Ultrafilters on words for a fragment of logic ⋆

Mai Gehrke[1], Andreas Krebs[2], and Jean-Éric Pin[1]
May 5, 2015

[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 ultrafilter equations characterising $\mathcal{B}$ and then projecting these to obtain profinite equations characterising $\mathcal{B} \cap \mathrm{Reg}$. This suffices to establish the decidability of the membership problem for $\mathcal{B} \cap \mathrm{Reg}$.

This paper is the third step of a programme aiming at widening the notion of recognisability using methods from topological duality theory. In two earlier papers, Gehrke, Grigorieff, and Pin proved the following results:

Result 1 [4] *Any Boolean algebra of regular languages can be defined by a set of equations of the form $u \leftrightarrow v$, where u and v are profinite words.*

Result 2 [5] *Any Boolean algebra of languages can be defined by a set of equations of the form $u \leftrightarrow v$, where u and v are ultrafilters on the set of words.*

These two results can be summarised by saying that Boolean algebras of languages can be defined by *ultrafilter equations* and by *profinite equations* in the regular case. When a Boolean algebra is closed under quotients, we use the notation $u = v$ instead of $u \leftrightarrow v$, for a reason that will be fully explained in Section 1.3.

Restricted instances of Result 1 have been obtained and applied very successfully long before the result was stated and proved 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

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

to obtain a new proof of the equality

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

where $[\![E]\!]$ denotes the class of languages defined by a set E of equations. This formula gives the profinite equations characterizing the regular languages in $\mathbf{FO}[\mathcal{N}]$, the class of languages defined by sentences of 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 generalise the powerful use in the regular setting of the ω-power? 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 5.16)

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

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 summarised as follows. First we find a set of ultrafilter equations characterising $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$ (Theorems 3.2, 3.3, and 4.7). Projecting these ultrafilter equations onto profinite words, we obtain profinite equations characterising $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u] \cap \mathrm{Reg}$ (Theorem 5.2). Finally we show that the simpler class (2) generates the full family of projections of our ultrafilter equations to obtain Theorem 5.16.

In the conference version of this paper [6], we had only proved the validity in $\mathcal{B}$ of the equations given in Section 3. Here we also prove their completeness in Section 4. As a consequence, we get a new completeness result for $\mathcal{B} \cap \mathrm{Reg}$ obtained by projection in Section 5.1. This leads to a new proof of decidability of membership in $\mathcal{B} \cap \mathrm{Reg}$ in Section 5.2. The completeness result expressed by equation (2) above is then obtained from the completeness result in Section 5.1 by rewriting in Section 5.3. In [6], the completeness part of (2) was proved by traditional automata theoretic means.

1 Stone duality and equations

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

1.1 Filters and ultrafilters

Let X be a set. A *Boolean algebra of subsets* of X is a subset of $\mathcal{P}(X)$ containing the empty set and closed under finite intersections, finite unions and complement. Let $\mathcal{B}$ be a Boolean algebra of subsets of X. An *ultrafilter* of $\mathcal{B}$ is a nonempty 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).

Nonempty subsets of $\mathcal{B}$ satisfying just conditions (2) and (3) above are called *filters*, while filters also satisfying (1) are said to be *proper*. A subset $\mathcal{S}$ of $\mathcal{B}$ is a *filter subbasis* if it has the *finite intersection property*: every finite intersection of elements of $\mathcal{S}$ is nonempty. In this case, the set of all supersets of finite intersections of elements of $\mathcal{S}$ is a proper filter, called the *filter generated by* $\mathcal{S}$.

A nonempty subset $\mathcal{S}$ of $\mathcal{B}$ is a *filter basis* if it does not not contain the empty set and if, for every $K, L \in \mathcal{S}$, there exists $M \in \mathcal{S}$ such that $M \subseteq K \cap L$. In this case, the filter generated by $\mathcal{S}$ is the set of all supersets of elements of $\mathcal{S}$. Note that if $\mathcal{S}$ and $\mathcal{T}$ are filter basis, then $\mathcal{S} \cup \mathcal{T}$ is a filter basis if and only if the intersection of any member of $\mathcal{S}$ with any member of $\mathcal{T}$ is nonempty.

In this paper, we will often need to show that ultrafilters with particular properties exist. The main tool for showing that ultrafilters exist is the *Stone Prime Filter Theorem*, which guarantees that any filter subbasis and in particular any proper filter extends to an ultrafilter.

1.2 Stone duality

Stone duality tells us that every Boolean algebra $\mathcal{B}$ has an associated compact Hausdorff space $S(\mathcal{B})$, called its *Stone space*. This space may be given by the set of ultrafilters of $\mathcal{B}$ with the topology generated by the basis of clopen sets of the form

$$\widehat{L} = \{\gamma \in S(\mathcal{B}) \mid L \in \gamma\},$$

where $L \in \mathcal{B}$.

[3] In other words, γ is an *upset*.

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. Viewing βX as the Stone space of $\mathcal{P}(X)$, we will consider elements of βX to be ultrafilters of $\mathcal{P}(X)$. Note that the map sending an element x of X to the principal ultrafilter generated by $\{x\}$ defines an injective map from X into βX.

An important property of Stone-Čech compactification is that every map $f : X \to K$, where K is a compact Hausdorff space, has a unique continuous extension $\beta f : \beta X \to K$. Furthermore, every map $f : X \to Y$ (where X and Y are discrete spaces) 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 and for each γ in βX. In particular, if $X = A^*$ and u is a word of A^*, the left translation $x \to ux$ extends to a continuous map from βA^* to βA^* and right translations can be extended in the same way. In other words, the product of a word with an element of βA^* is a well defined notion, but the product of two elements of βA^* is not.[4]

Our second example is the Stone space of the Boolean algebra Reg of all *regular* subsets of A^*. It is equal to the topological space underlying the free profinite monoid on A, denoted by $\widehat{A^*}$, see e.g. [1]. 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

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

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 $\widehat{A^*}$. In constrast with the case of βA^*, 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*. The following constructions of profinite words from given ones will play an important rôle in this paper. In a compact monoid, the smallest closed subsemigroup containing a given 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 convergent sequence $x^{n!}$. Moreover, the sequence $x^{n!-1}$ is also convergent and the element to which it converges is denoted by $x^{\omega-1}$. More details can be found in [2,8,9].

1.3 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

[4] The cognoscenti may object that in the literature, $\beta\mathbb{N}$ is routinely equipped with a monoid structure, but the multiplication is not continuous with respect to both of its arguments.

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 of $\mathcal{B}$ and let $L \in \mathcal{B}$. We say that L *satisfies the* $\mathcal{B}$-*equation* $\gamma_1 \leftrightarrow \gamma_2$ provided

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

By extension, we say that $\mathcal{B}'$ *satisfies the* $\mathcal{B}$-*equation* $\gamma_1 \leftrightarrow \gamma_2$ 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 $\mathcal{E}$ 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 $\mathcal{E}$. 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.

Let A be a finite alphabet and let $\mathcal{B}$ be a Boolean algebra of languages of A^*. We say that $\mathcal{B}$ 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 $Lu^{-1} = \{x \in A^* \mid xu \in L\}$.

If $\mathcal{B}$ is closed under quotients, then the set of all equations satisfied by $\mathcal{B}$ is a kind of congruence. More precisely, the following result holds:

Proposition 1.2. *Let $\mathcal{B}$ be a Boolean algebra of languages of A^* closed under quotients and let $\gamma_1, \gamma_2 \in \beta A^*$. If $\mathcal{B}$ satisfies the equation $\gamma_1 \leftrightarrow \gamma_2$, then it satisfies the equations $u\gamma_1 \leftrightarrow u\gamma_2$ and $\gamma_1 u \leftrightarrow \gamma_2 u$ for each word $u \in A^*$.*

For a Boolean algebra of regular languages closed under quotients, a stronger property holds.

Proposition 1.3. *Let $\mathcal{B}$ be a Boolean algebra of regular languages of A^* closed under quotients and let $w_1, w_2 \in \widehat{A^*}$. If $\mathcal{B}$ satisfies the profinite equation $w_1 \leftrightarrow w_2$, then it satisfies the profinite equations $uw_1 \leftrightarrow uw_2$ and $w_1 u \leftrightarrow w_2 u$ for each profinite word $u \in \widehat{A^*}$.*

In view of these two results, it is convenient to introduce the following notation. Given $\gamma_1, \gamma_2 \in \beta A^*$, we say that *a language satisfies the ultrafilter equation* $\gamma_1 = \gamma_2$ if it satisfies all the ultrafilter equations $u\gamma_1 \leftrightarrow u\gamma_2$ and $\gamma_1 u \leftrightarrow \gamma_2 u$ for all words $u \in A^*$. Similarly, given $w_1, w_2 \in \widehat{A^*}$, we say that *a regular language satisfies the profinite equation* $w_1 = w_2$ if it satisfies the profinite equations $uw_1 \leftrightarrow uw_2$ and $w_1 u \leftrightarrow w_2 u$ for each profinite word $u \in \widehat{A^*}$. The main interest of this notation is to allow one to produce smaller sets of defining equations for a Boolean algebra closed under quotients.

In the regular case, there is a convenient connection between profinite equations and syntactic morphisms. Let L be a regular language of A^* and let $\eta : A^* \to M$ be its syntactic morphism. We denote by $\widehat{\eta} : \widehat{A^*} \to M$ the unique continuous extension of η to $\widehat{A^*}$.

Proposition 1.4. *Let $u, v \in \widehat{A^*}$, let L be a regular language of A^* and let $\eta : A^* \to M$ be its syntactic morphism.*

(1) *L satisfies the profinite equation $u \leftrightarrow v$ if and only if $\widehat{\eta}(u) \in \eta(L)$ is equivalent to $\widehat{\eta}(v) \in \eta(L)$.*

(2) *L satisfies the profinite equation $u = v$ if and only if $\widehat{\eta}(u) = \widehat{\eta}(v)$.*

Proof. (1) follows from [4, Corollary 5.1].

We prove (2). By (1), L satisfies the profinite equation $u = v$ if and only if, for every $x, y \in A^*$, $\widehat{\eta}(xvy) \in \eta(L)$ is equivalent to $\widehat{\eta}(xvy) \in \eta(L)$. Since $\widehat{\eta}(xuy) = \widehat{\eta}(x)\widehat{\eta}(u)\widehat{\eta}(y)$ and since $\widehat{\eta}$ is surjective, this is equivalent to saying that, for all $s, t \in M$,

$$s\widehat{\eta}(u)t \in \eta(L) \iff s\widehat{\eta}(v)t \in \eta(L),$$

which means that $\widehat{\eta}(u) = \widehat{\eta}(v)$ by the definition of the syntactic morphism. □

Let $\mathcal{B}$ be a Boolean algebra of languages defined by a set $\mathcal{E}$ of ultrafilter equations. It follows from Result 1 that $\mathcal{B} \cap \mathrm{Reg}$ can be defined by a set of profinite equations. The following proposition, which follows immediately from Stone duality, explains how to obtain such a defining set of profinite equations for $\mathcal{B} \cap \mathrm{Reg}$ from $\mathcal{E}$. Let $\pi_{\mathrm{Reg}} : \beta A^* \to \widehat{A^*}$ be the projection defined by

$$\pi_{\mathrm{Reg}}(\mu) = \mu \cap \mathrm{Reg},$$

and let

$$\pi_{\mathrm{Reg}}(\mathcal{E}) = \{\pi_{\mathrm{Reg}}(\mu) \leftrightarrow \pi_{\mathrm{Reg}}(\nu) \mid \mu \leftrightarrow \nu \text{ is an equation in } \mathcal{E}\}.$$

By construction, $\pi_{\mathrm{Reg}}(\mathcal{E})$ is a set of profinite equations.

