# Dual spaces as completion of Pervin uniformities and their application to recognition of formal languages

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

[1]LIAFA, CNRS and University Paris Diderot

June 2010, Tbilisi

# Outline

(1) Uniform spaces and Pervin spaces
(2) Recognition
(3) Syntactic space
(4) Applications to logic

# Part I

## Uniform spaces

Uniform spaces are an abstraction of metric spaces that formalizes the notion of relative closeness.

# Relations

Let $X$ be a set. We denote by $UV$ the composition of two relations $U$ and $V$ on $X$.

$$UV = \{(x,y) \in X \times X \mid \text{there exists } z \in X, \\ (x,z) \in U \text{ and } (z,y) \in V\}$$

The transposed relation of $U$ is the relation

$${}^tU = \{(x,y) \in X \times X \mid (y,x) \in U\}$$

Finally, for $x \in X$, we set

$$U(x) = \{y \in X \mid (x,y) \in U\}$$

# Uniform spaces

A uniformity on a set $X$ is a nonempty set $\mathcal{U}$ of reflexive relations (entourages) on $X$ such that:

(1) if a relation $U$ on $X$ contains an element of $\mathcal{U}$, then $U \in \mathcal{U}$, (extension property),

(2) the intersection of any two elements of $\mathcal{U}$ is in $\mathcal{U}$, (intersection),

(3) for each $U \in \mathcal{U}$, there exists $V \in \mathcal{U}$ such that $VV \subseteq U$ (sort of transitivity).

(4) for each $U \in \mathcal{U}$, ${}^tU \in \mathcal{U}$ (symmetry).

# $U$-closeness

Two points $x$ and $y$ of $X$ are $U$-close if $(x, y) \in U$. $U$-closeness is a reflexive and symmetrical relation. Further,

(1) if $x$ and $y$ are $U$-close and if $U \subseteq V$, then $x$ and $y$ are $V$-close,

(2) if $x$ and $y$ are $U$-close and $V$-close, then they are $U \cap V$-close,

(3) for each entourage $U$, there is an entourage $V$ such that if $x$ and $z$ are $V$-close and if $z$ and $y$ are $V$-close, then $x$ and $y$ are $U$-close.

A set in which every two points are $U$-close is $U$-small.

# Hausdorff quotient of a uniform structure

The intersection of all entourages is an equivalence relation $\sim$ on $X$. Thus $x \sim y$ iff $x$ and $y$ are $U$-close for each entourage $U$.

The uniform structure on $X$ induces a uniform structure on $X/\!\sim$. The resulting uniform space is the Hausdorff quotient of $X$. Further, the map $\pi : X \to X/\!\sim$ is uniformly continuous.

The intersection of all entourages of $X/\!\sim$ is the diagonal: two points that are $U$-close for each entourage $U$ are equal.

# Completion of a uniform space

A uniform space is complete if every Cauchy filter is converging. Every Hausdorff uniform space $X$ admits a unique completion (up to isomorphism).

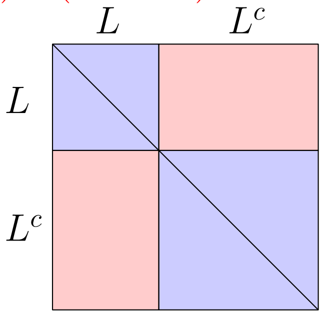
More precisely, $X$ is a dense subspace of a complete Hausdorff uniform space $\widehat{X}$ having the following universal property: every uniformly continuous mapping $\varphi : X \to Y$, where $Y$ is a complete Hausdorff uniform space, has a unique uniformly continuous extension $\widehat{\varphi} : \widehat{X} \to Y$.

# Pervin uniformities

Let $\mathcal{L}$ be a Boolean algebra of subsets of $X$. For each $L \in \mathcal{L}$, consider the entourage

$$V_L = (L \times L) \cup (L^c \times L^c)$$

Two elements $x$ and $y$ are $L$-close iff $x \in L \iff y \in L$.

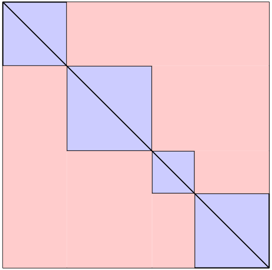



The uniformity generated by the $V_L$, for $L \in \mathcal{L}$, is the Pervin uniformity defined by $\mathcal{L}$.

# Examples

Let $X$ be a finite set. Then the Pervin uniformity defined by $\mathcal{P}(X)$ contains the diagonal and hence is equal to the discrete uniformity.

Let $X$ be an infinite set. The Pervin uniformity defined by $\mathcal{P}(X)$ does not contain the diagonal and hence is different from the discrete uniformity.



