Separation, duality, and profinite lambda-calculus

Sam van Gool

IRIF, Université Paris Cité

21 February 2025

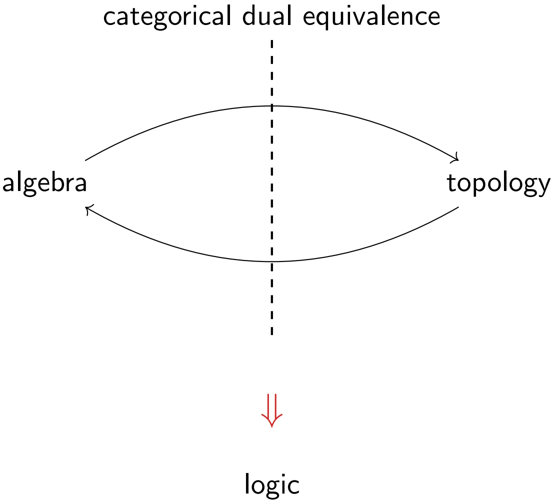
Plan

In this talk, I will present work on two recent projects:

1. separation problems in automata theory,
2. profinite semantics of lambda-calculus.

I will show how they both use a common methodology, based on topological duality for distributive lattices.

The duality approach



Overview

Part 1. Separability and profinite words

- Separability
- Profinite words and duality
- Computing pointlike sets

Part 2. Duality and profinite λ-calculus

- Definability and invariance
- Church encoding: words as λ-terms
- Regularity categorically

Conclusion

Overview

Part 1. Separability and profinite words

Separability

Profinite words and duality

Computing pointlike sets

Part 2. Duality and profinite λ-calculus

Definability and invariance

Church encoding: words as λ-terms

Regularity categorically

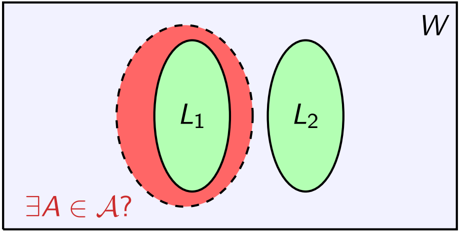
Conclusion

Separability

Let W be a set (e.g., $W = \Sigma^*$), let $\mathcal{R}$ and $\mathcal{A}$ be subsets of $\mathcal{P}(W)$.

Two sets L_1 and L_2 in $\mathcal{R}$ are called $\mathcal{A}$-separable if

there exists $A \in \mathcal{A}$ such that $L_1 \subseteq A$ and $L_2 \cap A = \emptyset$.



Note: Separability generalizes membership.

Example: $\mathcal{A}$:= FO-definable languages, $\mathcal{R}$:= regular languages.

→ Separability decidable: Henckell; Almeida; Place & Zeitoun.

Results

I will describe the following results about separation:

Theorem (v.G. & Steinberg, 2019)

FO[Mod]-*separability is decidable for regular languages of finite words.*

Here, FO[Mod] is the extension of first order logic with quantifiers of the form: 'there exist r mod p positions in the word such that'.

Theorem (Colcombet, v.G. & Morvan 2022)

FO-*separability is decidable for regular languages of ordinal words.*

Regular languages and finite monoids

Regular languages admit an algebraic approach:

Theorem (Myhill & Nerode)

A language $L \subseteq \Sigma^*$ *is regular if, and only if, there exist a finite monoid* M *and a homomorphism* $\varphi\colon \Sigma^* \to M$ *such that*

$$L = \varphi^{-1}(\varphi(L)).$$

The syntactic monoid of a regular language L is the minimal such.

Classical algorithms compute the syntactic monoid for a regular language L, given a regular expression or an automaton for L.