Proposition 1.5. *Let $\mathcal{B}$ be a Boolean algebra of languages defined by a set of ultrafilter equations $\mathcal{E}$. Then the Boolean algebra $\mathcal{B} \cap \mathrm{Reg}$ is defined by the set of profinite equations $\pi_{\mathrm{Reg}}(\mathcal{E})$.*

Proof. Since $\mathcal{E}$ is a complete set of ultrafilter equations for $\mathcal{B}$, one has, for each language L of A^*,

$$L \in \mathcal{B} \iff \big(\text{for all equations } \mu \leftrightarrow \nu \text{ in } \mathcal{E},\ L \in \mu \iff L \in \nu\big).$$

This holds in particular for each regular language L. However, if L is regular and $\mu \in \beta A^*$ we have

$$L \in \mu \iff L \in \mu \cap \mathrm{Reg} \iff L \in \pi_{\mathrm{Reg}}(\mu).$$

Thus we get, for each each regular language L,

$$L \in \mathcal{B} \cap \mathrm{Reg} \iff (\text{for all equations } \mu \leftrightarrow \nu \text{ in } \mathcal{E},\ L \in \pi_{\mathrm{Reg}}(\mu) \iff L \in \pi_{\mathrm{Reg}}(\nu)),$$

and thus the set $\pi_{\mathrm{Reg}}(\mathcal{E})$ defines $\mathcal{B} \cap \mathrm{Reg}$. □

Here is another useful result on equations.

Proposition 1.6. *Let $f : A^* \to B^*$ be a map and let L be a subset of B^*. Then $f^{-1}(L)$ satisfies $u \leftrightarrow v$ for some $u, v \in \beta A^*$, if and only if L satisfies $\beta f(u) \leftrightarrow \beta f(v)$.*

Proof. By definition, $f^{-1}(L)$ satisfies $u \leftrightarrow v$ if and only if

$$f^{-1}(L) \in u \iff f^{-1}(L) \in v. \tag{4}$$

The definition of βf tells us that $f^{-1}(L) \in u$ if and only if $L \in \beta f(u)$. Thus (4) is equivalent to

$$L \in \beta f(u) \iff L \in \beta f(v),$$

which means that L satisfies $\beta f(u) \leftrightarrow \beta f(v)$. □

The counterpart of Proposition 1.6 for regular languages can be stated as follows:

Proposition 1.7. *Let $f : A^* \to B^*$ be a function such that the inverse image of any regular language is regular and let L be a regular language of B^*. Then $f^{-1}(L)$ satisfies the profinite equation $u \leftrightarrow v$ for some $u, v \in \widehat{A^*}$, if and only if L satisfies the profinite equation $\hat{f}(u) \leftrightarrow \hat{f}(v)$.*

1.4 More results on ultrafilters

Consider the Stone space βX of a full power set $\mathcal{P}(X)$. If Y is a subset of X, then one can identify βY with the set

$$\widehat{Y} = \{\gamma \in \beta X \mid Y \in \gamma\}.$$

Indeed, the function which maps an element γ of βY to the filter of $\mathcal{P}(X)$ generated by γ yields an ultrafilter of $\mathcal{P}(X)$ that has Y as an element. Furthermore, one may show that this function is a homeomorphism from βY to $\widehat{Y}$. The inverse is the homeomorphism from $\widehat{Y}$ to βY, which maps an element γ of $\widehat{Y}$ to the set

$$\{S \cap Y \mid S \in \gamma\} = \{S \in \gamma \mid S \subseteq Y\},$$

which by construction is an ultrafilter on $\mathcal{P}(Y)$.

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

Proposition 1.8. *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) *$K \triangle L \notin \gamma$.*

Proof. It is a consequence of the following sequence of equivalent properties:

$$\begin{aligned}
& K \in \gamma \text{ if and only if } L \in \gamma \\
\iff & (K \in \gamma \text{ and } L \in \gamma) \text{ or } (K^c \in \gamma \text{ and } L^c \in \gamma) \\
\iff & K \cap L \in \gamma \text{ or } K^c \cap L^c \in \gamma && \text{since } \gamma \text{ is a filter} \\
\iff & (K \cap L) \cup (K^c \cap L^c) \in \gamma && \text{since } \gamma \text{ is an ultrafilter} \\
\iff & K \triangle L \notin \gamma && \text{since } K \triangle L = [(K \cap L) \cup (K^c \cap L^c)]^c.
\end{aligned}$$

□

2 A Boolean algebra and its logical description

The *length* of a word u is denoted by $|u|$ or by $\ell(u)$.

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 $\boldsymbol{a}_u$ defined by

$$\boldsymbol{a}_u = \{i \in \mathrm{Dom}(u) \mid a_i = a\}.$$

For instance, if $u = aabcbaba$, then $\boldsymbol{a}_u = \{0, 1, 5, 7\}$, $\boldsymbol{b}_u = \{2, 4, 6\}$, and $\boldsymbol{c}_u = \{3\}$.

We are now ready to introduce the Boolean algebra of languages for which we will obtain ultrafilter equations. For each letter a in A and for each subset P of $\mathbb{N}$, let

$$L_P = \{u \in A^* \mid |u| \in P\}$$

and

$$L_{a,P} = \{u \in A^* \mid \boldsymbol{a}_u \subseteq P\}.$$

Let $\mathcal{B}$ be the Boolean algebra generated by the languages L_P and $L_{a,P}$ for $P \subseteq \mathbb{N}$ and $a \in A$. We first establish some combinatorial properties of $\mathcal{B}$ and then provide a logical description for it.

2.1 Combinatorial properties of $\mathcal{B}$

Let us start with some elementary but useful relations. Note that Proposition 2.2 and Proposition 2.5 are not used in this paper. However Proposition 2.5 was instrumental in [6] but was not proved there.

Proposition 2.1. *The following formulas hold:*

$$L_P \cup L_Q = L_{P \cup Q} \qquad L_P \cap L_Q = L_{P \cap Q} \tag{5}$$

$$L_P^c = L_{P^c} \qquad L_{a,P}^c = \{u \in A^* \mid \boldsymbol{a}_u \cap P^c \neq \emptyset\} \tag{6}$$

$$L_{a,P} \cap L_{a,Q} = L_{a,P \cap Q} \qquad L_{a,P}^c \cup L_{a,Q}^c = L_{a,P \cap Q}^c. \tag{7}$$

Proof. Formulas (5) follow immediately from the equalities

$$L_P \cup L_Q = \{u \in A^* \mid |u| \in P \text{ or } |u| \in Q\} = L_{P \cup Q}$$
$$L_P \cap L_Q = \{u \in A^* \mid |u| \in P \text{ and } |u| \in Q\} = L_{P \cap Q}$$

To establish (6), it suffices to observe that

$$L_P^c = \{u \in A^* \mid |u| \notin P\} = \{u \in A^* \mid |u| \in P^c\} = L_{P^c}$$

Finally, (7) follows from the relations

$$\begin{aligned} L_{a,P} \cap L_{a,Q} &= \{u \in A^* \mid \boldsymbol{a}_u \subseteq P\} \cap \{u \in A^* \mid \boldsymbol{a}_u \subseteq Q\} \\ &= \{u \in A^* \mid \boldsymbol{a}_u \subseteq P \cap Q\} = L_{a,P \cap Q} \\ L_{a,P}^c \cup L_{a,Q}^c &= (L_{a,P} \cap L_{a,Q})^c = L_{a,P \cap Q}^c. \end{aligned}$$ □

Proposition 2.1 leads to a normal form for the languages in $\mathcal{B}$.

Proposition 2.2 (Normal form). *Each language of $\mathcal{B}$ can be written as a finite intersection of languages of the form*

$$L_P \cup \bigcup_{a \in A} \Big(L^c_{a,P_a} \cup \bigcup_{i \in I_a} L_{a,P_{a,i}} \Big) \tag{8}$$

where the sets I_a are finite and the sets P, P_a and $P_{a,i}$ are subsets of $\mathbb{N}$.

Proof. Since $\mathcal{B}$ is the Boolean algebra generated by the languages L_P and $L_{a,P}$, every language of $\mathcal{B}$ can be written as a finite intersection of finite unions of languages L_P, $L_{a,P}$ or their complement. Now a simple application of Proposition 2.1 leads to the desired normal form. $\square$

We now study the behaviour of $\mathcal{B}$ with respect to left and right quotients. The following notation will help us to formulate our results. Given $P \subseteq \mathbb{N}$ and $r \in \mathbb{N}$, we set

$$P + r = \{n \in \mathbb{N} \mid n - r \in P\}$$

and

$$P - r = \{n \in \mathbb{N} \mid n + r \in P\}.$$

We first consider the left and right quotients by a letter.

Lemma 2.3. *Let a and b be two distinct letters of A and let P be an arbitrary subset of $\mathbb{N}$. Then*

$$\begin{aligned}
a^{-1}L_P &= L_{P-1} & L_P a^{-1} &= L_{P-1}; \\
a^{-1}L_{a,P} &= \begin{cases} L_{a,P-1} & \textit{if } 0 \in P, \\ \emptyset & \textit{otherwise;} \end{cases} & L_{a,P} a^{-1} &= L_{a,P} \cap L_P; \\
b^{-1}L_{a,P} &= L_{a,P-1} & L_{a,P} b^{-1} &= L_{a,P}.
\end{aligned}$$

Proof. We first have

$$\begin{aligned}
a^{-1}L_P &= \{u \in A^* \mid au \in L_P\} = \{u \in A^* \mid |au| \in P\} \\
&= \{u \in A^* \mid |u| + 1 \in P\} = L_{P-1}; \\
L_P a^{-1} &= \{u \in A^* \mid ua \in L_P\} = \{u \in A^* \mid |ua| \in P\} \\
&= \{u \in A^* \mid |u| + 1 \in P\} = L_{P-1}.
\end{aligned}$$

Observing that $\boldsymbol{a}_{au} = \{0\} \cup (\boldsymbol{a}_u + 1)$ and $\boldsymbol{a}_{ua} = \boldsymbol{a}_u \cup \{|u|\}$, we get

$$\begin{aligned}
a^{-1}L_{a,P} &= \{u \in A^* \mid au \in L_{a,P}\} = \{u \in A^* \mid \{0\} \cup (\boldsymbol{a}_u + 1) \subseteq P\} \\
&= \begin{cases} L_{a,P-1} & \text{if } 0 \in P, \\ \emptyset & \text{otherwise;} \end{cases} \\
L_{a,P} a^{-1} &= \{u \in A^* \mid ua \in L_{a,P}\} = \{u \in A^* \mid \boldsymbol{a}_u \cup \{|u|\} \subseteq P\} \\
&= \{u \in A^* \mid \boldsymbol{a}_u \subseteq P \text{ and } |u| \in P\} = L_{a,P} \cap L_P.
\end{aligned}$$

Now, if $b \neq a$, $\boldsymbol{a}_{bu} = \boldsymbol{a}_u + 1$ and $\boldsymbol{a}_{ub} = \boldsymbol{a}_u$ and consequently,

$$b^{-1}L_{a,P} = \{u \in A^* \mid bu \in L_{a,P}\} = \{u \in A^* \mid \boldsymbol{a}_u + 1 \subseteq P\} = L_{a,P-1};$$
$$L_{a,P}b^{-1} = \{u \in A^* \mid ub \in L_{a,P}\} = \{u \in A^* \mid \boldsymbol{a}_u \subseteq P\} = L_{a,P}. \qquad \square$$

Corollary 2.4. *The Boolean algebras $\mathcal{B}$ and $\mathcal{B} \cap \mathrm{Reg}$ are closed under quotients.*

Proof. Lemma 2.3 shows that the quotients of the generators of $\mathcal{B}$ by a letter are still in $\mathcal{B}$. It follows by induction that the quotients of the generators of $\mathcal{B}$ by any word are still in $\mathcal{B}$. Since quotients commute with Boolean operations, it follows that $\mathcal{B}$ is closed under quotients. Since regular languages are closed under quotients, it also follows that $\mathcal{B} \cap \mathrm{Reg}$ is also closed under quotients. $\square$

