Some structural properties of the free profinite aperiodic semigroup

J. Almeida[1], J. C. Costa[2], M. Zeitoun[3]

[1] Univ. Porto, [2] Univ. Minho, [3] Univ. Bordeaux

Abstract. Profinite semigroups provide powerful tools to understand properties of classes of regular languages. Until very recently however, little was known on the structure of "large" relatively free profinite semigroups. In this paper, we present new results obtained for the class of all finite aperiodic (that is, group-free) semigroups. Given a finite alphabet X, we focus on the following problems: (1) the word problem for ω-terms on X evaluated on the free pro-aperiodic semigroup, and (2) the computation of closures of regular languages in the ω-subsemigroup of the free pro-aperiodic semigroup generated by X.

1 Introduction

Context. Profinite semigroups provide powerful tools to understand properties of classes of regular languages. However, until very recently, little was known on the structure of "large" relatively free profinite semigroups. In this paper, we present some results recently obtained for the class of all finite aperiodic (that is, group-free) semigroups. This class has been investigated for a long time: the first deep instance of an Eilenberg correspondence [13] goes back to Schützenberger [22], who related this class of semigroups with the class of star-free languages.

For a finite alphabet X and a pseudovariety V, the free pro-V semigroup on X is denoted $\overline{\Omega}_X\mathsf{V}$. We will recall its construction in Section 2, but for now, let us just state some of its most important properties: it is naturally equipped with a metric making it a topological semigroup (meaning that the multiplication is continuous). This topological semigroup is compact and totally disconnected. Moreover, it enjoys the following universal property: every mapping $\varphi : X \to S$ into a semigroup S of V can be extended in a unique way to a continuous homomorphism $\hat{\varphi} : \overline{\Omega}_X\mathsf{V} \to S$, when we endow the finite semigroup S with the discrete topology. If S denotes the pseudovariety of all finite semigroups, there is a unique continuous homomorphism $p_{\mathsf{V}} : \overline{\Omega}_X\mathsf{S} \to \overline{\Omega}_X\mathsf{V}$ extending the identity mapping on X.

Motivations. Consider a signature σ, whose elements have a natural interpretation on semigroups of V. The σ-word problem over V consists in determining whether two σ-terms are equal when evaluated on semigroups of V. Due to the universal property of $\overline{\Omega}_X\mathsf{V}$ mentioned above, the interpretation of elements of σ

may be lifted to $\overline{\Omega}_X\mathsf{V}$, and the σ-word problem is equivalent to testing that two σ-terms have the same interpretation in $\overline{\Omega}_X\mathsf{V}$.

An example of such a signature is denoted κ. It consists of the semigroup multiplication, and of the $(\omega-1)$-power, interpreted in $\overline{\Omega}_X\mathsf{V}$ as follows: $x^{\omega-1}$ is the limit of the sequence $(x^{n!-1})$. If V is a pseudovariety of aperiodic semigroups, the word problem for κ-terms is equivalent to the word problem over terms built from X using the multiplication and the ω-power, where we interpret x^ω as the limit of the sequence $(x^{n!})$, or equivalently as the unique idempotent of the closed subsemigroup generated by x. On finite semigroups, x^ω is the idempotent of the subsemigroup generated by X. We call such terms *ω-terms*.

When V is the pseudovariety G of finite groups, the κ-word problem can be solved easily: the algebra on the signature κ can be seen as the free group on X, and the term $x^{\omega-1}$ as the inverse x^{-1} of x. Testing that two terms are equal when interpreted in every finite group amounts to testing that they are equal in the free group, which can be done using rewriting rules $(xy)^{-1} \to y^{-1}x^{-1}$, $(x^{-1})^{-1} \to x$, $xx^{-1} \to 1$ and $x^{-1}x \to 1$. It is well known that this forms a confluent rewriting system, yielding a normal form which characterizes the value of a term in the free group (see e.g. [12]).

For the pseudovariety A of aperiodic semigroups, the κ-word problem is also decidable, but its solution is much more involved. It has been proposed by McCammond [18], and is again based on a rewriting system. However, the rewriting computation has to be guided in order to end at a normal form characterizing the value of the term on all finite aperiodic semigroups. In this paper, we sketch an alternative approach to prove that this normal form indeed characterizes the value of a term on all finite aperiodic semigroups.

Several other problems considered recently at the interface of the theories of finite semigroups and formal languages lead to natural formulations in terms of these profinite semigroups. One of them, the computation of pointlike sets, boils down to the computation of closures of regular languages in the subalgebra of $\overline{\Omega}_X\mathsf{V}$ on the signature κ, and to nonemptiness tests. For the pseudovariety G of all finite groups, Rhodes' type II conjecture asked for the computation of the *kernel* of an X-generated semigroup S, given by a surjective homomorphism $\varphi : X^+ \to S$. The kernel consists of the elements of S which have to be related to 1 through the composite relational morphism $S \xrightarrow{\varphi^{-1}} \overline{\Omega}_X\mathsf{S} \xrightarrow{p_\mathsf{G}} \overline{\Omega}_X\mathsf{G}$. The algorithm proposed by Rhodes has been validated by Ash's inevitability theorem [11] (in turn rediscovered itself by Herwig and Lascar [16], see [7, Thm. 8] for the connection between both results), and a formulation of Ash's theorem for any pseudovariety of semigroups has been proposed by the first author [2], and by the first author and Steinberg [8].

