

Equations for u^*

$$P_u = \bigcup_{p \text{ prefix of } u} u^* p \quad \text{and} \quad S_u = \bigcup_{s \text{ suffix of } u} s u^*$$

$$x^\omega y^\omega = 0 \text{ for } x, y \in A^* \text{ such that } xy \neq yx \tag{E_1}$$

$$x^\omega y = 0 \text{ for } x, y \in A^* \text{ such that } y \notin P_x \tag{E_2}$$

$$y x^\omega = 0 \text{ for } x, y \in A^* \text{ such that } y \notin S_x \tag{E_3}$$

$$x^\omega \leqslant 1 \text{ for } x \in A^* \tag{E_4}$$

$$x^\ell \leftrightarrow x^{\omega+\ell} \text{ for } x \in A^*,\ \ell > 0 \tag{E_5}$$

$$x \to x^\ell \text{ for } x \in A^*,\ \ell > 0 \tag{E_6}$$

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \tag{E_7}$$

DECIDABLE

Equations for u^*

$$P_u = \bigcup_{p \text{ prefix of } u} u^* p \quad \text{and} \quad S_u = \bigcup_{s \text{ suffix of } u} s u^*$$

$$x^\omega y^\omega = 0 \text{ for } x, y \in A^* \text{ such that } xy \neq yx \qquad (E_1)$$

$$x^\omega y = 0 \text{ for } x, y \in A^* \text{ such that } y \notin P_x \qquad (E_2)$$

$$y x^\omega = 0 \text{ for } x, y \in A^* \text{ such that } y \notin S_x \qquad (E_3)$$

$$x^\omega \leqslant 1 \text{ for } x \in A^* \qquad (E_4)$$

$$x^\ell \leftrightarrow x^{\omega+\ell} \text{ for } x \in A^*,\ \ell > 0 \qquad (E_5)$$

$$x \to x^\ell \text{ for } x \in A^*,\ \ell > 0 \qquad (E_6)$$

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

DECIDABLE Lattice

Equations for u^*

$$P_u = \bigcup_{p \text{ prefix of } u} u^* p \quad \text{and} \quad S_u = \bigcup_{s \text{ suffix of } u} s u^*$$

$$x^\omega y^\omega = 0 \text{ for } x, y \in A^* \text{ such that } xy \neq yx \tag{E_1}$$

$$x^\omega y = 0 \text{ for } x, y \in A^* \text{ such that } y \notin P_x \tag{E_2}$$

$$y x^\omega = 0 \text{ for } x, y \in A^* \text{ such that } y \notin S_x \tag{E_3}$$

$$x^\omega \leqslant 1 \text{ for } x \in A^* \tag{E_4}$$

$$x^\ell \leftrightarrow x^{\omega+\ell} \text{ for } x \in A^*,\ \ell > 0 \tag{E_5}$$

$$x \to x^\ell \text{ for } x \in A^*,\ \ell > 0 \tag{E_6}$$

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \tag{E_7}$$

DECIDABLE Lattice closed under quotients

Equations for u^*

$$P_u = \bigcup_{p \text{ prefix of } u} u^* p \quad \text{and} \quad S_u = \bigcup_{s \text{ suffix of } u} s u^*$$

$$x^\omega y^\omega = 0 \text{ for } x, y \in A^* \text{ such that } xy \neq yx \tag{E_1}$$

$$x^\omega y = 0 \text{ for } x, y \in A^* \text{ such that } y \notin P_x \tag{E_2}$$

$$y x^\omega = 0 \text{ for } x, y \in A^* \text{ such that } y \notin S_x \tag{E_3}$$

$$x^\omega \leqslant 1 \text{ for } x \in A^* \tag{E_4}$$

$$x^\ell \leftrightarrow x^{\omega+\ell} \text{ for } x \in A^*,\ \ell > 0 \tag{E_5}$$

$$x \to x^\ell \text{ for } x \in A^*,\ \ell > 0 \tag{E_6}$$

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \tag{E_7}$$

DECIDABLE Boolean algebra closed under quotients

Equations for u^*

$$P_u = \bigcup_{p \text{ prefix of } u} u^* p \quad \text{and} \quad S_u = \bigcup_{s \text{ suffix of } u} s u^*$$

$$x^\omega y^\omega = 0 \text{ for } x, y \in A^* \text{ such that } xy \neq yx \qquad (E_1)$$

$$x^\omega y = 0 \text{ for } x, y \in A^* \text{ such that } y \notin P_x \qquad (E_2)$$

$$y x^\omega = 0 \text{ for } x, y \in A^* \text{ such that } y \notin S_x \qquad (E_3)$$

$$x^\omega \leqslant 1 \text{ for } x \in A^* \qquad (E_4)$$

$$x^\ell \leftrightarrow x^{\omega+\ell} \text{ for } x \in A^*,\ \ell > 0 \qquad (E_5)$$

$$x \to x^\ell \text{ for } x \in A^*,\ \ell > 0 \qquad (E_6)$$

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

DECIDABLE Boolean algebra

The Boolean algebra

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

An example:

$$(a^2)^* - (a^6)^* = (a^6)^* a^2 \cup (a^6)^* a^4$$

$1 \quad a \quad a^2 \quad a^3 \quad a^4 \quad a^5 \quad a^6 \quad a^7 \quad a^8 \quad a^9 \quad a^{10} \quad a^{11} \quad a^{12} \quad a^{13} \quad a^{14} \quad \ldots$

