A topological approach to recognition — **Source link**

Mai Gehrke, Serge Grigorieff, Jean-Eric Pin

Institutions: Radboud University Nijmegen, Paris Diderot University

Topics: Syntactic monoid, Free monoid, Abstract family of languages, Trace theory and Monoid

Related papers:

- Duality and Equational Theory of Regular Languages

- Regular languages and stone duality

- Profinite methods in automata theory

- Finite Automata, Formal Logic, and Circuit Complexity

- Finite Semigroups and Universal Algebra

A Topological Approach to Recognition

Mai Gehrke, Serge Grigorieff, Jean-Eric Pin

A topological approach to recognition

Mai Gehrke[1], Serge Grigorieff[2], and Jean-Éric Pin[2][⋆]

[1] Radboud University Nijmegen, The Netherlands
[2] LIAFA, University Paris-Diderot and CNRS, France.

Abstract. We propose a new approach to the notion of recognition, which departs from the classical definitions by three specific features. First, it does not rely on automata. Secondly, it applies to any Boolean algebra (BA) of subsets rather than to individual subsets. Thirdly, topology is the key ingredient. We prove the existence of a *minimum recognizer* in a very general setting which applies in particular to any BA of subsets of a discrete space. Our main results show that this minimum recognizer is a uniform space whose completion is the dual of the original BA in Stone-Priestley duality; in the case of a BA of languages closed under quotients, this completion, called the *syntactic space* of the BA, is a compact monoid if and only if all the languages of the BA are regular. For regular languages, one recovers the notions of a syntactic monoid and of a free profinite monoid. For nonregular languages, the syntactic space is no longer a monoid but is still a compact space. Further, we give an equational characterization of BA of languages closed under quotients, which extends the known results on regular languages to nonregular languages. Finally, we generalize all these results from BAs to lattices, in which case the appropriate structures are partially ordered.

Recognizability is one of the most fruitful concepts in computer science [16, 3]. Originally introduced for finite words, it has been successfully extended to infinite words, general algebras, finite and infinite trees and to many other structures. Roughly speaking, a set is recognizable if it is saturated by a congruence of finite index. A desirable property for any notion of recognition is the existence of a minimum recognizer. Here, the word "minimum" does not refer to algorithmic properties, like the minimal number of states of an automaton, but rather to a final object in a suitable category. For instance, there is a well defined notion of morphism of deterministic automata. Final objects exist in this category and are just the usual minimal automata. But no such notion is known for automata on infinite words, even for ω-regular languages. This problem has been overcome by using another type of recognizer. For languages of finite words, automata can be replaced by monoids, for which there is a minimal recognizer, the syntactic monoid. For ω-regular languages, ω-semigroups are a category in which minimal recognizers do exist again [9, 17].

The aim of this paper is to propose a general definition of recognition for which each object has a *unique minimum recognizer*. Our approach departs from the classical definitions of recognition by three specific features:

(1) it does not rely at all on automata;

(2) it applies to Boolean algebras or more generally to lattices of subsets rather than to individual subsets;

(3) topology, and in particular Stone-Priestley duality, is the key ingredient.

Our most general definition is given in the category of uniform spaces, an abstraction of the notion of metric spaces well-known to topologists. We predominantly consider discrete spaces where the topology itself carries no valuable information. However, an appropriate uniformity gives rise by completion to a compact space, a common practice in mathematics, where it is often said that "compactness is the next best thing to finiteness".

We develop a notion of recognition in this general setting and prove that any Boolean algebra of subsets of a uniform space admits a *minimum recognizer*, which is again a uniform space, whose completion is the dual space of the original Boolean algebra in the sense of Stone-Priestley duality.

When the uniform space carries an algebraic structure for which the operations are at least separately uniformly continuous, it is natural to require that the recognizing maps be morphisms for the algebraic structure as well. In the case of a monoid, this amounts to working in the category of semiuniform monoids and imposes closure under quotients of the Boolean algebra. The minimum recognizer is then a semiuniform monoid whose completion is called the *syntactic space* of the Boolean algebra with quotients. We prove that this syntactic space is a compact monoid if and only if all the subsets of the Boolean algebra are recognizable. For a regular language, one recovers the classical notion of a syntactic monoid. For a variety of regular languages, one obtains the free profinite monoid of the corresponding variety of monoids. For nonregular languages, the syntactic space is no longer a monoid but is still a compact space. We also prove that any Boolean algebra of languages closed under quotient has an equational description. Finally, we generalize all these results from Boolean algebras to lattices and recover in this way the notion of a syntactic ordered monoid.

1 The topological setting

In this section, we recall the notions of topology needed to read this paper. We invite the reader to look at Wikipedia for an introduction and suitable references (notably the entries *uniform spaces* and *Stone-Čech compactification*).

Let X be a set. We denote by L^c the complement of a subset L of X. The subsets of $X \times X$ can be viewed as relations on X. Given a relation U, the *transposed relation* of U is the relation ${}^tU = \{(x, y) \in X \times X \mid (y, x) \in U\}$. We denote by UV the composition of two relations U and V on X. Thus

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

Finally, if $x \in X$ and $U \subseteq X \times X$, we write $U(x)$ for the set $\{y \in X \mid (x, y) \in U\}$.

