Formal languages and Pervin spaces

Jean-Éric Pin[1]
(joint work with Mai Gehrke and Serge Grigorieff)

[1] LIAFA, CNRS and University Paris Diderot

March 2013, Paris

Outline

(1) The algebraic approach to regular languages
(2) Duality
(3) Pervin spaces
(4) Perspectives and conclusion

Part I

The algebraic approach

Three levels:

- Local approach: attach to each regular language a finite (ordered) monoid.
- Global approach: attach to each variety of regular languages a variety of finite monoids.
- Profinite approach: Replace finite monoids by profinite monoids.

Syntactic ordered monoid of a subset L of A^*

Syntactic preorder [Schützenberger 1956]:
$u \leqslant_L v$ iff, for every $x, y \in A^*$,

$$xuy \in L \Rightarrow xvy \in L$$

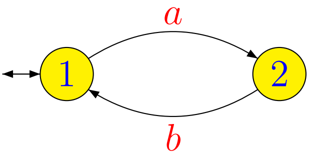
Syntactic (or context, observational) congruence

$$\begin{aligned} u \sim_L v \text{ iff } & u \leqslant_L v \text{ and } v \leqslant_L u \\ \text{iff } & xuy \in L \Leftrightarrow xvy \in L \end{aligned}$$

Syntactic monoid: $A^*/{\sim_L}$

Syntactic ordered monoid: $(A^*/{\sim_L}, \leqslant_L /{\sim_L})$

The syntactic ordered monoid of $(ab)^*$



Elements

	1	2
1	1	2
a	2	0
b	0	1
aa	0	0
ab	1	0
ba	0	2

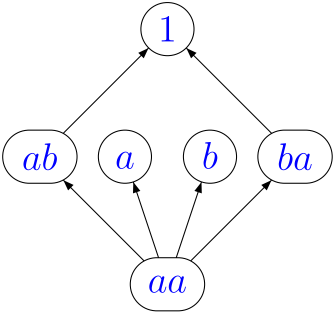
Relations

$bb = aa$

$aba = a$

$bab = b$

Syntactic order



Nonregular languages

Let $\textsc{Equality} = \{u \in \{a,b\}^* \mid |u|_a = |u|_b\}$.
Its ordered syntactic monoid is $(\mathbb{Z}, =)$.

Let $\textsc{Majority} = \{u \in \{a,b\}^* \mid |u|_a \geqslant |u|_b\}$.
Its syntactic monoid is $(\mathbb{Z}, \geqslant)$.

The local algebraic approach

Theorem (Kleene-Nerode)

Let L *be a subset of* A^**. Are equivalent:*

(1) L *is a regular language,*

(2) *The congruence* $\sim_L$ *has finite index,*

(3) *The syntactic monoid of* L *is finite.*

☞ Study regular languages through properties of their syntactic [ordered] monoids.

Star-free languages

Star-free languages = smallest class of languages containing the finite languages and closed under Boolean operations and product.

Theorem (Schützenberger 1965)

A language is star-free iff its syntactic monoid is finite and aperiodic.

A finite monoid M is aperiodic if there is an $n > 0$ such that for all $x \in M$, $x^n = x^{n+1}$.

Piecewise testable languages

Piecewise testable languages = Boolean combination of languages of the form $A^*a_1A^*a_2 \cdots A^*a_kA^*$, where $a_1, \ldots, a_k$ are letters:

Theorem (Simon 1972)

A language is piecewise testable iff its syntactic monoid is finite and division-free.

A monoid is division-free (or $\mathcal{J}$-trivial) if any two elements that divide each other are equal. (x divides y if $y = sxt$ for some $s, t \in M$).

The global approach: Varieties of languages

A variety of languages is a class of regular languages $\mathcal{V}$ such that:

(1) for each alphabet A, $\mathcal{V}(A^*)$ is closed under Boolean operations and quotients.

(2) if $L \in \mathcal{V}(B^*)$, then for each monoid morphism $\alpha : A^* \to B^*$, then $\alpha^{-1}(L) \in \mathcal{V}(A^*)$.

Quotients: $x^{-1}Ly^{-1} = \{u \in A^* \mid xuy \in L\}$

Regular languages, star-free languages, piecewise testable languages form varieties of languages.

The global approach: Varieties of finite monoids

Variety of finite monoids = class of finite monoids closed under taking submonoids, quotient monoids and finite direct products.

Examples:

- All finite monoids
- Commutative monoids
- Idempotent monoids
- Aperiodic monoids
- Division-free monoids
- Groups

Global approach: Eilenberg's variety theorem

Theorem (Eilenberg 1976)

There is bijection between varieties of monoids and varieties of languages.

Examples

Finite monoids	Regular languages
Aperiodic monoids	Star-free languages
Division-free monoids	Piecewise testable languages

Positive varieties

A positive variety of languages is a class of regular languages closed under union, intersection, quotients and inverses of morphisms.

A variety of finite ordered monoids is a class of finite ordered monoids closed under taking finite products, ordered submonoids and quotient monoids.

Theorem (Pin 1995)

There is a bijection between positive varieties of languages and varieties of finite ordered monoids.

Profinite metrics (1): Separating words

A monoid morphism $\varphi : A^* \to M$ separates two words u and v of A^* if $\varphi(u) \neq \varphi(v)$.

The morphism $u \to |u| \mod 2$ (from A^* into $\mathbb{Z}/2\mathbb{Z}$) separates $abaabaaba$ and $abaabaabab$.

Proposition

One can always separate two distinct words by a finite monoid [group, division-free monoid].

However, commutative monoids cannot separate ab from ba.

