From Ultrafilters on Words to the Expressive Power of a Fragment of Logic*

Mai Gehrke[1], Andreas Krebs[2], and Jean-Éric Pin[1]

[1] LIAFA, CNRS and Univ. Paris-Diderot, Case 7014, 75205 Paris Cedex 13, France
[2] Wilhelm-Schickard-Institut für Informatik, Universität Tübingen, Germany

Abstract. We give a method for specifying ultrafilter equations and identify their projections on the set of profinite words. Let $\mathcal{B}$ be the set of languages captured by first-order sentences using unary predicates for each letter, arbitrary uniform unary numerical predicates and a predicate for the length of a word. We illustrate our methods by giving profinite equations characterizing $\mathcal{B} \cap \mathrm{Reg}$ via ultrafilter equations satisfied by $\mathcal{B}$. This suffices to establish the decidability of the membership problem for $\mathcal{B} \cap \mathrm{Reg}$.

In two earlier papers, Gehrke, Grigorieff, and Pin proved the following results:

Result 1. [5] *Any Boolean algebra of regular languages can be defined by a set of equations of the form $u = v$, where u and v are profinite words.*[1]

Result 2. [6] *Any Boolean algebra of languages can be defined by a set of equations of the form $u = v$, where u and v are ultrafilters on the set of words.*

These two results can be summarized by saying that Boolean algebras of languages can be defined by *ultrafilter equations* and by *profinite equations* in the regular case. Restricted instances of Result 1 have proved to be very successful long before the result was stated in full generality. It is in particular a powerful tool for characterizing classes of regular languages or for determining the expressive power of various fragments of logic, see the book of Almeida [2] or the survey [9] for more information.

Result 2 however is still awaiting convincing applications and even an idea of how to apply it in a concrete situation. The main problem in putting it into practice is to cope with ultrafilters, a difficulty nicely illustrated by Jan van Mill, who cooked up the nickname *three headed monster* for the set of ultrafilters on $\mathbb{N}$. Facing this obstacle, the authors thought of using Results 1 and 2 simultaneously to obtain a new proof of the equality

$$\mathbf{FO}[\mathcal{N}] \cap \mathrm{Reg} = [\![\ (x^{\omega-1}y)^{\omega+1} = (x^{\omega-1}y)^{\omega} \quad \text{for } x, y \text{ words of the same length }]\!] \quad (1)$$

* Work supported by the project ANR 2010 BLAN 0202 02 FREC.

[1] In [5], these were denoted by $u \leftrightarrow v$.

H. Jürgensen et al. (Eds.): DCFS 2014, LNCS 8614, pp. 138–149, 2014.

This formula gives the profinite equations characterizing the regular languages captured by $\mathbf{FO}[\mathcal{N}]$, the first order logic using arbitrary numerical predicates and the usual letter predicates. This result follows from the work of Barrington, Straubing and Thérien [3] and Straubing [10] and is strongly related to circuit complexity. Indeed its proof makes use of the equality between $\mathbf{FO}[\mathcal{N}]$ and AC^0, the class of languages accepted by unbounded fan-in, polynomial size, constant-depth Boolean circuits [11, Theorem IX.2.1, p. 161]. See also [7] for similar results and problems.

However, before attacking this problem in earnest we have to tackle the following questions: how does one get hold of an ultrafilter equation given the non-constructibility of each one of them (save the trivial ones given by pairs of words)? In particular, how does one generalize the powerful use in the regular setting of x^ω? And how does one project such ultrafilter equations to the regular fragment? In answering these questions and facing these challenges, we have chosen to consider a smaller and simpler logic fragment first. Our choice was dictated by two parameters: we wanted to be able to handle the corresponding ultrafilters and we wished to obtain a reasonably understandable list of profinite equations. Finally, we opted for $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$, the restriction of $\mathbf{FO}[\mathcal{N}]$ to constant numerical predicates and to uniform unary numerical predicates. Here we obtain the following result (Theorem 4.7)

$$\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u] \cap \mathrm{Reg} = [\![(x^{\omega-1}s)(x^{\omega-1}t) = (x^{\omega-1}t)(x^{\omega-1}s), \\ (x^{\omega-1}s)^2 = (x^{\omega-1}s) \text{ for } x, s, t \text{ words of the same length }]\!] \quad (2)$$

which shows in particular that membership in $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$ is decidable for regular languages.

Although this result is of interest in itself, we claim that our *proof method* is more important than the result. Indeed, this case study demonstrates for the first time the workability of the ultrafilter approach.

This method can be summarized as follows. First we find a set of ultrafilter equations satisfied by $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$ (Theorem 3.2). These equations do not necessarily suffice to characterize $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$ [2], but projecting ultrafilters onto profinite words, we convert our ultrafilter equations to profinite equations for $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u] \cap \mathrm{Reg}$ (Theorems 3.3 and 3.4). The last step consists in verifying that the set of profinite equations thus obtained suffices to characterize $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u] \cap \mathrm{Reg}$ (Theorem 4.7).

Now, a closer look at our proof shows that we are far from making use of the potential power of ultrafilters. For instance, difficult combinatorial results like Szemeredi's theorem on arithmetic progressions can be formulated in terms of ultrafilters. Thus it is quite possible that more sophisticated arguments are required to extend our results to larger fragments of logic, including $\mathbf{FO}[\mathcal{N}]$.

[2] We recently proved that these equations actually do suffice to characterize $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$, but this will be the topic of another paper.