A generalization of the computation of the kernel of a finite semigroup is the computation of the so-called pointlike sets with respect to some pseudovariety V: these are the subsets T of S such that $\bigcap_{t\in T}\mu_\mathsf{V}(t) \neq \emptyset$, where μ_V is the composite relational morphism $S \xrightarrow{\varphi^{-1}} \overline{\Omega}_X\mathsf{S} \xrightarrow{p_\mathsf{V}} \overline{\Omega}_X\mathsf{V}$. It turns out that for the case of groups, being able to compute closures of regular languages in the free group and to test nonemptiness of the intersection of two such closures is sufficient for

solving Rhodes' conjecture. Pin and Reutenauer [19] proposed a nice algorithm to compute the topological closure of a regular language in the free group: the closure operator commutes with finite unions and products, while the closure of L^+, for some nonempty regular language L, is the subgroup generated by L.

Contributions. This paper presents structural properties of the relatively free aperiodic profinite semigroup $\overline{\Omega}_X\mathsf{A}$. Among the few known results concerning the profinite semigroup $\overline{\Omega}_X\mathsf{A}$, two are of particular importance: first, as stated above, McCammond [18] showed that the word problem for testing equality over $\overline{\Omega}_X\mathsf{A}$ of two terms built from letters using concatenation and ω-power is decidable. Second, one can effectively compute pointlike subsets of a semigroup with respect to $\overline{\Omega}_X\mathsf{A}$ (the first proof, due to Henckell[14], has been recently simplified and generalized by Henckell, Rhodes and Steinberg [15]).

The first contribution of the paper is an alternative proof of the word problem for ω-terms on aperiodic semigroups, which, unlike McCammond's original proof, does not use the solution of the word problem for certain Burnside semigroups. This method leads to several new applications: we prove in particular that all factors of an ω-term in the relatively free aperiodic profinite semigroup must also be representable by an ω-term. Next, we also generalize the algorithm of Pin and Reutenauer: we state and prove a similar algorithm for the pseudovariety A of aperiodic semigroups instead of that of groups. We show finally that this algorithm can be transferred from a pseudovariety to a subpseudovariety if both of them enjoy the property of being full. This makes it possible to show that the algorithm also holds for the pseudovariety R of $\mathcal{R}$-trivial semigroups.

The paper is organized as follows. We recall the basics of the theory of profinite semigroups in Section 2. We then present in Section 3 McCammond's normal form. We then introduce a family of star-free languages associated to an ω-term in Section 4, which we use to reprove McCammond's theorem. Some consequences of the star-freeness of these languages are drawn in Sections 5, 6 (which is devoted to the Pin-Reutenauer algorithm) and 7. Due to lack of space, all proofs are omitted, and will be available in [5,6].

2 Profinite semigroups

We briefly set some notation and recall the basics of the theory of profinite semigroups. See [3,4] for introductions to this theory, or [1,20] for comprehensive treatments.

A semigroup *pseudovariety* is a class of finite semigroups closed under *finite* direct product, subsemigroup and quotient. We call *star-free*, or sometimes *aperiodic*, a language recognized by an aperiodic semigroup (the terminology is justified by Schützenberger's theorem [22]). Fix a finite alphabet X. A semigroup S *separates* two words $u, v \in X^+$ if there exists a homomorphism $\varphi : X^+ \to S$ such that $\varphi(u) \neq \varphi(v)$. Given a pseudovariety V and $u, v \in X^+$, let $r_{\mathsf{V}}(u, v) = \min\{|S| : S \in \mathsf{V} \text{ and } S \text{ separates } u \text{ and } v\}$ (with $r_{\mathsf{V}}(u, v) = \infty$ if no semigroup of V separates u and v). Then $d_{\mathsf{V}}(u, v) = 2^{-r_{\mathsf{V}}(u,v)}$, with $2^{-\infty} = 0$,

defines a pseudo-metric on X^+. Further, the relation $\sim_{\mathsf{V}}$ defined by $u \sim_{\mathsf{V}} v$ if and only if $d_{\mathsf{V}}(u,v) = 0$ is a congruence, so that the quotient set $\Omega_X\mathsf{V} = X^+/\sim_{\mathsf{V}}$ inherits from X^+ a structure of semigroup.

We denote by $\overline{\Omega}_X\mathsf{V}$ the topological completion of the metric space $(\Omega_X\mathsf{V}, d_{\mathsf{V}})$. Elements of $\overline{\Omega}_X\mathsf{V}$ are called *pseudowords* over V. A topological semigroup is *pro-*V if it is compact and residually in V, where semigroups of V are endowed with the discrete topology. In particular, all semigroups of V are pro-V. It turns out that $\overline{\Omega}_X\mathsf{V}$ is the pro-V semigroup freely generated by X: every mapping $\varphi : X \to S$ into a pro-V semigroup S can be extended by a unique continuous homomorphism $\hat{\varphi} : \overline{\Omega}_X\mathsf{V} \to S$. This yields an interpretation of any pseudoword $u \in \overline{\Omega}_X\mathsf{V}$ in a pro-V-semigroup S, by the mapping $u_S : S^X \to S$ which associates to each function $\varphi : X \to S$ the element $\hat{\varphi}(u) \in S$. For instance, if $X = \{a, b\}$ and $u = ab$, the interpretation of u is the semigroup multiplication from $S \times S$ into S. For an element u of a pro-V semigroup, it is easy to check that the sequence $(u^{n!-1})_n$ converges, and we denote its limit by $u^{\omega-1}$. If $X = \{a\}$, the interpretation of $a^{\omega-1}$ in groups coincides with the usual inversion a^{-1}. For aperiodic semigroups, $u^{\omega-1}$ and $u.u^{\omega-1} = u^\omega$ have the same interpretation, namely the unique idempotent of the closed subsemigroup generated by u.