Profinite metrics: Step 1

Let **V** be a variety of finite monoids. The relation

$u \sim_{\mathbf{V}} v$ iff no monoid of **V** can separate u from v

is a congruence. The quotient $A^*/{\sim_{\mathbf{V}}}$ is a monoid.

Exemple. If **V** is the variety of all finite commutative monoids, $u \sim_{\mathbf{V}} v$ iff for all $a \in A$, $|u|_a = |v|_a$ and $A^*/{\sim_{\mathbf{V}}} = \mathbb{N}^A$.

Up to changing A^* to $A^*/{\sim_{\mathbf{V}}}$, we will assume that **V** separates the elements of A^*.

The profinite metric $d_{\mathbf{V}}$

Let **V** be a variety of finite monoids separating the elements of A^*. Let u and v be two words. Put

$$r_{\mathbf{V}}(u,v) = \min\Big\{|M| \mid M \in \mathbf{V} \text{ and separates } u \text{ from } v\Big\}$$

$$d_{\mathbf{V}}(u,v) = 2^{-r_{\mathbf{V}}(u,v)}$$

Then $d_{\mathbf{V}}$ is an ultrametric, that is

(1) $d(x,y) = 0 \iff x = y$,

(2) $d(x,y) = d(y,x)$,

(3) $d(x,z) \leqslant \max\{d(x,y), d(y,z)\}$

The free pro-**V** monoid

The completion of the metric space $(A^*, d_{\mathbf{V}})$ is denoted $\widehat{F}_{\mathbf{V}}(A)$. It is a compact space [nontrivial result].

Further, the product on A^* is uniformly continuous and has a unique uniformly continuous extension to the completion of A^*, making $\widehat{F}_{\mathbf{V}}(A)$ a compact monoid, called the free pro-**V** monoid.

For $\mathbf{V} = \{$all finite monoids$\}$, the completion is also denoted $\widehat{A^*}$ and its elements are called profinite words.

Properties of the free pro-**V** monoid

Theorem

(1) *A finite* A*-generated monoid belongs to* **V** *iff it is a continuous quotient of* $\widehat{F}_{\mathbf{V}}(A)$*.*

(2) *If* $\mathbf{V} \subseteq \mathbf{W}$ *then* $\widehat{F}_{\mathbf{V}}(A)$ *is a continuous quotient of* $\widehat{F}_{\mathbf{W}}(A)$*. In particular,* $\widehat{F}_{\mathbf{V}}(A)$ *is a continuous quotient of* $\widehat{A^*}$*.*

Reiterman's theorem

Let u and v be profinite words of $\widehat{A^*}$. A finite monoid satisfies the identity $u = v$ if, for all continuous morphisms $\varphi = \widehat{A^*} \to M$, one has $\varphi(u) = \varphi(v)$.

Theorem (Reiterman 1982)

A class of finite monoids is a variety of finite monoids iff can be defined by a set of profinite identities.

The operator ω

In a finite semigroup, every element x has a unique idempotent power, denoted x^ω.

In a compact semigroup, the closed subsemigroup generated by an element x contains a unique idempotent, denoted x^ω.

Equivalent definition in $\widehat{A^*}$:
For each profinite word u, the sequence $u^{n!}$ converges to an idempotent denoted u^ω.

Examples of identities

A finite monoid is aperiodic iff it satisfies the identity $x^{\omega+1} = x^{\omega}$.

A finite monoid is division-free iff it satisfies the identities

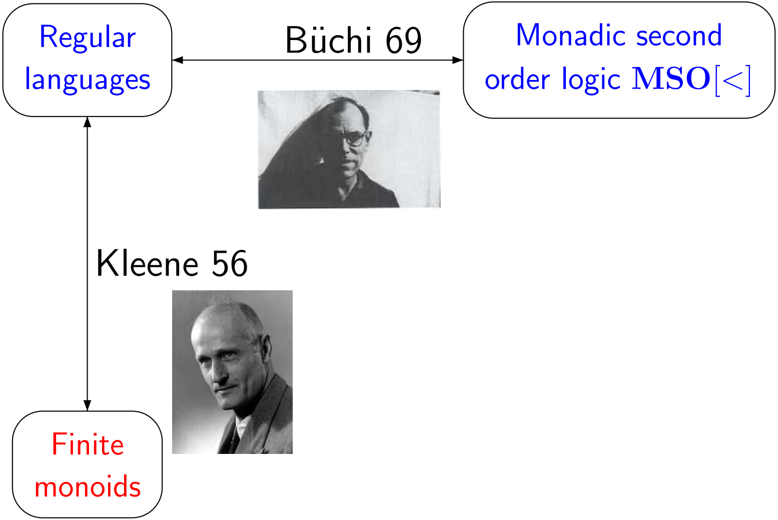
$$x^{\omega+1} = x^{\omega} \text{ and } (xy)^{\omega}x = (xy)^{\omega} = y(xy)^{\omega}$$

iff it satisfies the identities

$$x^{\omega+1} = x^{\omega} \text{ and } (xy)^{\omega} = (yx)^{\omega}.$$

A finite monoid is a group iff it satisfies the identity $x^{\omega} = 1$.