Decidable characterization of FO-membership

Theorem (Schützenberger; McNaughton & Papert)

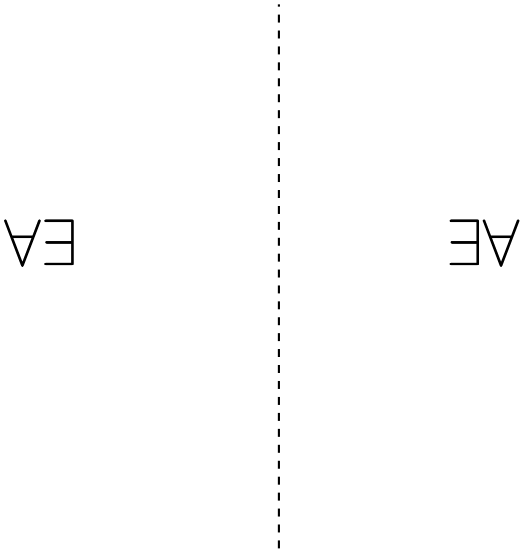
A regular language L is first-order definable if, and only if, the syntactic monoid is aperiodic, i.e., has no non-trivial subgroups.

In particular, FO-membership is decidable.

This prompted the development of a *variety theory* for regular languages and finite monoids, initially for deciding membership.

I will now show how this theory also applies for separation.

Separability reflected: intuition



Overview

Profinite words

Profinite words encode limiting behavior of finite words with respect to regular languages. Formally, they are the elements of $\widehat{\Sigma^*}$:

Proposition (Free profinite monoid over Σ)

There exists a unique topological monoid

$$\widehat{\Sigma^*}$$

which contains Σ^* as a dense submonoid, and such that, for any

$$\varphi\colon \Sigma^* \to M \quad \text{homomorphism, } M \text{ a finite monoid,}$$

there is a continuous homomorphism

$$\widehat{\varphi}\colon \widehat{\Sigma^*} \to M$$

extending φ, i.e., $\widehat{\varphi}|_{\Sigma^*} = \varphi$.

Profinite words as parametric families

The topological monoid $\widehat{\Sigma^*}$ can be realized as follows.

Let Φ be the set of homomorphisms $\varphi\colon \Sigma^* \to M$, with M finite.

Definition

A parametric family is a tuple

$$w = (w_\varphi)_{\varphi \in \Phi} \in \prod_{\varphi \in \Phi} \mathrm{cod}(\varphi)$$

such that, for any homomorphism $h\colon M \to M'$ of finite monoids,

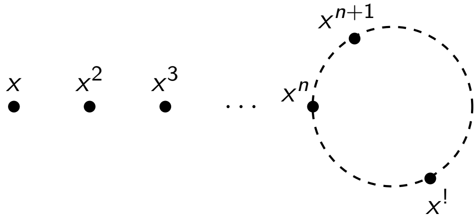
$$h(w_\varphi) = w_{h \circ \varphi} \ .$$

Fact

The topological monoid of parametric families is $\widehat{\Sigma^*}$.

Examples of profinite words

For any element x in a finite monoid M, the generated cyclic semigroup $\{x^n : n \geq 1\}$ contains a unique idempotent element, $x^!$.



For $w \in \Sigma^*$, the profinite word $w^!$ is the family $(\varphi(w)^!)_{\varphi \in \Phi}$.

When $\Sigma = \{a\}$, we have $\{a\}^* = \mathbb{N}$, and

$$\widehat{\{a\}^*} = \widehat{\mathbb{N}} \cong \mathbb{N} \uplus \widehat{\mathbb{Z}} = \mathbb{N} \uplus \prod_{p \text{ prime}} \mathbb{Z}_p,$$

where $\mathbb{Z}_p$ denotes the p-adic integers.

Duality: algebras of languages and quotients of $\widehat{\Sigma^*}$

For any Boolean algebra of languages $\mathcal{A} \leq \mathrm{Reg}(\Sigma^*)$, define the equivalence relation $\sim_{\mathcal{A}}$ on $\widehat{\Sigma^*}$ by:

$$w \sim_{\mathcal{A}} w' \overset{\text{def}}{\iff} \text{for all } A \in \mathcal{A},\ w \in \widehat{A} \text{ iff } w' \in \widehat{A}.$$

Theorem (Myhill-Nerode, profinite version)

For any $L \in \mathrm{Reg}(\Sigma^*)$,

$$L \in \mathcal{A} \qquad \textit{if, and only if,} \qquad \widehat{L} \textit{ is invariant under } \sim_{\mathcal{A}},$$

where $\widehat{L}$ *denotes the closure of* L *in* $\widehat{\Sigma^*}$.

