Some topological exercises around a Boolean algebra

Mai Gehrke

CNRS and Paris Diderot

Recognisable languages

Let A be an alphabet, $L \subseteq A^*$ a language.

The following conditions are equivalent:

- L is recognisable by an automaton;
- L is recognisable by a finite monoid;
- L is given by a rational expression;
- $L = \mathrm{Mod}(\varphi)$ for some φ In $\mathsf{MSO}[\leqslant, (\underline{a})_{a \in A}]$

For this talk, I want to explain the last formulation which depends on Büchi's logic on words

Logic on words

To each non-empty word u is associated a structure

$$\mathcal{M}_u = (\{1, 2, \ldots, |u|\}, <, (\mathbf{a})_{a \in A})$$

where **a** is interpreted as the set of integers i such that the i-th letter of u is an a, and $<$ as the usual order on integers.

Example:

Let $u = abbaab$ then

$$\mathcal{M}_u = (\{1, 2, 3, 4, 5, 6\}, <, (\mathbf{a}, \mathbf{b}))$$

where $\mathbf{a} = \{1, 4, 5\}$ and $\mathbf{b} = \{2, 3, 6\}$.

Some examples

The formula $\varphi = \exists x\ \mathbf{a}x$ interprets as:

> *There exists a position x in u such that the letter in position x is an a.*

This defines the language $L(\varphi) = A^* a A^*$.

The formula $\exists x\ \exists y\ (x < y) \wedge \mathbf{a}x \wedge \mathbf{b}y$ defines the language $A^* a A^* b A^*$.

The formula $\exists x\ \forall y\ [(x < y) \vee (x = y)] \wedge \mathbf{a}x$ defines the language aA^*.

Defining the set of words of even length

Macros:

$$(x < y) \vee (x = y) \quad \text{means } x \leqslant y$$

$$\forall y \; x \leqslant y \quad \text{means } x = 1$$

$$\forall y \; y \leqslant x \quad \text{means } x = |u|$$

$$x < y \;\wedge\; \forall z \; (x < z \to y \leqslant z) \quad \text{means } y = x + 1$$

Let $\varphi \;=\; \exists X \;\; (1 \notin X \;\wedge\; |u| \in X \;\wedge\; \forall x \; (x \in X \leftrightarrow x + 1 \notin X))$

Then $1 \notin X$, $2 \in X$, $3 \notin X$, $4 \in X$, $\ldots$, $|u| \in X$. Thus

$$L(\varphi) = \{u \mid |u| \text{ is even}\} = (A^2)^*$$

This language is often called PARITY.

Monadic second order logic

Only second order quantifiers over unary predicates are allowed.

Theorem: (Büchi '60, Elgot '61)

Monadic second order captures exactly the recognisable languages.

This is written as the equation

$$\mathrm{Rec}(A^*) = MSO[\leqslant, (\underline{a})_{a \in A}]$$

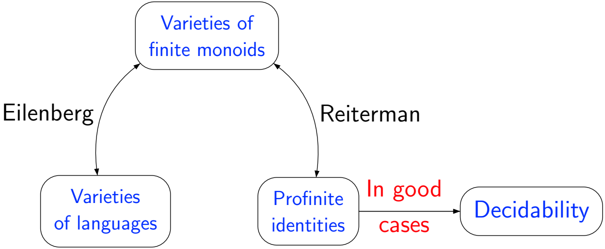
Basic problems in complexity theory

In complexity theory computing machines are studied, e.g., through corresponding formal languages

Typical problems that are studied are:

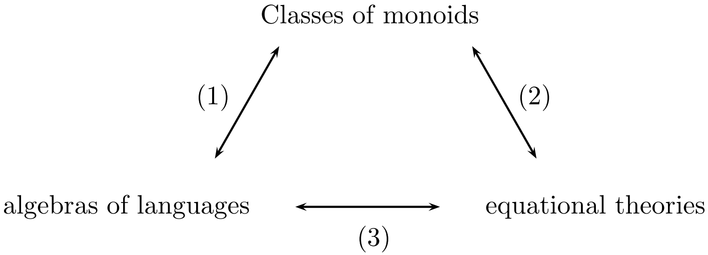
- decidability of membership in a class of languages
- separation of complexity classes
- comparison of complexity classes