Proposition 2.5. *For each word $u \in A^*$, the Boolean algebras $\mathcal{B}$ and $\mathcal{B} \cap \mathrm{Reg}$ are closed under the operation $L \to uL$.*

Proof. By induction, it suffices to prove that $\mathcal{B}$ is closed under the operation $L \to aL$ for each letter $a \in A$. But this is a consequence of Proposition 2.2 and of the following lemma:

Lemma 2.6. *Let a and b be two distinct letters of A and let P be an arbitrary subset of $\mathbb{N}$. Then*

$$aL_P = aA^* \cap L_{P+1} \qquad aA^* = \bigcap_{c \neq a} L_{c,\mathbb{N}-\{0\}} \tag{9}$$
$$aL_{a,P} = aA^* \cap L_{a,(P+1)\cup\{0\}} \qquad aL^c_{a,P} = aA^* \cap L^c_{a,(P+1)\cup\{0\}} \tag{10}$$
$$bL_{a,P} = bA^* \cap L_{a,P+1} \qquad bL^c_{a,P} = bA^* \cap L^c_{a,P+1}. \tag{11}$$

Proof. We first have

$$aL_P = \{au \mid |u| \in P\} = aA^* \cap L_{P+1}$$
$$\bigcap_{c \neq a} L_{c,\mathbb{N}-\{0\}} = \bigcap_{c \neq a} \{u \in A^* \mid \boldsymbol{c}_u \subseteq \mathbb{N} - \{0\}\}$$
$$= \bigcap_{c \neq a} \{u \in A^* \mid 0 \notin \boldsymbol{c}_u\} = aA^*$$

Furthermore we have

$$\begin{aligned}
aL_{a,P} &= \{au \mid \boldsymbol{a}_u \subseteq P\} = aA^* \cap L_{a,(P+1)\cup\{0\}};\\
aL^c_{a,P} &= aA^* \cap \{u \in A^* \mid \boldsymbol{a}_u \cap (P^c + 1) \neq \emptyset\}\\
&= aA^* \cap L^c_{a,(P+1)\cup\{0\}} \text{ since } (P^c+1)^c = (P+1) \cup \{0\};\\
bL_{a,P} &= \{bu \mid \boldsymbol{a}_u \subseteq P\} = bA^* \cap L_{a,P+1};\\
bL^c_{a,P} &= bA^* \cap \{u \in A^* \mid \boldsymbol{a}_u \cap (P^c + 1) \neq \emptyset\}\\
&= bA^* \cap \big\{u \in A^* \mid \boldsymbol{a}_u \cap \big((P^c+1) \cup \{0\}\big) \neq \emptyset\big\}\\
&= bA^* \cap L^c_{a,(P+1)} \text{ since } \big((P^c+1) \cup \{0\}\big)^c = P + 1. \qquad \square
\end{aligned}$$

So far, we have considered the languages L_P and $L_{a,P}$ as languages of A^*, where A was a fixed alphabet. But for the remainder of this section, we need to consider several alphabets simultaneously. Recall that a *class of languages* $\mathcal{C}$ assigns to each finite alphabet A a set $\mathcal{C}(A^*)$ of languages of A^*. In particular, we define the classes of languages $\mathcal{B}$ and $\mathcal{B} \cap \mathrm{Reg}$ as follows: for each alphabet A, $\mathcal{B}(A^*)$ is the Boolean algebra generated by the languages of the form L_P or $L_{a,P}$, where $P \subseteq \mathbb{N}$ and $a \in A$ and $\mathcal{B} \cap \mathrm{Reg}(A^*)$ is the Boolean algebra of all regular languages in $\mathcal{B}(A^*)$.

Actually, the definition of L_P and $L_{a,P}$ also depends on the alphabet, but in order to avoid cumbersome notation, we will keep the notation for L_P and $L_{a,P}$ regardless of the alphabet, the context making it clear whether these languages are considered as subsets of A^* or of B^*.

Recall that a monoid morphism from B^* to A^* is *length-multiplying* if there exists a positive integer k such that, for every $b \in B$, $|f(b)| = k$.

Lemma 2.7. *Let $f : B^* \to A^*$ be a morphism such that $|f(u)| = k|u|$ for all words $u \in B^*$. Then*

$$f^{-1}(L_P) = L_Q \qquad \text{where } Q = \{n \in \mathbb{N} \mid kn \in P\} \tag{12}$$

$$f^{-1}(L_{a,P}) = \bigcap_{b \in B} \bigcap_{s \in \boldsymbol{a}_{f(b)}} L_{b,Q_s} \qquad \text{where } Q_s = \{n \in \mathbb{N} \mid kn + s \in P\} \tag{13}$$

Proof. Formula (12) follows from the equalities

$$\begin{aligned} f^{-1}(L_P) &= \{u \in B^* \mid f(u) \in L_P\} = \{u \in B^* \mid |f(u)| \in P\} \\ &= \{u \in B^* \mid k|u| \in P\} = \{u \in B^* \mid |u| \in Q\} = L_Q. \end{aligned}$$

To establish Formula (13), first observe that for a word $u = u_0 u_1 \cdots u_{n-1}$ in B^*, the letter in position $kr + s$ in $f(u)$ is an a if and only if the letter in position s in $f(u_r)$ is an a. It follows that $\boldsymbol{a}_{f(u)}$ is the disjoint union of the sets $k\boldsymbol{b}_u + s$, where b runs over B and s runs over $\boldsymbol{a}_{f(b)}$. In particular, $\boldsymbol{a}_{f(u)}$ is a subset of P if and only if every set $k\boldsymbol{b}_u + s$ is a subset of P. Consequently, we get

$$\begin{aligned} f^{-1}(L_{a,P}) &= \{u \in B^* \mid f(u) \in L_{a,P}\} = \{u \in B^* \mid \boldsymbol{a}_{f(u)} \subseteq P\} \\ &= \bigcap_{b \in B} \bigcap_{s \in \boldsymbol{a}_{f(b)}} \{u \in B^* \mid k\boldsymbol{b}_u + s \subseteq P\} \\ &= \bigcap_{b \in B} \bigcap_{s \in \boldsymbol{a}_{f(b)}} \{u \in B^* \mid \boldsymbol{b}_u \subseteq Q_s\} = \bigcap_{b \in B} \bigcap_{s \in \boldsymbol{a}_{f(b)}} L_{b,Q_s}. \end{aligned} \qquad \square$$

Corollary 2.8. *The classes of languages $\mathcal{B}$ and $\mathcal{B} \cap \mathrm{Reg}$ are closed under the operation $L \to f^{-1}(L)$, for any length-multiplying morphism f.*

Proof. The result follows from Lemma 2.7 and from the fact that inverses of functions commute with Boolean operations. $\square$

It follows from Corollaries 2.4 and 2.8 that $\mathcal{B} \cap \mathrm{Reg}$ is a length-multiplying variety (or *lm*-variety) of languages, in the sense of Straubing [12].

2.2 Logical description of $\mathcal{B}$

Let us turn to the logical description of $\mathcal{B}$. For each subset P of $\mathbb{N}$, let us define two entities: a 0-ary predicate which is true on u if and only if $|u| \in P$ and a unary uniform numerical relation[5] defined by $P(n) = P \cap \{0, \dots, n-1\}$. Its interpretation on a word u is the subset $P(|u|)$ of $\{0, \dots, |u|-1\}$.

We denote by $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$ the class of languages defined by first-order sentences 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 at most one.

When defining the language given by a formula, it is preferable to avoid the empty word, as several problems arise when dealing with empty structures in logic.[6] Therefore, the language defined by a sentence φ is the set

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

For instance if $\varphi = \exists x \ \boldsymbol{a}x$, then $L(\varphi) = A^*aA^*$. We have the following logical description of our Boolean algebra $\mathcal{B}$.

Theorem 2.9. *A language L of A^+ belongs to $\mathcal{B}$ if and only it belongs to* $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$.

Proof. First we show that every language of $\mathcal{B}$ contained in A^+ belongs to $\mathbf{FO}[\mathcal{N}_0, \mathcal{N}_1^u]$. It suffices to do it for the generators of $\mathcal{B}$, namely the languages of the form L_P and $L_{a,P}$, where $P \subseteq \mathbb{N}$ and $a \in A$. The language L_P is defined by the atomic formula $|u| \in P$ and the language $L_{a,P}$ is defined by the formula $\forall x \ (\boldsymbol{a}x \to x \in P)$.

For the other direction we need to show that every language definable in the logic can be written as a Boolean combination of the L_P and $L_{a,P}$. Let us now have a closer look at the formulas of our logic fragment. Since we do not allow equality, the atomic formulas are $|u| \in P$, **true**, **false**, $\boldsymbol{a}x$ or $x \in P$ for some variable x and some subset P of $\mathbb{N}$ (viewed as a unary uniform numerical relation). Furthermore, $\neg \boldsymbol{a}x$ is equivalent to $\bigvee_{b \neq a} \boldsymbol{b}x$ and $\neg(x \in P)$ is equivalent to $x \in P^c$. Thus every quantifier-free formula can be written as a disjunction of conjunctions of atomic formulas.

Since all the predicates are 0-ary or unary, and since we do not allow equality, we cannot express any relationship between any two variables. Hence nested quantifiers can be pulled apart. It follows that every sentence is equivalent to a Boolean combination of existential formulas of depth at most one. Thus every sentence is a Boolean combination of sentences of the form

(1) $\varphi_P = |u| \in P$, where $P \subseteq \mathbb{N}$,

---

[5] Following the terminology of [11], a unary *numerical relation* R associates to each $n > 0$ a subset $R(n)$ of $\{0, \dots, n-1\}$. 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.

[6] See `http://en.wikipedia.org/wiki/First-order_logic#Empty_domains`.

(2) $\varphi_{a,P} = \exists x\ (\boldsymbol{a}x \wedge x \in P)$, where $P \subseteq \mathbb{N}$ and $a \in A$.

It only remains to show that the languages $L(\varphi_P)$ and $L(\varphi_{a,P})$ are in $\mathcal{B}$. Clearly, $L(\varphi_P) = L_P \in \mathcal{B}$. The language defined by $\varphi_{a,P}$ is

$$L(\varphi_{a,P}) = \{u \in A^+ \mid \boldsymbol{a}_u \cap P \neq \emptyset\} = \{u \in A^+ \mid \boldsymbol{a}_u \nsubseteq P\} = A^+ - L_{a,P^c},$$

and thus $L(\varphi_{a,P})$ belongs to $\mathcal{B}$. ◻

3 Some ultrafilter equations for $\mathcal{B}$

Let $\pi_0 : A^* \times \mathbb{N}^k \to A^*$ be the projection defined by $\pi_0(u, n_1, \ldots, n_k) = u$ and let, for $1 \leqslant i \leqslant k$, let $\pi_i : A^* \times \mathbb{N}^k \to \mathbb{N}$ be the projection on $\mathbb{N}$ defined by $\pi_i(u, n_1, \ldots, n_k) = n_i$.

We first characterise the ultrafilter of $\mathcal{P}(A^* \times \mathbb{N}^k)$ having the same projections under each π_i, for $1 \leqslant i \leqslant k$.

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.

