ω-TERMS OVER FINITE APERIODIC SEMIGROUPS

J. ALMEIDA[1], J. C. COSTA[2], AND M. ZEITOUN[3]

ABSTRACT. This paper provides a characterization of pseudowords over the pseudovariety of all finite aperiodic semigroups that are given by ω-terms, that is that can be obtained from the free generators using only multiplication and the ω-power. A necessary and sufficient condition for this property to hold turns out to be given by the conjunction of two rather simple finiteness conditions: the nonexistence of infinite anti-chains of factors and the rationality of the language of McCammond normal forms of ω-terms that define factors.

1. INTRODUCTION AND PRELIMINARIES

Since the mid nineteen seventies that the theory of finite semigroups has seen a significant boost thanks to its connections and applications in computer science. The now classical framework for the relationships between the two areas is provided by Eilenberg's correspondence between pseudovarieties of finite semigroups (classes of finite semigroups closed under taking homomorphic images, subsemigroups, and finite direct products) and so-called varieties of rational languages [10, 14]. The typical application consists in the solution (positive or negative) of an algorithmic problem about rational languages or finite automata by translating it to a membership problem of a computable finite semigroup in a suitable pseudovariety of semigroups.

Unlike varieties of algebras, pseudovarieties do not in general have free algebras. In the following decade, a substitute for relative free algebras emerged and found many applications in the context of pseudovarieties of semigroups: relative free profinite algebras. The common algebraic and combinatorial properties of the semigroups in a pseudovariety become encoded in the algebraic and topological properties of their free profinite semigroups. The difficulty, of course, is to obtain sufficient structural information about these profinite semigroups for the intended applications.

Throughout this paper, we assume that the reader is familiar with the general basic theory of pseudovarieties and specifically with the central role played by relatively free profinite semigroups. A quick introduction, both to the theory and to the applications, is found in [4]. For more comprehensive treatments, see [1, 15].

For a pseudovariety V of (finite) semigroups, a compact semigroup is said to be *pro*-V if continuous homomorphisms into members of V are sufficient to separate points. We adopt the following notation. A general finite alphabet is denoted X. The free semigroup on X, viewed as the set of all finite words in the letters of X, is denoted X^+; the set X^* is obtained by adding to X^+ the empty word. The pro-V semigroup freely generated by X is represented by $\overline{\Omega}_X\mathsf{V}$. There is a natural homomorphism $X^+ \to \overline{\Omega}_X\mathsf{V}$ determined by the choice of free generators. For the pseudovarieties in this paper, this is always injective, and so we view X^+ as a subsemigroup of $\overline{\Omega}_X\mathsf{V}$. The letters S and A denote, respectively, the pseudovarieties of all finite semigroups and of all finite *aperiodic* semigroups, that is semigroups all of whose subgroups are trivial. We represent by LSI

Key words and phrases. Profinite semigroup, aperiodic semigroup, pseudoword, omega-word, semigroupoid, pseudovariety, well quasi ordered set, uniform recurrent pseudoword.

[1] CMUP, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal.
[2]CMAT, Universidade do Minho, Campus de Gualtar, 4700-320 Braga, Portugal.
[3]LaBRI, Université Bordeaux 1 & CNRS UMR 5800. 351 cours de la Libération, 33405 Talence Cedex, France.

the pseudovariety of all finite semigroups S such that, for every idempotent $e \in S$, the monoid eSe is a semilattice.

A *profinite* semigroup is a pro-S semigroup. A profinite semigroup is *finitely generated* if it has a dense finitely generated subsemigroup (in the usual algebraic sense); in this case, it is well known that the topology is induced by a metric so that, in particular, it is characterized by the convergence of sequences.

Elements of $\overline{\Omega}_X\mathsf{V}$ are called *implicit operations* or *pseudowords* over V; if no reference to a pseudovariety is made, it is assumed to be S. In contrast, the elements of X^+ are called *finite words*. Implicit operations can be naturally interpreted in profinite semigroups. For a set σ of implicit operations, one may therefore speak of the σ-subalgebra of $\overline{\Omega}_X\mathsf{V}$ generated by X, which is denoted $\Omega_X^\sigma\mathsf{V}$. Formal terms over a finite alphabet X in the signature σ are called *σ-terms*. Since our multiplication is always associative, without further reference we identify terms that only differ by the order in which multiplications are to be carried out. The interest in such subalgebras arises from the fact that sometimes they are sufficiently rich to contain solutions of certain types of problems that admit solutions in $\overline{\Omega}_X\mathsf{V}$. The most famous example of this phenomenon is given by Ash's inevitability theorem for the pseudovariety of all finite groups [8]. Under relatively simple assumptions on the signature σ, which includes the word problem for $\Omega_X^\sigma\mathsf{V}$, such a property implies the decidability of the existence of solutions for the problems in question [11, 7].

Following [7], we denote by κ the set consisting of the operations of multiplication and pseudo-inversion $x \mapsto x^{\omega-1}$, where, for an element s of a finite semigroup, $s^{\omega-1}$ stands for the inverse of ss^ω in the maximal subgroup of the subsemigroup generated by s, whose idempotent is denoted s^ω. Note that, for pseudovarieties contained in A, the operations $x^{\omega-1}$ and x^ω coincide. Since this paper is concerned mainly with the pseudovariety A, our κ-terms will use the operation x^ω rather than $x^{\omega-1}$ and such terms are also, abusively, called *ω-terms*.

We will sometimes adopt the simplified notation of McCammond [13] for ω-terms under which the expression (α) stands for α^ω. This allows us to refer to ω-terms as (well parenthesized) words over an extended alphabet $X \cup \{(,)\}$, which is particularly useful for McCammond's solution of the word problem for $\Omega_X^\kappa\mathsf{A}$ and will also play an important role in our considerations. The *rank* of an ω-term is the maximum number of nested parentheses.

