# Stone duality in the theory of formal languages

Mai Gehrke[1]

CNRS, Paris[2], France

[2]Nice as of 01/07/2017

[1]The research discussed has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No.670624)

# Outline

1. Automata theory, BAOs, and Jónsson-Tarski
2. Equational theories for Boolean algebras of languages
3. Beyond regular languages
   - Boolean circuit complexity
   - Logic on words
   - Equations for circuit classes
4. Boolean spaces with internal monoids
5. Adding one layer of FO existential quantifiers
6. Luca Reggio's talk on Tuesday

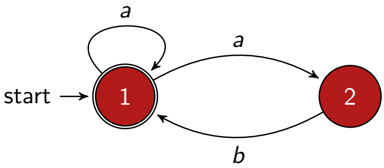
This talk concerns joint work with Bjarni Jónsson, Jean-Éric Pin, Serge Grigorieff, Andreas Krebs, Daniela Petrişan, Luca Reggio, Célia Borlido, Silke Czarnetski

# 1999 NMSU Holiday Mathematics Symposium
# Algebraic Structures for Logic


Summer 1998 at Bjarni and Harriett's cabin in Backus MN

# 1 – Automata theory

An automaton $\mathcal{A}$ is a set $Q$ of states with a non-deterministic right action $Q \times A^* \to \mathcal{P}(Q)$, a set of initial states $I \subseteq Q$, and a set of final states $F \subseteq Q$




$$L(\mathcal{A}) = \{w \in A^* \mid Iw \cap F \neq \emptyset\} = (a^*(ab)^*)^*$$

Computational difficulty is studied via language classes, e.g.

$$\mathrm{Reg}_A = \{L \subseteq A^* \mid \exists \mathcal{A}\ L = L(\mathcal{A})\} \subseteq \mathcal{P}(A^*)$$

Notice this is a Boolean algebra with operators (operations) (BAOs)

$$KL \subseteq M \iff L \subseteq K \backslash M \iff K \subseteq M/L$$

# 1 – Jónsson-Tarski 1951 and 1952 on BAOs

Canonical extension, Jónsson-Tarski duality, and canonicity for positive varieties led to my contribution to [G-Grigorieff-Pin 2008] and [G 2016]:
a duality between algebras and algebras

$$\left(\begin{array}{c}\text{Certain BAs}\\ \text{with residuation}\\ \text{operations}\end{array}\right) \leftrightsquigarrow \left(\begin{array}{c}\text{Topological}\\ \text{algebras based on}\\ \text{Boolean spaces}\end{array}\right)$$

and generalised Eilenberg-Reiterman theory given by duality

$$\left(\begin{array}{c}\text{BAO (DLO) subalgebras}\\ \text{of a given BA}\end{array}\right) \leftrightsquigarrow \left(\begin{array}{c}\text{(Ordered) quotients}\\ \text{of the dual of the BA}\end{array}\right)$$

In particular, $(\mathrm{Reg}_A, \backslash, /)$ is dual to $\widehat{A^*}$, the profinite completion of $A^*$

# 2 – Equational theories for BAs of languages

Equations arise from the duality between Boolean subalgebras and quotient spaces

$$\mathcal{C} \hookrightarrow \mathcal{B} \qquad \leftrightsquigarrow \qquad X_{\mathcal{C}} \twoheadrightarrow X_{\mathcal{B}}$$

For $x, y \in X_{\mathcal{B}}$ and $L \in \mathcal{B}$ define

$$L \models x \approx y \quad \text{iff} \quad L \in \mu_x \iff L \in \mu_y \quad \text{iff} \quad x \in \widehat{L} \iff y \in \widehat{L}$$

Then we get a Galois connection $\mathcal{P}(\mathcal{B}) \leftrightarrows \mathcal{P}(X_{\mathcal{B}} \times X_{\mathcal{B}})$ given by

$$\mathrm{Eq}(\mathcal{S}) = \{(x, y) \mid \forall L \in \mathcal{S} \quad L \models x \approx y\} \text{ for } \mathcal{S} \subseteq \mathcal{B}$$
$$\mathrm{Mod}(\Sigma) = \{L \mid \forall (x, y) \in \Sigma \quad L \models x \approx y\} \text{ for } \Sigma \subseteq X_{\mathcal{B}} \times X_{\mathcal{B}}$$

Theorem: The Galois closed sets are, respectively, the Boolean subalgebras of $\mathcal{B}$ and the Boolean equivalence relations on $X_{\mathcal{B}}$

For $\mathcal{B} = \mathrm{Reg}_A$ and $X_{\mathcal{B}} = \widehat{A^*}$, the equations are said to be profinite and the theorem generalises the Eilenberg-Reiterman theory