# Blocks

A subset $L$ of a uniform space $X$ is a block if $(L \times L) \cup (L^c \times L^c)$ is an entourage. They form a Boolean algebra.

In particular, the blocks of the Pervin uniformity defined by $\mathcal{L}$ are precisely the elements of $\mathcal{L}$.

More generally, if a uniformity $\mathcal{U}$ is generated by a basis $\mathcal{B}$, the blocks of $\mathcal{U}$ are the elements of the Boolean algebra generated by the blocks of $\mathcal{B}$.

# Blocks are the uniform counterpart of clopen sets

Recall that the characteristic function of a subset $L$ of $X$ is the function $\chi_L$ from $X$ to $\{0, 1\}$ defined by

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

## Proposition

*Let $X$ be a topological space. Then $L$ is clopen iff $\chi_L$ is continuous.*
*Let $X$ be a uniform space. Then $L$ is a block iff $\chi_L$ is uniformly continuous.*

# Compactness

A space $X$ is totally bounded if, for each entourage $U$, there is a finite cover of $X$ by $U$-small sets.

**Fact**. The completion of a uniform space is compact iff it is totally bounded.

## Proposition

*Any Pervin space is totally bounded.*

For instance, the completion of $(X, \mathcal{U}_{\mathcal{P}(X)})$ is compact even if $X$ is infinite.

# Uniform continuity

Let $X$ and $Y$ be uniform spaces. A function $\varphi\colon X \to Y$ is uniformly continuous if, for each entourage $V$ of $Y$, $(\varphi \times \varphi)^{-1}(V)$ is an entourage of $X$.

## Proposition

*Let $(X, \mathcal{U}_{\mathcal{K}})$ and $(Y, \mathcal{U}_{\mathcal{L}})$ be two Pervin spaces. A function $\varphi : X \to Y$ is uniformly continuous iff for each $L \in \mathcal{L}$, $\varphi^{-1}(L) \in \mathcal{K}$.*

# Pervin uniformities and duality

**Theorem**

*Let $\mathcal{L}$ be a Boolean algebra of subsets of $X$. The completion of a space $X$ for the Pervin uniformity defined by $\mathcal{L}$ is equal to the Stone dual of $\mathcal{L}$.*

In particular, the completion of $X$ for the Pervin uniformity defined by $\mathcal{P}(X)$ is equal to the Stone-Čech compactification of $X$.

# Part II

## Recognition

# Recognition

Let $M$ be a monoid and let $L$ be a subset of $M$. We say that $L$ is recognized by a surjective morphism of monoid $\varphi : M \to N$ if there is a subset $P$ of $N$ such that $L = \varphi^{-1}(P)$.

By extension, we say that $N$ recognizes $L$ if there exists a morphism $\varphi : M \to N$ that recognizes $L$.

A subset $L$ of $M$ is said to be recognizable if it is recognized by some finite monoid.

# Syntactic monoid of a subset

Let $L$ be a subset of a monoid $M$. The syntactic congruence of $L$ is the relation $\sim_L$ defined on $M$ by: $u \sim_L v$ iff, for all $x, y \in M$,

$$xuy \in L \iff xvy \in L$$

The quotient monoid $M/{\sim_L}$ is called the syntactic monoid of $L$.

**Universal property**. A monoid $N$ recognizes $L$ iff the syntactic monoid of $L$ is a homomorphic image of $N$.

# Subsets recognized by a morphism

Note that if $\varphi : M \to N$ is a morphism, the subsets of $M$ recognized by $\varphi$ form a Boolean algebra.

# Subsets recognized by a morphism

Note that if $\varphi : M \to N$ is a morphism, the subsets of $M$ recognized by $\varphi$ form a Boolean algebra.

Boolean algebra, you said Boolean algebra? Then it has a dual !

# Subsets recognized by a morphism

Note that if $\varphi : M \to N$ is a morphism, the subsets of $M$ recognized by $\varphi$ form a Boolean algebra.

Boolean algebra, you said Boolean algebra? Then it has a dual !

That's the way its all started...

# Duality comes in...

Let $L$ be a subset of a monoid $M$ and let $x, y \in M$. The quotient of $L$ by $x$ and $y$ is the subset

$$x^{-1}Ly^{-1} = \{u \in M \mid xuy \in L\}$$

It was shown in Mai's lecture than the syntactic monoid of a recognizable subset $L$ of $M$ is the dual space of the Boolean algebra generated by the sets $x^{-1}Ly^{-1}$, for $x, y \in M$.

One can actually consider the dual space of any Boolean algebra of subsets of $M$ closed under quotients.