1 Stone Duality and Equations

In this paper, we denote by S^c the complement of a subset S of a set E.

1.1 Stone Duality

Let A be a finite alphabet. A *Boolean algebra of languages* is a set $\mathcal{B}$ of languages of A^* closed under finite unions, finite intersections and complement. It is *closed under quotients* if, for each $L \in \mathcal{B}$ and $u \in A^*$, the languages $u^{-1}L$ and Lu^{-1} are also in $\mathcal{B}$. Recall that $u^{-1}L = \{x \in A^* \mid ux \in L\}$ and $Lu^{-1} = \{x \in A^* \mid xu \in L\}$.

Let $\mathcal{B}$ be a Boolean algebra of languages of A^*. An *ultrafilter* of $\mathcal{B}$ is a non-empty subset γ of $\mathcal{B}$ such that:

(1) the empty set does not belong to γ,
(2) if $K \in \gamma$ and $K \subseteq L$, then $L \in \gamma$ (closure under extension)[3],
(3) if $K, L \in \gamma$, then $K \cap L \in \gamma$ (closure under intersection),
(4) for every $L \in \mathcal{B}$, either $L \in \gamma$ or $L^c \in \gamma$ (ultrafilter condition).

Stone duality tells us that $\mathcal{B}$ has an associated compact Hausdorff space $S(\mathcal{B})$, called its *Stone space*. This space is given by the set of ultrafilters of $\mathcal{B}$ with the topology generated by the basis of clopen sets of the form $\{\gamma \in S(\mathcal{B}) \mid L \in \gamma\}$, where $L \in \mathcal{B}$.

Two Stone spaces are of special interest for this paper. The first one is the Stone space of the Boolean algebra of all the subsets of a set X. It is known as the *Stone-Čech compactification* of X and is usually denoted by βX. An important property is that every map $f : X \to Y$ has a unique continuous extension $\beta f : \beta X \to \beta Y$ defined by $L \in \beta f(\gamma)$ if and only if $f^{-1}(L) \in \gamma$ for each subset L of Y. Moreover, the map sending an element x of X to the principal ultrafilter generated by x defines an injective map from X into βX.

Our second example is the Stone space of the Boolean algebra Reg of all *regular* subsets of A^*. It was proved by Almeida [1] to be equal to the topological space underlying the free profinite monoid on A, denoted by $\widehat{A^*}$. We refer to [2,8,9] for more information on this space, but it can be seen as the completion of A^* for the *profinite metric* d defined as follows. A finite monoid M *separates* two words u and v of A^* if there is a monoid morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$. We set

$$r(u, v) = \min\{|M| \mid M \text{ is a finite monoid that separates } u \text{ and } v\}$$

and $d(u, v) = 2^{-r(u,v)}$, with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$. Then d is a *metric* on A^* and the completion of A^* for this metric is denoted by $\widehat{A^*}$. The product on A^* can be extended by continuity to $\widehat{A^*}$, making $\widehat{A^*}$ a compact topological monoid, called the *free profinite monoid*. Its elements are called *profinite words*. We will only use two types of profinite words in this paper. In a compact monoid, the smallest closed subsemigroup containing a given

[3] In other words, γ is an *upset*.

element x has a unique idempotent, denoted by x^ω. Thus if x is a (profinite) word, so is x^ω. In fact, one can show that x^ω is the limit of the converging sequence $x^{n!}$. Moreover, the sequence $x^{n!-1}$ is also converging to an element denoted by $x^{\omega-1}$. More details can be found in [2,8,9].

1.2 Equations

Assigning to a Boolean algebra its Stone space is a contravariant functor: if $\mathcal{B}'$ is a subalgebra of $\mathcal{B}$, then $S(\mathcal{B}')$ is a quotient of $S(\mathcal{B})$. More precisely, the function which maps an ultrafilter of $\mathcal{B}$ onto its trace on $\mathcal{B}'$ induces a surjective continuous map $\pi : S(\mathcal{B}) \to S(\mathcal{B}')$.

This leads to the notion of equation relative to $\mathcal{B}$ or $\mathcal{B}$-*equation*. Let γ_1, γ_2 be two ultrafilters on $\mathcal{B}$ and let $L \in \mathcal{B}$. We say that L satisfies the $\mathcal{B}$-equation $\gamma_1 = \gamma_2$ provided

$$L \in \gamma_1 \iff L \in \gamma_2. \tag{3}$$

By extension, we say that $\mathcal{B}'$ satisfies the $\mathcal{B}$-equation $\gamma_1 = \gamma_2$ provided (3) holds for all $L \in \mathcal{B}'$, or equivalently $\pi(\gamma_1) = \pi(\gamma_2)$. Note that if $\mathcal{B}'$ is generated as a Boolean algebra by a subset $\mathcal{C}$, then $\mathcal{B}'$ satisfies a $\mathcal{B}$-equation as soon as each $L \in \mathcal{C}$ does. Finally, we say that $\mathcal{B}'$ is defined by a set E of $\mathcal{B}$-equations if for each $L \in \mathcal{B}$, $L \in \mathcal{B}'$ if and only if L satisfies all the $\mathcal{B}$-equations in E. The following result is an immediate consequence of Stone duality.