McCammond described a normal form for ω-terms over A that we will use extensively in this paper. 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 as a power with integer exponent greater than 1. A *Lyndon word* is a primitive word that is lexicographically minimum in its conjugacy class.

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 the following conditions hold:

- (a) each β_k is a Lyndon word and a term of rank i;
- (b) no intermediate α_j is a prefix of a power of β_j or a suffix of a power of β_{j+1};
- (c) replacing each subterm (β_k) by $\beta_k\beta_k$, we obtain a rank i normal form term;
- (d) at least one of the preceding properties is lost by canceling from α_j a prefix β_j or a suffix β_{j+1} in case $0 < j < n$;
- (e) β_1 is not a suffix of α_0 and β_n is not a prefix of α_n.

McCammond also described an algorithm to transform an arbitrary ω-term into one in normal form that is equal to it over A. Moreover, he proved that if two terms in normal form are equal over A, then they are equal as words over the extended alphabet. The unique term in normal form that coincides with a given $w \in \Omega_X^\kappa\mathsf{A}$ is called its *normal form*.

McCammond's solution of the word problem may be formally stated as follows.

Theorem 1.1. *The normal form algorithm transforms any ω-term to one in normal form. If u and v are ω-terms in normal form which define the same pseudoword over A, then $u = v$ as parenthesized words.*

Section 2 presents some further preliminary results. Section 3 serves mostly to discuss examples that satisfy or violate certain chain conditions. In connection with ideas of symbolic dynamics, some of these examples are rather general. The section also serves to suggest the appropriate finiteness conditions. The main result is found at the end of Section 4, which also explains how it was reached and sketches the key ideas in the proof. The main theorem characterizes pseudowords over A which can be expressed by ω-terms. In other words, it characterizes iterated periodicity of pseudowords over finite aperiodic semigroups. The characterization is made in terms of two simple finiteness conditions, although the proof of correctness is technically rather involved.

Due to lack of space, results are presented without proofs in this extended abstract.

2. Factors of ω-terms over A

By a *quasi-order* on a set we mean a reflexive and transitive binary relation $\leq$. The associated strict order $<$ is defined by $x < y$ if $x \leq y$ and $y \not\leq x$. A quasi-order is a *well-quasi-order* (*wqo* for short) if it admits no infinite descending chains $x_1 > x_2 > \cdots$ and no infinite anti-chains, that is infinite sets in which, for any two distinct elements x and y, $x \not\leq y$. Equivalently, given any sequence $(x_n)_n$, there exist indices m and n such that $m < n$ and $x_m \leq x_n$.

For two elements s and t of a semigroup S, we write $s \leq_{\mathcal{J}} t$ if t is a factor of s, and then we also say that t lies *$\mathcal{J}$-above* s. In case s and t are factors of each other, we write $s \mathrel{\mathcal{J}} t$, which defines an equivalence relation on S. We say that a $\mathcal{J}$-class is *regular* if all its elements are regular. Equivalently, for compact semigroups, the $\mathcal{J}$-class contains some idempotent. We also say that s is a *prefix* of t and write $s \geq_{\mathcal{R}} t$ if it is a left factor in some factorization of t. If s and t are prefixes of each other, then we write $s \mathrel{\mathcal{R}} t$. The two relations $\mathcal{J}$ and $\mathcal{R}$ are related in the *left stability* property which states that $s \mathrel{\mathcal{J}} t$ and $s \leq_{\mathcal{R}} t$ implies $s \mathrel{\mathcal{R}} t$. *Right stability* is defined dually. *Stability*, which stands for both properties, holds for compact semigroups but not for semigroups in general.

The following is a key property of $\Omega_X^\kappa\mathsf{A}$ which does not seem to follow directly from Theorem 1.1. Our proof depends on a finer combinatorial analysis of ω-terms which we have been able to carry out through the study of certain star-free languages associated with their normal forms. These results actually also lead to a proof of the (hardest) second part of Theorem 1.1 which, unlike McCammond's original proof, does not depend on the solution of the word problem for Burnside semigroups [12].

Theorem 2.1. *If $v \in \Omega_X^\kappa\mathsf{A}$ and $u \in \overline{\Omega}_X\mathsf{A}$ is a factor of v, then $u \in \Omega_X^\kappa\mathsf{A}$.*

Once we know that factors of ω-terms must be ω-terms, we may use normal forms to construct them. This leads to the following result.

Corollary 2.2. *There are only finitely many regular $\mathcal{J}$-classes $\mathcal{J}$-above a given $w \in \Omega_X^\kappa\mathsf{A}$.*

The following is a special case of a much more general result [6, Lemma 4.7]. From another point of view, it is also a special case of a tighter property of the pseudovariety A [5, Lemma 4.8].

Lemma 2.3. *Let $w \in \overline{\Omega}_X\mathsf{A}$ and suppose that u and v are two prefixes of w. Then one of u and v is a prefix of the other.*

In the terminology of [9], Lemma 2.3 is expressed by stating that the semigroup $\overline{\Omega}_X\mathsf{A}$ has *unambiguous $\mathcal{R}$-order*. Dually, $\overline{\Omega}_X\mathsf{A}$ also has unambiguous $\mathcal{L}$-order.

Given a pseudoword $w \in \Omega_X^\kappa\mathsf{A}$, its *rank* is the rank of the only term in normal form that represents w (cf. Theorem 1.1). More generally, we define the *rank* of an element s of a semigroup S, denoted $r(s)$, to be the supremum of the cardinals of $<_{\mathcal{J}}$-chains of idempotents $\mathcal{J}$-above s. Note that $u \geq_{\mathcal{J}} v$ implies $r(u) \leq r(v)$.