An *implicit signature* is a set σ of pseudowords containing the semigroup multiplication. We consider a canonical such signature, namely $\kappa = \{ab,\ a^{\omega-1}\}$. A pro-$\mathsf{V}$ semigroup S has a structure of σ-semigroup, that is, a structure of σ-algebra in which each operation in σ receives its natural interpretation in S. Given an implicit signature σ, denote by $\Omega^\sigma_X\mathsf{V}$ the relatively V-free σ-semigroup generated by X, whose elements are called *σ-words*. Each σ-word has a representation by a formal term over X in the signature σ. These terms are called *σ-terms*. For instance, κ-terms are obtained from letters of X using multiplication and $(\omega-1)$-power. Since our multiplication is associative, we identify terms that only differ by the order in which multiplications are to be carried out.

For a subset L of a topological semigroup S, denote by $\mathrm{cl}_S(L)$ the closure of L in S. For convenience, we write $\mathrm{cl}(L)$ instead of $\mathrm{cl}_{\overline{\Omega}_X\mathsf{S}}(L)$, $\mathrm{cl}_\sigma(L)$ instead of $\mathrm{cl}_{\Omega^\sigma_X\mathsf{S}}(L)$, $\mathrm{cl}_{\mathsf{V}}(L)$ instead of $\mathrm{cl}_{\overline{\Omega}_X\mathsf{V}}(L)$, and $\mathrm{cl}_{\sigma,\mathsf{V}}(L)$ instead of $\mathrm{cl}_{\Omega^\sigma_X\mathsf{V}}(L)$.

Let $p_{\mathsf{V}} : \overline{\Omega}_X\mathsf{S} \to \overline{\Omega}_X\mathsf{V}$ be the only continuous homomorphism sending each free generator to itself. Slightly abusing notation, for $L \subseteq X^+$, we will write $\mathrm{cl}_{\sigma,\mathsf{V}}(L)$ to denote $\mathrm{cl}_{\sigma,\mathsf{V}}(p_{\mathsf{V}}(L))$. Since the pro-$\mathsf{V}$ topology of $\Omega^\sigma_X\mathsf{V}$ is its induced topology as a subspace of $\overline{\Omega}_X\mathsf{V}$, we note that, for every $L \subseteq \Omega^\sigma_X\mathsf{S}$,

$$\mathrm{cl}_{\sigma,\mathsf{V}}(L) = \mathrm{cl}_{\mathsf{V}}(L) \cap \Omega^\sigma_X\mathsf{V}. \tag{1}$$

If S is a σ-semigroup and $L \subseteq S$, we denote by $\langle L\rangle_\sigma$ the σ-subsemigroup of S generated by L. Finally, for $L \subseteq X^+$, we let $\langle L\rangle_{\sigma,\mathsf{V}} = \langle p_{\mathsf{V}}(L)\rangle_\sigma$.

3 McCammond's normal form

Recall that by an ω-term, we mean an expression constructed from letters by applying the operations of concatenation and (formal) ω-power. Such an expression

may be viewed naturally as an operation on finite semigroups. McCammond's solution of the word problem for $\Omega_X^\kappa \mathsf{A}$ [18] consists in the reduction of arbitrary ω-terms to a certain normal form. McCammond then goes on to show that different ω-terms in normal form cannot represent the same pseudoword over A, which he does by invoking his results on free Burnside semigroups [17]. In this section, we briefly describe the normal form and we associate with each ω-term w and positive integer n a regular language $L_n(w)$.

The first important result of this work is that, if w is in normal form and n is sufficiently large, then $L_n(w)$ is aperiodic. This result yields an alternative proof of the uniqueness of normal forms for ω-terms over A, which is independent of the theory on free Burnside semigroups, by showing that, if u and v are both ω-terms in normal form and n is sufficiently large such that $L_n(u) \cap L_n(v) \neq \emptyset$, then $u = v$.

Let us first recall the definition of McCammond's normal form. To simplify the notation, McCammond represents ω-terms over an alphabet X as well-parenthesized words in the alphabet $X \cup \{(,)\}$, for which the parentheses are thus viewed as letters. The ω-term associated with such a word is obtained by replacing each matching pair of parentheses $(*)$ by $(*)^\omega$. Conversely, every ω-term determines a unique correctly parenthesized word over $X \cup \{(,)\}$. We define the *length* of an ω-term w to be the length of the word over $X \cup \{(,)\}$ which it determines, and we denote it $|w|$. From hereon, in the absence of mention to the contrary, we will refer to an ω-term meaning its associated word over $X \cup \{(,)\}$. It is easy to check that the ω-subsemigroup of the free semigroup $(X \cup \{(,)\})^+$ generated by X, where the ω-power is interpreted as the operation $w \mapsto (w)$, is freely generated by X as a unary semigroup.