Theorem 1.1. *Every subalgebra of a Boolean algebra $\mathcal{B}$ can be defined by a set of $\mathcal{B}$-equations.*

Specializing this result to $\mathcal{B} = \mathrm{Reg}$ and to $\mathcal{B} = \mathcal{P}(A^*)$ yields Results 1 and 2 of the introduction. Another case of interest for this paper is to take $\mathcal{B} = \mathrm{Reg}$ and for $\mathcal{B}'$ a Boolean algebra closed under quotients. In this case, it is easier to reformulate Result 1 in terms of syntactic morphisms. Let L be a regular language and $\eta : A^* \to M$ be its syntactic morphism. We say that η *satisfies the profinite equation* $u = v$ or that L *syntactically satisfies the profinite equation* $u = v$ if $\widehat{\eta}(u) = \widehat{\eta}(v)$, where $\widehat{\eta} : \widehat{A^*} \to M$ is the unique continuous extension of η to $\widehat{A^*}$. It is easy to see that a regular language syntactically satisfies a profinite equation if and only if all of its quotients satisfy this equation. Therefore we have

Result 3. *Any Boolean algebra of regular languages closed under quotients can be syntactically defined by a set of profinite equations.*

When working with ultrafilter equations, the following two observations will be helpful. Let us denote by $K \triangle L$ the symmetric difference of the sets K and L.

Proposition 1.2. *Let γ be an ultrafilter of $\mathcal{B}$ and let $K, L \in \mathcal{B}$. Then the following statements are equivalent:*

(1) *$K \in \gamma$ if and only if $L \in \gamma$,*
(2) *$K \triangle L \notin \gamma$.*

Proposition 1.3. *Let $f : X \to Y$ be a map and let L be a subset of Y. If $f^{-1}(L)$ satisfies $u = v$ for some $u, v \in \beta X$, then L satisfies $\beta f(u) = \beta f(v)$.*

2 A Boolean Algebra and Its Logical Description

For each word $u = a_0 \dots a_{n-1}$ where $a_0, \dots, a_{n-1} \in A$ and each letter $a \in A$, let $\mathbf{a}_u = \{i \in \mathrm{Dom}(u) \mid a_i = a\}$. For instance, if $u = aabcbaba$, then $\mathbf{a}_u = \{0, 1, 5, 7\}$, $\mathbf{b}_u = \{2, 4, 6\}$, and $\mathbf{c}_u = \{3\}$. The *length* of u is denoted by $|u|$.

For each letter a in A and for each subset P of $\mathbb{N}$, let

$$L_P = \{u \in A^+ \mid |u| - 1 \in P\} \text{ and } L_{a,P} = \{u \in A^+ \mid \mathbf{a}_u \subseteq P\}.$$

In this paper, we are interested in the Boolean algebra $\mathcal{B}$ generated by the languages L_P and $L_{a,P}$ for $P \subseteq \mathbb{N}$ and $a \in A$. A little combinatorics on words leads to the following result:

Proposition 2.1. *The Boolean algebras $\mathcal{B}$ and $\mathcal{B} \cap \mathrm{Reg}$ are closed under quotients and under the operations $L \to uL$ for each word $u \in A^*$.*

Let us turn to the logical description of $\mathcal{B}$. Let $u = a_0 \dots a_{n-1}$ be a nonempty word where $a_0, \dots, a_{n-1}$ are letters of the alphabet A. Then u may be viewed as a first-order model whose *domain* is the set $\mathrm{Dom}(u) = \{0, \dots, |u|-1\}$, carrying, for each letter $a \in A$, the unary predicate $\mathbf{a}_u$ as defined above. For each subset P of $\mathbb{N}$, we also define two predicates: a 0-ary predicate which is true on u if and only if $|u|-1 \in P$ and a unary uniform predicate[4] defined by $P(n) = P \cap \{0, ..., n-1\}$. Its interpretation on a word u is the subset $P(|u|)$ of $\{0, ..., |u| - 1\}$.

We denote by $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$ the set of first-order formulas built on these predicates. Note that we do not consider $=$ as a logical symbol so that each formula is equivalent to one of quantifier depth one. The language defined by a sentence φ is the set[5]

$$L(\varphi) = \{u \in A^+ \mid u \text{ satisfies } \varphi\}$$

For instance if $\varphi = \exists x\ \mathbf{a}x$, then $L(\varphi) = A^*aA^*$. Let P be a subset of $\mathbb{N}$. If P is considered as a 0-ary numerical relation, then $L(P) = L_P$. If P is interpreted as a unary uniform numerical relation, then the formula $\forall x\ (\mathbf{a}x \to Px)$ defines the language $L_{a,P}$ since P is interpreted as $P(|u|)$. This proves one direction of the following logical description of $\mathcal{B}$.

Theorem 2.2. *A language L of A^+ belongs to $\mathcal{B}$ if and only it can be defined by a sentence of $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$.*

3 Some Equations of $\mathcal{B}$

For $1 \leqslant i \leqslant k$, let $\pi_i : A^* \times \mathbb{N}^k \to \mathbb{N}$ be the map defined by $\pi_i(u, n_1, \dots, n_k) = n_i$.

The following proposition shows that the classes of equations we will define subsequently contain at least one non-trivial equation for each $\alpha \in \beta\mathbb{N} - \mathbb{N}$.

[4] Following the terminology of [11], a unary *numerical relation* R associates to each $n > 0$ a subset $R(n)$ of $\{0, ..., n-1\}$. It is *uniform* if there exists a subset P of $\mathbb{N}$ such that, for all $n > 0$, $R(n) = P \cap \{0, \dots, n-1\}$. Not every numerical relation is uniform: for instance, the unary numerical relation R defined by $R(n) = \{n-1\}$ is not uniform.