The algebraic approach to regular languages



Regular languages

Büchi 69

Monadic second order logic $\mathbf{MSO}[<]$

Kleene 56

Finite monoids

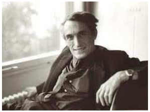
The algebraic approach to regular languages (2)

Star-free languages

Mc Naughton 71

First order logic $\mathbf{FO}[<]$

Schützenberger 65



Decidability

Aperiodic monoids

$x^\omega = x^{\omega+1}$

The algebraic approach to regular languages (3)

Piecewise testable languages

Thomas 87

$B\Sigma_1[<]$ formulas

I. Simon 75

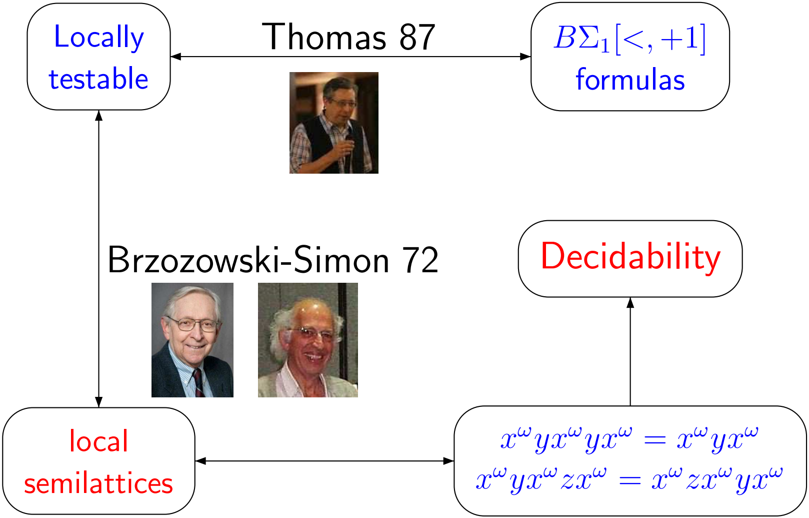
Decidability

$\mathcal{J}$-trivial monoids

$x^{\omega} = x^{\omega+1}$

$(xy)^{\omega} = (yx)^{\omega}$

The algebraic approach to regular languages (4)



Locally testable

Thomas 87

$B\Sigma_1[<, +1]$ formulas

Brzozowski-Simon 72

Decidability

local semilattices

$$x^\omega y x^\omega y x^\omega = x^\omega y x^\omega$$
$$x^\omega y x^\omega z x^\omega = x^\omega z x^\omega y x^\omega$$

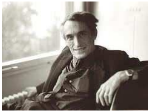
The algebraic approach to regular languages (6)

Unambiguous
star-free

SchThV 01, ThW 98
EVW 97, PW 97

$\mathbf{FO}^2[<]$
$\Sigma_2 \cap \Pi_2$

Schützenberger 76



Decidability

DA

$$x^\omega = x^{\omega+1}$$
$$(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$$

Summary of the first part

Local approach: study regular languages through properties of their syntactic monoid.

Global approach: characterize varieties of regular languages through profinite identities. Study the pro-**V** monoids.

Is it possible to use the local/global approach for more general classes of regular languages than varieties? And beyond regular languages?

Summary of the first part

Local approach: study regular languages through properties of their syntactic monoid.

Global approach: characterize varieties of regular languages through profinite identities. Study the pro-**V** monoids.

Is it possible to use the local/global approach for more general classes of regular languages than varieties? And beyond regular languages?

Yes we can...

Part II

Duality

Lattice? You mean bounded distributive lattice? Then it has a dual !

That's the way its all started...

Duality in a nutshell

Definition

The dual space of a bounded distributive lattice is the set of its prime filters.

Elements ⟷ Prime filters
Boolean algebras ⟷ Stone spaces
Distributive lattices ⟷ Ordered Stone spaces
Sublattices ⟷ Quotient spaces
n-ary operations ⟷ $(n+1)$-ary relations

Key idea: A lattice of subsets of A^* is completely characterized by its dual space.

Duality for regular languages

[Almeida ↑89] Duality between varieties of regular languages and clopen sets of profinite monoids.

[Pippenger 97] Slightly more general results: Stone duality explicitly mentioned.

[GGP 08] Extended Stone duality for any lattice of regular languages. The product is the dual of the residuation operations.

$$XY \subseteq Z \iff Y \subseteq X\backslash Z \iff X \subseteq Z/Y$$

Equational theory for lattices of regular languages.

Lattices, filters and ideals

A lattice of subsets of X is a set of subsets of X containing $\emptyset$ and X and closed under finite intersection and finite union. A Boolean algebra is a lattice closed under complement.

Let $\mathcal{L}$ be a lattice of subsets of X. A filter is a nonempty subset $\mathcal{F}$ of $\mathcal{L}$ such that:

(1) if $K \in \mathcal{F}$ and $K \subseteq L$, then $L \in \mathcal{F}$,
(2) if $K, L \in \mathcal{F}$, then $K \cap L \in \mathcal{F}$.

An ideal is a nonempty subset $\mathcal{I}$ of $\mathcal{L}$ such that:

(1) if $K \in \mathcal{I}$ and $L \subseteq K$, then $L \in \mathcal{I}$,
(2) if $K, L \in \mathcal{I}$, then $K \cup L \in \mathcal{I}$.

Prime filters

A prime filter on $\mathcal{L}$ is a subset $\mathcal{F}$ of $\mathcal{L}$ such that

(1) $X \in \mathcal{F}$, $\emptyset \notin \mathcal{F}$,

(2) if $K \in \mathcal{F}$ and $K \subseteq L$, then $L \in \mathcal{F}$,

(3) if $K, L \in \mathcal{F}$, then $K \cap L \in \mathcal{F}$,

(4) if $K \cup L \in \mathcal{F}$, then either $K \in \mathcal{F}$ or $L \in \mathcal{F}$.

In other words. . .

(1) A prime filter is nontrivial,

(2) closed under extension,

(3) closed under intersection,

(4) and has to choose. . .