In particular, there is a natural homomorphism of ω-semigroups $\epsilon : U_X \to \Omega_X^\kappa \mathsf{A}$, where U_X is the set of well-parenthesized words over X, that fixes each $x \in X$ (when we view X as a subset of U_X and $\Omega_X^\kappa \mathsf{A}$ in the natural way). To avoid ambiguities in the meaning of the parentheses, we write $\epsilon[w]$ for the image of $w \in U_X$ under ϵ. The elements of $\Omega_X^\kappa \mathsf{A}$ will sometimes be called *ω-words*. Whenever we say that an ω-term over the alphabet X is a factor of another, we mean that that is the case in the free semigroup $(X \cup \{(,)\})^+$.

The ω-word problem for A (over X) consists in deciding when two elements of U_X have the same image under ϵ. To solve this problem, McCammond described a normal form for ω-terms over A. For its description, a total order is fixed on the underlying alphabet X; on the extended alphabet, we set $(< x <)$ for every $x \in X$. A *primitive word* is a word that cannot be written in the form u^n with $n > 1$. Two words u and v are said to be *conjugate* if there are factorizations of the form $u = xy$ and $v = yx$, with the words x and y possibly empty. A *Lyndon word* is a primitive word that is lexicographically minimum in its conjugacy class. The *rank* of a word in the extended alphabet is the maximum number of nested parentheses in it.

A *rank 0 normal form* ω-term is simply a finite word. Assuming that rank i normal form terms have been defined, a *rank $i+1$ normal form term* is a term

of the form $\alpha_0(\beta_1)\alpha_1(\beta_2)\cdots\alpha_{n-1}(\beta_n)\alpha_n$, where the α_j and β_k are ω-terms such that

(1) each β_k is a Lyndon word;
(2) no intermediate α_j is a prefix of a power of β_j or a suffix of a power of β_{j+1};
(3) replacing each subterm (β_k) by $\beta_k\beta_k$, we obtain a rank i normal form term;
(4) at least one of the properties (2) and (3) is lost by canceling from α_j a prefix β_j (in case $j > 0$) or a suffix β_{j+1} (in case $j < n$).

These conditions yield a unique normal form. One can verify, for instance, that the normal form of $(aa)^\omega$ is a^ω, that the normal form of $a^\omega b^\omega$ is $a^\omega abb^\omega$, and that the normal form of $(a^\omega b^\omega)^\omega$ is $\left(a^\omega abb^\omega ba\right)^\omega a^\omega abb^\omega$, assuming the order $a < b$. McCammond also described a method to transform an arbitrary ω-term into one in normal form with the same image under ϵ. Moreover, he proved that if two ω-terms in normal form have the same image under ϵ, then they are equal.

We don't describe here McCammond's procedure to obtain the normal form. It consists in applying elementary changes that obviously retain the value of the ω-term under ϵ. The types of changes are given by the following rewriting rules, where $u \leftrightarrow v$ abbreviates $u \to v$ and $v \to u$, and where α and β stand for arbitrary ω-terms:

1. $((\alpha)) \leftrightarrow (\alpha)$
2. $(\alpha^k) \leftrightarrow (\alpha)$
3. $(\alpha)(\alpha) \leftrightarrow (\alpha)$
4. $(\alpha)\alpha \leftrightarrow (\alpha)$, $\alpha(\alpha) \leftrightarrow (\alpha)$
5. $(\alpha\beta)\alpha \leftrightarrow \alpha(\beta\alpha)$

Since all the rewriting rules are based on identities of ω-semigroups that are valid in A, every ω-term over X has the same image under ϵ as its normal form. To prove that distinct ω-terms in normal form have different images in $\Omega_X^\kappa \mathsf{A}$, McCammond used his solution of the word problem for certain free Burnside semigroups [17]. We have obtained a direct combinatorial proof of the same result which leads to many other applications.

We often use in our proofs the rank as an induction parameter for ω-terms. The rank of an ω-term is defined as the rank of its normal form, and it can be viewed as the maximal nesting of parentheses in this normal form. Thus, words have rank 0, a^ω and $(a^\omega)^\omega$ both have rank 1 (because their normal form is a^ω), and $(a^\omega b)^\omega$ has rank 2.

4 Languages $L_n(v)$ associated to an ω-term v

The main idea of the paper is to associate to each ω-term v a decreasing family of regular languages $L_n(v)$ characterizing the value of v over $\overline{\Omega}_X\mathsf{A}$ (or, which is equivalent by a standard compactness argument, over all finite aperiodic semigroups). In this section we present the main properties of the languages $L_n(v)$ and derive some applications. This gives, in particular, an alternative proof of McCammond's solution of the ω-word problem for A.

For $L \subseteq X^+$ and $n \geq 1$, let $L^{\geq n} = L^n L^*$. Given an ω-term w, we let $L_n(w)$ be the regular language obtained from w by replacing all ω-powers by $\geq n$. Since $L^{\geq n} \supseteq L^{\geq n+1}$, this clearly defines a decreasing family. For instance, $L_n(a^\omega) = a^n a^*$. The following statement is useful to compute inductively these languages.