Proposition 2.4. *If $w \in \Omega_X^\kappa \mathsf{A}$ then $r(w) = \operatorname{rank} w < \infty$.*

Using the normal form algorithm, it is now easy to establish the following result.

Corollary 2.5. *The set of prefixes of a given $w \in \Omega_X^\kappa \mathsf{A}$ is wqo under the prefix order.*

3. Chain conditions for pseudowords

The ascending chain conditions for $\overline{\Omega}_X \mathsf{V}$ with respect to the quasi-orders $\leq_{\mathcal{J}}$ and $\leq_{\mathcal{R}}$ (and its dual $\leq_{\mathcal{L}}$) have already been studied in [1, Chapter 12]. In this section, we consider these properties for $\overline{\Omega}_X \mathsf{A}$ under the hypothesis that the rank is bounded.

The following proposition shows that the set of factors of a given $w \in \overline{\Omega}_X \mathsf{A}$ may not be wqo under the prefix order, even under the restriction $r(w) < \infty$. The following example is taken from [1, Chapter 12] and satisfies $r(w) = 2$.

Let $(n_{1,k})_k$ be a strictly increasing sequence of integers such that the sequence $(b^{n_{1,k}} a \cdots b^2 aba)_k$ converges in $\overline{\Omega}_X \mathsf{A}$ and call x_1 its limit. Assuming $(n_{i,k})_k$ is a subsequence of $(n_{i-1,k})_k$ for $i = 2, \ldots, \ell$ such that the sequence $(b^{n_{i,k}} a \cdots b^{i+1} ab^i a)_k$ converges to x_i in $\overline{\Omega}_X \mathsf{A}$ for $i = 1, \ldots, \ell$, then there also exists a subsequence $(n_{\ell+1,k})_k$ of $(n_{\ell,k})_k$ such that $(b^{n_{\ell+1,k}} a \cdots b^{\ell+2} ab^{\ell+1} a)_k$ converges in $\overline{\Omega}_X \mathsf{A}$ and we denote by $x_{\ell+1}$ its limit. This recursive definition yields a sequence $(x_n)_n$. The following proposition states some of its properties.

Proposition 3.1. *The following properties hold for the above sequence $(x_n)_n$:*

(*a*) *$x_{n+1} >_{\mathcal{R}} x_n$ for every n;*
(*b*) *$\lim x_n = (b^\omega a)^\omega$ in $\overline{\Omega}_X \mathsf{A}$;*
(*c*) *the only regular $\mathcal{J}$-classes $\mathcal{J}$-above any given x_n are those of b^ω and $(b^\omega a)^\omega$ and so $r(x_n) = 2$.*

One may ask whether the stronger restriction $r(w) = 1$ is sufficient to guarantee that the set of prefixes of w is wqo. We first claim that, if there is a sequence $(x_n)_n$ in $\overline{\Omega}_X \mathsf{A}$ such that each $x_n <_{\mathcal{R}} x_{n+1}$ and $r(x_1) = 1$ then there is such a sequence that converges to an idempotent, which explains how we were led to a rather negative answer to the above question given by Theorem 3.4 below.

Indeed, for a strictly $\leq_{\mathcal{R}}$-ascending sequence $(x_n)_n$, no x_n belongs to X^+, whence each x_n has rank at least 1, that is, it has some idempotent factor. Without loss of generality, we may assume that $\lim x_n = x$. Then x is a prefix of every x_n and it also has some idempotent factor, whence $1 \leq r(x) \leq r(x_n) = 1$.

Since x is a prefix of x_n, there is a factorization $x_n = xy_n$. We claim that we may assume that $y_{n+1} >_{\mathcal{R}} y_n$ for all n. To prove it, we basically follow the same argument as for the construction of the example of Proposition 3.1. We first let z_n be such that $x_n = x_{n+1} z_n$, for all n. By taking subsequences, which corresponds to associating consecutive z_n's, we may arrange that the sequence $(z_n \cdots z_1)_n$ converges, and we call y_1 its limit, which is such that $x_1 = xy_1$. Then, again taking subsequences, we may assume that the sequence $(z_n \cdots z_2)_n$ converges, and we let y_2 be its limit, for which we have $y_1 = y_2 z_1$ and $x_2 = xy_2$. And so on. This proves the claim.

We may as well further assume that $\lim y_n = y$ in $\overline{\Omega}_X \mathsf{A}$. Taking limits in $x_n = xy_n$, we deduce that $x = xy = xy^\omega$. Since $r(y_1) = 1$, we also have $1 \leq r(y) \leq r(y_n) \leq 1$. If y is not idempotent, we deduce that $1 = r(x) \geq r(y^\omega) > r(y) \geq 1$, which is absurd. Hence y is an idempotent, which proves our claim on the existence of a special counter-example if any counter-example exists.

Note that, since $r(y) = 1$, there is no other idempotent $\mathcal{J}$-above y. By [3, Theorem 2.6], it follows that y is *uniformly recurrent*, meaning that, for every finite factor u of y, every sufficiently long finite factor of y contains u as a factor.

For a pseudovariety V, we let p_{V} denote the unique continuous homomorphism $\overline{\Omega}_X\mathsf{S} \to \overline{\Omega}_X\mathsf{V}$ that respects the choice of generators. Given $w \in \overline{\Omega}_X\mathsf{V}$, denote by $F_{\mathsf{V}}(w)$ the set of all $v \in X^+$ such that $v \geq_{\mathcal{J}} w$ and by $\mathcal{F}_{\mathsf{V}}(w)$ the set of all factors of w. We view $\mathcal{F}_{\mathsf{V}}(w)$ as a quasi-ordered set under the factor ordering $\geq_{\mathcal{J}}$. The following result is a slight improvement of [3, Corollary 2.10].