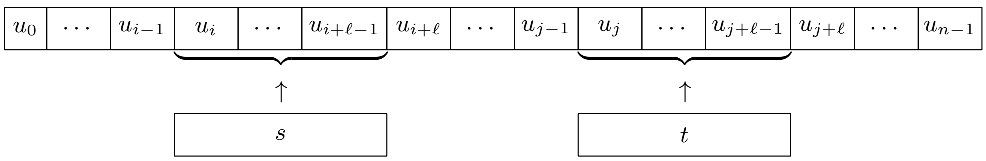
For the second assertion, suppose there is an $\alpha \in \beta\mathbb{N}$ such that (1) and (2) hold and $\{P_1, \ldots, P_n\}$ is a partition of $\mathbb{N}$. Then $\bigcup_{j=1}^{n} P_j = \mathbb{N}$ implies $P_\ell \in \alpha$ for some ℓ and thus $A^* \times P_\ell^k \in \gamma$ by (2). Since γ is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

Suppose now that γ satisfies (3) and let $\alpha = \{P \mid A^* \times P^k \in \gamma\}$. Then $\emptyset \notin \alpha$ and 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*

$$\beta f_{s,t}(\gamma) = \beta f_{t,s}(\gamma). \tag{14}$$

Proof. Let $a \in A$ and $P \subseteq \mathbb{N}$. By Proposition 1.2, it suffices to prove that $L_{a,P}$ and L_P satisfy the equations

$$\beta f_{s,t}(\gamma) \leftrightarrow \beta f_{t,s}(\gamma). \tag{15}$$

First we have

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

Thus (15) 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 1.8 this is equivalent to $S \notin \gamma$, where

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

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

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

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

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

If $S \in \gamma$, then $\bigcup_{i=0}^{\ell-1} \Big(\pi_1^{-1}(P-i) \bigtriangleup \pi_2^{-1}(P-i)\Big) \in \gamma$ and since γ is an ultrafilter, $\pi_1^{-1}(P-i) \bigtriangleup \pi_2^{-1}(P-i) \in \gamma$ 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 $\pi_1^{-1}(Q) \bigtriangleup \pi_2^{-1}(Q) \notin \gamma$, or equivalently, $(\pi_1^{-1}(Q) \bigtriangleup \pi_2^{-1}(Q))^c \in \gamma$. But this is a direct consequence of Proposition 3.1(3) since

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

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 (15) if and only if $f_{s,t}^{-1}(L_P) \bigtriangleup f_{t,s}^{-1}(L_P) \notin \gamma$. 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 $f_{s,t}^{-1}(L_P) \bigtriangleup f_{t,s}^{-1}(L_P) = \emptyset$ and therefore it does not belong to γ. □

The ultrafilter equations of Theorem 3.2 tell us that our Boolean algebra (or equivalently our logic fragment) cannot tell the order of occurrence of letters occurring in equivalent positions. We need another family of ultrafilter equations in order to characterise $\mathcal{B}$. These tell us that, though $\mathcal{B}$ can tell whether or not a letter occurs in a set of equivalent positions, it cannot tell how many times each such letter occurs. For this purpose, we need functions $f_{s_1,s_2,s_3} : A^* \times \mathbb{N}^3 \to A^*$ given by $s_1, s_2, s_3 \in A^*$ with $|s_1| = |s_2| = |s_3|$ and defined by

$$f_{s_1,s_2,s_3}(u, n_1, n_2, n_3) = u(s_1@n_1, s_2@n_2, s_3@n_3),$$

where $u(s_1@n_1, s_2@n_2, s_3@n_3)$ is the word obtained from u by putting s_i at position n_i when $n_1 + |s_1| \leqslant n_2$, $n_2 + |s_2| \leqslant n_3$ and $n_3 + |s_3| \leqslant |u|$ and as u otherwise. One can then prove the following theorem.

Theorem 3.3. *Let $s, t \in A^*$ with $|s| = |t|$. If $\gamma \in \beta(A^* \times \mathbb{N}^3)$ and $\beta\pi_1(\gamma) = \beta\pi_2(\gamma) = \beta\pi_3(\gamma)$, then $\mathcal{B}$ satisfies the equation $\beta f_{t,s,s}(\gamma) = \beta f_{t,t,s}(\gamma)$.*

Proof. The proof is very similar to the proof of Theorem 3.2 but is based on $f_{s_1,s_2,s_3} : A^* \times \mathbb{N}^3 \to A^*$. □

The ultrafilter equations introduced in this section can be used to prove separation results for nonregular languages. To illustrate this, we show that the set of words of odd length with an a in the middle position does not belong to $\mathcal{B}$. Let

$$\text{MIDDLE\_}a = \{uav \mid u, v \in \{a, b\}^* \text{ and } |u| = |v|\}$$

Proposition 3.4. *The language* MIDDLE_a *does not belong to $\mathcal{B}$.*

The proof relies on a technique that we will use again in Section 4. It consists in proving that adding certain sets to the filter subbasis

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

still yields a filter subbasis.

Proof. Let

$$S = \{(u, n_1, n_2) \in A^* \times \mathbb{N}^2 \mid n_1 < n_2 \leqslant 2n_1 + 1 = |u|\}$$

We show that adding the set S to the filter subbasis $\mathcal{F}$ yields again a filter subbasis. To this end, let $\{P_1, \ldots, P_n\}$ be a partition of $\mathbb{N}$. Then, for $m \in \mathbb{N}$ with $n \leqslant m$ there are $m+1$ natural numbers n_2 with $m < n_2 \leqslant 2m+1$. By the pigeonhole principle, there is an i with $1 \leqslant i \leqslant n$ and $n_1, n_2 \in \mathbb{N}$ such that $n_1, n_2 \in P_i$ and $m < n_1 < n_2 \leqslant 2m+1$. It follows that $n_1 < n_2 \leqslant 2n_1 + 1$ and thus, for any $u \in A^*$ with $|u| = 2n_1 + 1$, we have $(u, n_1, n_2) \in S \cap (A^* \times P_i^2)$ and thus $S \cap (A^* \times P_i^2)$ is nonempty and the union of the two families is a filter subbasis as required.

Now let $\gamma \in \mathcal{P}(A^* \times \mathbb{N}^2)$ be an ultrafilter containing this larger filter subbasis. Since $\mathcal{F} \subseteq \gamma$, it follows, by Proposition 3.1.3, that $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$ where $\pi_i \colon A^* \times \mathbb{N}^2 \to \mathbb{N}, (w, n_1, n_2) \mapsto n_i$ for $i = 1, 2$. Therefore, by Theorem 3.2, it follows that $\mathcal{B}$ satisfies $\beta f_{a,b}(\gamma) \leftrightarrow \beta f_{b,a}(\gamma)$. However, if $(u, n_1, n_2) \in S$, then $|u| = 2n_1 + 1$ and $u(a@n_1, b@n_2) \in \textsc{Middle\_}a$, but $u(b@n_1, a@n_2) \notin \textsc{Middle\_}a$. That is, $f_{a,b}(S) \subseteq \textsc{Middle\_}a$ and $f_{b,a}(S) \subseteq (\textsc{Middle\_}a)^c$. Now $f_{a,b}(S) \subseteq \textsc{Middle\_}a$ is equivalent to $S \subseteq f_{a,b}^{-1}(\textsc{Middle\_}a)$ and since $S \in \gamma$, it follows that $f_{a,b}^{-1}(\textsc{Middle\_}a) \in \gamma$, or equivalently that $\textsc{Middle\_}a \in \beta f_{a,b}(\gamma)$. Similarly, $f_{b,a}(S) \subseteq (\textsc{Middle\_}a)^c$ implies $(\textsc{Middle\_}a)^c \in \beta f_{b,a}(\gamma)$ or equivalently $\textsc{Middle\_}a \notin \beta f_{b,a}(\gamma)$. That is, $\textsc{Middle\_}a$ does not satisfy the equation $\beta f_{a,b}(\gamma) = \beta f_{b,a}(\gamma)$ and thus, by Theorem 3.2, $\textsc{Middle\_}a$ is not in $\mathcal{B}$. ◻

4 Completeness

In this section we show that the two families of ultrafilter equations introduced in the previous section are sufficient for characterising $\mathcal{B}$. For $a, b \in A$, let $\mathcal{E}_{ab=ba}$ denote the family of equations

$$\beta f_{ab}(\gamma) \leftrightarrow \beta f_{ba}(\gamma), \qquad (\mathcal{E}_{ab=ba})$$

where γ ranges over all elements of $\beta(A^* \times \mathbb{N}^2)$ satisfying $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$, and let $\mathcal{E}_{aab=abb}$ denote the family of equations

$$\beta f_{aab}(\gamma) \leftrightarrow \beta f_{abb}(\gamma), \qquad (\mathcal{E}_{aab=abb})$$

where γ ranges over all elements of $\beta(A^* \times \mathbb{N}^3)$ satisfying $\beta\pi_1(\gamma) = \beta\pi_2(\gamma) = \beta\pi_3(\gamma)$. We will show that any $L \in \mathcal{P}(A^*)$ which satisfies both $\mathcal{E}_{ab=ba}$ and

$\mathcal{E}_{aab=abb}$ for all $a, b \in A$ must belong to $\mathcal{B}$.

The proof may be divided into the following stages: First we define, for every language $L \in \mathcal{P}(A^*)$, a binary relation R_L on $\mathbb{N}$, which, roughly speaking, relates two positions provided L cannot differentiate between them. Next we prove that, if L satisfies the equations $\mathcal{E}_{ab=ba}$ for all $a, b \in A$, then R_L contains an equivalence relation of finite index. This allows us to concentrate on infinite subsets $P \subseteq \mathbb{N}$ such that P^2 is entirely contained in R_L. We then show that for such sets the equations $\mathcal{E}_{aab=abb}$, where a and b range over all letters of A allow us, for sufficiently long words, to decide membership in L based only on equality of words outside P and on the set of letters occurring within P. Finally, we show that the combination of the two families of equations allow us to prove completeness.

4.1 A binary relation on positions given by a language

The *support* of a permutation on $\mathbb{N}$ is the set of its non-fixpoints. Let σ be a permutation on $\mathbb{N}$ and $w \in A^*$. If the support of σ is contained in $\{0, \ldots, |w|-1\}$, we denote by $w \cdot \sigma$ the word defined by

$$(w \cdot \sigma)_k = w_{\sigma(k)}$$

for $0 \leqslant k \leqslant |w| - 1$. A permutation σ with finite support is said to be *compatible* with L provided that for all $w \in A^*$, if the support of σ is contained in $\{0, \ldots, |w| - 1\}$, then

$$w \in L \iff w \cdot \sigma \in L.$$

We denote the set of all permutations compatible with L by $\mathrm{Comp}(L)$. Note that $\mathrm{Comp}(L)$ contains the identity and is closed under inverses. While $\mathrm{Comp}(L)$ is not closed under composition in general, we do have that if the supports of σ and τ are both contained in the support of $\sigma \circ \tau$ (so that all words needed to be considered in checking compatibility of the composition are covered by the compatibility of each of σ and τ), then $\sigma, \tau \in \mathrm{Comp}(L)$ implies $\sigma \circ \tau \in \mathrm{Comp}(L)$. Let R_L be the binary relation on $\mathbb{N}$ defined by

$$i \; R_L \; j \iff i = j \text{ or the transposition } (i\, j) \text{ is compatible with } L.$$

Proposition 4.1. *For each language L of A^*, the relation R_L is reflexive and symmetric. Furthermore, if σ is a permutation with finite support satisfying $n \; R_L \; \sigma(n)$ for all n, then σ is compatible with L.*

Proof. The relation R_L is clearly reflexive and symmetric. For the second assertion, let σ be a permutation of finite support such that $n \; R_L \; \sigma(n)$ for all n. As any permutation with finite support, σ may be written as a finite product of disjoint finite cycles. Furthermore, any cycle $(n_1 n_2 \ldots n_k)$ may be written as a product of transpositions in the form

$$(n_1 n_2 \ldots n_k) = (n_2 n_3)(n_3 n_4) \ldots (n_{k-1} n_k)(n_k n_1),$$

and since each of these transpositions is compatible with L, it follows that the cycle $(n_1 n_2 \dots n_k)$ is compatible with L and σ is also compatible with L. $\square$

Note that for any word $w \in A^*$ with $k, l, m \leqslant |w|$, we have

$$w\cdot(k\,m) = [[w\cdot(k\,l)]\cdot(l\,m)]\cdot(k\,l),$$

so, if both $(k\,l)$ and $(l\,m)$ are compatible with L, then $w \in L$ if and only if $w\cdot(k\,m) \in L$. However, if $k, m < l$ it may happen that there is a word $w \in L$ with $k, m \leqslant |w| < l$ with $w\cdot(k\,m) \notin L$ even though both $(k\,l)$ and $(l\,m)$ are compatible with L. It follows that in general R_L is not transitive.