Lemma 1. *The following formulas hold:*

(1) *for ω-terms u and v, we have $L_n(uv) = L_n(u)L_n(v)$.*
(2) *if $v = u_0 v_1^\omega u_1 \cdots v_r^\omega u_r$ is a factorization of an ω-term such that all the v_j have the same rank i and all the u_j have rank at most i, then*

$$L_n(v) = L_n(u_0)L_n(v_1^\omega)L_n(u_1)\cdots L_n(v_r^\omega)L_n(u_r);$$

(3) *for an ω-term v, $L_n(v^\omega) = L_n(v)^n L_n(v)^*$.*

We now introduce another parameter for ω-terms. Let $v = u_0 v_1^\omega u_1 \cdots v_r^\omega u_r$ be a term of rank $i \geq 1$ where each v_j has rank $i-1$ and each u_j has rank at most $i-1$. Let $\mu(v)$ denote the integer

$$\mu(v) = 2(2 + \max\{|v_j u_j v_{j+1}|, |u_0 v_1|, |v_r u_r| : j = 1, \ldots, r-1\}).$$

In case v is a word, we let $\mu(v) = |v|$. It is easy to check that, if the above expression for v is its normal form, then $\mu(v) \geq \max\{\mu(u_j), \mu(v_j)\}$. The fundamental property of the languages $L_n(v)$ is the following.

Theorem 2. *Let v be a term in normal form and let $n \geq \mu(v)$. Then the language $L_n(v)$ is star-free.*

The theorem is proved by induction on the rank of the term in normal form. Using Schützenberger's characterization of star-free languages, it suffices to show that there exists a bound k such that $xy^k z \in L_n(v)$ if and only if $xy^{k+1}z \in L_n(v)$, for all words x, y, z. From $xy^k z \in L_n(v)$, we obtain two factorizations of the same word: one is given by $xy^k z$, and the other one by the expression of $L_n(v)$ that is obtained inductively using Lemma 1. One of the ingredients to control how these factorizations match is the following consequence of the well-known Theorem of Fine and Wilf on the relationship between the periods of a sufficiently long word and their synchronization.

Lemma 3. *Let u and v be Lyndon words and suppose that w is a word such that $|w| \geq |u| + |v|$ and w is a factor of both a power of u and a power of v. Then $u = v$. Moreover, for all factorizations $u^m = xwy$ and $v^n = zwt$, there is a factorization $w = w_1 w_2$ such that $xw_1, zw_1 \in u^*$.*

In the statement of Theorem 2, we do not know whether the bound $n \geq \mu(v)$ is optimal, but we do know that some bound is required, that is, that $L_n(v)$ may not be star-free for v in normal form. An example is obtained by taking $v = (a^\omega abb^\omega a^2 b^2)^\omega$. Then $L_1(v) \cap (a^2b^2)^* = ((a^2b^2)^2)^+$ so that $L_1(v)$ is not star-free since $(a^2b^2)^*$ is star-free and $((a^2b^2)^2)^+$ is not.

From Theorem 2, one can deduce several applications, yielding an alternative proof of the ω-word problem for A, that is, an algorithm to test equality under ϵ of two ω-terms. The proof is based on the following separation criterion.

Theorem 4. *Let u and v be two ω-terms in normal form and let n be an integer greater than* $\max\{|u|, |v|, \mu(u), \mu(v)\}$. *If $L_n(u) \cap L_n(v) \neq \emptyset$, then $u = v$.*

Corollary 5 (McCammond's solution of the word problem [18]). *If u and v are terms in normal form which define the same pseudoword over A, then $u = v$ as parenthesized words.*

5 First consequences of the star-freeness of $L_n(v)$

There are several consequences of the star-freeness of $L_n(v)$ for v in normal form and n large enough. We first prove an important property which, apparently, does not follow easily from McCammond's results. We say that a pseudovariety V is *σ-factorial* for an implicit signature σ if, for every $(u, v) \in \Omega_X^\sigma \mathsf{V} \times \overline{\Omega}_X \mathsf{V}$, if v is a factor of u, then also $v \in \Omega_X^\sigma \mathsf{V}$.

Theorem 6. *The pseudovariety A is κ-factorial.*

Recall now that a semigroup is *stable* if any two $\mathcal{J}$-equivalent elements which are comparable for the $\leq_{\mathcal{R}}$-ordering are also $\mathcal{R}$-equivalent, and dually for the $\leq_{\mathcal{L}}$-ordering.

Corollary 7. *The semigroup $\Omega_X^\kappa \mathsf{A}$ is stable, and the Green relations $\mathcal{J}$ and $\mathcal{D}$ coincide in $\Omega_X^\kappa \mathsf{A}$.*

Using the very definition of the languages $L_n(w)$ and the fact that for $u \in \overline{\Omega}_X \mathsf{A}$, the sequence $(u^n)_n$ converges to u^ω, one can show the following statement.

Proposition 8. *If u is an arbitrary ω-term, then*

$$p_{\mathsf{A}}\Big(\bigcap_n \mathrm{cl}(L_n(u))\Big) = \big\{p_{\mathsf{A}}(u)\big\} = \bigcap_n p_{\mathsf{A}}\big(\mathrm{cl}(L_n(u))\big).$$

Let us now formulate another direct consequence of the star-freeness of $L_n(w)$ and of Proposition 8.