Stone-Priestley duality

Let S be the dual space of $\mathcal{L}$. The map $e : \mathcal{L} \to 2^S$

$$e(L) = \{\ \text{prime filters containing } L\ \}$$

is an injective lattice morphism.

Topology on S:

Take as a basis of open sets the sets of the form $e(L)$. Then S is a compact, totally disconnected space (not necessarily Hausdorff). It is ordered by the specialization order: $x \leqslant y \iff \overline{\{x\}} \subseteq \overline{\{y\}}$.

Some examples of dual spaces

- Let L be a regular language and let $\mathcal{L}$ be the lattice generated by the languages $u^{-1}Lv^{-1}$ $(u, v \in A^*)$. Its dual space is the ordered syntactic monoid of L.
- Regular languages → the free profinite monoid on A. [Almeida]
- Star-free languages → the free pro-aperiodic monoid on A. [Almeida]

More examples of dual spaces

- Finite languages $\cup \{A^*\} \to A^* \cup \{\infty\}$, one-point compactification of A^*.
- All languages $\to \beta A^*$, the Stone-Čech compactification of A^*.

One can define βA^* as the closure of the range of A^* in $\prod_\varphi \varphi(A^*)$, where φ is any function from A^* into a compact space.

Reconciling duality and metric spaces

Back to the Boolean algebra of star-free languages. The profinite approach says that its dual space is the completion of a metric space.

Is it possible to define the dual space of any lattice as the completion of some suitable space?

Metric spaces do not suffice: βA^* is not metrizable.

Uniform spaces work well for Boolean algebras but quasi-uniform spaces are needed for lattices.

Reconciling duality and metric spaces (2)

Potential problem: completion is well-defined for uniform spaces but is quite messy for quasi-uniform spaces (three competing definitions).

What kind of (quasi)-uniform spaces do we get?

- transitive (→ ultrametric)
- totally bounded (→ The completion of the space is compact.)

This leads to consider transitive, totally bounded quasi-uniform spaces.

Reconciling duality and metric spaces (2)

Potential problem: completion is well-defined for uniform spaces but is quite messy for quasi-uniform spaces (three competing definitions).

What kind of (quasi)-uniform spaces do we get?

- transitive (→ ultrametric)
- totally bounded (→ The completion of the space is compact.)

This leads to consider transitive, totally bounded quasi-uniform spaces.

And then a miracle occurs...

The miracle

Theorem

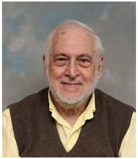
A quasi-uniform space is a Pervin space iff it is transitive and totally bounded.

The miracle is that the definition of Pervin spaces is really very simple !

Part III

Pervin spaces

- W. J. Pervin, Quasi-uniformization of topological spaces, *Math. Ann.* **147** (1962), 316–317.



Pervin spaces

Definition

A Pervin space is a pair $(X, \mathcal{L})$ where $\mathcal{L}$ is a lattice of subsets of X.

The elements of $\mathcal{L}$ form a basis of a topology in which each open set is a (possibly infinite) union of elements of $\mathcal{L}$.

Examples of Pervin spaces

Examples on $X = \{0, 1\}$

- The Boolean space: $\mathcal{L} = \Big\{\emptyset, \{0\}, \{1\}, \{0, 1\}\Big\}$
- The Sierpinski space: $\mathcal{L} = \Big\{\emptyset, \{1\}, \{0, 1\}\Big\}$.

Examples on $X = \mathbb{N}$

- $\mathcal{L} = \{\emptyset\} \cup \{$cofinite subsets of $\mathbb{N}\}$.
- $\mathcal{L} = \{\mathbb{N}\} \cup \{$finite subsets of $\mathbb{N}\}$.
- $\mathcal{L} = \{$finite/cofinite subsets of $\mathbb{N}\}$.

Examples on $X = A^*$

- $\mathcal{L} = \left\{\text{Regular languages}\right\}$
- $\mathcal{L} = \left\{\text{Star-free languages}\right\}$
- $\mathcal{L} = \left\{\text{Group languages}\right\}$
- $\mathcal{L} = \left\{\text{All languages}\right\}$
- $\mathcal{L} = \left\{\text{Commutative languages}\right\}$
- $\mathcal{L} = \{A^*\} \cup \left\{\text{Finite languages}\right\}$

The preorder of a Pervin space $(X, \mathcal{L})$

Preorder on X: $x \leqslant_{\mathcal{L}} y$ iff, for all $L \in \mathcal{L}$,

$$x \in L \Rightarrow y \in L$$

Equivalence relation: $x \sim_{\mathcal{L}} y$ iff, for all $L \in \mathcal{L}$,

$$x \in L \Leftrightarrow y \in L$$

We often write $\leqslant$ for $\leqslant_{\mathcal{L}}$ and $\sim$ for $\sim_{\mathcal{L}}$.

This preorder coincides with the specialisation preorder.

Examples

Let L be a language and let $\mathcal{L}$ be the lattice generated by the quotients of L. Then $\sim$ is the syntactic congruence of L and $\leqslant_L$ is its syntactic order.

Let $X = A^*$ and $\mathcal{L} = \{$commutative languages$\}$. Then $\sim$ is the commutative equivalence:

$$u \sim v \quad \text{iff} \quad |u|_a = |v|_a \text{ for all } a \in A$$

and $A^*/{\sim} = \mathbb{N}^A$.

The quotient space $(X/{\sim}, \mathcal{L}/{\sim})$

Let $L \in \mathcal{L}$. If $x \in L$ and $x \sim y$, then $y \in L$.