4.2 R_L and $\mathcal{E}_{ab=ba}$

If L satisfies the equations $\mathcal{E}_{ab=ba}$, we get close to having that R_L is an equivalence relation in the following sense.

Lemma 4.2. *If a language L of A^* satisfies the equations $\mathcal{E}_{ab=ba}$ for all $a, b \in A$, then R_L contains an equivalence relation of finite index.*

Proof. For $(a, b) \in A^2$, let

$$S_{ab} = \{(u, k, \ell) \in A^* \times \mathbb{N}^2 \mid k < \ell < |u|, u_k = a, u_\ell = b,\\ u \in L \text{ but } u\cdot(k\,\ell) \notin L\}$$

and

$$M_{ab} = \{(k, \ell) \in \mathbb{N}^2 \mid \text{there exists } u \in A^* \text{ such that}\\ (u, k, \ell) \in S_{ab} \text{ or } (u, \ell, k) \in S_{ab}\}.$$

Then we have

$$R_L^c = \bigcup_{(a,b)\in A^2} M_{ab}.$$

We show that for all $(a, b) \in A^2$ there is a finite partition $\{P_1, \dots, P_n\}$ of $\mathbb{N}$ such that the corresponding equivalence relation θ_{ab} is disjoint from M_{ab}. To see this, suppose that, for each finite partition $\{P_1, \dots, P_n\}$ of $\mathbb{N}$,

$$M_{ab} \cap (\bigcup_{i=1}^{n} P_i^2) \neq \emptyset.$$

Then adding S_{ab} to the filter subbasis $\mathcal{F}$ introduced on page 16 yields a filter subbasis, and thus there is an ultrafilter $\gamma \in \beta(A^* \times \mathbb{N}^2)$ containing $\mathcal{F}$ and having S_{ab} as an element. Now it follows by the definition of S_{ab} that $f_{ab}(S_{ab}) \subseteq L$ or equivalently that $S_{ab} \subseteq f_{ab}^{-1}(L)$. Thus $f_{ab}^{-1}(L) \in \gamma$ and thus $L \in \beta f_{ab}(\gamma)$. Also by definition of S_{ab} we have $f_{ba}(S_{ab}) \subseteq L^c$ and thus $L \notin \beta f_{ba}(\gamma)$. By contraposition, if L satisfies $\mathcal{E}_{ab=ba}$, then there is an equivalence relation θ_{ab} of finite index which is disjoint from M_{ab}. Setting $\theta = \bigcap_{a,b\in A} \theta_{ab}$, we see that θ is an equivalence relation of finite index contained in R_L since

$$\theta = \bigcap_{(a,b)\in A^2} \theta_{ab} \subseteq \bigcap_{(a,b)\in A^2} M_{ab}^c = R_L. \qquad \square$$

Corollary 4.3. *If a language L of A^* satisfies the equations $\mathcal{E}_{ab=ba}$ for all $a, b \in A$, then R_L contains an equivalence relation of finite index for which each finite equivalence class is a singleton.*

Proof. By Lemma 4.2, if L satisfies the equations $\mathcal{E}_{ab=ba}$ for all $a, b \in A$, then R_L contains an equivalence relation θ which is of finite index. It follows that θ has only finitely many finite equivalence classes. By splitting each of these finite equivalence classes into singleton classes, we obtain an equivalence relation θ', still of finite index, which is contained in θ, and thus also in R_L, with the required property. $\square$

We will use the following notation. For $w \in A^*$, $a \in A$, and $P \subseteq \mathbb{N}$, we set

$$|w|_{a,P} = |\boldsymbol{a}_w \cap P| = |\{n \in P \mid w_n = a\}|.$$

Proposition 4.4. *Let L be a language of A^* and let θ be an equivalence relation of finite index contained in R_L. Let u and v be two words such that $|u| = |v|$ and $|u|_{a,P} = |v|_{a,P}$ for each $a \in A$ and each equivalence class P of θ. Then*

$$u \in L \iff v \in L.$$

Proof. Let $n = |u| = |v|$ and let P be an equivalence class of θ. For each $a \in A$, the sets $\boldsymbol{a}_u \cap P$ and $\boldsymbol{a}_v \cap P$ have the same cardinality and thus there exists a bijection

$$\sigma_{a,P} \colon \boldsymbol{a}_u \cap P \to \boldsymbol{a}_v \cap P.$$

Observe that the sets $\boldsymbol{a}_u \cap P$ (respectively $\boldsymbol{a}_v \cap P$), where $a \in A$, are pairwise disjoint and their union is $P \cap \{0, \ldots, n-1\}$. Therefore one can define a permutation σ_P on $\mathbb{N}$ of support contained in $P \cap \{0, \ldots, n-1\}$ by setting

$$\sigma_P(k) = \begin{cases} \sigma_{a,P}(k) & \text{if } k \in P \cap \{0, \ldots, n-1\} \text{ and } u_k = a \\ k & \text{if } k \notin P \cap \{0, \ldots, n-1\}. \end{cases}$$

Since $P \times P$ is contained in R_L, one has $k \ R_L \ \sigma_P(k)$ for all k, and thus by Proposition 4.1, σ_P is compatible with L. Let $P_1, \ldots, P_r$ be the equivalence classes of θ. Then the permutations $\sigma_{P_1}, \ldots, \sigma_{P_r}$ have pairwise disjoint support and hence pairwise commute. Their product (in any order) is a permutation σ of support $\{0, \ldots, n-1\}$ which is also compatible with L. Finally, since $u \cdot \sigma = v$ by construction, we get that $u \in L$ if and only if $v \in L$. $\square$

4.3 Infinite sets of R_L-equivalent positions and $\mathcal{E}_{aab=abb}$

We will need the following notation. For $w \in A^*$ and $P \subseteq \mathbb{N}$, we let

$$c_P(w) = \{a \in A \mid \text{there exists } n \in P \text{ such that } w_n = a\}.$$

Lemma 4.5. *Let L be a language of A^* satisfying the equations $\mathcal{E}_{aab=abb}$ for all $a, b \in A$. Then there exists $n \in \mathbb{N}$ such that for all $u, v \in A^*$, if*

(i) $n \leqslant |u| = |v|$,

(ii) $u_i = v_i$ *for all* $i \notin P$,

(iii) $c_P(u) = c_P(v)$,

then

$$u \in L \iff v \in L.$$

Proof. By way of contraposition, we suppose that for each $n \in \mathbb{N}$ there exist two words of A^*, $u(n)$ and $v(n)$ satisfying (i)–(iii) and $u(n) \in L$ but $v(n) \notin L$.

As a first step, we prove that we may assume in addition that for each n, there exist $(a_n, b_n) \in A^2$ such that $u(n)$ and $v(n)$ satisfy

(iv)

$$\begin{aligned} |u(n)|_{a_n,P} &= |v(n)|_{a_n,P} + 1 \\ |v(n)|_{b_n,P} &= |u(n)|_{b_n,P} + 1 \\ |u(n)|_{c,P} &= |v(n)|_{c,P} \text{ for all } c \in A \text{ with } a_n \neq c \neq b_n. \end{aligned}$$

If for each $a \in A$ we have $|u(n)|_{a,P} = |v(n)|_{a,P}$, then by (ii) this would be true for each θ equivalence class and thus by Proposition 4.4 we would have $u(n) \in L$ if and only if $v(n) \in L$. Thus there exists $a \in A$ with $|u(n)|_{a,P} \neq |v(n)|_{a,P}$. Consider the graph $G = (V, E)$ on

$$V = \{w \in A^* \mid |w| = |u(n)|, w_i = u_i \text{ for all } i \notin P, \text{ and } c_P(w) = c_P(u(n))\}$$

given by $(w, w') \in E$ if and only if there exist $a, b \in A$ with $|w|_{a,P} = |w'|_{a,P} + 1$, $|w'|_{b,P} = |w|_{b,P} + 1$, and $|w|_{c,P} = |w'|_{c,P}$ for all $c \in A$ with $a \neq c \neq b$. It is not hard to see that G is connected and that $u(n), v(n) \in V$. Thus there is a path in G from $u(n)$ to $v(n)$ and there must be an edge (w, w') on this path such that $w \in L$ and $w' \notin L$. By picking w for $u(n)$ and w' for $v(n)$ it follows that we may assume that (i)-(iv) hold for $u(n)$ and $v(n)$.

Now let $p : \mathbb{N} \to A^2$ be the map defined by $p(n) = (a_n, b_n)$. By the Pigeonhole Principle there is a pair $(a, b) \in A^2$ such that the set

$$M = p^{-1}(a, b)$$

is infinite.

We claim that for all $i, j, k \in P$ with $i < j < k$ there is $x \in A^*$ with $f_{aab}(x, i, j, k) \in L$ and $f_{abb}(x, i, j, k) \notin L$. To show this, let $i, j, k \in P$ with $i < j < k$. Let $n \in M$ with $k < n$. Then the words $u = u(n)$ and $v = v(n)$ satisfy Conditions (i)–(iv). Also note that by definition of M we have $a_n = a$ and $b_n = b$. Conditions (iii) and (iv) imply that u contains at least two occurrences of a, say in positions $i' \neq j'$ both in P, and at least one b, say in position k' also in P. Now let σ be any permutation of support contained in $P \cap \{0, \ldots, |u| - 1\}$ which maps i', j' and k' to i, j and k, respectively and let $x = u \cdot \sigma$. Since P is an equivalence class contained in R_L, one has $p \; R_L \; \sigma(p)$ for all $p \in \mathbb{N}$. It follows by Proposition 4.1 that $x \in L$. Furthermore, the equality $x = f_{aab}(x, i, j, k)$ holds by construction. The word $x' = f_{abb}(x, i, j, k)$ satisfies $|x'|_{c,P} = |v|_{c,P}$ for all

$c \in A$ and $x'_i = v_i$ for all $i \notin P$, so by Proposition 4.4 we have $f_{abb}(x, i, j, k) \notin L$, which proves the claim.

Finally, we let

$$S = \{(x, i, j, k) \in A^* \times \mathbb{N}^3 \mid f_{aab}(x, i, j, k) \in L \text{ and } f_{abb}(x, i, j, k) \notin L\}.$$

For any partition $\{P_1, \ldots, P_r\}$ of $\mathbb{N}$ there is an $i \in \{1, \ldots, r\}$ such that $P \cap P_i$ is infinite. Now picking $i < j < k$ in $P \cap P_i$, the claim shows that there exists $x \in A^*$ such that $(x, i, j, k) \in S$ and thus $(A^* \times P_i^3) \cap S$ is nonempty. As in the proof of Lemma 4.2 it now follows that there exists $\gamma \in \beta(A^* \times \mathbb{N}^3)$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma) = \beta\pi_3(\gamma)$ and $L \in \beta f_{aab}(\gamma)$ but $L \notin \beta f_{abb}(\gamma)$.