1.1 Uniform spaces

Uniform spaces generalize metric spaces and enable the formalization of the notion of relative closeness. See [4] for a thorough discussion of these notions and [12] for a connection with Eilenberg's variety theory.

A *uniformity* on a set X is a nonempty set $\mathcal{U}$ of subsets of $X \times X$ satisfying the following properties:

(1) if a subset U of $X \times X$ contains an element of $\mathcal{U}$, then $U \in \mathcal{U}$,

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

(3) each element of $\mathcal{U}$ contains the diagonal of $X \times X$,

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

(5) for each $U \in \mathcal{U}$, ${}^t U \in \mathcal{U}$.

The elements of a uniformity are called *entourages*. The pair $(X, \mathcal{U})$ (or the set X if $\mathcal{U}$ is understood) is called a *uniform space*. The *discrete uniformity* on X is the unique uniformity which contains the diagonal of $X \times X$.

A *basis* of a uniformity $\mathcal{U}$ is a subset $\mathcal{B}$ of $\mathcal{U}$ such that each element of $\mathcal{U}$ contains an element of $\mathcal{B}$. In particular, $\mathcal{U}$ consists of all the relations on X containing an element of $\mathcal{B}$. We say that $\mathcal{U}$ is *generated* by $\mathcal{B}$.

A *subbasis* of a uniformity $\mathcal{U}$ is a subset $\mathcal{B}$ of $\mathcal{U}$ such that the finite intersections of members of $\mathcal{B}$ form a basis of $\mathcal{U}$.

If X and Y are uniform spaces, a function $\varphi \colon X \to Y$ is said to be *uniformly continuous* if, for each entourage V of Y, $(\varphi \times \varphi)^{-1}(V)$ is an entourage of X, or, equivalently, if for each entourage V of Y, there exists an entourage U of X such that $(\varphi \times \varphi)(U) \subseteq V$.

1.2 Pervin uniformities

Pervin uniformities were first introduced in [10]. They play a key role in our definition of recognition given in Section 2.

For each subset L of a set X, consider the equivalence relation U_L on X:

$$U_L = (L \times L) \cup (L^c \times L^c) = \{(x,y) \in X \times X \mid x \in L \iff y \in L\}$$

Let $\mathcal{S}$ be a collection of subsets of X. The sets of the form U_L, for $L \in \mathcal{S}$, form the subbasis of a uniformity of X, called the *Pervin uniformity defined by $\mathcal{S}$* and denoted by $\mathcal{U}_{\mathcal{S}}$. The space $(X, \mathcal{U}_{\mathcal{S}})$ is called a *Pervin space*.

1.3 Induced topology

Let $\mathcal{U}$ be a uniformity on X. For each $x \in X$, let $\mathcal{U}(x) = \{U(x) \mid U \in \mathcal{U}\}$. There exists a unique topology on X, called the *topology induced* by $\mathcal{U}$, for which $\mathcal{U}(x)$ is the set of neighborhoods of x for each $x \in X$.

If $\mathcal{U}$ is a uniformity on X, the intersection of all its entourages is an equivalence relation $\sim$ on X, which is equal to the diagonal if and only if the topology induced by $\mathcal{U}$ is Hausdorff. Further, if $\pi : X \to X/\sim$ denotes the quotient map,

then the sets of the form $(\pi \times \pi)(U)$, where U is an entourage of X, form a subbasis of a uniformity on $X/\sim$. This map is uniformly continuous and the Hausdorff uniform space $X/\sim$ is called the *Hausdorff quotient* of X.

Let $\mathcal{L}$ be a Boolean algebra of subsets of X. In the topology induced by the Pervin uniformity $\mathcal{U}_{\mathcal{L}}$, the neighborhoods of x are the supersets of the sets of $\mathcal{L}$ containing x.

1.4 Filters, ultrafilters and Stone duality

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

(1) the empty set does not belong to $\mathcal{F}$,

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

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

An *ultrafilter* is a filter which is maximal for the inclusion. Recall that a filter is an ultrafilter if and only if, for every $L \in \mathcal{L}$, either $L \in \mathcal{F}$ or $L^c \in \mathcal{F}$. Let $S(\mathcal{L})$ be the set of ultrafilters of $\mathcal{L}$. For each $L \in \mathcal{L}$, let ω_L be the set of ultrafilters containing L. The set $\Omega = \{\omega_L \mid L \in \mathcal{L}\}$ is a Boolean algebra of subsets of $S(\mathcal{L})$, which defines a Pervin uniformity $\mathcal{U}_\Omega$. The uniform space $(S(\mathcal{L}), \mathcal{U}_\Omega)$ is called the *Stone dual space* of $\mathcal{L}$ and $S(\mathcal{L})$ with the induced topology is the Stone dual space of $\mathcal{L}$ in the topological sense. Note that Ω is a basis of clopen sets for the topology of $S(\mathcal{L})$ (recall that a set is *clopen* if it is both open and closed).