# Translations and quotients

Let $\mathcal{L}$ be a Boolean algebra of subsets of a monoid $M$. Let $\mathcal{U}_{\mathcal{L}}$ be the Pervin uniformity defined by $\mathcal{L}$.

## Proposition

*The translations $x \mapsto xs$ and $x \mapsto sx$ are $\mathcal{U}_{\mathcal{L}}$-uniformly continuous for each $s \in M$ iff $\mathcal{L}$ is closed under quotients.*

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

# Syntactic congruence

Let $\sim_{\mathcal{L}}$ be the relation on $M$ defined by $u \sim_{\mathcal{L}} v$ iff, for all $L \in \mathcal{L}$, for all $x, y \in M$,

$$xuy \in L \iff xvy \in L$$

Then $\sim_{\mathcal{L}}$ is the intersection of all entourages of $\mathcal{L}$.

Thus $M/{\sim_{\mathcal{L}}}$ is the Hausdorff quotient of the Pervin space $(M, \mathcal{U}_{\mathcal{L}})$ and the canonical map $M \to M/{\sim_{\mathcal{L}}}$ is uniformly continuous. But $\sim_{\mathcal{L}}$ is also a monoid congruence and the monoids $M$ and $M/{\sim_{\mathcal{L}}}$ are both semiuniform.

# Minimum recognizer

The semiuniform monoid $M/{\sim_{\mathcal{L}}}$ is called the minimum recognizer of $\mathcal{L}$ in $M$.

To state the universal property of this object, we have to define the category of Pervin monoids. Let $(M, \mathcal{K})$ and $(N, \mathcal{L})$ be two Pervin monoids. A morphism from $M$ to $N$ is a uniformly continuous monoid morphism such that $\varphi(\mathcal{K}) = \mathcal{L}$.

The latter condition is mandatory: uniformly continuous maps do not suffice in this theory.

# Universal property of the minimum recognizer

**Definition**. A semiuniform monoid $N$ recognizes a Boolean algebra $\mathcal{L}$ of subsets of $M$ closed under quotients if there is a surjective morphism of Pervin monoids $\varphi : (M, \mathcal{L}) \to (N, \varphi(\mathcal{L}))$.

This condition implies that the lattices $\mathcal{L}$ and $\varphi(\mathcal{L})$ are isomorphic.

**Universal property**. A semiuniform monoid $N$ recognizes $\mathcal{L}$ iff the minimum recognizer of $\mathcal{L}$ is a homomorphic image of $N$.

# Syntactic space

**Definition**. The syntactic space of a Boolean algebra closed under quotient is the completion of its minimum recognizer.

## Theorem

*The syntactic space of a Boolean algebra closed under quotient is isomorphic to its Stone dual.*

In particular, the syntactic space of a Boolean algebra closed under quotient is always compact.

# The case of a single recognizable set

If $L$ is a recognizable subset of $M$, its syntactic space is finite and its uniform structure is discrete (and hence useless!). It is equal to its completion.

This explains why, for recognizable sets, only the algebraic properties of the syntactic monoid are important.

# Comparing the two approaches

In the classical theory, we just have the algebraic notion of a syntactic monoid, which applies to a single subset of $M$. In our new approach,

- the notion of a minimum recognizer extends that of a syntactic monoid. It can be applied to any Boolean algebra of subsets of $M$.
- It is a topological notion. The minimum recognizer is a Pervin space and its completion, the syntactic space, is always compact.

The minimum recognizer is a semi-uniform monoid, but in general, the product $(u, v) \to uv$ is not uniformly continuous.

# When is the product uniformly continuous?

## Theorem

*Let $\mathcal{L}$ be a Boolean algebra of subsets of $M$ closed under quotients. TFCAE:*

1. *its minimum recognizer is a uniform monoid,*
2. *the closure of the product of its minimum recognizer is functional,*
3. *its syntactic space is a compact monoid,*
4. *the elements of $\mathcal{L}$ are all recognizable.*

# Two examples

## Theorem

*The syntactic space of* $\mathrm{Rec}(M)$ *is the profinite monoid on* $M$*.*

One can define the profinite monoid on $M$ as the projective limit of the directed system all morphisms from $M$ to a finite monoid.

## Theorem

*The syntactic space of* $\mathcal{P}(M)$ *is* $\beta M$*, the Stone-Čech compactification of* $M$*.*

# Another example

Let $M = (\mathbb{Z}, +)$ and let $\mathcal{L}$ be the Boolean algebra of finite or cofinite subsets of $\mathbb{Z}$. The associated Pervin completion of $\mathbb{Z}$ is $\widehat{\mathbb{Z}} = \mathbb{Z} \cup \{\infty\}$: the $\mathbb{Z}$ part corresponds to the principal ultrafilters on $\mathcal{L}$ and $\infty$ is the ultrafilter of cofinite subsets of $\mathbb{Z}$.