[5] The empty word is excluded to avoid any problem related to empty models.

Proposition 3.1. *Let $\gamma \in \beta(A^* \times \mathbb{N}^k)$ with $k \geqslant 1$. Then, for each $\alpha \in \beta\mathbb{N}$, the following conditions are equivalent:*

(1) *$\beta\pi_i(\gamma) = \alpha$ for each $i \in \{1, \ldots, k\}$;*

(2) *$\{A^* \times P^k \mid P \in \alpha\} \subseteq \gamma$.*

Furthermore, these conditions hold for γ with respect to some α if and only if

(3) *For each partition $\{P_1, \ldots, P_n\}$ of $\mathbb{N}$, we have $\bigcup_{j=1}^{n}(A^* \times P_j^k) \in \gamma$.*

Proof. (1) implies (2) since $A^* \times P^k = \bigcap_{i=1}^{k} \pi_i^{-1}(P)$ and γ is closed under finite intersections.

(2) implies (1). Let $P \in \alpha$ and $i \in \{1, \ldots, k\}$. Then by (2), $A^* \times P^k \in \gamma$ and thus $\pi_i^{-1}(\pi_i(A^* \times P^k)) \in \gamma$ so that $P = \pi_i(A^* \times P^k) \in \beta\pi_i(\gamma)$. It follows that $\alpha \subseteq \beta\pi_i(\gamma)$ and thus $\alpha = \beta\pi_i(\gamma)$ since ultrafilters are maximal.

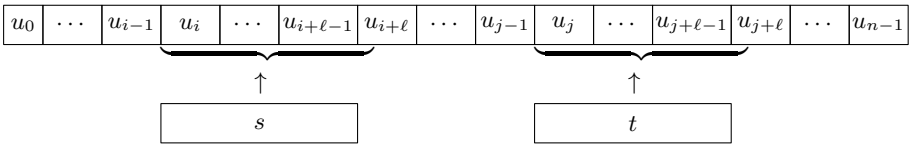
For the second assertion, suppose there is an $\alpha \in \beta\mathbb{N}$ such that (1) and (2) hold and $\{P_1, \ldots, P_n\}$ is a partition of $\mathbb{N}$. Then $\bigcup_{j=1}^{n} P_j = \mathbb{N}$ implies $P_\ell \in \alpha$ for some ℓ and thus $A^* \times P_\ell^k \in \gamma$ by (2). Since γ is an upset, condition (3) holds.

Suppose now that γ satisfies (3) and let $\alpha = \{P \mid A^* \times P^k \in \gamma\}$. Then α is an upset closed under intersection. Furthermore, for each $P \subseteq \mathbb{N}$, the partition $\{P, P^c\}$ forces $A^* \times P^k \in \gamma$ or $A^* \times (P^c)^k \in \gamma$ so that α is an ultrafilter. It follows by the equivalence of (1) and (2) that $\beta\pi_i(\gamma) = \alpha$ for each $i \in \{1, \ldots, k\}$. □

We are now ready to introduce the first class of equations pertinent to the languages treated in this paper. For this purpose, given $u, s, t \in A^*$, where $u = u_0 \cdots u_{n-1}$ with each $u_k \in A$ and $|s| = |t| = \ell$, and $i, j \in \mathbb{N}$, define

$$u(s@i, t@j) = \begin{cases} u_0 \ldots u_{i-1} s u_{i+\ell} \ldots u_{j-1} t u_{j+\ell} \ldots u_{n-1} & \text{if } i + \ell \leqslant j \text{ and } j + \ell \leqslant n \\ u & \text{otherwise} \end{cases}$$

Informally, we put s at position i and t at position j.



For each pair (s, t) of words of the same length, let $f_{s,t} : A^* \times \mathbb{N}^2 \to A^*$ be the function defined by $f_{s,t}(u, i, j) = u(s@i, t@j)$.

Theorem 3.2. *Let $s, t \in A^*$ with $|s| = |t|$. If $\gamma \in \beta(A^* \times \mathbb{N}^2)$ and $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$, then $\mathcal{B}$ satisfies the equation*

$$\beta f_{s,t}(\gamma) = \beta f_{t,s}(\gamma). \tag{4}$$

Proof. Let $a \in A$ and $P \subseteq \mathbb{N}$. We show that $L_{a,P}$ and L_P satisfy the equations (4). First we have

$$L_{a,P} \in \beta f(\gamma) \iff f^{-1}(L_{a,P}) \in \gamma.$$

Thus (4) holds for $L_{a,P}$ if and only if

$$f_{s,t}^{-1}(L_{a,P}) \in \gamma \iff f_{t,s}^{-1}(L_{a,P}) \in \gamma$$

and by Proposition 1.2 this is equivalent to $S \notin \gamma$, where

$$S = f_{s,t}^{-1}(L_{a,P}) \triangle f_{t,s}^{-1}(L_{a,P}).$$

Let ℓ be the common length of s and t. If an element $(u, n_1, n_2) \in A^* \times \mathbb{N}^2$ is in S then $n_1 + 2\ell \leqslant n_2 + \ell \leqslant |u|$ since otherwise $f_{s,t}(u, n_1, n_2) = f_{t,s}(u, n_1, n_2) = u$. Suppose that $(u, n_1, n_2) \in f_{s,t}^{-1}(L_{a,P}) \setminus f_{t,s}^{-1}(L_{a,P})$, that is, $f_{s,t}(u, n_1, n_2) \in L_{a,P}$ and $f_{t,s}(u, n_1, n_2) \notin L_{a,P}$. Then all the positions of a in $f_{s,t}(u, n_1, n_2)$ are in P and some position of a in $f_{t,s}(u, n_1, n_2)$ is not in P. This latter position necessarily occurs inside one of the factors s or t of $f_{s,t}(u, n_1, n_2)$. Consequently, there is an $i \in \{0, \ldots, \ell - 1\}$ such that one of the two following possibilities occurs:

(1) the letter in position $n_1 + i$ in $f_{t,s}(u, n_1, n_2)$ is an a but $n_1 + i \notin P$,

(2) the letter in position $n_2 + i$ in $f_{t,s}(u, n_1, n_2)$ is an a but $n_2 + i \notin P$.