If X is a space, a filter of $\mathcal{P}(X)$ is simply called *a filter on X*. A filter $\mathcal{F}$ *converges* to a point x of X if, for each neighborhood U of x, there is a set B of $\mathcal{F}$ such that $B \subseteq U$. In this case, x is called a *limit* of $\mathcal{F}$.

1.5 Blocks

A subset L of a uniform space X is a *block* if U_L is an entourage. Intuitively, blocks are to uniformities what clopen sets are to topologies. Recall that the *characteristic function* of a subset L of X is the function χ_L from X to $\{0, 1\}$ defined by $\chi_L(x) = 1$ if $x \in L$ and $\chi_L(x) = 0$ if $x \in L^c$.

Proposition 1.1. *Let X be a topological [uniform] space. A subset of X is clopen [a block] if and only if its characteristic function is a [uniformly] continuous function from X to the discrete [uniform] space $\{0, 1\}$.*

In the same way, Pervin uniformities are the uniform analogs of zero-dimensional topologies: a topology is zero-dimensional [a uniformity is Pervin] if and only if it has a basis of clopen sets [blocks].

Proposition 1.2. *The blocks of a uniform space form a Boolean subalgebra of the Boolean algebra formed by the clopen sets. These two Boolean algebras coincide if the space is compact.*

The next result gives a simple description of the blocks of a Pervin uniformity.

Proposition 1.3. *Let $\mathcal{S}$ be a collection of subsets of X. The blocks of the Pervin uniformity defined by $\mathcal{S}$ form a Boolean algebra, equal to the Boolean algebra $\mathcal{L}$ generated by $\mathcal{S}$. Further, $\mathcal{S}$ and $\mathcal{L}$ define the same Pervin uniformity.*

1.6 Hausdorff completion of a uniform space

Let X be a uniform space. A filter $\mathcal{F}$ on X is a *Cauchy filter* if, for every entourage U, there exists $B \in \mathcal{F}$ with $B \times B \subseteq U$. In Pervin spaces, Cauchy filters have a simple description.

Proposition 1.4. *Let $\mathcal{L}$ be a Boolean algebra of subsets of X. A filter $\mathcal{F}$ is Cauchy for the Pervin uniformity defined by $\mathcal{L}$ if and only if, for every $L \in \mathcal{L}$, either $L \in \mathcal{F}$ or $L^c \in \mathcal{F}$.*

A uniform space X is *complete* if every Cauchy filter converges. Every uniform space admits a unique *Hausdorff completion* [4]. More precisely, let X be a uniform space. Then there exist a complete Hausdorff uniform space $\widehat{X}$ and a uniformly continuous mapping $\imath \colon X \to \widehat{X}$ such that $\imath(X)$ is dense in $\widehat{X}$ and the following universal property holds: for each uniformly continuous mapping φ from X into a complete Hausdorff uniform space Y, there exists a unique uniformly continuous mapping $\hat{\varphi} \colon \widehat{X} \to Y$ such that $\hat{\varphi} \circ \imath = \varphi$. The image of X under $\imath$ is the Hausdorff quotient of X and $\imath$ is thus injective if and only if X is Hausdorff.

A uniform space is said to be *totally bounded* if, for each entourage U, there exists a finite cover $(B_i)_{i \in F}$ of X such that, for all $i \in F$, $B_i \times B_i$ is a subset of U. The interest of totally bounded uniformities lies in the following result [4, TG.II.29, Thm. 3]: a uniform space is totally bounded if and only if its completion is compact. This result applies in particular to Pervin spaces. Indeed, the subbasic entourages of the form U_L contain the sets $L \times L$ and $L^c \times L^c$ and $\{L, L^c\}$ is a finite cover of X. Thus we obtain

Proposition 1.5. *Any Pervin space is totally bounded.*

1.7 Pervin completions

We now show that complete Pervin spaces are the uniform analogs of topological Stone spaces. Let $\mathcal{L}$ be a Boolean algebra of subsets of X and let $\mathcal{U}_{\mathcal{L}}$ be the Pervin uniformity defined by $\mathcal{L}$. By Proposition 1.5, the completion $\widehat{X}$ of X for this uniformity is compact. It consists of all *Cauchy filters* on X that are *minimal* for the inclusion order on filters. For each $x \in X$, $\imath(x)$ is the filter of neighborhoods of x in X and $\widehat{X}$ is equipped with the Pervin uniformity defined by the sets $\{\mathcal{F} \in \widehat{X} \mid L \in \mathcal{F}\}$, for each $L \in \mathcal{L}$. For this reason, we call $\widehat{X}$ the *Pervin completion of X with respect to $\mathcal{L}$.*

We now give an alternative description of $\widehat{X}$. If $\mathcal{G}$ is a filter on the Boolean algebra $\mathcal{L}$, we denote by $\uparrow\mathcal{G}$ the filter on X generated by $\mathcal{G}$.