Corollary 9. *Let w be an ω-term in normal form. Then*

(1) *the set $\bar{\mathcal{N}}_w$ of all $\mathrm{cl}_{\mathsf{A}}(L_n(w))$ $(n \geq \mu(w))$ is a basis of open neighborhoods of w in $\overline{\Omega}_X \mathsf{A}$;*
(2) *the set $\mathcal{N}_w$ of all $\mathrm{cl}_{\kappa,\mathsf{A}}(L_n(w))$ $(n \geq \mu(w))$ is a basis of open neighborhoods of w in $\Omega_X^\kappa \mathsf{A}$.*

However, one can show that the closures of these aperiodic languages do not form a basis for the topology of the whole pro-A semigroup.

Proposition 10. *Let X be a finite alphabet with at least two letters. Then the set of all open subsets of the form $\mathrm{cl}_{\mathsf{A}}(L_n(w))$, with w an ω-term in normal form and $n \geq \mu(w)$, is not a basis of the topology of $\overline{\Omega}_X \mathsf{A}$.*

We next want to test whether some given ω-word lies in the closure of a given regular language. An answer to this question is given by the following statement. Given a semigroup S, we denote by $\mathcal{P}(S)$ the semigroup whose elements are the subsets of S, with the multiplication given by $XY = \{xy : x \in X \text{ and } y \in Y\}$. The *index* of a finite semigroup S is the least integer i such that S satisfies $x^{i+p} = x^i$ for some positive integer p.

Proposition 11. *Let $u \in \Omega_X^\kappa \mathsf{A}$, $L \subseteq X^+$ be a regular language, and let i be the index of $\mathcal{P}(\mathrm{Synt}(L))$. If $L \cap L_i(u) \neq \emptyset$ then, for every k, $L \cap L_k(u) \neq \emptyset$.* □

Theorem 12. *Given a regular language $L \subseteq X^+$ and $w \in \Omega_X^\kappa \mathsf{A}$, let $\bar{w}$ denote the normal form of w. Then $w \in \mathrm{cl}_{\kappa,\mathsf{A}}(L)$ if and only if $L \cap L_i(\bar{w}) \neq \emptyset$, where i is the index of $\mathcal{P}(\mathrm{Synt}(L))$.*

Corollary 13. *Given a regular language $L \subseteq X^+$ and $u \in \Omega_X^\kappa \mathsf{A}$, one can decide whether $u \in \mathrm{cl}_{\kappa,\mathsf{A}}(L)$.*

6 The Pin-Reutenauer algorithm for A

We say that the *Pin-Reutenauer procedure holds for a pseudovariety V in the implicit signature σ* if, for all nonempty regular languages $K, L \subseteq X^+$, the following equations hold:

$$\mathrm{cl}_{\sigma,\mathsf{V}}(KL) = \mathrm{cl}_{\sigma,\mathsf{V}}(K) \cdot \mathrm{cl}_{\sigma,\mathsf{V}}(L), \tag{2}$$

$$\mathrm{cl}_{\sigma,\mathsf{V}}(L^+) = \langle \mathrm{cl}_{\sigma,\mathsf{V}}(L) \rangle_\sigma. \tag{3}$$

Note that, given subsets $K, L \subseteq \Omega_X^\sigma \mathsf{V}$, the inclusion $\mathrm{cl}_{\sigma,\mathsf{V}}(K)\mathrm{cl}_{\sigma,\mathsf{V}}(L) \subseteq \mathrm{cl}_{\sigma,\mathsf{V}}(KL)$ is always true: it follows directly from the continuity of the multiplication.

The name is justified by the fact that the above formulas hold for the pseudovariety G of all finite groups and the implicit signature κ. This was conjectured by Pin and Reutenauer [19] and the conjecture was reduced to another conjecture concerning the profinite topology of the free group $\Omega_X^\kappa \mathsf{G}$, which in turn was proved by Ribes and Zalesskiĭ [21]. The Pin-Reutenauer conjecture was in turn introduced as an approach and later shown to be formally equivalent to the Rhodes "type II conjecture", which was settled independently by Ash [11]. The situation for groups is somewhat simpler since the right side of formula (3) reduces to the subgroup H generated by L. To prove that indeed $H = \langle \mathrm{cl}_{\kappa,\mathsf{G}}(L) \rangle_\kappa$, it suffices to observe that L^+ is contained in H and that H is finitely generated by a theorem of Anissimow and Seifert [10], and therefore it is closed by a theorem of M. Hall (see [19] for details).

For more general pseudovarieties, the topological closure on the right side of (3) cannot be dropped, as the following example shows. Consider the regular language $L = a^+b^+$ and its closure $\mathrm{cl}_\kappa(L)$ in $\Omega_X^\kappa \mathsf{S}$. We claim that $\langle L \rangle_\kappa$ is not closed in $\Omega_X^\kappa \mathsf{S}$. Indeed, since $L^+ \subseteq \langle L \rangle_\kappa$, if $\langle L \rangle_\kappa$ were closed then we would have $\mathrm{cl}_\kappa(L^+) \subseteq \langle L \rangle_\kappa$. Since one can check that $\langle L \rangle_\kappa \subseteq \langle \mathrm{cl}_\kappa(L) \rangle_\kappa \subseteq \mathrm{cl}_\kappa(L^+)$, it follows that $\langle L \rangle_\kappa$ closed implies $\langle L \rangle_\kappa = \langle \mathrm{cl}_\kappa(L) \rangle_\kappa \supseteq \mathrm{cl}_\kappa(L)$. Now, clearly $a^\omega b \in \mathrm{cl}_\kappa(L)$

while one can show that for every element of $\langle L\rangle_\kappa$, the exponents of its factors of the form a^n are all finite. Hence $\langle L\rangle_\kappa$ is not closed in $\Omega_X^\kappa \mathsf{S}$.

