Journal of Mathematical Sciences, Vol. 144, No. 2, 2007

PROFINITE GROUPS ASSOCIATED WITH WEAKLY PRIMITIVE SUBSTITUTIONS

J. Almeida

UDC 512.53

Abstract. A uniformly recurrent pseudoword is an element of a free profinite semigroup in which every finite factor appears in every sufficiently long finite factor. An alternative characterization is as a pseudoword that is a factor of all its infinite factors, i.e., one that lies in a $\mathcal{J}$-class with only finite words strictly $\mathcal{J}$-above it. Such a $\mathcal{J}$-class is regular, and therefore it has an associated profinite group, namely, any of its maximal subgroups. One way to produce such $\mathcal{J}$-classes is to iterate finite weakly primitive substitutions. This paper is a contribution to the computation of the profinite group associated with the $\mathcal{J}$-class that is generated by the infinite iteration of a finite weakly primitive substitution. The main result implies that the group is a free profinite group provided the substitution induced on the free group on the letters that appear in the images of all of its sufficiently long iterates is invertible.

1. Introduction

The theory of profinite semigroups, particularly those that are free relative to a pseudovariety, has been given considerable attention since the mid-1980s. The main interest in the theory stems from its connections with language and automata theory via Eilenberg's correspondence between certain classes of rational languages and pseudovarieties of semigroups. In particular, structural knowledge about a relatively free profinite semigroup can often lead to important applications [2, 4, 5, 10, 11, 23]. Yet, very little is known to date about the structure of absolutely free profinite semigroups, whose elements are called *pseudowords*. This paper further develops a connection between the structure of finitely generated free profinite semigroups with symbolic dynamics that first emerged in [3] and that has also been used in [10].

One of the key ideas in the symbolic dynamics approach is to produce pseudowords by iterating continuous endomorphisms of the free profinite semigroup. The setting for this iteration is established by a result that states that if a profinite semigroup is finitely generated, then its monoid of continuous endomorphisms is itself a profinite monoid under the pointwise convergence topology [5]. Thus, there is a natural infinite iterate of a continuous endomorphism, namely, the unique one that is idempotent, also known as its *ω-power*. This technique has been used in [3] to show that the pseudovariety of finite p-groups is tame (see [4, 8, 9] for the significance of this property) and in [10] to construct group-generic sets of pseudowords, which can then be used to study pseudovarieties characterized by the fact that the subgroups of its members lie in a given pseudovariety of groups.

The specific idea adopted here is to iterate finite substitutions to produce uniform recurrence phenomena. This paper is a contribution to the understanding of the $\mathcal{J}$-classes of uniformly recurrent pseudowords. We show in Sec. 2 that the $\mathcal{J}$-classes that consist entirely of uniformly recurrent pseudowords are precisely those that only have finite words strictly $\mathcal{J}$-above them and, therefore, they are regular. In a further connection with symbolic dynamics that is not used or explored here but can be found in [6, 7], it can be shown that there is a one-to-one correspondence with minimal symbolic dynamical systems, namely, through the language of finite factors.

In particular, to each uniformly recurrent pseudoword we associate the profinite group that is obtained by taking any of the maximal subgroups in the $\mathcal{J}$-class of the pseudoword (it is well known that they are all isomorphic). Sections 3–5 develop methods that, at least in a wide class of cases, allow us to compute such profinite groups. The main result is Theorem 5.3, whose statement is too technical to be reproduced

Translated from Fundamentalnaya i Prikladnaya Matematika, Vol. 11, No. 3, pp. 13–48, 2005.

1072–3374/07/1442–3881 © 2007 Springer Science+Business Media, Inc.

here. The wide class of applications is given by Corollary 5.7, which establishes that if the finite weakly primitive substitution acts as an automorphism on the free group with generating set consisting of the letters that appear in the image of the infinite iterates of the substitution, then the associated profinite group is a finitely generated, free profinite group. Moreover, Theorem 5.3 provides techniques to exhibit free generators of this profinite group.

Section 6 shows that, within our wide class of uniformly recurrent $\mathcal{J}$-classes, one can find the Sturmian and Arnoux–Rauzy $\mathcal{J}$-classes, corresponding respectively to Sturmian and Arnoux–Rauzy symbolic dynamical systems generated by substitutions (see [14,18] for further information on such systems). More specifically, here the groups in question are free profinite groups on n generators if the system involves precisely n letters. The extension of this result to systems that are not necessarily generated by substitutions has been sketched in [7].

Finally, Sec. 7 presents a few examples, including a maximal subgroup of a uniformly recurrent $\mathcal{J}$-class that is not a free profinite group.

The combinatorial tools that play a crucial role in the paper include elementary combinatorial group theory along with some topics from the algebraic theory of codes, such as circular codes and codes with bounded delay. All of the required results are established in the paper, which should make it self-contained in this respect. Some important coding theorems from [15,19] are also used.

Preliminary versions of the main results of this paper have been announced at various meetings and seminars. They have also been announced or sketched, without proof, in [6,7].

2. Uniform Recurrence as an Algebraic Property

Throughout this paper, A denotes a finite set, which is called an *alphabet*. The free monoid on A is denoted by A^*.