Thus L does not satisfy the equations $\mathcal{E}_{aab=abb}$, which proves the lemma by contraposition. □

4.4 Proof of completeness

Lemma 4.6. *Let L be a language of A^* satisfying the equations $\mathcal{E}_{ab=ba}$ and $\mathcal{E}_{aab=abb}$ for all $a, b \in A$. Then there exists a finite index equivalence relation θ contained in R_L and an $n \in \mathbb{N}$ such that for all $u, v \in A^*$, if $n \leqslant |u| = |v|$ and*

$$c_P(u) = c_P(v) \text{ for each } \theta \text{ equivalence class } P,$$

then

$$u \in L \iff v \in L.$$

Proof. If L satisfies the equations $\mathcal{E}_{ab=ba}$ then by Corollary 4.3, R_L contains an equivalence relation θ of finite index, for which each finite equivalence class is a singleton. Let $P_1, \ldots, P_r$ be the equivalence classes of θ. For each $i \in \{1, \ldots, r\}$ with P_i infinite, we define n_i as in Lemma 4.5 and we let

$$n = \max\{n_i \mid P_i \text{ is infinite}\}.$$

Now let $u, v \in A^*$, with $n \leqslant |u| = |v|$ and $c_P(u) = c_P(v)$ for each θ equivalence class P. We define words $w_i \in A^*$ for $i = 0, \ldots, r$ by

$$(w_i)_j = \begin{cases} u_j & \text{if } j \in P_k \text{ and } i < k \\ v_j & \text{otherwise.} \end{cases}$$

By construction we have $w_0 = u$, $w_r = v$ and Lemma 4.5 applies to each pair w_{i-1}, w_i with $i \in \{1, \ldots, r\}$ and thus

$$w_{i-1} \in L \iff w_i \in L,$$

and it follows that

$$u \in L \iff v \in L.$$

□

Theorem 4.7. *If $L \in \mathcal{P}(A^*)$ satisfies the equations $\mathcal{E}_{ab=ba}$ and $\mathcal{E}_{aab=abb}$ for all $a, b \in A$, then $L \in \mathcal{B}$.*

Proof. First notice that for $P \subseteq \mathbb{N}$ and $B \subseteq A$, the set

$$L_{P,B} = \{u \in A^* \mid c_P(u) = B\}$$

belongs to $\mathcal{B}$ since

$$L_{P,B} = \Big(\bigcap_{a \in A \setminus B} L_{a,P^c}\Big) \cap \Big(\bigcap_{a \in B} L^c_{a,P^c}\Big).$$

By Corollary 4.3, the relation R_L contains an equivalence relation θ of finite index for which each finite equivalence class is a singleton. Let $P_1, \ldots, P_r$ be the corresponding partition of $\mathbb{N}$. By Lemma 4.6, there is an $n \in \mathbb{N}$ such that for each $m \in \mathbb{N}$ with $m \geqslant n$, there exists a subset S_m of $\mathcal{P}(A)^r$ such that

$$A^m \cap L = A^m \cap \Big(\bigcup_{(B_1,\ldots,B_r) \in S_m} \bigcap_{i=1}^{r} L_{P_i,B_i}\Big).$$

Now let $f : [n, +\infty[\to \mathcal{P}((\mathcal{P}(A))^r)$ be defined by $f(m) = S_m$, and, for each $S \in \mathcal{P}((\mathcal{P}(A))^r)$, define the shorthand

$$L(S) = \bigcup_{(B_1,\ldots,B_r) \in S} \bigcap_{i=1}^{r} L_{P_i,B_i}.$$

Then by Lemma 4.6 the following equality holds

$$L = \Big(L \cap (1 \cup A)^n\Big) \cup \Big(\bigcup_{S \in \mathcal{P}((\mathcal{P}(A))^r)} L(S) \cap L_{f^{-1}(S)}\Big),$$

and since $\mathcal{B}$ contains all finite languages, this formula shows that $L \in \mathcal{B}$. □

5 The regular case

Proposition 1.5 shows that in order to obtain a set of profinite equations defining the Boolean algebra $\mathcal{B} \cap \mathrm{Reg}$, it suffices to project, for all $a, b \in A$, the families $\mathcal{E}_{ab=ba}$ and $\mathcal{E}_{aab=abb}$ introduced above onto the free profinite monoid. The resulting set of profinite equations will then be used to prove that membership in $\mathcal{B} \cap \mathrm{Reg}$ is decidable.

However, these equations obtained by projection are not in a form that is familiar to researchers working on regular languages. As a last step, we show by purely classical rewriting methods that our first set of equations is equivalent to a set of equations in a more familiar form.

5.1 The profinite projections of the ultrafilter equations for $\mathcal{B}$

The length homomorphism $\ell\colon A^* \to \mathbb{N}$ given by $\ell(a) = 1$ for each $a \in A$ and its extension $\widehat{\ell}\colon \widehat{A^*} \to \widehat{\mathbb{N}}$ will play an essential role in this subsection. It is important to note that $\widehat{\ell}$ is a homomorphism of profinite monoids. We denote by ω the unique idempotent of $\widehat{\mathbb{N}} - \mathbb{N}$. It is the limit of the sequence $n!$. It then follows that $n! - 1$ is also a convergent sequence in $\widehat{\mathbb{N}}$ and its limit, which we denote by $\omega - 1$, is the unique solution of the equation $x + 1 = \omega$.

We begin with the following partial description of the equations obtained by projection. For this purpose, we will need the following notation: Given a word $u = a_0 \cdots a_{n-1} \in A^*$ where $a_i \in A$, and k and ℓ with $0 \leqslant k \leqslant \ell < n$, we let $u[k, \ell] = a_k \cdots a_\ell$.

Proposition 5.1. *Let $a, b \in A$. Every non-trivial equation in the set $\pi_{\mathrm{Reg}}(\mathcal{E}_{ab=ba})$ is of the form*

$$xaybz \leftrightarrow xbyaz, \tag{16}$$

where $x, y, z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \omega - 1$.

Similarly, every non-trivial equation in the set $\pi_{\mathrm{Reg}}(\mathcal{E}_{aab=abb})$ is of the form

$$xayay'bz \leftrightarrow xayby'bz, \tag{17}$$

where $x, y, y', z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \widehat{\ell}(y') = \omega - 1$.

Proof. Let a, b be two fixed letters. We give a detailed proof for $\mathcal{E}_{ab=ba}$, the proof for $\mathcal{E}_{aab=abb}$ being similar. Let $\gamma \in \beta(A^* \times \mathbb{N}^2)$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$. We first note that we may assume that the set

$$D = \{(u, i, j) \mid u \in A^* \text{ and } i < j < |u|\}$$

belongs to γ. Otherwise, D^c belongs to γ and thus $f_{ab}^{-1}(L) \in \gamma$ if and only if $f_{ab}^{-1}(L) \cap D^c \in \gamma$. Similarly, $f_{ba}^{-1}(L) \in \gamma$ if and only if $f_{ba}^{-1}(L) \cap D^c \in \gamma$. Now, observing that $f_{ab} = f_{ba} = \pi_0$ on D^c, we get

$$D^c \cap f_{ab}^{-1}(L) = D^c \cap \pi_0^{-1}(L) = D^c \cap f_{ba}^{-1}(L)$$

Thus $\beta f_{ab}(\gamma)$ and $\beta f_{ba}(\gamma)$ are one and the same ultrafilter, namely $\beta\pi_0(\gamma)$. It thus follows that in this case the equation $\beta f_{ab}(\gamma) \leftrightarrow \beta f_{ba}(\gamma)$ is trivially satisfied by all languages in A^*. Thus we may restrict our attention to the equations $\beta f_{ab}(\gamma) \leftrightarrow \beta f_{ba}(\gamma)$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$ and $D \in \gamma$.

As explained in Section 1.3, we will identify $\widehat{D}$ with βD. In order to prove the proposition, we will show that given $\gamma \in \beta D$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$, there exist $x, y, z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \omega - 1$ such that

$$\pi_{\mathrm{Reg}}(\beta f_{ab}(\gamma)) = xaybz \text{ and } \pi_{\mathrm{Reg}}(\beta f_{ba}(\gamma)) = xbyaz.$$

Let $q\colon D \to (A^*)^3$ and $g_{ab} : (A^*)^3 \to A^*$ be the maps given by

$$\begin{aligned} q(w, i, j) &= (w[0, i-1], w[i+1, j-1], w[j+1, |w|-1]) \\ g_{ab}(x, y, z) &= xaybz. \end{aligned}$$

Since $\widehat{A^*}^3$ is compact, q has a unique continuous extension $\beta q : \beta D \to \widehat{A^*}^3$. Similarly, g_{ab} has a unique continuous extension $\widehat{g}_{ab} : \widehat{A^*}^3 \to \widehat{A^*}$. Consider the following diagram, in which all the functions are continuous:

$$\begin{tikzcd} \beta D \arrow[r, "\beta f_{ab}"] \arrow[d, "\beta q"'] & \beta A^* \arrow[d, "\pi_{\mathrm{Reg}}"] \\ \widehat{A^*}^3 \arrow[r, "\widehat{g}_{ab}"] & \widehat{A^*} \end{tikzcd}$$

Since, for all $(w, i, j) \in D$,

$$(g_{ab} \circ q)(w, i, j)) = w[0, i-1]aw[i+1, j-1]bw[j+1, |w|-1] = f_{ab}(w),$$

and since D is dense in βD, the diagram commutes.

Let now $\gamma \in \beta D$ be an ultrafilter such that $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$. Setting $(x, y, z) = \beta q(\gamma)$, we get $\pi_{\mathrm{Reg}}(\beta f_{ab}(\gamma)) = xaybz$ and the same argument applied to βf_{ba} and $\widehat{g}_{ba}$ yields the equality $\pi_{\mathrm{Reg}}(\beta f_{ba}(\gamma)) = xbyaz$.

In order to show that $x \notin A^*$ and that $\widehat{\ell}(y) = \omega - 1$, consider the following diagrams, where $p_1(x, y, z) = \widehat{\ell}(x)$ and $p_2(x, y, z) = \widehat{\ell}(x) + \widehat{\ell}(y) + 1$.

$$\begin{tikzcd} \beta D \arrow[r, "\beta \pi_1"] \arrow[d, "\beta q"'] & \beta \mathbb{N} \arrow[d, "\pi_{\mathrm{Reg}}"] \\ \widehat{A^*}^3 \arrow[r, "p_1"] & \widehat{\mathbb{N}} \end{tikzcd} \qquad \begin{tikzcd} \beta D \arrow[r, "\beta \pi_2"] \arrow[d, "\beta q"'] & \beta \mathbb{N} \arrow[d, "\pi_{\mathrm{Reg}}"] \\ \widehat{A^*}^3 \arrow[r, "p_2"] & \widehat{\mathbb{N}} \end{tikzcd}$$

Since each diagram commutes on D, they both commute. Thus, for $\gamma \in \beta D$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$, we have $p_1 \circ \beta q(\gamma) = p_2 \circ \beta q(\gamma)$. That is, the projection of the equation $\beta f_{ab}(\gamma) \leftrightarrow \beta f_{ba}(\gamma)$ is of the form

$$xaybz \leftrightarrow xbyaz,$$

where $(x, y, z) \in \widehat{A^*}^3$ satisfies $\widehat{\ell}(x) = \widehat{\ell}(x) + \widehat{\ell}(y) + 1$ or equivalently $\widehat{\ell}(x) \notin \mathbb{N}$ and $\widehat{\ell}(y) + 1 = \omega$ and thus $x \notin A^*$ and $\widehat{\ell}(y) = \omega - 1$. $\square$

We are now ready to identify the projections of our ultrafilter equations precisely.

Theorem 5.2. *The Boolean algebra $\mathcal{B} \cap \mathrm{Reg}$ is defined by the set of profinite equations of the form*

$$xaybz \leftrightarrow xbyaz \quad \text{and} \quad xayay'bz \leftrightarrow xayby'bz, \tag{18}$$

where $a, b \in A$, $x, y, y', z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \widehat{\ell}(y') = \omega - 1$.

Proof. Again, we just treat the case of the equations in $\mathcal{E}_{ab=ba}$, the one of $\mathcal{E}_{aab=abb}$ being similar. All that remains to show is that for each choice of $x, y, z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \omega - 1$, there exists $\gamma \in \beta D$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$ such that

$$\pi_{\mathrm{Reg}}(\beta f_{ab}(\gamma)) = xaybz \text{ and } \pi_{\mathrm{Reg}}(\beta f_{ba}(\gamma)) = xbyaz.$$

To this end, let $x, y, z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \omega - 1$. We think of x, y, and z as ultrafilters of $\mathrm{Reg}(A^*)$.