Proposition 3.2. *Let V be a pseudovariety containing LSI and let $v, w \in \overline{\Omega}_A\mathsf{V}$ be such that $F_{\mathsf{V}}(v) \subseteq F_{\mathsf{V}}(w)$ and w is uniformly recurrent and $v \notin X^+$. Then v and w are $\mathcal{J}$-equivalent.*

We first eliminate the possibility of finding an example for which y is actually a periodic idempotent, that is of the form u^ω for some $u \in X^+$. For this purpose the following observation of combinatorics on words is useful.

Lemma 3.3. *Let u be a primitive word and v be a word such that every factor of v of length at most $2|u|$ is a factor of u^3. Then v is a factor of some power of u.*

Lemma 3.3 implies that all finite factors of w are factors of u^ω. Let $(w_n)_n$ be a sequence in X^+ such that $\lim w_n = w$. We may as well assume that the factors of w_n of length at most $2|u|$ are factors of w and, therefore, by hypothesis, also factors of u^3. By Lemma 3.3, it follows that every w_n is a factor of u^ω. Hence so is the limit w of the sequence $(w_n)_n$.

This implies that the idempotent y must be non-periodic. Indeed, otherwise, $y = u^\omega$ for some primitive word u. Since $\lim y_n = y$, for every sufficiently large n all factors of y_n of length at most $2|u|$ are factors of y, hence factors of u^3. By the above, this implies that y_n is a factor of $y = u^\omega$ for all sufficiently large n. Hence, for such n, the pseudowords y, y_n and y_{n+1} are $\mathcal{J}$-equivalent, which contradicts $y_{n+1} >_{\mathcal{R}} y_n$ since $\overline{\Omega}_X\mathsf{A}$ is stable.

Another way to put our conclusion in the periodic case is to state that the set $\mathcal{F}_{\mathsf{V}}(u^\omega)$ is open in case $u \in X^+$ and $\mathsf{V} \supseteq \mathsf{LI}$. In fact, we actually showed that

$$\mathcal{F}_{\mathsf{V}}(u^\omega) = \bigcap_{|v| \leq 2|u|,\, v \notin F(u^3)} \overline{X^+ \setminus X^* v X^*}$$

where the overline denotes closure in $\overline{\Omega}_X\mathsf{V}$.

More generally, we have Theorem 3.4 below, which is an extension and generalization of [3, Theorem 2.11]. Before stating it, we introduce a more general notion of periodicity. We say that $w \in \overline{\Omega}_X\mathsf{V}$ *has period* u if $u >_{\mathcal{J}} w \geq_{\mathcal{J}} u^\omega$. We say that w is *periodic* if w has some period $u \in X^+$.

Theorem 3.4. *Let V be any pseudovariety containing LSI and let $w \in \overline{\Omega}_X\mathsf{V}$ be uniformly recurrent. Then the following conditions are equivalent:*

(*a*) *w is periodic;*
(*b*) *the language of finite factors of w is rational;*
(*c*) *the set $\mathcal{F}_{\mathsf{V}}(w)$ is open;*
(*d*) *w is not the limit of a sequence of words w_n, none of which is a factor of w;*
(*e*) *there is no infinite strictly ascending $\mathcal{J}$-chain in $\overline{\Omega}_X\mathsf{V}$ converging to w;*
(*f*) *there is no infinite strictly ascending $\mathcal{R}$-chain $(w_n)_n$ in $\overline{\Omega}_X\mathsf{V}$ converging to w such that, for all n, all idempotents $\mathcal{J}$-above w_n are $\mathcal{J}$-equivalent to w.*

We note that, in view of [3, Theorem 2.6], which uses only the consideration of finite factors, and thus applies to any pseudovariety containing LSI, the uniformly recurrent elements of $\overline{\Omega}_X\mathsf{V}$ are precisely the regular pseudowords of rank 1. There exist such pseudowords that are non-periodic provided the alphabet X has at least two letters.

Corollary 3.5. *Let V be a pseudovariety containing LSI and let X be an alphabet with at least two letters. Then every non-periodic regular element of $\overline{\Omega}_X\mathsf{V}$ of rank 1 is the limit of a strictly $\leq_{\mathcal{R}}$-ascending chain of pseudowords of rank 1.* □

It is easy to show that, if the pseudovariety V contains LSI and $w \in \overline{\Omega}_X\mathsf{S}$, then $F_{\mathsf{V}}(p_{\mathsf{V}}(w)) = F_{\mathsf{LSI}}(p_{\mathsf{LSI}}(w))$. From hereon, we will write $F(w)$ or $F(p_{\mathsf{V}}(w))$ for $F_{\mathsf{S}}(w)$. We also denote $\mathcal{F}_{\mathsf{A}}(w)$ simply by $\mathcal{F}(w)$.

In contrast with Corollary 3.5, we have the following result for ω-terms over A.

Theorem 3.6. *If $w \in \Omega^{\kappa}_X\mathsf{A}$ then the following conditions hold:*
(a) for every $v \in \mathcal{F}(w)$, $F(v)$ is a rational language;
(b) $\mathcal{F}(w)$ is wqo.

4. Good pseudowords over A

We say that an infinite pseudoword $w \in \overline{\Omega}_X\mathsf{A}$ is *good* if $F(w)$ is a rational language and $\mathcal{F}(w)$ contains no infinite $\leq_{\mathcal{J}}$-anti-chains.

The following is a crucial result in the sequel.

Proposition 4.1. *Every factor of a good pseudoword is also good.*

We say that $v \in \mathcal{F}(w)$ is a *special ω-factor* of w if $v = u^\omega$ for some Lyndon word u. In this case, the Lyndon word u is called a *special base* of w. A factor v of w is called a *special factor* if it is of the form $u_1^\omega z u_2^\omega$, where $z \in X^+$ and u_1^ω, u_2^ω are special ω-factors and v is not itself a special ω-factor. We say that a prefix v of w is *special* if v is of the form zu^ω, where $z \in X^*$ and u^ω is a special ω-factor. The definition of *special suffix* of w is dual. Good pseudowords cannot have many special factors.