Thus the quotient space $X/{\sim}$, the quotient lattice

$$\mathcal{L}/{\sim} = \{L/{\sim} \mid L \in \mathcal{L}\}$$

and the Pervin space $(X/{\sim}, \mathcal{L}/{\sim})$ are well-defined notions.

Proposition

If $\mathcal{L}$ is the lattice generated by the quotients of a language L of A^*, then $A^*/{\sim}$ is the syntactic monoid of L.

The Kolmogorov quotient of a Pervin space $(X, \mathcal{L})$

Proposition (trivial)

The following conditions are equivalent:

(1) $\leqslant$ *is an order,*

(2) $\sim$ *is the equality relation,*

(3) X *is a Kolmogorov* (T_0) *space.*

Kolmogorov: open sets distinguish points.

In particular, the quotient space $(X/{\sim}, \mathcal{L}/{\sim})$ is Kolmogorov.

Uniform continuity

Let $(X, \mathcal{K})$ and $(Y, \mathcal{L})$ be Pervin spaces.

Definition

A function $\varphi\colon X \to Y$ is uniformly continuous if, for each $L \in \mathcal{L}$, $\varphi^{-1}(L) \in \mathcal{K}$.

Remark. Any uniformly continuous map is continuous and order preserving, but the converse is not true.

Completion of a Pervin space

Step one: Take the quotient $X/\sim$.

Step two: All the completions available for quasi-uniform spaces are equivalent for Pervin spaces. Take the one you prefer!

A. Császár, D-completions of Pervin-type quasi-uniformities., *Acta Sci. Math.* **57**,1-4 (1993), 329–335.

Valuations on a lattice of subsets $\mathcal{L}$

Definition

A valuation on $\mathcal{L}$ is a lattice morphism from $\mathcal{L}$ into the Boolean algebra $\{0, 1\}$.

Thus it is a map $v : \mathcal{L} \to \{0, 1\}$ such that, for all $L_1, L_2 \in \mathcal{L}$,

$$\begin{aligned}
&(1)\quad v(\emptyset) = 0,\ v(X) = 1,\\
&(2)\quad v(L_1 \cap L_2) = v(L_1)v(L_2),\\
&(3)\quad v(L_1 \cup L_2) = v(L_1) + v(L_2).
\end{aligned}$$

where the sum and the product are the Boolean operations.

Valuations and prime filters are the same

If $v : \mathcal{L} \to \{0, 1\}$ is a valuation, then the set $v^{-1}(1)$ is a prime filter.

If p is a prime filter, the map

$$v(L) = \begin{cases} 1 & \text{if } L \in p \\ 0 & \text{otherwise} \end{cases}$$

is a valuation.

Completion of a Pervin space $(X, \mathcal{L})$

From now on, we assume that $\leqslant$ is an order on X.

For each $L \in \mathcal{L}$, let

$$\widehat{L} = \{v \text{ is a valuation such that } v(L) = 1\}$$

In particular, $\widehat{X}$ is the set of all valuations on $\mathcal{L}$.

Definition

The completion of a Pervin space $(X, \mathcal{L})$ is the Pervin space $(\widehat{X}, \widehat{\mathcal{L}})$, where $\widehat{\mathcal{L}}$ is the lattice of subsets of $\widehat{X}$ defined by $\widehat{\mathcal{L}} = \{\widehat{L} \mid L \in \mathcal{L}\}$.

Embedding $(X, \mathcal{L})$ into $(\widehat{X}, \widehat{\mathcal{L}})$

For each $x \in X$, let v_x be the valuation defined by

$$v_x(L) = \begin{cases} 1 & \text{if } x \in L \\ 0 & \text{if } x \notin L \end{cases}$$

Proposition

The map $x \to v_x$ defines an injective and uniformly continuous embedding from $(X, \mathcal{L})$ into $(\widehat{X}, \widehat{\mathcal{L}})$. Further, X is dense in $\widehat{X}$.

Properties of the completion of a Pervin space

$\widehat{L} = \{v$ is a valuation such that $v(L) = 1\}$
$\widehat{X} = \{$all valuations$\}$ $\qquad \widehat{\mathcal{L}} = \{\widehat{L} \mid L \in \mathcal{L}\}$

Theorem

The lattice $\widehat{\mathcal{L}}$ is the set of all compact open subsets of $\widehat{X}$. In particular, $\widehat{X}$ is compact.

Pervin spaces and duality

Theorem (Duality theorem)

The maps $L \mapsto \widehat{L}$ *and* $K \mapsto K \cap X$ *are mutually inverse lattice isomorphisms between* $\mathcal{L}$ *and* $\widehat{\mathcal{L}}$*.*

Corollary

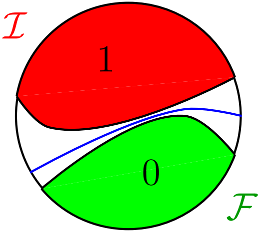
The completion of the Pervin space $(X, \mathcal{L})$ *is equal to the dual of* $\mathcal{L}$*.*

For instance, the completion of $(X, \mathcal{P}(X))$ is the Stone-Čech compactification of X.

The prime filter theorem

Theorem

Let $\mathcal{I}$ *be an ideal and let* $\mathcal{F}$ *be a filter disjoint from* $\mathcal{I}$*. Then there is a valuation* v *on* $\mathcal{L}$ *such that* $v(L) = 1$ *for all* $L \in \mathcal{F}$ *and* $v(L) = 0$ *for all* $L \in \mathcal{I}$*.*



Example 1

Let $X = \mathbb{N}$ and $\mathcal{L} = \{\emptyset\} \cup \{\text{cofinite subsets of } \mathbb{N}\}$.