Eilenberg-Reiterman theory



Various generalisations: [Pin 1995], [Pin-Weil 1996], [Pippenger 1997], [Polák 2001], [Esik 2002], [Straubing 2002], [Kunc 2003]

Eilenberg, Reiterman, and Stone



(1) Eilenberg theorems
(2) Reiterman theorems
(3) extended Stone/Priestley duality

(3) allows generalisation to non-varieties and even to non-regular languages

Most general form of the Eilenberg-Reiterman theorem

Lattices of recognisable languages are given by profinite equations

This is a spacial case of the duality between

subalgebras $\longleftrightarrow$ quotient structures

$$B \hookrightarrow Rec(A^*)$$

dually

$$X_B \twoheadleftarrow \widehat{A^*}$$

That is, B is described dually by equating elements of $\widehat{A^*}$.

A Galois connection for subalgebras and quotient spaces

Let B be a Boolean algebra, X the dual space of B.

The maps $\mathcal{P}(B) \leftrightarrows \mathcal{P}(X \times X)$ given by

$$S \mapsto \approx_S = \{(x, y) \in X \mid \forall b \in S \quad (b \in y \iff b \in x)\}$$

and

$$E \mapsto B_E = \{b \in B \mid \forall (x, y) \in E \quad (b \in y \iff b \in x)\}$$

establish a Galois connection whose Galois closed sets are the Boolean equivalence relations and the Boolean subalgebras, respectively.

Example: the star free languages

The star free languages are those recognisable languages that are generated by $\{a\}$ for $a \in A$ using the Boolean operations and concatenation product

In logic terms,

$$\text{Star free} = \mathrm{FO}[\leqslant, (\underline{a})_{a \in A}]$$

Schützenberger-Simon theorem

$$\text{Star free} = [\![\, x^{\omega+1} = x^{\omega}]\!]$$

Here x^ω is the unique idempotent in the closed subsemigroup generated by x, and the theorem means that the class of star free languages is given by the one pair, $(x^{\omega+1}, x^\omega)$, when closing under:

- substitution
- monoid congruence
- Stone duality subalgebra-quotient adjunction

That is the class of star free languages is B_E where

$$E = \{(ux^{\omega+1}v, ux^{\omega}v) \mid x, u, v \in \widehat{A^*}\}$$

Beyond recognisable languages

$$B \hookrightarrow \mathcal{P}(A^*)$$

dually

$$X_B \twoheadleftarrow \beta(A^*)$$

That is, lattices of languages are given by "β-equations"

A case with some handle

Joint work with Andreas Krebs and Jean-Éric Pin. Idea of the project: start with a relatively small lattice for which some connection with $Rec(A^*)$ is known

AC^0 consists of all families of circuits of bounded depth and polynomial size, with negation on inputs and unlimited fanin AND and OR gates
$= FO[\mathcal{N}, (\underline{a})_{a \in A}]$ where $\mathcal{N}$ is the class of all predicates on $\mathbb{N}$

By a deep result of Barrington, Straubing, and Thérien

$$FO[\mathcal{N}, (\underline{a})_{a \in A}] \cap \mathrm{Rec}(A^*) = [\![\, (x^{\omega-1}y)^{\omega+1} = (x^{\omega-1}y)^{\omega} \text{ for } x, y \text{ words of the same length }]\!]$$

An even simpler case

We start by investigating the fragment given by nullary and unary numerical predicates (in FO without equality)

$$\mathcal{B} = FO[\mathcal{N}_1, \mathcal{N}_0, (\underline{a})_{a \in A}]$$

$$= <L_P^a, L_P \mid a \in A, P \subseteq \mathbb{N}>_{BA}$$

where

$$L_P^a = \{u \in A^* \mid u_i = a \implies i \in P\}$$

$$L_P = \{u \in A^* \mid |u| \in P\}$$

Problem: Find $E \subseteq \beta(A^*) \times \beta(A^*)$ so that $B_E = \mathcal{B}$