Theorem 1.6. *The space $\widehat{X}$ is the dual space of $\mathcal{L}$ as defined in Section 1.4. Further, the maps $\mathcal{F} \mapsto \mathcal{F} \cap \mathcal{L}$ and $\mathcal{G} \mapsto\, \uparrow\! \mathcal{G}$ define mutually inverse uniformly continuous bijections between the set of minimal Cauchy filters on X and the set of ultrafilters of $\mathcal{L}$.*

The next series of results gives another natural correspondence between the blocks of X and the clopen sets of $\widehat{X}$. We shall use freely the following notation. Let χ_L be the characteristic function of some block L. By Proposition 1.1, χ_L is uniformly continuous. The universal property of the completion ensures that there is a unique uniformly continuous map $\widehat{\chi}_L : \widehat{X} \to \{0, 1\}$ such that $\widehat{\chi}_L \circ \imath = \chi_L$. If L is a subset of X, we set $\tilde{L} = \imath^{-1}(\imath(L))$.

Theorem 1.7. *One has $\tilde{L} \in \mathcal{L}$ if and only if $\overline{\imath(L)}$ is clopen and $\overline{\imath(L)} \cap \imath(X) = \imath(L)$. If these conditions are satisfied, then $\imath(L) = \widehat{\chi}_{\tilde{L}}^{-1}(1)$.*

If the space $(X, \mathcal{U}_{\mathcal{L}})$ is Hausdorff, then $\imath$ is the identity map and Theorem 1.7 can be simplified as follows:

Corollary 1.8. *Suppose that $(X, \mathcal{U}_{\mathcal{L}})$ is Hausdorff. Then $L \in \mathcal{L}$ if and only if $\overline{L}$ is clopen and $\overline{L} \cap X = L$. If these conditions are satisfied, then $\overline{L} = \widehat{\chi}_L^{-1}(1)$.*

Let us denote by $\mathrm{Clopen}(\widehat{X})$ the Boolean algebra of all clopen sets of $\widehat{X}$.

Theorem 1.9. *The maps $L \mapsto \overline{\imath(L)}$ and $K \mapsto \imath^{-1}(K \cap \imath(X))$ define mutually inverse isomorphisms between the Boolean algebras $\mathcal{L}$ and $\mathrm{Clopen}(\widehat{X})$. In particular, the following formulas hold, for all $L, L_1, L_2 \in \mathcal{L}$:*

(1) $\overline{\imath(L^c)} = \overline{\imath(L)}^c$,
(2) $\overline{\imath(L_1 \cup L_2)} = \overline{\imath(L_1)} \cup \overline{\imath(L_2)}$,
(3) $\overline{\imath(L_1 \cap L_2)} = \overline{\imath(L_1)} \cap \overline{\imath(L_2)}$,

2 Recognition in a uniform space

In this section, we define a notion of recognition for Boolean algebras of blocks of a uniform space. In spite of the analogy between clopen sets and blocks illustrated by Proposition 1.1, our definition can only be reformulated in terms of clopen sets and continuous maps if one moves to the Hausdorff completions. Indeed, there are nontrivial Pervin uniformities inducing the discrete topology.

2.1 Recognition of a Boolean algebra

Let $(X, \mathcal{U})$ be a uniform space and let $\mathcal{L}$ be a Boolean algebra of blocks of X.

Definition 1. A [surjective] uniformly continuous map φ from X into a uniform space Y [fully] *recognizes* $\mathcal{L}$ if, for each $L \in \mathcal{L}$, $\varphi(L)$ is a block of Y and $\varphi^{-1}(\varphi(L)) = L$.

Let us state two important consequences of this definition. The first one shows that recognition preserves the Boolean structure and the second one is a transitivity property.

Proposition 2.1. *If φ recognizes $\mathcal{L}$, then φ induces an isomorphism of Boolean algebras between $\mathcal{L}$ and $\varphi(\mathcal{L})$.*

Proposition 2.2. *Let X, Y and Z be three uniform spaces and let $\mathcal{L}$ be a Boolean algebra of blocks of X. If $\mathcal{L}$ is [fully] recognized by $\varphi : X \to Y$ and $\varphi(\mathcal{L})$ is [fully] recognized by $\gamma : Y \to Z$, then $\mathcal{L}$ is [fully] recognized by $\gamma \circ \varphi$.*

2.2 Minimum recognizer of a Boolean algebra

If $\mathcal{L}$ is a Boolean algebra of blocks of X, the Pervin uniformity $\mathcal{U}_{\mathcal{L}}$ is contained in $\mathcal{U}$ and therefore, the identity on X is a uniformly continuous map from $(X, \mathcal{U})$ onto $(X, \mathcal{U}_{\mathcal{L}})$. The canonical map from $(X, \mathcal{U}_{\mathcal{L}})$ onto its Hausdorff quotient $X_{\mathcal{L}}$ is surjective and uniformly continuous. The composition of these two maps yields a surjective and uniformly continuous $\eta_{\mathcal{L}}$ from X onto $X_{\mathcal{L}}$, called the *minimum recognizing map* of $\mathcal{L}$. The space $X_{\mathcal{L}}$ is called the *minimum recognizer* of $\mathcal{L}$. This terminology is justified by the following universal property:

Proposition 2.3. *The map $\eta_{\mathcal{L}} : X \to X_{\mathcal{L}}$ fully recognizes $\mathcal{L}$. Further, if $\varphi : X \to Y$ fully recognizes $\mathcal{L}$, there exists a unique surjective uniformly continuous map $\pi : Y \to X_{\mathcal{L}}$ fully recognizing the Boolean algebra $\varphi(\mathcal{L})$ and such that $\pi \circ \varphi = \eta_{\mathcal{L}}$. Further, $\eta(\mathcal{L})$ is the set of blocks of $X_{\mathcal{L}}$.*

If X is a discrete space, then every subset is a block. It means that every Boolean algebra $\mathcal{L}$ of subsets of X admits a minimum recognizer. In this case, the space $X_{\mathcal{L}}$ is simply the quotient of X under the equivalence relation $\sim_{\mathcal{L}}$ defined by $u \sim_{\mathcal{L}} v$ if and only if, for each $L \in \mathcal{L}$, the conditions $u \in \mathcal{L}$ and $v \in \mathcal{L}$ are equivalent. But of course, the interesting part is the uniform structure of $X_{\mathcal{L}}$, inherited from the Pervin uniformity $\mathcal{U}_{\mathcal{L}}$ on X, which is in general nontrivial.

2.3 Recognition with additional algebraic structure

The notion of recognition originates in the setting of monoids and automata. Recall that a subset of a monoid is *recognizable* if its syntactic monoid is finite. Equivalently, a subset is recognizable if and only if it has only finitely many distinct quotients. We now consider additional algebraic structure on uniform spaces. For this purpose, we require the operations to be separately uniformly continuous for each variable and the recognizing maps be morphisms for the algebraic structure. Theorem 2.4 below shows that this condition forces some structural conditions upon the Boolean algebra being recognized. Due to the lack of space, we only treat the case of monoids, which is sufficient to illustrate our ideas.

Let us define a *semiuniform monoid* as a monoid M equipped with a uniform structure for which the translations $x \mapsto xs$ and $x \mapsto sx$ are uniformly continuous for each $s \in M$. If the multiplication is jointly uniformly continuous, that is, if the map $(x, y) \mapsto xy$ is uniformly continuous, M is a *uniform monoid*.

Let L be a subset of M and let s and t be elements of M. The *quotient* $s^{-1}Lt^{-1}$ of L by s and t is defined by the formula $s^{-1}Lt^{-1} = \{x \in M \mid sxt \in L\}$. We can now state:

Theorem 2.4. *Let $\mathcal{L}$ be a Boolean algebra of blocks of M. Then $\mathcal{L}$ is closed under quotients if and only if $(M, \mathcal{U}_{\mathcal{L}})$ is a semiuniform monoid. If these conditions are satisfied, then the relation $\sim_{\mathcal{L}}$ is a congruence of monoids.*

Under the conditions of Theorem 2.4, the minimum recognizer $M_{\mathcal{L}}$ of $\mathcal{L}$ is a semiuniform monoid. This allows us to link our definition with the standard definition of a syntactic monoid. Suppose that M is a discrete monoid and let L be a subset of M. Consider the smallest Boolean algebra $\mathcal{L}$ containing L and closed under quotients. Since quotients commute with Boolean operations, this is also the Boolean algebra generated by the quotients of L. Now, the equivalence $\sim_{\mathcal{L}}$ defined at the end of Section 2.2 is the *syntactic congruence* of L, defined by $u \sim_L v$ if and only if, for all $x, y \in M$, the conditions $xuy \in L$ and $xvy \in L$ are equivalent. In summary, we obtain:

Proposition 2.5. *The minimum recognizer of the Boolean algebra generated by a set L and its quotients is the usual syntactic monoid of L enriched with a uniform structure which makes the translations uniformly continuous.*

3 Syntactic monoid and syntactic space

We are now ready to introduce our second main definition.

Definition 2. Let $\mathcal{L}$ be a Boolean algebra of blocks of X. The completion of its minimum recognizer is called the *minimum space* of $\mathcal{L}$. If $\mathcal{L}$ is closed under quotients the minimum recognizer will be called the *syntactic monoid*, the minimum space the *syntactic space*, and the minimal recognizing map the *syntatic morphism*.

Since $X_{\mathcal{L}}$ is Hausdorff, it can be identified with a subset of $\widehat{X}_{\mathcal{L}}$. The minimum space has a universal property similar to that of the minimum recognizer. Let us say that a map $\varphi : X \to Y$ is *dense* if $\varphi(X)$ is dense in Y. A map *densely recognizes* $\mathcal{L}$ if it recognizes $\mathcal{L}$ and is dense. Then $\eta_{\mathcal{L}} : X \to \widehat{X}_{\mathcal{L}}$ is minimum among Hausdorff complete recognizers, in the following sense:

(1) $\eta_{\mathcal{L}} : X \to \widehat{X}_{\mathcal{L}}$ densely recognizes $\mathcal{L}$;

(2) if Y is an Hausdorff complete space and if $\varphi : X \to Y$ densely recognizes $\mathcal{L}$, there is a unique surjective uniformly continuous map $\pi : Y \to \widehat{X}_{\mathcal{L}}$ such that $\pi \circ \varphi = \eta_{\mathcal{L}}$.