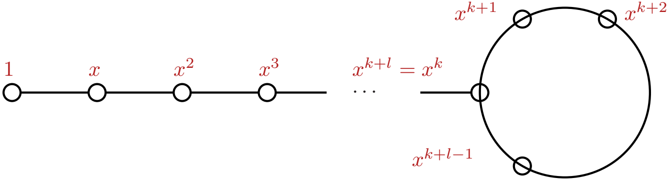
# 2 – Example: profinite equations for the star-free languages

A language is Star-free provided it is in the BA closed under concatenation product generated by the singletons

[Schützenberger 1965] and [McNaughton-Papert 1971]

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

Here $x^{\omega}$ is a profinite term giving the unique idempotent in the principal closed subsemigroup generated by $x$




NB! This makes star-freeness decidable

# 3 – Beyond regular languages: Boolean circuit complexity

Boolean circuit classes have members that are specified by sequences of Boolean circuits, one for each input length, that identify which words of the given length are accepted.

For example, $\mathrm{AC}^0$ consists of *constant depth* and polynomial size circuit sequences and $\mathrm{ACC}^0$, is obtained from $\mathrm{AC}^0$ by adding modular gates.

[Furst, Saxe, and Sipser 1981] separated these two classes:

$$\mathrm{NP} \geqslant \mathrm{P} \geqslant \mathrm{AC} \ldots \geqslant \mathrm{AC}^2 \geqslant \mathrm{AC}^1 \geqslant \mathrm{NL} \geqslant \mathrm{L} \geqslant \mathrm{ACC}^0 > \mathrm{AC}^0$$

by showing that the regular language

$$\text{PARITY} = \{w \in \{0,1\}^* \mid w \text{ has an odd number of 1's}\}$$

is not in $\mathrm{AC}^0$

# 3 – Logic on words

To each non-empty word $w$ is associated a structure

$$(\{0, 1, \ldots, |w| - 1\}, (\mathbf{a}^w)_{a \in A}) \text{ where } \mathbf{a}^w = \{i < |w| \mid w_i = a\}$$

In addition, these structures inherit any predicates on $\mathbb{N}$ by restriction (numerical predicates).

Theorem: [Büchi 1960; Elgot '61; Trakhtenbrot '62] $\mathbf{MSO}[\leqslant] = \mathrm{Reg}$

Meaning that the model classes of monadic second order sentences in the language augmented by $\leqslant$ are precisely the languages recognisable by automata

Theorem: [McNaughton-Papert 1971] $\mathbf{FO}[\leqslant] = \text{Star-Free}$

(Star-Free = languages generated by the singleton letters using the Boolean operations and binary concatenation)

# 3 – Logic on words for circuit classes

As with classes of regular languages, many computational complexity classes have been given characterisations as model classes of appropriate logic fragments on finite words [Immerman 1999]

For example,

$$\mathrm{AC}^0 = \mathbf{FO}[\mathcal{N}] \quad \mathrm{ACC}^0 = (\mathbf{FO} + \mathbf{MOD})[\mathcal{N}] \quad \mathrm{TC}^0 = \mathbf{MAJ}[\mathcal{N}]$$

$\mathcal{N}$ = all predicates on the positions of a word
**FO** = first-order logic
**MOD** and **MAJ** = modular and majority quantifiers, respectively.

The presence of arbitrary (numerical) predicates, and of the majority quantifier is what brings one far beyond the scope of the profinite algebraic theory of regular languages.

# 3 – Connection to algebraic automata theory

PARITY is a regular language, so the separation result of Furst, Saxe, and Sipser is witnessed at this level.

$$
\begin{aligned}
\mathbf{FO}[\mathcal{N}] \cap \mathrm{Reg} &= \text{languages given by quasi-aperiodic stamps} \\
&= [\![ \, (x^{\omega-1}y)^{\omega+1} = (x^{\omega-1}y)^{\omega} \\
&\qquad\qquad \text{for } x, y \text{ words of the same length} ]\!]
\end{aligned}
$$

[Barington, Compton, Straubing, Thérien 1992]
[Kunc 2003]

In the regular setting, the algebraic theory of monoids, including decomposition results in terms of semidirect products, plays a central rôle.
We want to generalise the algebraic theory to treat classes of languages that are not necessarily regular

# The dual space of a powerset

Let $S$ be an infinite set, then $\mathcal{P}(S)$ is a Boolean algebra

$\mathrm{St}(\mathcal{P}(S))$:

- (principal filters) For each $s \in S$

$$\mu_s = \{T \subseteq S \mid s \in T\} \text{ is an ultrafilter of } \mathcal{P}(S)$$