Lemma 4.2. *If w is good, then it has only finitely many special ω-factors, finitely many special factors, and precisely one special prefix and one special suffix.*

We say that a good pseudoword is *periodic at the ends* if its special prefix and suffix are special ω-factors. Up to finite prefix and suffix, every good pseudoword is obtained from one of this form.

Corollary 4.3. *Let $w \in \overline{\Omega}_X\mathsf{A}$ be good. Then there is a factorization $w = xw'y$, where $x, y \in X^*$ and $w' \in \overline{\Omega}_X\mathsf{A}$ is periodic at the ends.* $\square$

The following lemma explains where the finite factors of a good pseudoword are found.

Lemma 4.4. *Let $w \in \overline{\Omega}_X\mathsf{A}$ be good and periodic at the ends. If w has no special factor, then w is its own special ω-factor. Otherwise, every finite factor of w is a factor of some special factor of w.*

In order to understand the structure of good pseudowords, we need to figure out how they are built up from their special factors. For this purpose, we develop a number of technical tools and lemmas which we proceed to present.

Let $w \in \overline{\Omega}_X\mathsf{A}$. We denote by $\Gamma(w)$ the directed multigraph whose vertices are the special ω-factors of w and which has an edge $ezf : e \to f$ if $z \in X^+$ is such that ezf is a special factor of w. By Lemma 4.2, $\Gamma(w)$ is finite if w is good.

By a *semigroupoid* we mean a directed graph with an associative multiplication of edges with matching ends, that is a small category with the requirement for local identities dropped. Given a semigroupoid S, denote by S^c the category that is obtained from S by adding the missing local identities. A semigroup T is viewed as a semigroupoid with T as the set of edges and only one (virtual) vertex. Homomorphisms between semigroupoids are functions that respect the partial operations of edge composition and taking the ends of an edge. A *quotient homomorphism* is a surjective homomorphism whose restriction to vertices in injective. A homomorphism is *faithful* if its restriction to every hom-set is injective.

A *pseudovariety of semigroupoids* is a class of finite semigroupoids containing the one-vertex one-edge semigroupoid that is closed under taking quotient images, faithfully embedded semigroupoids and finite direct products [16, 5]. For a pseudovariety V of semigroups, the smallest pseudovariety of semigroupoids that contains V is denoted gV and the largest pseudovariety of semigroupoids all of whose semigroups belong to V is denoted $\ell\mathsf{V}$.

A *topological semigroupoid* is a semigroupoid endowed with a topology such the semigroupoid operations are continuous. A topological semigroupoid S is said to be *generated* by a finite graph Γ if the subsemigroupoid generated by Γ is dense; if there is such a graph, then we also say that S is *finitely generated*.

We say that a finitely generated compact semigroupoid is *pro*-V for a pseudovariety V of semigroupoids if its edges can be separated by continuous homomorphisms into members of V. For a finite graph Γ, there is always a free pro-V semigroupoid on Γ, denoted $\overline{\Omega}_\Gamma\mathsf{V}$. In analogy with the semigroup case, elements of $\overline{\Omega}_\Gamma\mathsf{V}$ may be called *pseudopaths* over the graph Γ. Given a finite graph Γ with set of edges $E(\Gamma)$ and a pseudovariety V, the unique continuous homomorphism $\gamma_{\mathsf{V}} : \overline{\Omega}_\Gamma\mathsf{gV} \to \overline{\Omega}_{E(\Gamma)}\mathsf{V}$ that sends each edge to itself is faithful [2].

Recall that the pseudovariety of two-testable semigroups, which is generated by the syntactic semigroups of the languages whose membership is characterized by the first letter, the last letter and the two-letter factors is also generated by the real matrix (aperiodic) semigroup

$$A_2 = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}.$$

The two-testability condition is what characterizes the pseudowords on edges of a graph that are determined by pseudopaths. More formally, we have the following lemma.

Lemma 4.5. *Let Γ be a finite graph and V a pseudovariety containing A_2. Suppose that $v \in \overline{\Omega}_\Gamma\mathsf{gV}$ and $u \in \overline{\Omega}_{E(\Gamma)}\mathsf{V}$ are such that u is a factor of $\gamma_{\mathsf{V}}(v)$. Then there exists $u' \in \overline{\Omega}_\Gamma\mathsf{gV}$ such that $\gamma_{\mathsf{V}}(u') = u$ and u' is a factor of v.*

Let V and W be pseudovarieties. We denote by $\mathsf{V}*\mathsf{W}$ the pseudovariety generated by all semidirect products of the form $S * T$ with $S \in \mathsf{V}$ and $T \in \mathsf{W}$. For $n \geq 1$, we denote D_n the pseudovariety consisting of all finite semigroups in which products of length n act as zeros on the right. The pseudovariety K_n is defined dually. Denote by A_k the pseudovariety defined by the pseudoidentity $x^{k+1} = x^k$. Note that $\mathsf{A}_\omega = \mathsf{A} = \mathsf{A} * \mathsf{D}_n$.

Denote by B_n the *de Bruijn graph of order* n, whose set of vertices is X^n and whose set of edges is X^{n+1}, where $axb : ax \to xb$ whenever $a, b \in X$ and $x \in X^{n-1}$. If both vertices and edges are restricted to be factors of a pseudoword $w \in \overline{\Omega}_X\mathsf{V}$ then we obtain a subgraph of B_n, denoted $G_n(w)$, which is also called the *Rauzy graph* of w.