The results of Section 1.7 show that the syntactic space is precisely the dual space of $\mathcal{L}$. Although the minimum recognizer $M_{\mathcal{L}}$ is a monoid, the product on $M_{\mathcal{L}}$ is not in general uniformly continuous and the completion of $M_{\mathcal{L}}$ is not in general a monoid. More precisely, the closure of the product is in general a ternary relation, as shown in Example 3.2.

Let us give a few examples of syntactic spaces.

Example 3.1. Let $\mathcal{P}(M)$ be the Boolean algebra of all subsets of a monoid M. Its minimum recognizer is M and its syntactic space is the Stone-Čech compactification of M, traditionally denoted by βM, and studied in detail in [7].

As foretold by Theorem 4.1 below, the closure in βM of the product on M is not a binary operation any longer. As a function from $\beta M \times \beta M$ into $\mathcal{P}(\beta M)$ it maps a pair of ultrafilters (p, q) to the family of ultrafilters extending the filter of supersets of $\{XY \mid X \in p, Y \in q\}$.

Example 3.2. The minimal recognizer of the language $\{u \in \{a, b\}^* \mid |u|_a = |u|_b\}$ is $(\mathbb{Z}, +)$. Its syntactic image is $\{0\}$ and its quotients are the sets $\{n\}$, where $n \in \mathbb{Z}$. The Boolean algebra generated by these quotients is the set of finite or cofinite subsets of $\mathbb{Z}$. The associated Pervin completion of $\mathbb{Z}$ is $\mathbb{Z} \cup \{\infty\}$: $\mathbb{Z}$ corresponds to the principal ultrafilters on $\mathcal{L}$ and ∞ is the ultrafilter of cofinite subsets of $\mathbb{Z}$.

The minimal recognizer of the language $\mathrm{MAJORITY} = \{u \in \{a, b\}^* \mid |u|_a \geqslant |u|_b\}$ is also $(\mathbb{Z}, +)$ and its syntactic image is $[0, +\infty[$. Its quotients are $[n, +\infty[$, with $n \in \mathbb{Z}$, and they generate the Boolean algebra of all subsets of $\mathbb{Z}$ having a finite symmetric difference with one of the sets $\emptyset$, $\mathbb{Z}$, $-\mathbb{N}$ or $\mathbb{N}$. The associated Pervin completion of $\mathbb{Z}$ is $\mathbb{Z} \cup \{-\infty, +\infty\}$: again $\mathbb{Z}$ corresponds to the principal ultrafilters, and $+\infty$ $[-\infty]$ is the ultrafilter containing the cofinite subsets of $\mathbb{N}$ $[-\mathbb{N}]$.

Let us denote by $\widehat{\mathbb{Z}}$ the completion of $\mathbb{Z}$. Thus $\widehat{\mathbb{Z}} = \mathbb{Z} \cup \{\infty\}$ in the first case, and $\widehat{\mathbb{Z}} = \mathbb{Z} \cup \{-\infty, +\infty\}$ in the second case. The tables below describe the closure $\widehat{+}$ of the addition on $\mathbb{Z}$ for these two completions. These are not binary operations any longer. We give them as functions from $\widehat{\mathbb{Z}} \times \widehat{\mathbb{Z}}$ into $\mathcal{P}(\widehat{\mathbb{Z}})$.

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

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

4 Recognizable Boolean algebras

When $\mathcal{L}$ is a Boolean algebra of regular languages of A^* its syntactic space is the *profinite monoid* attached to $\mathcal{L}$, in the sense of [1, 6]. In particular, this syntactic space is a *compact monoid*, that is, a compact space equipped with a monoid operation which is jointly continuous. This is in fact characteristic, as we show in the following theorem. Let us say that a Boolean algebra of blocks of M is *recognizable* if all its members are recognizable.

Theorem 4.1. *Let $\mathcal{L}$ be a Boolean algebra of blocks of M closed under quotients. The following conditions are equivalent:*
  (1) *the syntactic monoid of $\mathcal{L}$ is a uniform monoid,*
  (2) *the syntactic space of $\mathcal{L}$ is a compact monoid,*
  (3) *the closure of the operation on the syntactic monoid of $\mathcal{L}$ is functional,*
  (4) *all the languages of $\mathcal{L}$ are recognizable.*

Example 4.1. In this example, we consider three Boolean algebras on $\mathbb{Z}$.

Let $\mathrm{Rec}(\mathbb{Z})$ be the Boolean algebra of all recognizable subsets of $\mathbb{Z}$, that is, the family of finite unions of congruence classes $\{a + n\mathbb{Z} \mid n \geqslant 1, 0 \leqslant a < n\}$. Its minimal recognizer is $\mathbb{Z}$ and its syntactic space $\widehat{\mathbb{Z}}$ is the free profinite group on one generator. Note that in this case, the closure of the addition is the usual addition on $\widehat{\mathbb{Z}}$. In the sequel, we denote by $i \mapsto \underline{i}$ the natural embedding from $\mathbb{Z}$ into $\widehat{\mathbb{Z}}$ and by $+$ the addition on $\widehat{\mathbb{Z}}$.