For $K \in x$, $L \in y$ and $M \in z$, let

$$\Gamma(K, L, M) = \{(uavaw, \ell(u), \ell(u) + \ell(v) + 1) \mid u \in K, v \in L, w \in M\},$$

and

$$\mathcal{F}(x, y, z) = \{\Gamma(K, L, M) \mid K \in x, L \in y, M \in z\}.$$

Note that, being elements of an ultrafilter, the sets K, L and M are nonempty and thus $\Gamma(K, L, M)$ is also nonempty. Furthermore, for $K_1, K_2 \in x$, $L_1, L_2 \in y$, and $M_1, M_2 \in z$, we have

$$\Gamma(K_1 \cap K_2, L_1 \cap L_2, M_1 \cap M_2) \subseteq \Gamma(K_1, L_1, M_1) \cap \Gamma(K_2, L_2, M_2)$$

so that $\mathcal{F}(x, y, z)$ is a filter basis.

We claim that any ultrafilter γ extending $\mathcal{F}(x, y, z)$ satisfies $\beta q(\gamma) = (x, y, z)$. First of all, since each $\Gamma(K, L, M)$ is contained in D, γ belongs to βD. We show that the first coordinate of $\beta q(\gamma)$ is x, the other arguments being similar. To this end, let $q_1 = \pi_0 \circ q$. Thus $q_1 \colon D \to A^*$ is the map defined by

$$q_1((u, i, j)) = u[0, i - 1].$$

Then $\beta q_1 = \beta\pi_0 \circ \beta q$ and thus we just need to show that $\beta q_1(\gamma) = x$. If $K \in x$, then

$$q_1^{-1}(K) \supseteq \Gamma(K, A^*, A^*) \in \mathcal{F}(x, y, z) \subseteq \gamma,$$

and thus $q_1^{-1}(K) \in \gamma$ or equivalently $K \in \beta q_1(\gamma)$. Thus $x \subseteq \beta q_1(\gamma)$ and as x and $\beta q_1(\gamma)$ are ultrafilters it follows that $x = \beta q_1(\gamma)$, which proves the claim.

Now suppose that $x \notin A^*$ and $\widehat{\ell}(y) = \omega - 1$. We show that there is an ultrafilter γ extending $\mathcal{F}(x, y, z)$ with $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$. By Proposition 3.1, $\beta\pi_1(\gamma) = \beta\pi_2(\gamma)$ if and only if γ extends the filter basis $\mathcal{F}$. It thus suffices to show that $\mathcal{F}(x, y, z) \cap \mathcal{F}$ is a filter subbasis. Let $K \in x, L \in y$, and $M \in z$, and let $\{P_1, \ldots, P_n\}$ be a partition of $\mathbb{N}$. We need to show that

$$\Gamma(K, L, M) \cap (A^* \times P_i^2) \neq \emptyset$$

for some $i \in \{1, \ldots, n\}$. Since x is nonprincipal, the regular language K is infinite, and thus $\ell(K)$ contains an infinite arithmetic progression, say $r + p\mathbb{N}$ with $p > 0$. Furthermore, for $L \in y$ we have $aL \in ay$, and since $\widehat{\ell}(ay) = \omega$, there is $q \geqslant 1$ and $N \in \mathbb{N}$ such that

$$[N, +\infty[\cap q\mathbb{N} \subseteq \ell(aL).$$

Now let m be a common multiple of p and q. Then there is $i \in \{1, \dots, n\}$ such that the set

$$P_i \cap r + m\mathbb{N}$$

is infinite. Now let $n_1, n_2 \in P_i \cap r + m\mathbb{N}$ with $n_2 - n_1 > N$. Then $n_1 \in r + m\mathbb{N} \subseteq r + p\mathbb{N}$ implies that there is $u \in K$ with $\ell(u) = n_1$. Also, $n_2 - n_1 \in m\mathbb{N} \cap [N, +\infty[$ so there is $v \in L$ with $\ell(av) = n_2 - n_1$. Taking now any word $w \in M$, we get

$$(uavaw, n_1, n_2) \in \Gamma(K, L, M) \cap (A^* \times P_i^2),$$

which shows that $\Gamma(K, L, M) \cap (A^* \times P_i^2)$ is nonempty as required. $\square$

One can slightly simplify the equations given in Theorem 5.2.

Corollary 5.3. *The Boolean algebra $\mathcal{B} \cap \mathrm{Reg}$ is defined by the set of profinite equations of the form*

$$xayb = xbya \tag{19}$$

and

$$xayay'b = xayby'b, \tag{20}$$

where $a, b \in A$, $x, y, y' \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \widehat{\ell}(y') = \omega - 1$.

Proof. First of all, since $\mathcal{B} \cap \mathrm{Reg}$ is closed under quotients, one can freely replace the equations (18) by

$$xaybz = xbyaz \quad \text{and} \quad xayay'bz = xayby'bz, \tag{21}$$

where $a, b \in A$, $x, y, y', z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \widehat{\ell}(y') = \omega - 1$.

The equations (19) and (20) correspond to the equations (21) with $z = 1$. Furthermore, Proposition 1.3 shows that if $\mathcal{B} \cap \mathrm{Reg}$ satisfies an equation of the form $xayb = xbya$, then it also satisfies the equations $xaybz = xbyaz$ for all $z \in \widehat{A^*}$. A similar argument works for an equation of the form $xayay'b = xayby'b$, which proves that the set of equations (18) on the one hand and (19) and (20) on the other hand define the same Boolean algebra closed under quotients. $\square$

5.2 Membership in $\mathcal{B} \cap \mathrm{Reg}$

The aim of this section is to prove that membership in $\mathcal{B} \cap \mathrm{Reg}$ is decidable. By Theorem 5.2 it suffices to effectively decide whether a given regular language satisfies the equations (19) and (20). These equations involve two types of profinite words that require a separate study: the nonfinite profinite words and the profinite words of length $\omega - 1$.

Let L be a regular language of A^*. Let $\eta : A^* \to M$ be its syntactic morphism and let $\widehat{\eta} : \widehat{A^*} \to M$ be the continuous extension of η to $\widehat{A^*}$.

Let us first compute the image by $\widehat{\eta}$ of a nonfinite profinite word. Let E be the set of idempotents of the semigroup $\eta(A^+)$. The following lemma is a direct consequence of [2, Corollary 5.6.2 (c)]:

Lemma 5.4. *The following formula holds:* $\widehat{\eta}(\widehat{A^*} - A^*) = MEM$.

Next we compute the image by $\widehat{\eta}$ of the set of profinite words of length $\omega - 1$. This requires to work with the monoid $\mathcal{P}(M)$, equipped with the subset multiplication defined as follows. For every $X, Y \in \mathcal{P}(M)$,

$$XY = \{xy \mid x \in X, y \in Y\}.$$

Let $R = \eta(A)$. Then R generates a cyclic submonoid of $\mathcal{P}(M)$, whose minimal ideal is a group G. The map $n \to R^n$ defines a monoid morphism from the additive monoid $\mathbb{N}$ to $\mathcal{P}(M)$. This morphism has a unique continuous extension to $\widehat{\mathbb{N}}$ and since ω is an idempotent of $\widehat{\mathbb{N}}$, R^ω is an idempotent of $\mathcal{P}(M)$. Consequently, R^ω is the identity of G and $R^{\omega-1}$ is the inverse of $R^{\omega+1}$ in G. The following lemma shows how $R^{\omega-1}$ is related to the profinite words of length $\omega - 1$.

Lemma 5.5. *An element m of M belongs to $R^{\omega-1}$ if and only if there exists a profinite word $y \in \widehat{A^*}$ such that $\widehat{\eta}(y) = m$ and $\widehat{\ell}(y) = \omega - 1$.*

Proof. If $y \in \widehat{A^*}$ is a profinite word such that $\widehat{\ell}(y) = \omega - 1$, then $\widehat{\eta}(y) \in R^{\omega-1}$. Let n be an integer such that $R^\omega = R^n$. Then for all $k > n$, $R^{k!} = R^\omega$ and $R^{k!-1} = R^{\omega-1}$. Therefore, if $m \in R^{\omega-1}$, there exists a word y_k such that $\eta(y_k) = m$ and $|y_k| = k! - 1$. Since $\widehat{A^*}$ is compact, there is a subsequence of the sequence $(y_k)_{k>n}$ converging to a profinite word y. By construction, one has $\widehat{\eta}(y) = m$ and $\widehat{\ell}(y) = \omega - 1$, which proves the lemma. $\square$

We are now ready to prove the decidability of the membership in $\mathcal{B} \cap \mathrm{Reg}$. More precisely, we get the following result.

Proposition 5.6. *A regular language L satisfies the equations* (19) *and* (20) *if and only if the equalities*

$$xayb = xbya \quad \textit{and} \quad xayay'b = xayby'b$$

hold for all $x \in MEM$, $a, b \in A$ and $y, y' \in R^{\omega-1}$.

Proof. This is an immediate consequence of the structure of the equations (19) and (20), of the definition of R, of Theorem 5.2 and of Lemmas 5.4 and 5.5. $\square$

Corollary 5.7. *Membership in* $\mathcal{B} \cap \mathrm{Reg}$ *is decidable.*

5.3 An alternative set of equations for $\mathcal{B} \cap$ Reg

Though our work in the previous subsection provides a set of profinite equations for $\mathcal{B} \cap \mathrm{Reg}$ and establishes the decidability of membership in this Boolean algebra, we proceed to give an alternative set of profinite equations, which is closer in spirit to the profinite equations usually given in the theory of regular languages. We begin by identifying certain families of projections of the equations introduced in Section 3.

Theorem 5.8. *The Boolean algebra $\mathcal{B} \cap \mathrm{Reg}$ satisfies the profinite equations of the form*

$$(x^{\omega-1}s)(x^{\omega-1}t) = (x^{\omega-1}t)(x^{\omega-1}s), \tag{22}$$

where $x, s, t \in A^*$ and $|s| = |t| = |x|$.

Proof. Let $B = \{a, b, c\}$ be a three letter alphabet. It follows from Corollary 5.3 that $(\mathcal{B} \cap \mathrm{Reg})(B^*)$ satisfies the profinite equations (19)

$$xayb = xbya$$

where $x, y \in \widehat{B^*} - B^*$ and $\widehat{\ell}(y) = \omega - 1$. Taking $x = y = c^{\omega-1}$ in (19), we get the profinite equation

$$(c^{\omega-1}a)(c^{\omega-1}b) = (c^{\omega-1}b)(c^{\omega-1}a). \tag{23}$$

Let now x, s and t be words of A^* of the same length and let $f : B^* \to A^*$ be the length-multiplying morphism defined by $f(a) = s$, $f(b) = t$ and $f(c) = x$. Corollary 2.8 shows that if $L \in \mathcal{B} \cap \mathrm{Reg}(A^*)$, then $f^{-1}(L) \in \mathcal{B} \cap \mathrm{Reg}(B^*)$. In particular, $f^{-1}(L)$ satisfies (23). It follows now from Proposition 1.7 that L satisfies the profinite equation $\hat{f}((c^{\omega-1}a)(c^{\omega-1}b)) = \hat{f}((c^{\omega-1}b)(c^{\omega-1}a))$ which is exactly Equation (22). $\square$

A similar argument using (20) instead of (19) as a starting point yields the following theorem.

Theorem 5.9. *The Boolean algebra $\mathcal{B} \cap \mathrm{Reg}$ satisfies the profinite equations of the form*

$$(x^{\omega-1}s)(x^{\omega-1}s)(x^{\omega-1}t) = (x^{\omega-1}s)(x^{\omega-1}t)(x^{\omega-1}t), \tag{24}$$

where $x, s, t \in A^*$ and $|s| = |t| = |x|$.