- (free ultrafilters) All ultrafilters extending the Frechet filter

$$\mathcal{F} = \{T \subseteq S \mid S - T \text{ is finite}\}$$

(these are all non-constructive)

Theorem: $S \hookrightarrow \mathrm{St}(\mathcal{P}(S))$ is the Stone-Čech compactification of $S$ equipped with the discrete topology

We denote it $\beta(S)$

# 4 – The rôle of monoids

The reason monoids enter the picture, is that most classes of interest are closed under the quotient operations, that is, if $L \subseteq A^*$ is in the class, then all of

$\mathcal{B}(L)$= the BA generated by the languages $u^{-1}Lv^{-1}$ for $u, v \in A^*$

is contained in the class, where

$$u^{-1}Lv^{-1} = \{w \in A^* \mid uwv \in L\}$$

This is a Bi-action of $A^*$ on $\mathcal{B}(L)$, $\Gamma_{uv} : K \mapsto u^{-1}Kv^{-1}$, for all $u, v \in A^*$

Duality gives us

$$\begin{tikzcd}
\mathcal{P}(A^*) \arrow[r, "u^{-1}(_)v^{-1}"] & \mathcal{P}(A^*) \\
\mathcal{B}(L) \arrow[u, hook] \arrow[r, dashed, "\Gamma_{uv}"] & \mathcal{B}(L) \arrow[u, hook]
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
\beta(A^*) \arrow[r, "\beta(u\,(_)\,v)"] \arrow[d] & \beta(A^*) \arrow[d] \\
X_L \arrow[r, dashed, "\gamma_{uv}"] & X_L
\end{tikzcd}$$

# The syntactic space of a language

For $L \subseteq A^*$, let

$$\mathcal{B}(L) := \langle u^{-1}Lv^{-1} \mid u, v \in A^* \rangle_{\mathrm{BA}}$$

Since the embedding $\mathcal{B}(L) \hookrightarrow \mathcal{P}(A^*)$ preserves the bi-action of $A^*$, dually, we obtain

$$\begin{tikzcd}
A^* \arrow[r, hook] \arrow[d, dashed, "\psi_L"] & \beta(A^*) \arrow[r, "\beta(u\,(_)\,v)"] \arrow[d] & \beta(A^*) \arrow[d] \\
M_L \arrow[r, hook, dashed] & X_L \arrow[r, "\gamma_{uv}"] & X_L
\end{tikzcd}$$

where $M_L$ is the image of $A^*$ in $X_L$.

It is not hard to see that since the quotient map is a morphism of biactions, $M_L$ carries a monoid structure. It is what is known in language theory as the syntactic monoid of $L$

# 4 – Boolean spaces with internal monoids

Let $L \subseteq A^*$. Then $\iota\colon M_L \hookrightarrow X_L$ satisfies:

- $X_L$ is a Boolean Stone space
- $M_L$ is a monoid
- $X_L$ is equipped with a continuous bi-action of $M_L$
- The map $\iota$ satisfies:
  - $\iota$ is injective
  - the image of $\iota$ is dense in $X_L$
  - $\iota$ is a morphism of sets with bi-actions of $M_L$

We denote such an object by $(X_L, M_L)$ and call it a Boolean space with an internal monoid or BM (or BiM) for short

NB! Sometimes it is convenient to drop the requirement that $\iota$ is injective in the definition of BMs

# 4 – Recognition

Let $(X, M)$ be a BM, a BM morphism

$$\psi\colon (\beta(A^*), A^*) \to (X, M)$$

is a commuting diagram

$$\begin{tikzcd}
A^* \arrow[r, hook] \arrow[d, "\psi"] & \beta(A^*) \arrow[d, "\psi'"] \\
M \arrow[r, hook, "\iota"] & X
\end{tikzcd}$$

NB! $\psi'$ is uniquely determined by $\psi$

Now $\psi$ recognises $L \subseteq A^*$ provided $(\iota \circ \psi)^{-1}(C) = L$ for some clopen $C \subseteq X$

<u>Theorem:</u> [G-Petrișan-Reggio 2016]
$\psi$ recognises $L$ iff $\psi_L\colon (\beta(A^*), A^*) \to (X_L, M_L)$ factors through $\psi$

# 4 – Example: The syntactic space of MAJORITY

Let $A = \{a, b\}$ and $L = \{w \in A^* \mid |w|_a > |w|_b\}$ where $|w|_a$ is the number of $a$'s in $w$. Then

$$h_L \colon A^* \longrightarrow \mathbb{Z}, w \mapsto |w|_a - |w|_b$$

is the syntactic morphism of $L$ and $\mathbb{Z}^+$ is the syntactic image, i.e. $L = h_L^{-1}(\mathbb{Z}^+)$ and

$$\begin{aligned}\mathcal{B}(L) &\cong \langle \mathbb{Z}^+ + k \mid k \in \mathbb{Z}\rangle_{\mathrm{BA}} \\ &= \{K \mid K \cap \mathbb{Z}^+, K \cap \mathbb{Z}^- \text{ are each finite or cofinite}\}\end{aligned}$$