Consider now the Boolean algebra $\mathcal{L}$ generated by $\mathrm{Rec}(\mathbb{Z})$ and by the finite subsets of $\mathbb{Z}$. Its minimal recognizer is $\mathbb{Z}$ and its syntactic space is the disjoint union $\mathbb{Z} \cup \widehat{\mathbb{Z}}$: $\mathbb{Z}$ consists of the principal ultrafilters of $\mathcal{L}$ and the profinite group $\widehat{\mathbb{Z}}$ corresponds to the nonprincipal ones. The closure $\widehat{+}$ of the addition on $\mathbb{Z}$ is commutative but nonfunctional. It extends $+$ on $\mathbb{Z}$ and is such that, for $i \in \mathbb{Z}$ and $u, v \in \widehat{\mathbb{Z}}$, $i \mathbin{\widehat{+}} u = \underline{i} + u$ and $u \mathbin{\widehat{+}} v$ is $\{k, \underline{k}\}$ if $u + v = \underline{k}$ for some $k \in \mathbb{Z}$ and is $\{u + v\}$ otherwise.

Finally, let $\mathrm{Rat}(\mathbb{Z})$ be the Boolean algebra of rational subsets of $\mathbb{Z}$. Its minimal recognizer is again $\mathbb{Z}$ and its syntactic space is the disjoint union $(\widehat{\mathbb{Z}} \times \{-\}) \cup \mathbb{Z} \cup (\widehat{\mathbb{Z}} \times \{+\})$: as before $\mathbb{Z}$ stands for the principal ultrafilters and, for $\varepsilon \in \{+, -\}$, $\widehat{\mathbb{Z}} \times \{\varepsilon\}$ consists of the nonprincipal ones which contain $\varepsilon\mathbb{N}$. The closure $\widehat{+}$ of the addition on $\mathbb{Z}$ is commutative but nonfunctional. It extends the addition on $\mathbb{Z}$ and on each copy of $\widehat{\mathbb{Z}}$. Further, for $i \in \mathbb{Z}$ and $u, v \in \widehat{\mathbb{Z}}$, one has $i \mathbin{\widehat{+}} (u, \varepsilon) = (\underline{i} + u, \varepsilon)$ and

$$(u, +) \mathbin{\widehat{+}} (v, -) = \begin{cases} \{(\underline{k}, -), k, (\underline{k}, +)\} & \text{if } u + v = \underline{k} \text{ for some } k \in \mathbb{Z} \\ \{(u + v, -), (u + v, +)\} & \text{otherwise} \end{cases}$$

5 Equational theory

Let A^* be a free monoid. We consider A^* as a uniform space, endowed with the Pervin uniformity defined by $\mathcal{P}(A^*)$. As we have seen, its completion is βA^*, the Stone-Čech compactification of A^*. Let $\mathcal{L}$ be a Boolean algebra of languages of A^* closed under quotients and let $\eta : A^* \to M_{\mathcal{L}}$ be its syntactic morphism. Then η extends uniquely to a uniformly continuous map $\widehat{\eta} : \beta A^* \to \widehat{M_{\mathcal{L}}}$. We denote by $\overline{L}$ the closure in βA^* of a language L of A^*.

Formally, an *equation* is a pair (u, v) of elements of βA^*. 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 stating that $\widehat{\eta}(u) = \widehat{\eta}(v)$. Given a set E of equations, the set of languages *defined by* E is the set of all languages of A^* satisfying all the equations of E. We are now ready to state our equational characterization of Boolean algebras closed under quotients, which extends the results of [6] to nonregular languages.

Theorem 5.1. *A set of languages of A^* is a Boolean algebra of languages closed under quotients if and only if it can be defined by a set of equations of the form $u = v$, where $u, v \in \beta A^*$.*

6 Lattices and ordered structures

Let us briefly indicate how to generalize these definitions and results to lattices. It consists mainly in extending the ideas developed in [11, 12]. We treat directly the case of a monoid M.

Let $\mathcal{L}$ be a lattice of subsets of M. Define the *specialization preorder* on M given by $\mathcal{L}$ as the relation $\preccurlyeq$ defined by $x \preccurlyeq y$ if and only if for all $L \in \mathcal{L}$,

$$y \in L \implies x \in L$$

We define the preordered Pervin uniformity given by $\mathcal{L}$ to be $(M, \preccurlyeq, \mathcal{U}_\mathcal{L})$. The minimum recognizer of $\mathcal{L}$ is the Hausdorff quotient $(M_\mathcal{L}, \leqslant)$ with $\leqslant \, = \, \preccurlyeq /{\sim}$, where $\sim$ is the Hausdorff equivalence defined in Section 1.3 and the minimum space of $\mathcal{L}$ is the ordered completion $(\widehat{M}_\mathcal{L}, \leqslant)$ of $(M_\mathcal{L}, \leqslant)$. The latter is the uniform space version of the Priestley space of $\mathcal{L}$ [13]. If the syntactic space is a monoid, we obtain an ordered monoid. For instance, if L is a regular language, one gets the syntactic ordered monoid of L as defined in [11]. The order of the syntactic monoid of the language MAJORITY considered in Example 3.2 is the usual order on $\mathbb{Z}$.