Now, in the first case, the letter in position $n_2 + i$ in $f_{s,t}(u, n_1, n_2)$ is an a. Thus $n_2 + i \in P$ since $f_{s,t}(u, n_1, n_2) \in L_{a,P}$. Similarly, we conclude that $n_1 + i \in P$ in the second case. In summary, we have either $n_1 + i \notin P$ and $n_2 + i \in P$ (first case) or $n_1 + i \in P$ and $n_2 + i \notin P$ (second case). In both cases we conclude that

$$(u, n_1, n_2) \in \bigcup_{i=0}^{\ell-1} \Big(\pi_1^{-1}(P - i) \triangle \pi_2^{-1}(P - i)\Big).$$

The case $(u, n_1, n_2) \in f_{t,s}^{-1}(L_{a,P}) \setminus f_{s,t}^{-1}(L_{a,P})$ leads to the same conclusion and thus we have shown that

$$S \subseteq \bigcup_{i=0}^{\ell-1} \Big(\pi_1^{-1}(P - i) \triangle \pi_2^{-1}(P - i)\Big).$$

If $S \in \gamma$, then $\bigcup_{i=0}^{\ell-1} \Big(\pi_1^{-1}(P - i) \triangle \pi_2^{-1}(P - i)\Big) \in \gamma$ and since γ is an ultrafilter, $\pi_1^{-1}(P - i) \triangle \pi_2^{-1}(P - i) \in \gamma$ for some $i \in \{0, \ldots, \ell - 1\}$. We complete the proof that $S \notin \gamma$ by showing that, for every $Q \subseteq \mathbb{N}$ we have $\pi_1^{-1}(Q) \triangle \pi_2^{-1}(Q) \notin \gamma$, or equivalently, $(\pi_1^{-1}(Q) \triangle \pi_2^{-1}(Q))^c \in \gamma$. But this is a direct consequence of Proposition 3.1(3) since

$$(\pi_1^{-1}(Q) \triangle \pi_2^{-1}(Q))^c = A^* \times \big((Q \times Q) \cup (Q^c \times Q^c)\big).$$

Thus $S \notin \gamma$ and $L_{a,P}$ satisfies the equation $\beta f_{s,t}(\gamma) = \beta f_{t,s}(\gamma)$.

By the same argument as applied above, L_P satisfies the equations (4) if and only if $f_{s,t}^{-1}(L_P) \triangle f_{t,s}^{-1}(L_P) \notin \gamma$. However, since $|f_{s,t}(u, n_1, n_2)| = |f_{t,s}(u, n_1, n_2)|$ and since $x \in L_P$ implies $y \in L_P$ if $|y| = |x|$, we have $f_{s,t}^{-1}(L_P) = f_{t,s}^{-1}(L_P)$ and thus $f_{s,t}^{-1}(L_P) \triangle f_{s,t}^{-1}(L_P) = \emptyset$ and therefore it does not belong to γ. □

We now consider the projection of the equations introduced above on the Stone space of the regular fragment of the Boolean algebra $\mathcal{B}$.

Theorem 3.3. *Let $x, s, t, \in A^*$ with $|s| = |t| = |x|$. Then $\mathcal{B} \cap \mathrm{Reg}$ satisfies the profinite equation $x^{\omega-1}sx^{\omega-1}t = x^{\omega-1}tx^{\omega-1}s$.*

Proof. It suffices to show that there is a $\gamma \in \beta(A^* \times \mathbb{N}^2)$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$ such that the projection $\pi_{\mathrm{Reg}} : \beta A^* \to \widehat{A^*}$ defined by

$$\pi_{\mathrm{Reg}}(\gamma) = \gamma \cap \mathrm{Reg}$$

maps $\beta f_{s,t}(\gamma)$ to $x^{\omega-1}sx^{\omega-1}t$ and $\beta f_{t,s}(\gamma)$ to $x^{\omega-1}tx^{\omega-1}s$.

Proposition 3.1 shows that in order for γ to satisfy $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$, we just need γ to contain the down-directed filter base

$$\Big\{ \bigcup_{j=1}^{n} (A^* \times P_j^2) \mid \{P_1, \ldots, P_n\} \text{ is a partition of } \mathbb{N} \Big\}.$$

We now show that for $\ell = |x|$, adding the sets $W_N = \{(x^{m!}, (k!-1)\ell, (m!-1)\ell) \mid N \leqslant k < m\}$ for each $N \in \mathbb{N}$ to this filter base still yields a filter base. To this end we just need to show that for each partition $\{P_1, \ldots, P_n\}$ of $\mathbb{N}$ and $N \in \mathbb{N}$, the set

$$W_N \cap (\bigcup_{j=1}^{n} (A^* \times P_j^2))$$

is non-empty. But since $\{P_1, \ldots, P_n\}$ is a partition of $\mathbb{N}$, there is $j \in \{1, \ldots, n\}$ with $P_j \cap \{(k!-1)\ell \mid k \geqslant N\}$ infinite. It readily follows that $W_N \cap (A^* \times P_j^2)$ is infinite and thus the bigger set $W_N \cap (\bigcup_{j=1}^{n} (A^* \times P_j^2))$ is non-empty.