Separability reflected: formal statement

Let $\mathcal{A} \leq \text{Reg}(\Sigma^*)$ be a Boolean algebra.

Proposition (Almeida, 1999)

For any regular Σ-languages L_1 and L_2:

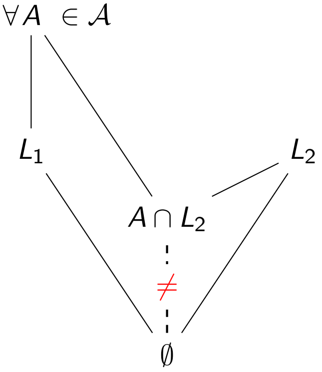
L_1 and L_2 are not $\mathcal{A}$-separable

if, and only if,

there exist profinite Σ-words $u_1 \in \widehat{L_1}$ and $u_2 \in \widehat{L_2}$ such that $u_1 \sim_{\mathcal{A}} u_2$.

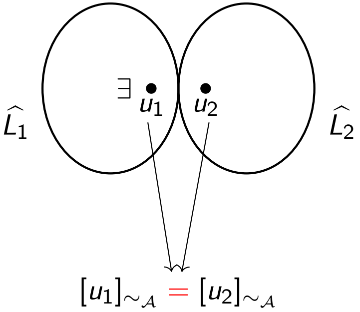
Separability reflected: intuition

Boolean algebra $\mathcal{A} \leq \mathrm{Reg}(\Sigma^*)$



non-$\mathcal{A}$-separable (L_1, L_2)

Profinite monoid $\widehat{\Sigma^*} \twoheadrightarrow \widehat{\Sigma^*}/\sim_{\mathcal{A}}$



$\mathcal{A}$-pointlike pair (u_1, u_2)

Overview

Pointlike sets

For any finite monoid M, we have the interpretation function $\pi_M \colon \widehat{M^*} \to M$, which sends a parametric family w to w_{id_M}.

Definition

A subset X of a finite monoid M is $\mathcal{A}$-pointlike if there exists a function $u \colon X \to \widehat{M^*}$ such that, for every $x, x' \in X$,

$$u(x) \sim_{\mathcal{A}} u(x') \quad \text{and} \quad \pi_M(u(x)) = x \ .$$

Almeida's Proposition $\Rightarrow$ Deciding $\mathcal{A}$-separability is equivalent to computing the $\mathcal{A}$-pointlike sets of size two in any finite monoid.

Separating excluded-subgroup languages

When $\mathbf{H}$ is a *variety of finite groups*, write $\bar{\mathbf{H}}$ for the class of finite monoids M such that every subgroup of M is in $\mathbf{H}$.

Theorem (v.G. & Steinberg, 2019)

For any decidable variety of finite groups $\mathbf{H}$*, the* $\bar{\mathbf{H}}$*-pointlike sets are computable.*

Consequences:

- For $\mathbf{H} = \{\mathbf{1}\}$, we recover: FO-separability is decidable.
- For $\mathbf{H}$ = solvable groups, decidability of FO[Mod]-separability.

FO-separation for ordinal words

An ordinal word over Σ is a function $\alpha \to \Sigma$, where α is a countable ordinal. Write Σ^{ord} for the set of ordinal words.

An ordinal monoid is a pair (M, π) where M is a set and $\pi \colon M^{\mathrm{ord}} \to M$ is an $(-)^{\mathrm{ord}}$-*algebra*.

Bedon, 1998: A finite ordinal monoid can be viewed as a tuple $(M, 1, \cdot, (-)^{\omega})$, subject to certain axioms.

This leads to a theory of regular languages of ordinal words.

Theorem (Colcombet, v.G. & Morvan 2022)

The FO-pointlike sets of a finite ordinal monoid are computable.
FO-separation is decidable for regular languages of ordinal words.

Computing pointlike sets: a recipe

Consider $\mathcal{A}$:= first-order definable languages.

In any finite monoid M:

- singletons are pointlike, subsets of pointlike sets are pointlike;
- pointlike sets are stable under product;
- for any subgroup G of the semigroup of pointlike sets of M, its union $\bigcup G$ is pointlike.

Thus, an under-approximation of the pointlike sets M is obtained by saturating under these rules, giving a subsemigroup S of $\mathcal{P}(M)$.