Denote by $X^{\leq n}$ the set of all words $u \in X^+$ such that $|u| \leq n$. We define a continuous mapping $\bar{\Phi}_n^{\mathsf{V}}$ using the following diagram:

$$\begin{tikzcd}
\overline{\Omega}_X(\mathsf{V} * \mathsf{D}_n) \setminus X^{\leq n} \arrow[rr, "\bar{\Phi}_n^{\mathsf{V}}"] \arrow[dr, "\iota_{\mathsf{V}}"'] & & E(\overline{\Omega}_{B_n}\mathsf{gV}) \arrow[dl, "\eta_{\mathsf{V}}"] \\
& S_n &
\end{tikzcd}$$

where

- S_n is the subsemigroup of the semigroup $M_n(\overline{\Omega}_{X^{n+1}}\mathsf{V}, \Phi_n^{\mathsf{V}})$ of [1, Section 10.6], whose universe is $X^n \times \overline{\Omega}_{X^{n+1}}\mathsf{V} \times X^n$ and whose operation is given by

$$(u_1, w_1, v_1)\,(u_2, w_2, v_2) = (u_1, w_1\Phi_n^{\mathsf{V}}(v_1u_2)w_2, v_2),$$

  where, for a word t of length at least $n+1$, $\Phi_n^{\mathsf{V}}(t)$ is the value in $\overline{\Omega}_{X^{n+1}}\mathsf{V}$ of the word over the alphabet X^{n+1} which reads the successive factors of length $n+1$ of t;

- the arrow η_{V} is the continuous mapping that sends each edge $u : \alpha(u) \to \omega(u)$ of the relatively free semigroupoid $\overline{\Omega}_{B_n}\mathsf{gV}$ to the triple $(\alpha(u), \gamma_{\mathsf{V}}(u), \omega(u))$;
- the arrow ι_{V} is the continuous homomorphism given by [1, Theorem 10.6.12].

Since η_{V} is continuous and the image of the subsemigroupoid generated by B_n is dense in $\operatorname{Im}\iota_{\mathsf{V}}$, we have $\operatorname{Im}\eta_{\mathsf{V}} = \operatorname{Im}\iota_{\mathsf{V}}$. On the other hand, since γ_{V} is faithful, η_{V} is injective. Hence η_{V} is a homeomorphism of $E(\overline{\Omega}_{B_n}\mathsf{gV})$ with $\operatorname{Im}\iota_{\mathsf{V}}$, and we may define the continuous mapping $\bar{\Phi}_n^{\mathsf{V}}$ to be the composite $\eta_{\mathsf{V}}^{-1} \circ \iota_{\mathsf{V}}$. It is therefore just a reinterpretation of the mapping ι_{V}. Note that each finite word w is mapped by $\bar{\Phi}_n^{\mathsf{V}}$ to the path which starts at the prefix $i_n(w)$ of length n, ends at the suffix $t_n(w)$ of length n, and goes through the edges given by the successive factors of length $n+1$ of w.

In case $w \in \overline{\Omega}_X\mathsf{A}$ is good, we consider the unique continuous homomorphism $\lambda_w : \overline{\Omega}_{\Gamma(w)}\mathsf{gA} \to \overline{\Omega}_X\mathsf{A}$ that sends each edge $ezf : e \to f$ to the pseudoword ezf. The following lemma provides a convenient way of approximating good pseudowords by finite words.

Lemma 4.6. *Let $w \in \overline{\Omega}_X\mathsf{A}$ be good and periodic at the ends, and let n be a multiple of the lengths of all special bases of w. Let k be a positive integer or ω. Then there is a unique mapping τ_k such that the following diagram commutes:*

$$\begin{tikzcd}
\overline{\Omega}_{\Gamma(w)}\mathsf{gA} \arrow[rr, "\lambda_w"] \arrow[d, "\tau_k"'] & & \overline{\Omega}_X\mathsf{A} \setminus X^{\leq n} \arrow[r, hook] \arrow[d, "p_k|\ldots"] & \overline{\Omega}_X\mathsf{A} \arrow[d, "p_k"] \\
\overline{\Omega}_{G_n(w)}\mathsf{gA}_k \arrow[r, hook] & \overline{\Omega}_{B_n}\mathsf{gA}_k & \overline{\Omega}_X(\mathsf{A}_k * \mathsf{D}_n) \setminus X^{\leq n} \arrow[l, "\bar{\Phi}_n^{\mathsf{A}_k}"] \arrow[r, hook] & \overline{\Omega}_X(\mathsf{A}_k * \mathsf{D}_n)
\end{tikzcd}$$

where p_k is the natural continuous homomorphism. Moreover the mapping τ_k is a continuous homomorphism.

As a consequence of Lemma 4.2, we may associate to each good pseudoword w a positive integer $s(w) = m'm''$ where m' is a multiple of the length of each special base of w and $m'' > |xyz|$ for every special factor $x^\omega yz^\omega$ of w in normal form. This parameter is useful to locate more precisely where sufficiently long finite factors of w may occur, which allows us to lift approximations.

Lemma 4.7. *Let $w \in \overline{\Omega}_X\mathsf{A}$ be good and periodic at the ends, let $n = s(w)$, and let k be a positive integer. Let w_k be a finite word that has the same factors, prefixes and suffixes of length $(k+1)n$ as w does. Then there is some path w_k' in $\Gamma(w)$, starting and ending respectively at the special prefix and special suffix of w such that $\tau_k(w_k') = \bar{\Phi}_n^{\mathsf{A}_k}(w_k)$.*

This leads to the following encoding of good pseudowords in terms of "pseudopaths" in their special factors.

Proposition 4.8. *If $w \in \overline{\Omega}_X\mathsf{A}$ is good and has at least one special factor, then there are some edge $w' \in \overline{\Omega}_{\Gamma(w)}\mathsf{gA}$ and some words $x, y \in X^+$ such that $w = x\lambda_w(w')y$.*

To proceed, we need a rather technical but unsurprising lemma, which requires some further notation to state.