Let $\gamma \in \beta(A^* \times \mathbb{N}^2)$ be an ultrafilter containing the extended filter base. Then clearly $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$ so that, by Theorem 3.2, the Boolean algebra $\mathcal{B}$ satisfies the equation $\beta f_{s,t}(\gamma) = \beta f_{t,s}(\gamma)$.

Now let $L \in \beta f_{s,t}(\gamma) \cap \mathrm{Reg}$. Then $f_{s,t}^{-1}(L) \in \gamma$. Also, since $W_N \in \gamma$ for each $N \in \mathbb{N}$, it follows that $f_{s,t}^{-1}(L) \cap W_1$ is infinite, or equivalently $L \cap f_{s,t}(W_1)$ is infinite. But

$$f_{s,t}(W_1) = \{x^{n!-1}sx^{(m!-n!)-1}t \mid 1 \leqslant n < m\}$$

and $m! - n! = (m!/n! - 1)n!$ where $(m!/n! - 1) \geqslant 1$. Since any sequence in this set with $n \to \infty$ converges to $x^{\omega-1}sx^{\omega-1}t$ in $\widehat{A^*}$, and since $L \cap f_{s,t}(W_1) \subseteq \widehat{L}$ with the latter closed, we must have $x^{\omega-1}sx^{\omega-1}t \in \widehat{L}$. But as $\widehat{A^*}$ is Hausdorff,

$$\bigcap \{\widehat{L} \mid L \in \beta f_{s,t}(\gamma) \text{ and } L \in \mathrm{Reg}\} = \pi_{\mathrm{Reg}}(\beta f_{s,t}(\gamma))$$

so $x^{\omega-1}sx^{\omega-1}t = \pi_{\mathrm{Reg}}(\beta f_{s,t}(\gamma))$. Similarly $x^{\omega-1}tx^{\omega-1}s = \pi_{\mathrm{Reg}}(\beta f_{t,s}(\gamma))$. □

A similar argument using the ultrafilter equations $\beta f_{tss}(\gamma) = \beta f_{tts}(\gamma)$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma) = \beta\pi_3(\gamma)$ and projecting yields the profinite equation

$$(x^{\omega-1}t)(x^{\omega-1}s)(x^{\omega-1}s) = (x^{\omega-1}t)(x^{\omega-1}t)(x^{\omega-1}s).$$

Specializing to $x = t$ we get $(x^{\omega-1}s)(x^{\omega-1}s) = (x^{\omega-1}s)$, which proves the following result.

Theorem 3.4. *Let $x, s \in A^*$ with $|s| = |x|$. Then $\mathcal{B} \cap \mathrm{Reg}$ satisfies the profinite equation $(x^{\omega-1}s)(x^{\omega-1}s) = (x^{\omega-1}s)$.*

Applications of the ultrafilter equations introduced in this section are not limited to the interplay with regular languages and they can also be used to prove separation results for nonregular languages. For instance, it is easy to find an ultrafilter equation of $\mathcal{B}$ not satisfied by the language $\{uav \mid u, v \in \{a, b\}^* \text{ and } |u| = |v|\}$.

4 The Regular Case

Consider the two profinite equations introduced in the previous sections, where x, s and t are words of the same length

$$(x^{\omega-1}s)(x^{\omega-1}t) = (x^{\omega-1}t)(x^{\omega-1}s) \tag{5}$$

$$(x^{\omega-1}s)(x^{\omega-1}s) = (x^{\omega-1}s) \tag{6}$$

We will show that the regular languages of our class $\mathcal{B}$ are exactly the languages whose syntactic morphism satisfies the equations (5) and (6) for all words x, s and t of the same length. Before we do this, it is useful to introduce some further notation.

Let $k, r, d \in \mathbb{N}$ with $d > 0$. Given a word $u = a_0 \cdots a_n \in A^*$ where $a_i \in A$, let $p_k(u) = a_0 \cdots a_{k-1}$ be the prefix of length k of u and let

$$C_{d,r}(u) = \{a_i \mid i \geqslant d \text{ and } i \equiv r \bmod d\}$$

be the *content* of u at r modulo d. For instance, if $u = ccbbacabac$, then $p_5(u) = ccbba$, $C_{3,0} = \{a, b, c\}$, $C_{3,1} = \{a, b\}$ and $C_{3,2} = \{a, c\}$.

For each positive integer d, let $\sim_d$ be the equivalence on A^* defined as follows. Given $u, v \in A^*$, $u \sim_d v$ if and only if the three following conditions are satisfied:

(1) for $0 < k \leqslant d$, $p_k(u) = p_k(v)$,
(2) $|u| \equiv |v| \bmod d$,
(3) for $0 \leqslant r < d$, $C_{d,r}(u) = C_{d,r}(v)$.