Crucial point to prove: S contains all the pointlike sets of M.

For any X not in S, one needs to construct FO-separators.

To adapt this to other $\mathcal{A}$, or to ordinal monoids, we need to modify the last saturation rule and the separator construction.

Overview

Profinite words and Stone duality

A different characterization of the free profinite monoid $\widehat{\Sigma^*}$:

Theorem (Gehrke, Grigorieff, Pin 2008)

The topological monoid $\widehat{\Sigma^*}$ is the Stone dual of the Boolean algebra of regular languages, with derivatives $L \mapsto a^{-1}L$.

That is:

- *points* of $\widehat{\Sigma^*}$ are ultrafilters of regular languages;
- the *clopen* subsets of $\widehat{\Sigma^*}$ are the sets $\widehat{L}$, for L regular;
- the *multiplication* of $\widehat{\Sigma^*}$ is a Kripke relation in the sense of modal and temporal logic, dual to the modality a^{-1}.

Morally: $\widehat{\Sigma^*}$ is the canonical frame for the modal logic of $\mathrm{Reg}(\Sigma^*)$.

Definability and invariance

By Stone-Priestley duality theory, the correspondence

$$\text{Reg}(\Sigma^*) \quad \longleftrightarrow \quad \widehat{\Sigma^*}$$

extends to an isomorphism

$$\text{sublattices of Reg}(\Sigma^*) \quad \longleftrightarrow \quad \text{ordered quotients of } \widehat{\Sigma^*} \ ,$$

giving an equivalence between definability of a regular language and invariance under profinite inequalities. (Gehrke, Grigorieff, Pin, 2008)

We work this out further and give applications in chapter 8 of:

Gehrke & v.G., *Topological duality for distributive lattices: Theory and applications.* 296pp. Cambridge University Press (2024).

Overview

Words and λ-terms

A finite word over an alphabet Σ:

$$w = a_1 a_2 \ldots a_n \quad (a_1, \ldots, a_n \in \Sigma)$$

represents a composition of basic actions on a set of states Q:

$$a_1 : Q \to Q, \ldots, a_n : Q \to Q \vdash w : Q \to Q$$

and may therefore be thought of as a λ-term:

$$W = \lambda a_1. \ \ldots \ \lambda a_n. \lambda q. \ a_n(\cdots a_2(a_1(q)) \cdots) \ .$$

This is a special case of the Church encoding of data structures (in this case, lists) as higher-order functionals.

Types

We work in a simply typed λ-calculus over a single base type $\mathbb{o}$, with one binary type constructor $\Rightarrow$.

For instance, the Church-encoded word

$$W = \lambda a_1. \ \dots \ \lambda a_n. \lambda q. \ a_n(\cdots a_2(a_1(q))\cdots)$$

has the following type:

$$\mathrm{Church}_\Sigma := \underbrace{(\mathbb{o} \Rightarrow \mathbb{o})}_{\text{type of } a_1} \Rightarrow \cdots \Rightarrow \underbrace{(\mathbb{o} \Rightarrow \mathbb{o})}_{\text{type of } a_n} \Rightarrow \underbrace{\mathbb{o}}_{\text{type of } q} \Rightarrow \mathbb{o} \ .$$

Words as equivalence classes of λ-terms

The Church encoding induces a monoid isomorphism

$$\Sigma^* \cong \{\lambda\text{-terms of type } \mathrm{Church}_\Sigma\}/\beta\eta \ ,$$

where $\beta\eta$ denotes the equivalence relation generated by the rules:

(β) $(\lambda x.f)M \approx f\ M$ and

(η) $\lambda x.fx \approx f$.

Overview

Categorical semantics of λ-calculus

Suppose that $\mathbf{C}$ is a category with finite products and exponentials:

$$\mathbf{C}(X \times Y, Z) \quad \cong \quad \mathbf{C}(Y, X \Rightarrow Z)\ .$$

Any object Q of $\mathbf{C}$ uniquely determines a model of our λ-calculus:

- each type A is interpreted as a $\mathbf{C}$-object $[\![A]\!]_Q$;
- each λ-term-in-context $x_1 : A_1, \ldots, x_n : A_n \vdash M : B$ is interpreted as a $\mathbf{C}$-morphism $[\![M]\!]_Q \colon A_1 \times \cdots \times A_n \to B$.