We denote by π_k the natural continuous homomorphism $\overline{\Omega}_X\mathsf{A} \to \overline{\Omega}_X\mathsf{A}_k$. Note that $\Omega_X^\kappa\mathsf{A}_k = \Omega_X\mathsf{A}_k$ is the subsemigroup of $\overline{\Omega}_X\mathsf{A}_k$ generated by X. For $u, v \in \Omega_X^\kappa\mathsf{A}$, we write $u \sim_{w,k} v$ if it is possible to transform u into v by changing the exponents of factors of the form z^p, keeping them at least k, where z is a primitive word such that z^ω is a factor of w. Note that $\sim_{w,k}$ is a congruence on $\Omega_X^\kappa\mathsf{A}$ such that $u \sim_{w,k} v$ implies $\pi_k(u) = \pi_k(v)$ but the converse is false.

Noting that A is local, that is $\mathsf{gA} = \ell\mathsf{A}$ (cf. [16]), there is a natural continuous homomorphism, $\varpi_k : \overline{\Omega}_{\Gamma(w)}\mathsf{gA} \to \overline{\Omega}_{\Gamma(w)}\ell\mathsf{A}_k$. For a finite semigroup S, denote by $\operatorname{ind}(S)$ the smallest $k \geq 0$ such that there exists some $\ell \geq 1$ such that S satisfies the identity $x^{k+\ell} = x^k$.

Lemma 4.9. *Let $w \in \overline{\Omega}_X\mathsf{A}$ be good and let $\varphi : \overline{\Omega}_{\Gamma(w)}\mathsf{gA} \to S$ be a continuous homomorphism onto some $S \in \mathsf{A}$ such that the natural continuous homomorphisms $\overline{\Omega}_{\Gamma(w)}\mathsf{gA} \to \overline{\Omega}_{E(\Gamma(w))}\mathsf{K}_1$ and $\overline{\Omega}_{\Gamma(w)}\mathsf{gA} \to \overline{\Omega}_{E(\Gamma(w))}\mathsf{D}_1$ factor through φ. Let $n = 3s(w)$ and let $k = \max\{\mathrm{ind}(S), n\}$. Denote by μ_w the restriction of λ_w to $\Omega_{\Gamma(w)}\mathsf{gA}$, the subsemigroupoid of $\overline{\Omega}_{\Gamma(w)}\mathsf{gA}$ generated by the graph $\Gamma(w)$. Let m be the supremum of the lengths of the special bases of w. Then the following properties hold:*

- (a) *μ_w is injective on edges;*
- (b) *if $u \in \Omega^\kappa_{\Gamma(w)}\mathsf{gA}$ is an edge of rank at most 1 and $v \in \Omega^\kappa_X\mathsf{A}$ is such that $\pi_k\big(\lambda_w(u)\big) = \pi_k(v)$, then there exists an edge $u' \in \Omega^\kappa_{\Gamma(w)}\mathsf{gA}$ of rank at most 1 such that $\lambda_w(u') \sim_{w,k} v$ and $\varpi_k(u) = \varpi_k(u')$;*
- (c) *$\ker(\pi_k \circ \mu_w) \subseteq \ker\varphi$ so that φ induces a homomorphism of partial semigroups $\psi : \pi_k(\mathrm{Im}\,\mu_w) \to S$ such that $\psi \circ \pi_k \circ \mu_w = \varphi|_{\Omega_{\Gamma(w)}\mathsf{gA}}$;*
- (d) *let θ denote the congruence on $\Omega_X\mathsf{A}_k$ generated by the binary relation $\ker\psi$; then $\pi_k(\mathrm{Im}\,\mu_w)$ is a union of θ-classes;*
- (e) *the restriction of θ to $\pi_k(\mathrm{Im}\,\mu_w)$ coincides with $\ker\psi$;*
- (f) *denote by $\hat\psi$ the natural homomorphism $\Omega_X\mathsf{A}_k \to \Omega_X\mathsf{A}_k/\theta$; then, whenever $u \in \Omega_X\mathsf{A}_k$, $\hat\psi(u)$ is a factor of some element of $\hat\psi \circ \pi_k(\mathrm{Im}\,\mu_w)$ if and only if u is a factor of some element of $\pi_k(\mathrm{Im}\,\mu_w)$;*
- (g) *if $u \in \Omega^\kappa_A\mathsf{A}$ and $x, y \in X^*$ are such that $\pi_k(xuy) \in \pi_k(\mathrm{Im}\,\mu_w)$ and $\pi_k(x'uy') \notin \pi_k(\mathrm{Im}\,\mu_w)$ whenever x' is a suffix of x and y' is a prefix of y such that $|x'y'| < |xy|$, then $(k+1)m > \max\{|x|, |y|\}$; in particular, there are only finitely many such pairs (x, y).*

The following diagram may help to keep track of the mappings involved in Lemma 4.9.

$$\begin{tikzcd}
\overline{\Omega}_{\Gamma(w)}\mathsf{gA} \arrow[rrrrr, "\lambda_w"] \arrow[d, "\varpi_k"'] \arrow[ddrr, bend right, "\varphi"] & & & & & \overline{\Omega}_A\mathsf{A} \arrow[d, "\pi_k"] \\
\overline{\Omega}_{\Gamma(w)}\ell\mathsf{A}_k \arrow[ddrr, dashed] & \Omega^\kappa_{\Gamma(w)}\mathsf{gA} \arrow[ul, hook] \arrow[rrr] & & & \Omega^\kappa_A\mathsf{A} \arrow[ur, hook] \arrow[d] & \overline{\Omega}_A\mathsf{A}_k \\
& & \Omega_{\Gamma(w)}\mathsf{gA} \arrow[ul, hook] \arrow[r, "\mu_w"] \arrow[d] & \mathrm{Im}\,\mu_w \arrow[ur, hook] \arrow[d] & \Omega_A\mathsf{A}_k \arrow[ur, hook] \arrow[d, "\hat\psi"] & \\
& & S & \pi_k(\mathrm{Im}\,\mu_w) \arrow[l, dashed, "\psi"'] \arrow[ur, hook] & \Omega_A\mathsf{A}_k/\theta &
\end{tikzcd}$$

Lemma 4.9 provides the technicalities required to prove the following rather natural result but for which we do not know of any simpler proof.