The dual space of $\mathcal{B}(L)$ is $\mathbb{Z}_{-\infty}^{+\infty} = \mathbb{Z} \cup \{-\infty, +\infty\}$, where

$$\mu_{+\infty} = \{K \mid K \triangle \mathbb{Z}^+ \text{ is finite}\} \quad \text{and} \quad \mu_{-\infty} = \{K \mid K \triangle \mathbb{Z}^- \text{ is finite}\}$$

with topology making it the ‘two point compactification’ of $\mathbb{Z}$

# Equations for MAJORITY

The dual of $e\colon \mathcal{B}(L) \hookrightarrow \mathcal{P}(\mathbb{Z})$ is

$$\beta e\colon \beta(\mathbb{Z}) \longrightarrow \mathbb{Z}_{-\infty}^{+\infty}, \; w \mapsto w \text{ for } w \in A^*$$

For $\mu \in \beta(\mathbb{Z}) - \mathbb{Z}$

$$\mu \mapsto \begin{cases} +\infty & \text{if MAJORITY} \in \mu \\ -\infty & \text{otherwise} \end{cases}$$

Proposition: $\mathcal{B}(L)$ is characterised relative to $\mathcal{P}(\mathbb{Z})$ by the equations

$$\Sigma = \{\mu \approx \mu + 1 \mid \mu \in \beta(\mathbb{Z}) - \mathbb{Z}\}$$

For a proof, see the complexity column of SIGLOG News, April 2017 (a survey article written jointly with Andreas Krebs)

# 5 – Adding a layer of existential quantifier

Let $\varphi(x)$ be a formula of the logic on words with one free variable

<u>Problem:</u> Given a recogniser for $L = \mathrm{Mod}(\varphi(x))$, construct a recogniser for $L_\exists = \mathrm{Mod}(\exists x \varphi(x))$

NB! $L$ consists of $x$-models based on words, i.e., elements of

$$A^* \otimes \mathbb{N} = \{(w, i) \mid w \in A^* \text{ and } i \leqslant |w|\}$$

We can embed these $x$-models in the free monoid $(A \times 2)^*$ via

$$(w, i) \mapsto w^i \text{ given by } (w^i)_j = \begin{cases} (w_j, 0) & \text{if } j \neq i \\ (w_j, 1) & \text{if } j = i \end{cases}$$

We say that $\psi \colon (\beta((A \times 2)^*), (A \times 2)^*) \to (X, M)$ recognises $\mathrm{Mod}(\varphi(x))$ if it is the preimage of a clopen of $X$ under the composition

$$A^* \otimes \mathbb{N} \hookrightarrow (A \times 2)^* \xrightarrow{\psi} M \hookrightarrow X$$

# The Vietoris space

Let $\mathcal{V}(X)$ be the Vietoris space of $X$. That is,

$$\mathcal{V}(X) = \{C \subseteq X \mid C \text{ is closed in } X\}$$

with the topology generated by

$$\Diamond U = \{C \mid C \cap U \neq \emptyset\} \text{ and } \Box U = \{C \mid C \subseteq U\} \text{ for } U \in \mathcal{O}(X)$$

NB! $\mathcal{P}_{fin}(M) \hookrightarrow \mathcal{V}(X)$ with dense image

$\mathcal{V}(X)$ recognises the quantified languages, but not as a monoid

# A recogniser for $L_\exists$ from one for $L$

Definition: $\Diamond(X, M) = (\mathcal{V}(X) \times X, \mathcal{P}_{fin}(M) \times M)$ with left action

$$\begin{aligned}(F, m)(C, x) &= (Fx \cup mC, mx) \\ &= (\{m'x \mid m' \in F\} \cup \{mx' \mid x' \in C\}, mx)\end{aligned}$$

Theorem: [G-Petrișan-Reggio 2016]
If $\psi\colon (\beta((A \times 2)^*), (A \times 2)^*) \to (X, M)$ recognises $\mathrm{Mod}(\varphi(x))$, then

$$\begin{aligned}\Diamond\psi\colon (\beta(A^*), A^*) &\to \Diamond(X, M) \\ w &\mapsto (\{\psi(w^i) \mid i \leqslant |w|\}, \psi(w))\end{aligned}$$

recognises $\mathrm{Mod}(\exists\varphi(x))$

Go to Luca's talk on Tuesday for more on a generalisation of this!


Back in Backus