Dual space of $\mathcal{B}$

It is not necessary to compute the dual of $\mathcal{B}$, but, when this is possible it tends to be useful in language theory

In addition, we thought it might help us in the difficult task of coming up with a method for picking pairs in $\beta(A^*)$

Even though $\mathcal{B}$ is quite small and simple, computing its ultrafilters directly is not easy

To solve this problem we have devised a method based on duality which I think is interesting in its own right

Some observations

By Priestley (Nerode), it suffices to compute the dual of a sublattice of $\mathcal{B}$ which generates $\mathcal{B}$ as a Boolean algebra

We pick

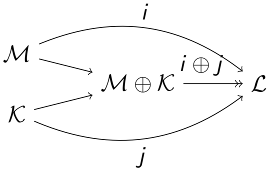
$$\mathcal{L} = < L_P^a, L_P \mid a \in A, P \subseteq \mathbb{N} >_{DL}$$

Let $\mathcal{L}_a = < L_P^a \mid P \subseteq \mathbb{N} >_{DL}$ and $\mathcal{K} = < L_P \mid P \subseteq \mathbb{N} >_{DL}$, then

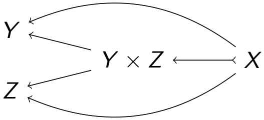
$$\mathcal{L} = (\bigvee_{a \in A} \mathcal{L}_a) \vee \mathcal{K}$$

The dual space of the join of two lattices I

If $i : \mathcal{K} \to \mathcal{L}$ and $j : \mathcal{M} \to \mathcal{L}$ are sublattices with $\mathcal{L} = \mathcal{K} \vee \mathcal{M}$ then by the universal property of coproducts, we have the following diagram:



The map $i \oplus j$ is surjective because the union of $\mathcal{M}$ and $\mathcal{K}$ generates $\mathcal{L}$. Accordingly, by duality, we obtain the following diagram:



The dual space of the join of two lattices II

Let $i : \mathcal{K} \to \mathcal{L}$ and $j : \mathcal{M} \to \mathcal{L}$ be sublattices with $\mathcal{L} = \mathcal{K} \vee \mathcal{M}$, and let X, Y, and Z be the dual spaces of $\mathcal{L}, \mathcal{M}$, and $\mathcal{K}$, respectively.

Then X is the (closed) subspace of $Y \times Z$ consisting of the points (y, z) satisfying, for all $U_1, U_2 \in \mathcal{M} \oplus \mathcal{K}$

$$(i \oplus j)(U_1) \leqslant (i \oplus j)(U_2) \implies ((y,z) \in \widehat{U}_1 \implies (y,z) \in \widehat{U}_2)$$

or equivalently

$$[(y,z) \in \widehat{U}_1 \ \text{ and } \ (i \oplus j)(U_1) \leqslant (i \oplus j)(U_2)] \implies (y,z) \in \widehat{U}_2$$

That is, for all $M, M_1, \ldots, M_k \in \mathcal{M}$ and $K, K_1, \ldots, K_k \in \mathcal{M}$

$$[M \in y, K \in z, \ \text{ and } \ M \cap K \subseteq \bigcup_{i=1}^{k} (M_i \cap K_i)] \implies \exists i (M_i \in y, K_i \in z)$$

The duals of the $\mathcal{L}_a$s

Recall $\mathcal{L}_a = \langle L^a_P \subseteq A^* \mid P \subseteq \mathbb{N} \rangle_{DL}$

Theorem: The dual space of $\mathcal{L}_a$ is (homeomorphic) to $Filt(\mathcal{P}(\mathbb{N}))$ with the topology generated by the sets

$$\widehat{P} = \{F \in Filt(\mathcal{P}(\mathbb{N})) \mid P \in F\}$$

and the Stone embedding given by $L^a_P \mapsto \widehat{P}$

Consider $c_a \colon A^* \to \mathcal{P}(\mathbb{N}), u \mapsto \{i \in \mathbb{N} \mid u_i = a\}$, then $L^a_P = c_a^{-1}(P)$