In this section, we generalize the Pin-Reutenauer procedure to pseudovarieties of finite aperiodic semigroups.

Theorem 14. *The Pin-Reutenauer procedure holds for the pseudovariety* A *with respect to the signature* κ*.*

We first present general results concerning closures of regular languages. It shows in particular that the Pin-Reutenauer procedure can be transferred from a pseudovariety to a subpseudovariety, provided both pseudovarieties are σ-full (as defined below). We shall use this to deduce that the Pin-Reutenauer procedure also holds for R, the pseudovariety of all $\mathcal{R}$-trivial semigroups.

Given a finite X-generated semigroup S and an onto continuous homomorphism $\varphi : \overline{\Omega}_X\mathsf{S} \to S$, we denote by μ_V the relational morphism $S \to \overline{\Omega}_X\mathsf{V}$ given by $\mu_\mathsf{V} = p_\mathsf{V} \circ \varphi^{-1}$, by μ_V^σ the relational morphism $S \to \Omega_X^\sigma\mathsf{V}$ given by $\mu_\mathsf{V}^\sigma = p_\mathsf{V} \circ (\varphi|_{\Omega_X^\sigma \mathsf{S}})^{-1}$, and by $\bar{\mu}_\mathsf{V}^\sigma$ the relational morphism given by $\bar{\mu}_\mathsf{V}^\sigma = \mu_\mathsf{V} \cap (S \times \Omega_X^\sigma \mathsf{V})$. Following [8,9], we say that V is *σ-full* if $\bar{\mu}_\mathsf{V}^\sigma = \mu_\mathsf{V}^\sigma$ for every such homomorphism $\varphi : \overline{\Omega}_X\mathsf{S} \to S$ into a finite semigroup S.

Proposition 15. *A pseudovariety* V *is σ-full with respect to an implicit signature σ if and only if, for every regular language $L \subseteq X^+$, we have the equality* $\mathrm{cl}_{\sigma,\mathsf{V}}(L) = p_\mathsf{V}(\mathrm{cl}(L) \cap \Omega_X^\sigma \mathsf{S})$.

The following statement allows us to transfer properties (2) and (3) to subpseudovarieties, assuming fullness.

Proposition 16. *Let* V*,* W *be two σ-full pseudovarieties such that* $\mathsf{V} \subseteq \mathsf{W}$*. Let* $K, L \subseteq X^+$ *be regular languages.*

(*a*) *If* $\mathrm{cl}_{\sigma,\mathsf{W}}(KL) = \mathrm{cl}_{\sigma,\mathsf{W}}(K) \cdot \mathrm{cl}_{\sigma,\mathsf{W}}(L)$*, then* $\mathrm{cl}_{\sigma,\mathsf{V}}(KL) = \mathrm{cl}_{\sigma,\mathsf{V}}(K) \cdot \mathrm{cl}_{\sigma,\mathsf{V}}(L)$.
(*b*) *If* $\mathrm{cl}_{\sigma,\mathsf{W}}(L^+) = \langle \mathrm{cl}_{\sigma,\mathsf{W}}(L)\rangle_\sigma$*, then* $\mathrm{cl}_{\sigma,\mathsf{V}}(L^+) = \langle \mathrm{cl}_{\sigma,\mathsf{V}}(L)\rangle_\sigma$.

We state for the record the following consequence of Proposition 16.

Corollary 17. *Let* V *and* W *be σ-full pseudovarieties such that* $\mathsf{V} \subseteq \mathsf{W}$*. If the Pin-Reutenauer procedure holds for* W *with respect to σ then it also holds for* V*.* □

Note that an immediate consequence of Corollary 17 is that if the Pin-Reutenauer procedure holds for the pseudovariety S of all finite semigroups with respect to the implicit signature σ then it also holds for every σ-full pseudovariety. Indeed, S is trivially σ-full for every implicit signature σ. This motivates the problem of determining for which implicit signatures the Pin-Reutenauer procedure holds for S. In view of the above and later results in this paper, it would be particularly interesting to consider this problem for the signature κ.

Proposition 18. *Let $K, L \subseteq X^+$ be regular languages. Let σ be an implicit signature, and* V *be a σ-factorial pseudovariety. Then* $\mathrm{cl}_{\sigma,\mathsf{V}}(KL) = \mathrm{cl}_{\sigma,\mathsf{V}}(K)\mathrm{cl}_{\sigma,\mathsf{V}}(L)$.

For the closure of regular languages of the form L^+ in $\Omega_X^\kappa \mathsf{A}$, the most difficult inclusion to prove is the following.

Proposition 19. *Let w be a κ-term in normal form over the finite alphabet X, let $L \subseteq X^+$ be a regular language, and suppose that $w \in \mathrm{cl}_{\kappa,\mathsf{A}}(L^+)$. Then $w \in \langle \mathrm{cl}_{\kappa,\mathsf{A}}(L) \rangle_\kappa$.*

7 κ-fullness of A and an application

In order to be able to apply Theorem 14 and Corollary 17 to deduce that κ-full subpseudovarieties of A are also such that the Pin-Reutenauer procedure holds for them, we need the following result.