In the setting of Boolean algebras of regular languages closed under quotients, the equations of Theorem 5.9 are equivalent to a simpler family.

Proposition 5.10. *A Boolean algebra of regular languages closed under quotients satisfies the set of profinite equations (24) if and only if it satisfies the set of profinite equations*

$$(x^{\omega-1}s)(x^{\omega-1}s) = x^{\omega-1}s, \tag{25}$$

where $x, s \in A^*$ and $|s| = |x|$.

Proof. Let $\mathcal{L}$ be a Boolean algebra of regular languages closed under quotients. Suppose that the equations (24) hold for $\mathcal{L}$ and let $x, s \in A^*$ with $|s| = |x|$. Then (24) with x substituted for s and s substituted for t yields

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

which gives (25) since $x^{\omega}x^{\omega-1} = x^{\omega-1}$.

Conversely, if (25) holds for $\mathcal{L}$, and $x, s, t \in A^*$ are three words of the same length, then the equations $(x^{\omega-1}s)(x^{\omega-1}s) = x^{\omega-1}s$ and $x^{\omega-1}t = (x^{\omega-1}t)(x^{\omega-1}t)$ hold for $\mathcal{L}$. Since $\mathcal{L}$ is closed under quotients, Proposition 1.3 shows that

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

holds for $\mathcal{L}$. $\square$

We will now show that any regular language satisfying the equations (22) and (25) also satisfies the profinite equations (19) and (20).

The *circular shift operator* $\sigma : A^* \to A^*$ maps a word $x = a_0 \dots a_{n-1}$ to $\sigma(x) = a_1 \dots a_{n-1}a_0$. As in Section 5.2, $\eta : A^* \to M$ denotes the syntactic morphism of a regular language L. For the remainder of the paper, we define d as the smallest multiple of $|M|!$ such that for all $R \in \mathcal{P}(M)$, R^d is idempotent. In particular, s^d is idempotent for all $s \in M$. For any $r \in \mathbb{N}$, we denote by $[r]$ the remainder after division of r by d. Furthermore, we use the notation $u =_\eta v$ for $\eta(u) = \eta(v)$.

Lemma 5.11. *Suppose that L satisfies the equations* (22)*. Let $p, x \in A^*$ with $|x| = d$ and $px^\omega =_\eta p$. If $q \in A^*$ is of length n, then $pq =_\eta pq\,(\sigma^n(x))^\omega$.*

Proof. The result may be proved by induction on the length of q. We give the proof in the case $n = 1$ in order to simplify notation. The inductive step is then an easy consequence. Let $a \in A$. Setting $x = b_0 \dots b_{d-1}$ with $b_i \in A$, we get

$$\begin{aligned} pa(\sigma(x))^\omega &=_\eta p\,x^\omega a(\sigma(x))^\omega =_\eta p\,x^d a(\sigma(x))^d \\ &= p\,b_0(\sigma(x))^{d-1}(b_1 \dots b_{d-1}a)(\sigma(x))^{d-1}\sigma(x) \\ &=_\eta p\,b_0(\sigma(x))^{\omega-1}(b_1 \dots b_{d-1}a)(\sigma(x))^{\omega-1}\sigma(x). \end{aligned}$$

It follows from (22) that for $s = b_1 \dots b_{d-1}a$ we have $|s| = |x| = |\sigma(x)|$ and thus

$$(\sigma(x))^{\omega-1}s(\sigma(x))^{\omega-1}\sigma(x) =_\eta (\sigma(x))^{\omega-1}\sigma(x)(\sigma(x))^{\omega-1}s.$$

Using that $=_\eta$ is a congruence, the properties of ω and some rewriting we obtain

$$\begin{aligned} pa(\sigma(x))^\omega &=_\eta p\,b_0(\sigma(x))^{\omega-1}\sigma(x)(\sigma(x))^{\omega-1}(b_1 \dots b_{d-1}a) \\ &=_\eta p\,b_0(\sigma(x))^{d-1}\sigma(x)(\sigma(x))^{d-1}(b_1 \dots b_{d-1}a) \\ &= px^{2d}a =_\eta pa. \end{aligned}$$

Now for the proof by induction, if the length of q is 0, then the result simply follows from the relation $px^\omega =_\eta p$. Suppose by induction that the result holds for a word of length less than or equal to n. A word of length $n+1$ is of the form qa where q is of length n. Thus, by the induction hypothesis, we have

$$pq =_\eta pq\,(\sigma^n(x))^\omega.$$

By the case $n = 1$ with pq in the place of p and $\sigma^n(x)$ in place of x, we obtain

$$pqa(\sigma(\sigma^n(x)))^\omega =_\eta pqa.$$

Since $\sigma(\sigma^n(x)) = \sigma^{n+1}(x)$, the desired result follows. $\square$

In the next corollary, the notation $x(a_i@[i])$ stands for the word obtained by replacing in x the letter in position i by a_i.

Corollary 5.12. *Suppose that L satisfies the equations* (22)*. Let $p, x \in A^*$ with $|x| = d$ and $px =_\eta p$. If $q = a_0 \dots a_{n-1}$ with $a_i \in A$, then*

$$pq =_\eta p\big(x^{\omega-1}\, x(a_0@[0])\big) \cdots \big(x^{\omega-1}\, x(a_{n-1}@[n-1])\big)x^{\omega-1}\, x^n[0,[n-1]].$$

Proof. By the assumption on x we have $p =_\eta px^{\omega-1} ==_\eta px^\omega$. Now applying Lemma 5.11 after each letter of q, we obtain

$$pq =_\eta p\, x^{\omega-1} a_0(\sigma(x))^\omega a_1(\sigma^2(x))^\omega \cdots (\sigma^{n-1}(x))^\omega a_{n-1}(\sigma^n(x))^\omega.$$

Setting $x = b_0 \cdots b_{d-1}$, we have

$$a_0(\sigma(x))^\omega =_\eta a_0(\sigma(x))^d = a_0 b_1 \cdots b_{d-1}\, x^{d-1} b_0 =_\eta x(a_0@[0])x^{\omega-1}b_0,$$

and similarly

$$b_0 a_1(\sigma^2(x))^\omega =_\eta x(a_1@[1])x^{\omega-1}b_0 b_1,$$

and so on up through

$$b_0 \cdots b_{[n-2]} a_{n-1}(\sigma^n(x))^\omega =_\eta x(a_{n-1}@[n-1])x^{\omega-1}\, x[0,[n-1]],$$

and the conclusion now follows. □

We will need a small combinatorial lemma:

Lemma 5.13. *Let u be a word of length at least $|M|$. Then there exist a prefix p of u of length less than $|M|$ and a word v of length d such that $pv =_\eta p$.*

Proof. For each $k \geqslant 0$, let $s_k = \eta(u[0, k-1])$. If $s_0, \dots, s_{|M|-1}$ are all distinct, one of them, say s_i, is idempotent. Then since p divides d, $p = u[0, i]$ and $v = p^{d/|p|}$ give the result. On the other hand, if $s_i = s_j$ with $i < j < |M|$, let $p = u[0, i]$, $z = u[i + 1, j]$ and $v = z^{d/|z|}$. Then $pz =_\eta p$ and thus $pv =_\eta p$. □

Proposition 5.14. *If L satisfies the equations* (22)*, then it satisfies the equations* (19)*.*

Proof. Let $x, y, z \in \widehat{A^*}$ with $x \notin A^*$ and $\widehat{\ell}(y) = \omega - 1$. Since $x \notin A^*$, there is $u \in A^*$ with $|u| > |M|$ and $x =_\eta u$. Now by Lemma 5.13 there exist $p, q, v \in A^*$ such that

$$u = pq, \;\; pv =_\eta p \;\text{ and } |v| = d.$$

Since $\widehat{\ell}(y) = \omega - 1$, it follows by Lemma 5.5 that $\widehat{\eta}(y) \in R^{d-1}$ and hence there exists $r \in A^{d-1}$ such that $y =_\eta r$. Therefore

$$xayb =_\eta pqarb,$$

and applying Corollary 5.12 to the word $qarb = a_0 \cdots a_{n-1}$, we get

$$pqarb =_\eta p(v^{d-1}\, v(a_0@[0])) \cdots (v^{d-1}\, v(a_{n-1}@[n-1]))v^{d-1}\, v[0,[n-1]]. \quad (26)$$

Note that

$$a_{|q|} = a \quad \text{and} \quad a_{|q|+d} = b \quad \text{and} \quad [|q|] = [|q|+d].$$

Since the words v and the $v(a_i@[i])$ all have the same length, one can apply (22) to permute the $v(a_i@[i]))$ as one wishes. In particular, applying the transposition $(|q|\ |q|+d)$ will permute the letters a and b. Since $[|q|] = [|q|+d]$, one can apply Corollary 5.12 again and remove all the inserted copies of shifts of v to obtain $pqbra$. Therefore

$$xayb =_\eta xbya$$

as required. □

A similar argument would lead to the following proposition.

Proposition 5.15. *If L satisfies the equations* (24)*, then it satisfies the equations* (20)*.*

We can now state our final result.

Theorem 5.16. *The Boolean algebra $\mathcal{B} \cap \mathrm{Reg}$ is defined by the profinite equations*

$$(x^{\omega-1}s)(x^{\omega-1}t) = (x^{\omega-1}t)(x^{\omega-1}s) \quad \textit{and} \quad (x^{\omega-1}s)(x^{\omega-1}s) = x^{\omega-1}s, \quad (27)$$

where $x, s, t \in A^*$ and $|s| = |t| = |x|$.

Proof. It suffices to apply Theorem 5.2 and Propositions 5.10, 5.14 and 5.15. □

Acknowledgements

We would like to thank Jorge Almeida, Pascal Weil and the anonymous referees for their valuable comments.

References

1. J. Almeida, Residually finite congruences and quasi-regular subsets in uniform algebras, *Portugaliæ Mathematica* **46** (1989), 313–328.
2. J. Almeida, *Finite semigroups and universal algebra*, World Scientific Publishing Co. Inc., River Edge, NJ, 1994. Translated from the 1992 Portuguese original and revised by the author.
3. D. A. M. Barrington, H. Straubing and D. Thérien, Non-uniform automata over groups, *Information and Computation* **89** (1990), 109–132.

4. 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.
5. M. Gehrke, S. Grigorieff and J.-E. Pin, A topological approach to recognition,, in *ICALP 2010, Part II*, S. e. a. Abramsky (ed.), Berlin, 2010, pp. 151–162, *Lect. Notes Comp. Sci.* vol. 6199, Springer.
6. M. Gehrke, A. Krebs and J.-É. Pin, From ultrafilters on words to the expressive power of a fragment of logic, in *Descriptional Complexity of Formal Systems*, A. O. H. Jürgensen, J. Karhumäki (ed.), Berlin, 2014, pp. 138–149, *Lect. Notes Comp. Sci.* vol. 8614, Springer.
7. P. McKenzie, M. Thomas and H. Vollmer, Extensional uniformity for Boolean circuits, *SIAM J. Comput.* **39**,7 (2010), 3186–3206.
8. J.-É. Pin, Profinite methods in automata theory, in *26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009)*, S. Albers and J.-Y. Marion (eds.), pp. 31–50, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2009.
9. J.-É. Pin, Equational descriptions of languages, *Int. J. Found. Comput. S.* **23** (2012), 1227–1240.
10. H. Straubing, Constant-depth periodic circuits, *Internat. J. Algebra Comput.* **1**,1 (1991), 49–87.
11. H. Straubing, *Finite automata, formal logic, and circuit complexity*, Birkhäuser Boston Inc., Boston, MA, 1994.
12. H. Straubing, On logical descriptions of regular languages, in *LATIN 2002*, Berlin, 2002, pp. 528–538, *Lect. Notes Comp. Sci.* n° 2286, Springer.