We may consider $c_a \colon A^* \to \mathcal{V}(\beta(\mathbb{N}))$ and the unique extension $\beta(c_a) \colon \beta(A^*) \twoheadrightarrow \mathcal{V}(\beta(\mathbb{N}))$ is then the dual of $\mathcal{L}_a \to \mathcal{P}(A^*)$

Putting $\mathcal{L}_a$s together

Let $B \subsetneq A$ with $|B| = m$. For $F \in Filt(\mathcal{P}(\mathbb{N})) = X$, define

$$C(F) = \bigcap F$$

<u>Theorem</u>: The dual space of $\mathcal{B}_B = \bigvee_{b \in B}$ consists of all those $\overline{F} = (F_1, \ldots, F_m) \in X^m$ such that

the sets $\{C(F_i)\}_{i=1}^m$ are pairwise disjoint

<u>Theorem</u>: Let $\mathcal{L}_A = \bigvee_{a \in A} \mathcal{L}_a$ and $X = \mathcal{V}(\beta(\mathbb{N}))$. Denote by X_A the dual of $\mathcal{B}_A$ viewed as a subspace of $X^{|A|}$. For $\overline{F} \in X^{|A|}$, we have $\overline{F} \in X_A$ if and only if *either one* of the following two conditions is satisfied:

1. Each $F_i = \uparrow P_i$ and $\{P_i\}_{i=1}^{|A|}$ is a decomposition of $\downarrow n$ for some $n \in \mathbb{N}$
2. $\{C(F_i)\}_{i=1}^{|A|}$ is a decomposition of $\mathbb{N}$.

That is,

$$X_A = \{\overline{F}_w \mid w \in A^*\} \cup \{\overline{F} \mid \{C(F_i)\}_{i=1}^{|A|} \text{is a decomposition of } \mathbb{N}\}$$

The dual of $\mathcal{L}$

Recall $\mathcal{K} = <L_P \subseteq A^* \mid P \subseteq \mathbb{N}>_{DL}$. It is easy to see that $\mathcal{K} \cong \mathcal{P}(\mathbb{N})$ and thus its dual space is $\beta(\mathbb{N})$

Putting together $\mathcal{L}_A$ and $\mathcal{K}$ we get

Theorem: The dual space of $\mathcal{B}$ is the subspace of

$$\mathcal{V}(\beta(\mathbb{N}))^{|A|} \times \beta(\mathbb{N})$$

given by

$$X = \{(\overline{F}_w, \uparrow|w|) \mid w \in A^*\}$$
$$\cup \{(\overline{F}, \mu) \mid \mu \in \beta(\mathbb{N}) - \mathbb{N} \text{ and } \{C(F_i)\}_{i=1}^{|A|} \text{is a decomposition of } \mathbb{N}\}$$

Equations for $\mathcal{B}$

Intuition: If in a set of spots, both as and bs are allowed, then the Boolean algebra $\mathcal{B}$ can't count how many of each there are, nor can it say which order they are in

We make equations expressing this fact in the following way:

To be continued on the whiteboard!

References

The following are the papers from which the research discussed at the two talks I gave at AAA88 came:

1. Mai Gehrke, Serge Grigorieff, and Jean-Éric Pin, Duality and Equational Theory of Regular Languages, *LNCS* (ICALP) **5125** (2008), 246–257.
2. Mai Gehrke, Stone duality, topological algebra, and recognition, preprint. See, `http://hal.archives-ouvertes.fr/hal-00859717`
3. Mai Gehrke, Andreas Krebs, and Jean-Éric Pin, From ultrafilters on words to the expressive power of a fragment of logic, to appear in *Proceedings of the 16th International Workshop on Descriptional Complexity of Formal Systems*, 2014.