Lemma 4.10. *Let $w \in \overline{\Omega}_X\mathsf{A}$ be good. Then the homomorphism λ_w is injective on edges of $\overline{\Omega}_{\Gamma(w)}\mathsf{gA}$ and preserves $<_{\mathcal{J}}$.*

For $w \in \overline{\Omega}_X\mathsf{A}$, let $P(w)$ be the set of all normal forms, viewed as well-parenthesized words $u \in (X \cup \{(,)\})^+$, that are factors of w. Note that, since $F(w) = P(w) \cap X^+$, if $P(w)$ is rational then so is $F(w)$.

Putting together the main conclusions of the technical lemmas, we can prove the following result.

Proposition 4.11. *Let $w \in \overline{\Omega}_X\mathsf{A}$ be good and periodic at the ends. Then there is a unique $w' \in \overline{\Omega}_{\Gamma(w)}\mathsf{gA}$ such that $\lambda_w(w') = w$. Moreover, w' enjoys the following properties:*

- (a) *if $P(w)$ is a rational language then so is $P(w')$;*
- (b) *$\mathcal{F}(w')$ has no infinite $\leq_{\mathcal{J}}$-anti-chains;*

(c) $r(w') \leq r(w)$;
(d) *if* $r(w)$ *is finite then* $r(w') = r(w) - 1$;
(e) *if* $r(w)$ *is infinite then so is* $r(w')$.

In particular, w' *is good.*

Iterating the application of Proposition 4.11, we obtain the following characterization of the elements of $\overline{\Omega}_X \mathsf{A}$ that are given by ω-terms in terms of properties of their sets of factors and which constitutes the main result presented in this paper. The guarantee of termination of the encoding routine comes from the observation that if a subset of the Dyck language (of well-parenthesized words in the alphabet $\{(,)\}$) is rational then there is a bound on the maximum number of nested parentheses of its members.

Theorem 4.12. *Let* $w \in \overline{\Omega}_X \mathsf{A}$ *be an arbitrary pseudoword. Then* $w \in \Omega^{\kappa}_X \mathsf{A}$ *if and only if* $\mathcal{F}(w)$ *contains no infinite* $\leq_{\mathcal{J}}$*-anti-chains and* $P(w)$ *is a rational language.*

Acknowledgements. This work was partly supported by the Pessoa French-Portuguese 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*, which is partly funded by the European Community Fund FEDER. The work of the second author was supported, in part, by FCT through the *Centro de Matemática da Universidade do Minho*. The work of the first two authors was also supported by the project PTDC/MAT/65481/2006.

References

[1] J. Almeida, *Finite semigroups and universal algebra*, World Scientific, Singapore, 1995, English translation.
[2] ———, *A syntactical proof of locality of DA*, Int. J. Algebra Comput. **6** (1996), 165–177.
[3] ———, *Profinite groups associated with weakly primitive substitutions*, Fundamentalnaya i Prikladnaya Matematika (Fundamental and Applied Mathematics) **11** (2005), no. 3, 13–48, In Russian. English version in J. Math. Sciences **144**, No. 2 (2007) 3881–3903.
[4] ———, *Profinite semigroups and applications*, Structural Theory of Automata, Semigroups, and Universal Algebra (New York) (Valery B. Kudryavtsev and Ivo G. Rosenberg, eds.), NATO Science Series II: Mathematics, Physics and Chemistry, vol. 207, Springer, 2005, 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 and A. Costa, *Infinite-vertex free profinite semigroupoids and symbolic dynamics*, Tech. Report CMUP 2007-20, Univ. Porto, 2007.
[6] J. Almeida, J. C. Costa, and M. Zeitoun, *Complete reducibility of systems of equations with respect to R*, Portugal. Math. **64** (2007), 445–508.
[7] J. Almeida and B. Steinberg, *On the decidability of iterated semidirect products and applications to complexity*, Proc. London Math. Soc. **80** (2000), 50–74.
[8] C. J. Ash, *Inevitable graphs: a proof of the type II conjecture and some related decision procedures*, Int. J. Algebra Comput. **1** (1991), 127–146.
[9] J.-C. Birget, *Iteration of expansions—unambiguous semigroups*, J. Pure Appl. Algebra **34** (1984), 1–55.
[10] S. Eilenberg, *Automata, languages and machines*, vol. B, Academic Press, New York, 1976.
[11] K. Henckell, S. Margolis, J.-E. Pin, and J. Rhodes, *Ash's type II theorem, profinite topology and Malcev products. Part I*, Int. J. Algebra Comput. **1** (1991), 411–436.
[12] 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** (1991), 1–32.
[13] ———, *Normal forms for free aperiodic semigroups*, Int. J. Algebra Comput. **11** (2001), 581–625.
[14] J.-E. Pin, *Varieties of formal languages*, Plenum, London, 1986, English translation.
[15] J. Rhodes and B. Steinberg, *The q-theory of finite semigroups*, 2001–2007, Book under preparation. Preliminary versions available through `http://mathstat.math.carleton.ca/~bsteinbg/qtheor.html`.
[16] B. Tilson, *Categories as algebra: an essential ingredient in the theory of monoids*, J. Pure Appl. Algebra **48** (1987), 83–198.