For instance, the type Church_Σ is interpreted as

$$[\![\mathrm{Church}_\Sigma]\!]_Q = (Q \Rightarrow Q)^\Sigma \Rightarrow (Q \Rightarrow Q)\ .$$

Examples of such *cartesian closed* categories $\mathbf{C}$:

Lam := types with λ-terms-in-context; **FinSet** := finite sets.

Runs are semantic evaluations

A run of a deterministic finite automaton on the word w becomes an evaluation of the corresponding λ-term W in the model **FinSet**:

$$\text{the interpretation } [\![W]\!]_Q \colon [\![\mathrm{Church}_\Sigma]\!]_Q$$

is

$$\begin{aligned} \mathsf{eval}(w) \colon (Q \Rightarrow Q)^\Sigma &\Rightarrow (Q \Rightarrow Q) \\ (\delta_\sigma)_{\sigma \in \Sigma} &\mapsto \delta_{a_n} \circ \cdots \circ \delta_{a_1} \end{aligned}$$

For fixed $\delta \in (Q \Rightarrow Q)^\Sigma$ and q_0, now letting w vary, we obtain

$$\mathsf{eval}_{\delta, q_0} = w \mapsto \mathsf{eval}(w)(\delta)(q_0) \colon \Sigma^* \to Q \ .$$

Observation. A language $L \subseteq \Sigma^*$ is regular if, and only if, there exist a finite set Q, $\delta \in (Q \to Q)^\Sigma$, $q_0 \in Q$, and $F \subseteq Q$ such that

$$\mathsf{eval}_{\delta, q_0}^{-1}(F) = L \ .$$

Regular languages of λ-terms

This re-framing of the concept of regular languages leads to:

Definition (Salvati)

For A a type, let $\Lambda(A)$ be the set of λ-terms of type A, up to $\beta\eta$.

A subset L of $\Lambda(A)$ is regular if there exist a finite set Q and a subset F of $[\![A]\!]_{\mathbb{o}\mapsto Q}$ such that

$$L = \{M \in \Lambda(A) \mid [\![M]\!]_A \in F\} \ .$$

We write $\mathrm{Reg}\langle A\rangle$ for the Boolean algebra of regular languages of λ-terms of type A.

A profinite model of λ-calculus

In joint work with P.-A. Melliès and our PhD student V. Moreau, we use Stone duality to construct a model of profinite λ-terms:

Definition

A profinite λ-term of type A is a point of the dual space of $\mathrm{Reg}\langle A \rangle$.

The category **ProLam** has as objects the types, and the morphisms from A to B are the profinite λ-terms of type $A \Rightarrow B$.

We get a profinite model of λ-calculus, extending profinite words:

Theorem (v.G., Melliès, Moreau 2023)

The category **ProLam** *is Cartesian closed, and, for any finite alphabet* Σ*, we have* $\mathbf{ProLam}(1, \mathrm{Church}_\Sigma) \cong \widehat{\Sigma^*}$.

Summary

- Separation of regular languages by logic-defined subclasses.
  - A combination of logic and combinatorics of finite monoids.
- Duality and profinite words as a topological foundation.
  - A common mathematical theory with a large scope.
- Extending profinite words to profinite λ-terms.
  - A new model of the simply typed λ-calculus.

Some current and future directions

- Connections between profinite λ-calculus and fixed points:
  - We construct a profinite λ-term $\Omega_A : (A \Rightarrow A) \Rightarrow (A \Rightarrow A)$ for any type A, sending f to $f^!$.
  - The specification logic MSO on trees has close connections to μ-calculus. How about for λ-terms?
- Duality methods for higher order logics.
  - Current work in progress with a PhD student and postdoc in my ANR JCJC project "Topology for types and terms".
- Categorical automata theory: How does it apply to, e.g., transducers and weighted automata?
  - Recent work with Aristote, Petrișan, and Shirmohammadi.
- Interactions of separation, and its logical counterpart interpolation, with verification.
  - Formalization works with Férée, van der Giessen, Shillito.