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

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

# Extension to lattices

This theory extends from Boolean algebras to lattices of subsets. One needs quasi-uniformities. In particular, the Pervin quasi-uniformity associated with a lattice of subsets $\mathcal{L}$ is generated by the sets

$$V_L = (L^c \times X) \cup (X \times L)$$

for each $L \in \mathcal{L}$.

If $\mathcal{L}$ is closed under quotients, the minimal recognizer is an ordered monoid. For a recognizable language, one gets the syntactic ordered monoid.

# Equations

**Theorem**

*A set of recognizable languages of $A^*$ is a Boolean algebra closed under quotients iff it can be defined by a set of equations of the form $u = v$, where $u, v$ are profinite words.*

**Theorem**

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

# Equations: the recognizable case

If $\mathcal{L}$ is a Boolean algebra of recognizable languages, its syntactic space is by duality a quotient of the syntactic space of $\mathrm{Rec}(A^*)$, i.e. the free profinite monoid $\widehat{A^*}$.

A quotient space is defined by identifying points. Let $(u, v)$ be a pair of elements of $\widehat{A^*}$. We say that $\mathcal{L}$ satisfies the equation $u = v$ if, for all $L \in \mathcal{L}$ and for all $x, y \in A^*$, the conditions $xuy \in \overline{L}$ and $xvy \in \overline{L}$ are equivalent.

This is equivalent to state that $\widehat{\eta}(u) = \widehat{\eta}(v)$, where $\eta : A^* \to M_{\mathcal{L}}$ is the minimum recognizing map of $\mathcal{L}$.

# Equations: the general case

If $\mathcal{L}$ is a Boolean algebra of languages, then its syntactic space is a quotient of the syntactic space of $\mathcal{P}(A^*)$, namely $\beta A^*$.

Let $(u, v)$ be a pair of elements of $\beta A^*$. We say that $\mathcal{L}$ satisfies the equation $u = v$ if, for all $L \in \mathcal{L}$ and for all $x, y \in A^*$ the conditions $xuy \in \overline{L}$ and $xvy \in \overline{L}$ are equivalent.

This is equivalent to state that $\widehat{\eta}(u) = \widehat{\eta}(v)$, where $\eta : A^* \to M_{\mathcal{L}}$ is the minimum recognizing map of $\mathcal{L}$.

# Part III

# Applications to logic

# Logic on words

To each nonempty word $u = a_1 \cdots a_n$ is associated a structure

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

where $\mathbf{a}$ is a predicate symbol interpreted as the set of positions $i$ such that the $i$-th letter of $u$ is an $a$.

If $u = abbaab$, then $\mathrm{Dom}(u) = \{1, 2, 3, 4, 5, 6\}$, $\mathbf{a} = \{1, 4, 5\}$ and $\mathbf{b} = \{2, 3, 6\}$.

We also use the relation symbol $<$ with its usual interpretation on the integers.

# Some examples

The language defined by a sentence $\varphi$ is

$$L(\varphi) = \{u \in A^* \mid \mathcal{M}_u \text{ satisfies } \varphi\}$$

For instance the sentence $\exists x\ \mathbf{a}x$ defines the language $A^*aA^*$.

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

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

# Characterization of some logical fragments

**Theorem** [Büchi 1960, Elgot 1961]
A language is $\mathbf{MSO}[<, \mathbf{a}]$-definable iff it is recognizable.

**Def**. If $x$ is a profinite word, then the sequence $x^{n!}$ is Cauchy and converges to a profinite word $x^{\omega}$.

**Theorem** [Schützenberger 65 + McNaughton 71]
A language is $\mathbf{FO}[<, \mathbf{a}]$-definable iff its syntactic monoid satisfies the profinite equation $x^{\omega+1} = x^{\omega}$.

**Corollary**. One can effectively decide whether a given recognizable language is $\mathbf{FO}[<, \mathbf{a}]$-definable.

# Aperiodic and quasi-aperiodic monoids

Let $\mathbf{MOD}$ be the set of modular predicates, e.g. $x \equiv 1 \bmod 6$.

**Theorem** [Barrington et al. 1992]
A language is $\mathbf{FO}[<, \mathbf{MOD}, \mathbf{a}]$-definable iff its syntactic monoid satisfies the profinite equation $(x^{\omega-1}y)^{\omega} = (x^{\omega-1}y)^{\omega+1}$ for all words $x, y$ of the same length.

# 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 $\mathsf{AC}^0$, 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

It is also known that

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

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

This would permit to attack difficult conjectures in circuit complexity.