Theorem 20. *The pseudovariety A is κ-full.*

In turn, the following is a simple application of well-developed techniques concerning the pseudovariety R of all finite R-trivial semigroups.

Theorem 21. *The pseudovariety R is κ-full.*

We may now apply Corollary 17 to obtain the following result.

Theorem 22. *The Pin-Reutenauer procedure holds for the pseudovariety R with respect to the signature κ.* □

Acknowledgements This work was partly supported by the Pessoa Portuguese-French project Egide-Grices 11113YM *Automata, profinite semigroups and symbolic dynamics*. The work of the first author was supported, in part, by *Fundação para a Ciência e a Tecnologia* (FCT) through the *Centro de Matemática da Universidade do Porto*, by FCT. The work of the second author was supported, in part, by FCT through the *Centro de Matemática da Universidade do Minho*. Both the first and second authors were also supported by the project PTDC/MAT/65481/2006, which is partly funded by the European Community Fund FEDER. The work leading to this paper has also been carried out within the framework of the ESF programme "Automata: from Mathematics to Applications (AutoMathA)", whose support is gratefully acknowledged.

References

1. J. Almeida. *Finite Semigroups and Universal Algebra*. World Scientific, Singapore, 1995. English translation.
2. J. Almeida. Hyperdecidable pseudovarieties and the calculation of semidirect products. *Int. J. Algebra Comput.*, 9:241–261, 1999.
3. J. Almeida. Finite semigroups: an introduction to a unified theory of pseudovarieties. In G. M. S. Gomes, J.-É. Pin, and P. V. Silva, editors, *Semigroups, Algorithms, Automata and Languages*, pages 3–64, Singapore, 2002. World Scientific.

4. J. Almeida. Profinite semigroups and applications. In V. B. Kudryavtsev and I. G. Rosenberg, editors, *Structural Theory of Automata, Semigroups, and Universal Algebra*, volume 207 of *NATO Science Series II: Mathematics, Physics and Chemistry*, New York, 2005. Springer. Proceedings of the NATO Advanced Study Institute on Structural Theory of Automata, Semigroups and Universal Algebra, Montréal, Québec, Canada, 7-18 July 2003.
5. J. Almeida, J. C. Costa, and M. Zeitoun. McCammond normal forms for free aperiodic semigroups revisited. In preparation.
6. J. Almeida, J. C. Costa, and M. Zeitoun. The Pin-Reutenauer algorithm for aperiodic semigroups. In preparation.
7. J. Almeida and M. Delgado. Sur certains systèmes d'équations avec contraintes dans un groupe libre. *Portugal. Math.*, 56(4):409–417, 1999.
8. J. Almeida and B. Steinberg. On the decidability of iterated semidirect products and applications to complexity. *Proc. London Math. Soc.*, 80:50–74, 2000.
9. J. Almeida and B. Steinberg. Syntactic and global semigroup theory, a synthesis approach. In J. C. Birget, S. W. Margolis, J. Meakin, and M. V. Sapir, editors, *Algorithmic Problems in Groups and Semigroups*, pages 1–23. Birkhäuser, 2000.
10. A. W. Anissimow and F. D. Seifert. Zur algebraischen Charakteristik der durch kontext-freie Sprachen definierten Gruppen. *Elektron. Informationsverarbeit. Kybernetik*, 11(10–12):695–702, 1975.
11. C. J. Ash. Inevitable graphs: a proof of the type II conjecture and some related decision procedures. *Int. J. Algebra Comput.*, 1:127–146, 1991.
12. J. Berstel. *Transductions and context-free languages*, volume 38 of *Leitfäden der Angewandten Mathematik und Mechanik [Guides to Applied Mathematics and Mechanics]*. B. G. Teubner, Stuttgart, 1979.
13. S. Eilenberg. *Automata, Languages and Machines*, volume B. Academic Press, New York, 1976.
14. K. Henckell. Pointlike sets: the finest aperiodic cover of a finite semigroup. *J. Pure Appl. Algebra*, 55(1-2):85–126, 1988.
15. K. Henckell, J. Rhodes, and B. Steinberg. Aperiodic Pointlikes and Beyond. *ArXiv e-prints*, June 2007.
16. B. Herwig and D. Lascar. Extending partial automorphisms and the profinite topology on free groups. *Trans. Amer. Math. Soc.*, 352(5):1985–2021, 2000.
17. J. McCammond. The solution to the word problem for the relatively free semigroups satisfying $t^a = t^{a+b}$ with $a \geq 6$. *Int. J. Algebra Comput.*, 1(1):1–32, 1991.
18. J. McCammond. Normal forms for free aperiodic semigroups. *Int. J. Algebra Comput.*, 11(5):581–625, 2001.
19. J.-E. Pin and C. Reutenauer. A conjecture on the Hall topology for the free group. *Bull. London Math. Soc.*, 23:356–362, 1991.
20. J. Rhodes and B. Steinberg. *The q-theory of finite semigroups*. Springer Monographs in Mathematics. Springer, 2009.
21. L. Ribes and P. A. Zalesskiĭ. On the profinite topology on a free group. *Bull. London Math. Soc.*, 25:37–43, 1993.
22. M. P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.