Equivalence relation over the integers

$r \equiv_m s$ if and only if $\gcd(r, m) = \gcd(s, m)$
$(u^m)^* u^r \subseteq L$ if and only if $(u^m)^* u^s \subseteq L$

The Boolean algebra

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

An example:

$$(a^2)^* - (a^6)^* = (a^6)^* a^2 \cup (a^6)^* a^4$$

$1 \; a \; a^2 \; a^3 \; a^4 \; a^5 \; a^6 \; a^7 \; a^8 \; a^9 \; a^{10} \; a^{11} \; a^{12} \; a^{13} \; a^{14} \; \ldots$

Equivalence relation over the integers

$r \equiv_m s$ if and only if $\gcd(r, m) = \gcd(s, m)$
$(u^m)^* u^r \subseteq L$ if and only if $(u^m)^* u^s \subseteq L$

$$2 \equiv_6 4 \text{ since } \gcd(2, 6) = 2 = \gcd(4, 6)$$
$$(u^6)^* u^2 \subseteq L \text{ if and only if } (u^6)^* u^4 \subseteq L$$

The Boolean algebra

$$x^\alpha \leftrightarrow x^\beta \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

An example:

$$(a^2)^* - (a^6)^* = (a^6)^* a^2 \cup (a^6)^* a^4$$

$1 \quad a \quad a^2 \quad a^3 \quad a^4 \quad a^5 \quad a^6 \quad a^7 \quad a^8 \quad a^9 \quad a^{10} \quad a^{11} \quad a^{12} \quad a^{13} \quad a^{14} \quad \ldots$

Equivalence relation over the integers

$r \equiv_m s$ if and only if $\gcd(r, m) = \gcd(s, m)$
$(u^m)^* u^r \subseteq L$ if and only if $(u^m)^* u^s \subseteq L$

$x^\alpha \leftrightarrow x^\beta$ for α and β representing sequences of integers $(km + r)_k$ and $(km + s)_k$ with $r \equiv_m s$...

The Boolean algebra

$$x^{\alpha} \leftrightarrow x^{\beta} \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

An example:

$$(a^2)^* - (a^6)^* = (a^6)^* a^2 \cup (a^6)^* a^4$$

$1 \quad a \quad a^2 \quad a^3 \quad a^4 \quad a^5 \quad a^6 \quad a^7 \quad a^8 \quad a^9 \quad a^{10} \quad a^{11} \quad a^{12} \quad a^{13} \quad a^{14} \quad \ldots$

Equivalence relation over the integers

$r \equiv_m s$ if and only if $\gcd(r, m) = \gcd(s, m)$
$(u^m)^* u^r \subseteq L$ if and only if $(u^m)^* u^s \subseteq L$

$x^{\alpha} \leftrightarrow x^{\beta}$ for α and β profinite numbers in $\widehat{\mathbb{N}} = \widehat{\{a\}^*}$
satisfying some specific conditions...

The Boolean algebra

$$x^{\alpha} \leftrightarrow x^{\beta} \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

An example:

$$(a^2)^* - (a^6)^* = (a^6)^* a^2 \cup (a^6)^* a^4$$

$1 \quad a \quad a^2 \quad a^3 \quad a^4 \quad a^5 \quad a^6 \quad a^7 \quad a^8 \quad a^9 \quad a^{10} \quad a^{11} \quad a^{12} \quad a^{13} \quad a^{14} \quad \ldots$

Γ is the set of all the pairs of profinite numbers $(dz^{\mathcal{P}}, dpz^{\mathcal{P}})$ s.t.:

- $\mathcal{P}$ is a cofinite sequence of prime numbers $\{p_1, p_2, \ldots\}$
- $z^{\mathcal{P}} = \lim_n (p_1 p_2 \ldots p_n)^{n!}$
- $p \in \mathcal{P}$
- if q divides d then $q \notin \mathcal{P}$

$$x^{\alpha} \leftrightarrow x^{\beta} \text{ for all } (\alpha, \beta) \in \Gamma \qquad (E_7)$$

Conclusion and further questions

- Equational description of classes of rational languages generated by the languages u^*
- Decidability

- Variety of languages generated by the languages u^*
- Classes generated by F^* for F a finite set of words