7 Conclusion and perspectives

We have developed a topological approach to the notion of recognition which seems general enough to be applied not only to monoids but to more general algebras, notably those having finitely determined syntactic congruences in the sense of [5], including vector spaces, Boolean algebras, groups and rings.

Let us come back to finite words. The notion of a syntactic monoid has been extremely successful for regular languages and has developed into a rich theory. However, besides the noticeable exception of the theory of context-free groups [8], this notion is not doing so well beyond regular languages. The reason is that the syntactic monoid does not capture enough information. To work around this difficulty, Sakarovitch [14] proposed to use pointed monoids, a category which also admits minimal recognizers. The *pointed syntactic monoid* of a language is the pair formed by its syntactic monoid and by the image of the language in its syntactic monoid. Our new definition is an extension of this idea. However, instead of adding a "point" to the syntactic monoid, we attach to it a uniform structure (which also depends on the original language) and we consider its completion as a uniform space.

The power of topological methods opens new perspectives for the solution of problems beyond regular languages. Let us give a concrete example. Let AC^0 be the class of languages accepted by unbounded fan-in, polynomial size, constant-depth Boolean circuits. A famous result [2] states that a regular language belongs to AC^0 if and only if its syntactic monoid is quasi-aperiodic. It would be nice to prove this result (and the more general conjectures of this nature proposed by Straubing in his book [15]) by finding some suitable property of the syntactic space of AC^0.

Acknowledgments

The authors would like to thank the referees for many useful suggestions.

References

1. J. ALMEIDA, Profinite semigroups and applications, in *Structural theory of automata, semigroups, and universal algebra*, pp. 1–45, *NATO Sci. Ser. II Math. Phys. Chem.* vol. 207, Lect. Notes Comp. Sci., Dordrecht, 2005.
2. D. A. M. BARRINGTON, K. COMPTON, H. STRAUBING AND D. THÉRIEN, Regular languages in NC^1, *J. Comput. System Sci.* **44**,3 (1992), 478–499.
3. S. L. BLOOM AND Z. ÉSIK, A Mezei-Wright theorem for categorical algebras, *Theoret. Comput. Sci.* **411**,2 (2010), 341–359.
4. N. BOURBAKI, *General topology. Chapters 1–4*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1998.
5. D. M. CLARK, B. A. DAVEY, R. S. FREESE AND M. JACKSON, Standard topological algebras: syntactic and principal congruences and profiniteness, *Algebra Universalis* **52**,2-3 (2004), 343–376.
6. M. GEHRKE, S. GRIGORIEFF AND J.-E. PIN, Duality and equational theory of regular languages, in *ICALP 2008, Part II*, L. Aceto and al. (eds.), Berlin, 2008, pp. 246–257, *Lect. Notes Comp. Sci.* vol. 5126, Springer.
7. N. HINDMAN AND D. STRAUSS, *Algebra in the Stone-Čech compactification*, de Gruyter Expositions in Mathematics vol. 27, Walter De Gruyter & Co., Berlin, 1998. Theory and applications.
8. D. E. MULLER AND P. E. SCHUPP, Groups, the theory of ends, and context-free languages, *J. Comput. System Sci.* **26**,3 (1983), 295–310.
9. D. PERRIN AND J.-E. PIN, *Infinite Words*, Pure and Applied Mathematics vol. 141, Elsevier, 2004. ISBN 0-12-532111-2.
10. W. J. PERVIN, Quasi-uniformization of topological spaces, *Math. Ann.* **147** (1962), 316–317.
11. J.-E. PIN, A variety theorem without complementation, *Russian Mathematics (Iz. VUZ)* **39** (1995), 80–90.
12. J.-E. PIN AND P. WEIL, Uniformities on free semigroups, *International Journal of Algebra and Computation* **9** (1999), 431–453.
13. H. A. PRIESTLEY, Representation of distributive lattices by means of ordered Stone spaces, *Bull. London Math. Soc.* **2** (1970), 186–190.
14. J. SAKAROVITCH, An Algebraic Framework for the Study of the Syntactic Monoids Application to the Group Languages, in *MFCS*, A. W. Mazurkiewicz (ed.), pp. 510–516, *Lect. Notes Comp. Sci.* vol. 45, Springer, 1976.
15. H. STRAUBING, *Finite automata, formal logic, and circuit complexity*, Progress in Theoretical Computer Science, Birkhäuser Boston Inc., Boston, MA, 1994.
16. W. THOMAS, Uniform and nonuniform recognizability, *Theoret. Comput. Sci.* **292**,1 (2003), 299–316. Selected papers in honor of Jean Berstel.
17. T. WILKE, An algebraic theory for regular languages of finite and infinite words, *Int. J. Alg. Comput.* **3** (1993), 447–489.