This space is compact, but not complete. Indeed, the valuation v given by $v(L) = 1$ for each cofinite set L defines a new element, denoted ∞.

Thus $\widehat{X} = X \cup \{\infty\}$ and
$\widehat{\mathcal{L}} = \{\emptyset\} \cup \{\text{cofinite subsets of } \mathbb{N} \text{ containing } \infty\}$.

The order on X is the equality relation, but in $\widehat{X}$, $x \leqslant \infty$ for all $x \in \widehat{X}$.

Example 2

Let $X = \mathbb{N}$ and $\mathcal{L} = \{\mathbb{N}\} \cup \{\text{finite subsets of } \mathbb{N}\}$.

This space is neither compact nor complete. Indeed, the valuation v given by $v(X) = 1$ and $v(L) = 0$ for each finite set L defines a new element $-\infty$.

Thus $\widehat{X} = X \cup \{-\infty\}$ and
$\widehat{\mathcal{L}} = \{\mathbb{N}\} \cup \{\text{finite subsets of } \mathbb{N}\}$.

The order on X is the equality relation, but in $\widehat{X}$, $-\infty \leqslant x$ for all $x \in \widehat{X}$.

Example 3

Let $X = \mathbb{N}$ and $\mathcal{L} = \{\text{finite/cofinite subsets of } \mathbb{N}\}$.

This space is neither compact nor complete. Indeed, the valuation v given by

$$v(L) = \begin{cases} 1 & \text{if } L \text{ is cofinite} \\ 0 & \text{if } L \text{ is finite} \end{cases}$$

defines a new element ∞. Thus $\widehat{X} = X \cup \{\infty\}$ and

$$\widehat{\mathcal{L}} = \{\text{finite subsets of } \mathbb{N}\} \cup \{\text{cofinite subsets of } \mathbb{N} \text{ containing } \infty\}.$$

Example 4

Let $L_n = \{\frac{1}{k} \mid 0 < k \leqslant n\}$ and

$$X = \left\{\frac{1}{n} \mid n \text{ is a positive integer}\right\}$$
$$\mathcal{L} = \{X\} \cup \{L_n \mid n > 0\}$$

This space is compact, but not complete. Indeed, the valuation v given by $v(X) = 1$ and $v(L_n) = 0$ defines a new element 0. Thus $\widehat{X} = X \cup \{0\}$ and $\widehat{\mathcal{L}} = \{\emptyset\} \cup \{\text{finite subsets of } \widehat{X} \text{ containing } 0\}$.

The order on $\widehat{X}$ is $0 \leqslant \cdots \leqslant \frac{1}{n} \leqslant \cdots \leqslant \frac{1}{2} \leqslant 1$.

Syntactic monoid and syntactic space

Let $\mathcal{L}$ be a lattice closed under quotients.

Definition

The syntactic monoid of $\mathcal{L}$ is the monoid $(A^*/{\sim})$. The syntactic space of $\mathcal{L}$ is the completion of the Pervin space $(A^*/{\sim}, \mathcal{L}/{\sim})$.

By construction, the syntactic space of $\mathcal{L}$ is the dual of $\mathcal{L}$. It is compact but it is not always a monoid because the product in $A^*/{\sim}$ might not be uniformly continuous.

Some examples

The syntactic space of ...

- the lattice of all regular languages of A^* is the free profinite monoid on A.
- the set of all star-free languages of A^* is the free pro-aperiodic monoid on A.
- the lattice of finite or full languages is the one-point compactification of A^*.
- the set of all languages of A^* is the Stone-Čech compactification of A^*.

The case of a single language

Let L be a language. The syntactic space of L is obtained as follows:

(1) Compute the syntactic monoid M of L and the image P of L in M the usual way).

(2) Let $\mathcal{L}$ be the lattice generated by the quotients of P in M.

(3) The syntactic space of L is the completion of $(M, \mathcal{L})$.

If L is a regular language, its syntactic monoid is finite and equal to its completion. Thus, for regular languages, only the algebraic properties of the syntactic monoid are important.

Syntactic space of MAJORITY

The completion is $\widehat{\mathbb{Z}} = \mathbb{Z} \cup \{-\infty, +\infty\}$. The closure of the addition (considered as a subset of $\widehat{\mathbb{Z}}^3$) is the relation $\widehat{+}$ given in the following table:

$\widehat{+}$	i	$-\infty$	$+\infty$
j	$\{i + j\}$	$\{-\infty\}$	$\{+\infty\}$
$-\infty$	$\{-\infty\}$	$\{-\infty\}$	$\widehat{\mathbb{Z}}$
$+\infty$	$\{+\infty\}$	$\widehat{\mathbb{Z}}$	$\{+\infty\}$

A nonregular example

Let $M = (\mathbb{Z}, +)$ and let $\mathcal{L}$ be the Boolean algebra of finite or cofinite subsets of $\mathbb{Z}$. The syntactic space of $\mathcal{L}$ is $\widehat{\mathbb{Z}} = \mathbb{Z} \cup \{\infty\}$.

The closure of the addition on $\mathbb{Z}$ is the relation $\widehat{+}$

$\widehat{+}$	i	∞
j	$\{i + j\}$	$\{\infty\}$
∞	$\{\infty\}$	$\widehat{\mathbb{Z}}$

The Boolean algebra $\mathrm{Rec}(\mathbb{Z})$

Let $\mathrm{Rec}(\mathbb{Z})$ be the set of recognizable subsets of $\mathbb{Z}$, that is, the finite unions of subsets of the form $\{a + n\mathbb{Z} \mid n \geqslant 1, 0 \leqslant a < n\}$.