Proposition 4.1. *The relation $\sim_d$ is a congruence of finite index on A^*.*

We now consider a regular language L and we denote by $\eta : A^* \to M$ its syntactic morphism. We also let $d = |M|!$. It is well known that, for each $x \in M$, x^d is idempotent, that is, $x^{2d} = x^d$. For the remainder of the paper, we use the notation $u =_\eta v$ for $\eta(u) = \eta(v)$, and, for any $r \in \mathbb{N}$, we denote by $[r]$ the remainder after division of r by d. We will need a small combinatorial lemma:

Lemma 4.2. *Let u be a word of length at least $|M|$. Then there exist a prefix p of u of length lesser than $|M|$ and a word x of length $|M|!$ such that $px =_\eta p$.*

Proof. For each $k \geqslant 0$, let $s_k = \eta(p_k(u))$. If $s_0, \ldots, s_{|M|-1}$ are all distinct, one of them, say s_i, is idempotent. Then $p = p_i(u)$ and $x = p^{|M|!/(i+1)}$ give the result. On the other hand, if $s_i = s_j$ with $i < j < |M|$, then $p = p_i(u)$ and $x = z^{|M|!/|z|}$ where $p_j(u) = pz$ yield the result. □

Let $\mathcal{B}_{\mathrm{Reg}}$ be the Boolean algebra generated by the languages L_P or $L_{a,P}$ where P is a regular subset of $\mathbb{N}$, that is, a finite union of languages of the form $r + d\mathbb{N}$ for $r, d \in \mathbb{N}$. Clearly $\mathcal{B}_{\mathrm{Reg}} \subseteq \mathcal{B} \cap \mathrm{Reg}$. Our aim is to show that if η satisfies the equations (5) and (6), then L is a union of $\sim_d$-classes. In view of the following proposition, it then follows that $L \in \mathcal{B}_{\mathrm{Reg}}$.

Proposition 4.3. *For every $d \geqslant 1$, every $\sim_d$-class is a language of $\mathcal{B}_{\mathrm{Reg}}$.*

We now suppose that η satisfies equations (5) and (6) for all words x, s and t of the same length.

Lemma 4.4. *Let a_0, a_1, …, a_r be letters and let p and x be two words such that $|x| = d$ and $px =_\eta p$. Setting $x = b_0 \cdots b_{d-1}$ where b_0, b_1, …, b_{d-1} are letters, we have $pa_0 \cdots a_r =_\eta pa_0 \cdots a_r (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^d$.*

Proof. We prove the result by induction on the length of the word $a_0 a_1 \cdots a_r$. If the length is 0, the result simply follows from the relation $px =_\eta p$. Suppose by induction that the result holds for a word of length $\leqslant r$, that is

$$pa_0 \cdots a_{r-1} =_\eta pa_0 \cdots a_{r-1} (b_{[r]} \cdots b_{d-1} b_0 \cdots b_{[r-1]})^d \tag{7}$$

Then we get by (7)

$$\begin{aligned}
&pa_0 \cdots a_{r-1} a_r (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^d \\
&\quad =_\eta pa_0 \cdots a_{r-1} (b_{[r]} \cdots b_{d-1} b_0 \cdots b_{[r-1]})^d a_r (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^d \\
&\quad =_\eta pa_0 \cdots a_{r-1} b_{[r]} (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^{d-1} \underbrace{b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r-1]} a_r}_{s} \\
&\qquad\qquad (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^{d-1} \underbrace{b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]}}_{t}
\end{aligned}$$

Equation (5) allows one to swap s and t and consequently we obtain

$$\begin{aligned}
&pa_0 \cdots a_r (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^d \\
&\quad =_\eta pa_0 \cdots a_{r-1} b_{[r]} (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^{d-1} \underbrace{b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]}}_{t} \\
&\qquad\qquad (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^{d-1} \underbrace{b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r-1]} a_r}_{s} \\
&\quad =_\eta pa_0 \cdots a_{r-1} (b_{[r]} \cdots b_{d-1} b_0 \cdots b_{[r-1]})^{2d} a_r =_\eta pa_0 \cdots a_{r-1} a_r \quad \text{by (7)},
\end{aligned}$$

which concludes the induction step. □

Lemma 4.5. *Let a_0, a_1, …, a_r be letters and let p and $x = b_0 \cdots b_{d-1}$ be two words such that $px =_\eta p$. Setting for each $k \geqslant 0$*

$$z_k = b_0 b_1 \cdots b_{[k-1]} a_{[k]} b_{[k+1]} \cdots b_{d-1}$$

the following relation holds

$$pa_0 \cdots a_r =_\eta px^{d-1} z_0 x^{d-1} z_1 \cdots x^{d-1} z_{[r]} x^{d-1} b_0 \cdots b_{[r]} \tag{8}$$

Proof. Applying Lemma 4.4 repeatedly yields the formula

$$pa_0 \cdots a_r =_\eta p(b_0 \cdots b_{d-1})^d a_0 (b_1 \cdots b_{d-1} b_0)^d a_1 \cdots \\ (b_{[r]} \cdots b_{d-1} b_0 \cdots b_{[r-1]})^d a_r (b_{[r+1]} \cdots b_{d-1} b_0 \cdots b_{[r]})^d \quad (9)$$

It suffices now to observe that the right hand sides of (9) and of (8) are the same word. □

Proposition 4.6. *If $u \sim_d v$, then $u =_\eta v$.*

Proof. Let $u \in L$ and let v be a word such that $u \sim_d v$. We claim that $u =_\eta v$. If $|u| < d$ or $|v| < d$, then $u = v$ and the result is trivial. Thus we may assume that $|u|, |v| \geqslant d$ and by the definition of $\sim_d$, $p_d(u) = p_d(v)$.