Its syntactic monoid is $\mathbb{Z}$ and its syntactic space $\widehat{\mathbb{Z}}$ is the one-generated profinite free group. The closure of the addition is here the addition of $\widehat{\mathbb{Z}}$.

We denote by $i \mapsto \underline{i}$ the natural embedding of $\mathbb{Z}$ into $\widehat{\mathbb{Z}}$ and by $+$ the addition on $\widehat{\mathbb{Z}}$.

Recognizable + finite subsets of $\mathbb{Z}$

Let $\mathcal{L}$ be the Boolean algebra generated by the finite subsets and the recognizable subsets of $\mathbb{Z}$. Its syntactic monoid is $\mathbb{Z}$ and its syntactic space is the disjoint union $\mathbb{Z} \cup \widehat{\mathbb{Z}}$: $\mathbb{Z}$ corresponds to the principal ultrafilters of $\mathcal{L}$ and the profinite group $\widehat{\mathbb{Z}}$ corresponds to the nonprincipal ultrafilters.

The closure $\widehat{+}$ of the addition on $\mathbb{Z}$ is commutative but nonfonctional. It extends $+$ on $\mathbb{Z}$ and, for $i \in \mathbb{Z}$ and $u, v \in \widehat{\mathbb{Z}}$, one has $i \mathbin{\widehat{+}} u = \underline{i} + u$ and

$$u \mathbin{\widehat{+}} v = \begin{cases} \{k, \underline{k}\} & \text{if } u + v = \underline{k} \text{ with } k \in \mathbb{Z} \\ \{u + v\} & \text{otherwise.} \end{cases}$$

Duality versus Pervin spaces

A subtle difference. Duality deals with abstract lattices (pointless topology). Pervin spaces only deals with lattices of subsets.

Pros of duality.

- Duality for lattices with operations.
- Existing results in duality ready for use.

Pros of Pervin spaces.

- Completion in two steps.
- Definitions are simpler than for spectral spaces.
- Intuition from (ultra)metric spaces available, but most results need to be rewritten.

Uniformly continuous extensions

Theorem

Every uniformly continuous map $\varphi : (X, \mathcal{L}_X) \to (Y, \mathcal{L}_Y)$ admits a unique uniformly continuous extension $\widehat{\varphi} : (\widehat{X}, \widehat{\mathcal{L}}_X) \to (\widehat{Y}, \widehat{\mathcal{L}}_Y)$.

Corollary

Let φ_1 and φ_2 be two uniformly continuous maps from $(X, \mathcal{L}_X)$ to $(Y, \mathcal{L}_Y)$ and let $\widehat{\varphi}_1$ and $\widehat{\varphi}_2$ be their uniformly continuous extensions from $(\widehat{X}, \widehat{\mathcal{L}}_X)$ to $(\widehat{Y}, \widehat{\mathcal{L}}_Y)$. If $\varphi_1 \leqslant \varphi_2$, then $\widehat{\varphi}_1 \leqslant \widehat{\varphi}_2$.

Principe du prolongement des identités

Proposition

Let φ_1 and φ_2 be two uniformly continuous maps from $(\widehat{X}, \widehat{\mathcal{L}}_X)$ to $(\widehat{Y}, \widehat{\mathcal{L}}_Y)$. If, for all $x \in X$, $\varphi_1(x) \leqslant \varphi_2(x)$, then $\varphi_1 \leqslant \varphi_2$. In particular, if φ_1 and φ_2 coincide on X, then they are equal.

Equations: the profinite case

Let $\mathcal{L}$ be a lattice of regular languages closed under quotients. Let $(M_{\mathcal{L}}, \leqslant) = (A^*/{\sim_L}, \leqslant_{\mathcal{L}})$ be the syntactic ordered monoid of $\mathcal{L}$ and let $\eta : A^* \to M_{\mathcal{L}}$ be the quotient map.

Then η is a uniformly continuous function from $(A^*, \mathrm{Reg}(A^*))$ to $(M_{\mathcal{L}}, \mathcal{L}/{\sim})$. It extends to a uniformly continuous function from $\widehat{A^*}$ to $\widehat{M_{\mathcal{L}}}$.

Let (u, v) be a pair of profinite words. We say that $\mathcal{L}$ satisfies the equation $u \leqslant v$ if $\widehat{\eta}(u) \leqslant \widehat{\eta}(v)$.

Equations: the general case

Let $\mathcal{L}$ be a lattice of languages closed under quotients. Let $(M_{\mathcal{L}}, \leqslant) = (A^*/{\sim_L}, \leqslant_{\mathcal{L}})$ be the syntactic ordered monoid of $\mathcal{L}$ and let $\eta : A^* \to M_{\mathcal{L}}$ be the quotient map.

Then η is a uniformly continuous function from $(A^*, \mathcal{P}(A^*))$ to $(M_{\mathcal{L}}, \mathcal{L}/{\sim})$. It extends to a uniformly continuous function from βA^* to $\widehat{M_{\mathcal{L}}}$.

Let (u, v) be a pair of elements of βA^*. We say that $\mathcal{L}$ satisfies the equation $u \leqslant v$ if $\widehat{\eta}(u) \leqslant \widehat{\eta}(v)$, where $\eta : A^* \to A^*/{\sim}$ is the syntactic map.

Equations

Theorem

A set of regular languages of A^* is a lattice closed under quotients iff it can be defined by a set of equations of the form $u \leqslant v$, where u, v are profinite words.

Theorem

A set of languages of A^* is a lattice closed under quotients iff it can be defined by a set of equations of the form $u \leqslant v$, where $u, v \in \beta A^*$.

Open maps

Theorem

Let $\varphi : (X, \mathcal{L}_X) \to (Y, \mathcal{L}_Y)$ *be a uniformly continuous map and let* $L \in \mathcal{L}_X$*. Are equivalent:*

(1) *There is a smallest* $K \in \mathcal{L}_Y$ *such that* $\varphi(L) \subseteq K$*,*

(2) *The upper set generated by* $\widehat{\varphi}(\widehat{L})$ *is open.*

Best approximation property

Best approximation property: for all $L_1, L_2 \in \mathcal{L}$, there is a smallest L such that $L_1L_2 \subseteq L$.

Proposition

Let $\mathcal{L}$ be a Boolean algebra of regular languages of A^* closed under quotients. Are equivalent:

(1) *$\mathcal{L}$ has the best approximation property,*

(2) *$\widehat{\mathcal{L}}$ is closed under product,*

(3) *The product on the completion of $(A^*, \mathcal{L})$ is an open map.*

Closure under product

Proposition

Suppose that $\mathcal{L}$ contains the finite languages. Then $\mathcal{L}$ has the best approximation property iff it is closed under product.

A group language is a language whose syntactic monoid is a finite group. The group languages of A^* form a Boolean algebra closed under quotient, which has the best approximation property but is not closed under product.

Metrizability

Proposition

Let $\mathcal{L}$ be a Boolean algebra of subsets of A^*. Are equivalent:

(1) *The associated Pervin space is metrizable,*

(2) *The uniformity has a countable basis,*

(3) *$\mathcal{L}$ is countable.*

Similar results for lattices/quasi-metrizability.

Translations and quotients

Let $\mathcal{L}$ be a lattice of subsets of A^*.

Proposition

The translations $u \mapsto xu$ and $u \mapsto uy$ are all uniformly continuous iff $\mathcal{L}$ is closed under quotients.

A monoid in which the translations are uniformly continuous is called a semiuniform monoid.

Uniform continuity of the product

Let $(M, \mathcal{L})$ be a Pervin space (with $\leqslant$ an order).

Proposition

Suppose that M is a monoid and that the product is uniformly continuous. Then the product admits a unique uniformly continuous extension to $\widehat{M}$, which defines a structure of monoid on $\widehat{M}$.

When is the product uniformly continuous?

Let $\mathcal{L}$ be a Boolean algebra of subsets of A^* closed under quotients and let M be its syntactic monoid.

Theorem (GGP 2010)

Are equivalent:

(1) *the product on M uniformly continuous,*
(2) *the completion of M is a compact monoid,*
(3) *the closure of the product of M is functional,*
(4) *the elements of $\mathcal{L}$ are all regular languages.*

Part IV

Perspectives and conclusion

Related notions

Syntactic monoids are very successful for regular languages, but are not doing so well beyond regular languages (noticeable exception: the theory of context-free groups [Muller Schupp 83]).

Sakarovitch [1976] proposed to use the pair formed by a syntactic monoid and the image of the language in its syntactic monoid (pointed monoid).

Our new definition is an extension of this idea. The image of the language is used to define a lattice on the syntactic monoid, and then consider the completion of the resulting Pervin space.

Dreams

Any lattice closed under quotients can be defined by a set of equations of the form $u \leqslant v$, where $u, v \in \beta A^*$.

Further, one has $\mathcal{L}_1 \subseteq \mathcal{L}_2$ iff $\mathcal{L}_1$ satisfies the equations defining $\mathcal{L}_2$ iff the syntactic space of $\mathcal{L}_1$ is a quotient of the syntactic space of $\mathcal{L}_2$.

Thus in principle, one could separate two classes of languages $\mathcal{L}_1$ and $\mathcal{L}_2$ by finding an equation satisfied by one of the classes and not by the other.

Home work. Find a set of equations defining your own favorite complexity class or logical fragment...

Logic and circuit complexity

Let $\mathcal{N}$ be the class of all numerical predicates. Then the $\mathbf{FO}[\mathcal{N}]$-definable languages of A^* form a Boolean algebra, whose syntactic space is $\beta\mathbb{N}$.

It is known that $\mathbf{FO}[\mathcal{N}, \mathbf{a}]$ defines AC^0, the class of languages computed by unbounded fanin, polynomial size, constant-depth Boolean circuits.

What is the syntactic space of the Boolean algebra of all $\mathbf{FO}[\mathcal{N}, \mathbf{a}]$-definable languages?

Beyond recognizable languages

[Barrington, Compton, Straubing, Thérien 1992] proved that

$$\mathbf{FO}[\mathcal{N}, \mathbf{a}] \cap \mathrm{Reg}(A^*) = \mathbf{FO}[<, \mathbf{MOD}, \mathbf{a}]$$

Is it possible to prove this result by using syntactic spaces?

This would permit to attack difficult conjectures in circuit complexity.

Why to be pessimistic

- Eilenberg's variety theorem doesn't help much for proving Schützenberger's theorem.
- $\beta\mathbb{N}$ is a very complex object (its nickname is the three-headed monster).
- $\beta\mathbb{N}$ is not a compact monoid. Chasing idempotents does not work as in the finite or compact case.
- Duality gives a nice encoding, but intrinsic difficulties may just remain.
- Known results on $\beta\mathbb{N}$ indicate that set theoretic problems may occur on the way.

Why to be optimistic

- This approach is very successful for lattices of regular languages.
- It is a global approach and syntactic spaces contain a lot of information.
- It gives access to the power of topology and to higher mathematics.
- Stone duality has been successful in other areas of mathematics, notably in algebraic geometry.
- For circuit complexity classes, known partial results may guide intuition.