Let p and $x = b_0 b_1 \cdots b_{d-1}$ be the words given by Proposition 4.2. Then p is a common prefix of length $< |M|$ of u and v and x is a word of length d such that $px =_\eta p$.

Let $u = pa_0 \cdots a_m$ and $v = pc_0 \cdots c_n$. Since $u \sim_d v$, $|u| \equiv |v| \bmod d$ and thus $[n] = [m]$. Setting

$$y_k = b_0 b_1 \cdots b_{[k-1]} a_{[k]} b_{[k+1]} \cdots b_{d-1}$$
$$z_k = b_0 b_1 \cdots b_{[k-1]} c_{[k]} b_{[k+1]} \cdots b_{d-1}$$

we get by Lemma 4.5

$$u =_\eta px^{d-1} y_0 x^{d-1} y_1 \cdots x^{d-1} y_{[m]} x^{d-1} b_0 \cdots b_{[m]}$$
$$v =_\eta px^{d-1} z_0 x^{d-1} z_1 \cdots x^{d-1} z_{[n]} x^{d-1} b_0 \cdots b_{[n]}$$

Since L satisfies the equations (5) and (6), one has for each i, j

$$x^{d-1} y_i x^{d-1} y_i =_\eta x^{d-1} y_i \qquad x^{d-1} z_i x^{d-1} z_i =_\eta x^{d-1} z_i$$
$$x^{d-1} y_i x^{d-1} y_j =_\eta x^{d-1} y_j x^{d-1} y_i \qquad x^{d-1} z_i x^{d-1} z_j =_\eta x^{d-1} z_j x^{d-1} z_i$$

We can now conclude the proof of Proposition 4.6. Since $u \sim_d v$, for each $i \geqslant d$ there is a j such that $j \equiv i \bmod d$ and $a_i = c_j$. Therefore, for each i there is a j such that $y_i = z_j$. Similarly, for each j there is an i such that $z_j = y_i$. It follows that $u =_\eta v$. □

We are now ready to prove the main result of this section.

Theorem 4.7. *Let L be regular language, let $\eta : A^* \to M$ be its syntactic morphism and let $d = |M|!$. Then the following conditions are equivalent:*

(1) *η satisfies the profinite equations (5) and (6) for all words x, s and t of the same length,*
(2) *L is a finite union of $\sim_d$-classes,*
(3) *$L \in \mathcal{B}_{\mathrm{Reg}}$,*
(4) *$L \in \mathcal{B} \cap \mathrm{Reg}$.*

Proof. Proposition 4.6 proves that (1) implies (2). Proposition 4.3 shows that (2) implies (3), (3) implies (4) is trivial and (4) implies (1) follows from Theorems 3.3 and 3.4. □

Corollary 4.8. *One can effectively decide whether or not a given a regular language belongs to* $\mathcal{B}$.

Coming back to logic, one could derive the following characterization, in which $=_c$ stands for the set of unary predicates of the form $\{c\}$, for $c \in \mathbb{N}$ and **MOD** stands for the set of modulo predicates, as defined in [4].

Theorem 4.9. *A language belongs to* $\mathcal{B} \cap \mathrm{Reg}$ *if and only if it can be defined by a sentence of* $\mathbf{FO}[\mathbf{MOD}, =c]$.

Acknowledgement. The authors would like to thank Charles Paperman for his useful suggestions.

References

1. Almeida, J.: Residually finite congruences and quasi-regular subsets in uniform algebras. Portugaliæ Mathematica 46, 313–328 (1989)
2. Almeida, J.: Finite semigroups and universal algebra. World Scientific Publishing Co. Inc., River Edge (1994), Translated from the 1992 Portuguese original and revised by the author
3. Barrington, D.A.M., Straubing, H., Thérien, D.: Non-uniform automata over groups. Information and Computation 89, 109–132 (1990)
4. Chaubard, L., Pin, J.-É., Straubing, H.: First order formulas with modular predicates. In: 21st Annual IEEE Symposium on Logic in Computer Science (LICS 2006), pp. 211–220. IEEE (2006)
5. Gehrke, M., Grigorieff, S., Pin, J.-É.: Duality and equational theory of regular languages. In: Aceto, L., Damgård, I., Goldberg, L.A., Halldórsson, M.M., Ingólfsdóttir, A., Walukiewicz, I. (eds.) ICALP 2008, Part II. LNCS, vol. 5126, pp. 246–257. Springer, Heidelberg (2008)
6. Gehrke, M., Grigorieff, S., Pin, J.-É.: A topological approach to recognition. In: Abramsky, S., Gavoille, C., Kirchner, C., Meyer auf der Heide, F., Spirakis, P.G. (eds.) ICALP 2010. Part II. LNCS, vol. 6199, pp. 151–162. Springer, Heidelberg (2010)
7. McKenzie, P., Thomas, M., Vollmer, H.: Extensional uniformity for Boolean circuits. SIAM J. Comput. 39(7), 3186–3206 (2010)
8. Pin, J.-É.: Profinite methods in automata theory. In: Albers, S., Marion, J.-Y. (eds.) 26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009), pp. 31–50. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl (2009)
9. Pin, J.-É.: Equational descriptions of languages. Int. J. Found. Comput. S. 23, 1227–1240 (2012)
10. Straubing, H.: Constant-depth periodic circuits. Internat. J. Algebra Comput. 1(1), 49–87 (1991)
11. Straubing, H.: Finite automata, formal logic, and circuit complexity. Birkhäuser Boston Inc., Boston (1994)