Recall that a *profinite monoid* is a compact zero-dimensional monoid. Equivalently, a profinite monoid is a projective limit of finite monoids, which are viewed as topological monoids under the discrete topology. Let $\widehat{A^*}$ denote the *free profinite monoid*, which is obtained by profinite completion of the free monoid A^* [5]. The elements of $\widehat{A^*}$ will be called *pseudowords* in this paper. While the elements of A^* are called *finite words*, the elements of $\widehat{A^*} \setminus A^*$ are said to be *infinite pseudowords*. We also consider the subsemigroup $\widehat{A^+}$, which is obtained from $\widehat{A^*}$ by dropping the neutral element, namely the empty word, and is the free profinite semigroup on A.

The reader may wish to consult [2, 4, 5, 11, 17, 23] for general background on finite and profinite semigroups.

An example of an elementary but useful observation that is an immediate consequence of compactness is the following lemma.

Lemma 2.1. *The set of factors of an element of a compact monoid is closed.*

We may use freely the fact that the closure $\overline{L}$ in $\widehat{A^*}$ of a rational language $L \subseteq A^*$ is a clopen subset, which, moreover, satisfies $\overline{L} \cap A^* = L$ (see [2, Sec. 3.6] or [5]). Of particular relevance are the sets of the form $\widehat{A^*}\,u\widehat{A^*} = \overline{A^*\,uA^*}$, which are therefore clopen for every $u \in A^*$. In other words, all but finitely many elements of a convergent sequence of finite words have a given finite word as a factor if and only if the limit does. Similarly, considering the languages of the form $u\widehat{A^*} = \overline{uA^*}$, with $u \in A^+$, we conclude that every infinite pseudoword has a well-defined finite prefix (and, dually, a suffix) for each finite length. In fact, canceling such a prefix, the remainder is also uniquely determined [2], but we will not be using this result here. Moreover, every clopen subset of $\widehat{A^*}$ is the closure of a rational language, so that the topology of the zero-dimensional space $\widehat{A^*}$ closely reflects the combinatorics of rational languages, but again this property will not be so relevant here.

For a pseudoword $w \in \widehat{A^*}$, denote by $F(w)$ the set of all finite factors of w and by $F_n(w)$ the set of all factors of w of length n.

Let w be an infinite pseudoword. We say that w is *recurrent* if, for every $u \in F(w)$, there is a word v such that $uvu \in F(w)$. We say that w is *uniformly recurrent* if, for every $u \in F(w)$, there is a positive integer N such that every $v \in F_N(w)$ admits u as a factor. Note that every uniformly recurrent pseudoword is recurrent and so is every member of the minimal ideal I_A of $\widehat{A^*}$. On the other hand, it follows from the following lemma that, if $|A| > 1$, then no element of I_A is uniformly recurrent.

There is an alternative definition of recurrent pseudoword that may be imported from symbolic dynamics: we could call an infinite pseudoword recurrent if all of its infinite factors have the same finite factors. This alternative notion turns out to be equivalent to uniform recurrence, as the following result shows.

Lemma 2.2. *An infinite pseudoword is uniformly recurrent if and only if all of its infinite factors have the same finite factors.*

Proof. Let w be an infinite pseudoword and suppose first that it is uniformly recurrent. Let u be an infinite factor of w. Then, of course, every finite factor of u is also a factor of w. Conversely, since u is infinite, it has arbitrarily long finite prefixes and so it has arbitrarily long finite factors that are also factors of w. Since w is uniformly recurrent, every finite factor of w is a factor of any sufficiently long finite factor and therefore it is also a factor of u. This shows that u has the same finite factors as w does.

Suppose next that all infinite factors of w have the same finite factors. Let v be a finite factor of w. Arguing by contradiction, suppose that there are arbitrarily long factors of w that do not admit v as a factor. Then there is a sequence of factors of w that do not admit v as a factor converging to an infinite pseudoword u, which is itself also a factor of w by Lemma 2.1. It follows that u also does not admit v as a factor, in contradiction with the assumption. Hence w is uniformly recurrent. □

The aim of this section is to characterize uniform recurrence in algebraic terms. For a pseudoword w, denote by $X(w)$ the set of all infinite pseudowords that are limits of sequences of finite factors of w, i.e., $X(w) = \overline{F(w)} \setminus A^*$.

Recall that, for two elements s and t of a semigroup S, the element s is said to *lie $\mathcal{J}$-above t* and we write $s \geq_{\mathcal{J}} t$ if s is a factor in some factorization of t. We further say that s and t are *$\mathcal{J}$-equivalent* if each of them is a factor of the other. We shall write $s >_{\mathcal{J}} t$ if $s \geq_{\mathcal{J}} t$ but s and t are not $\mathcal{J}$-equivalent. An element s of a semigroup S is said to be *regular* if $sxs = s$ for some $x \in S$. It is well known that in a compact semigroup, a $\mathcal{J}$-class consists entirely of regular elements if and only if one of its elements is regular, if and only if it contains an idempotent. The equivalence relation $\mathcal{J}$ is one of Green's relations on a semigroup S.

Replacing "factor" by "factor on the left" (or *prefix*), we obtain the quasi-orders $\geq_{\mathcal{R}}$ and $>_{\mathcal{R}}$ and the equivalence relation $\mathcal{R}$. Dually, replacing "factor" by "factor on the right" (or *suffix*), we obtain the quasi-orders $\geq_{\mathcal{L}}$ and $>_{\mathcal{L}}$ and the equivalence relation $\mathcal{L}$. The intersection of $\mathcal{R}$ and $\mathcal{L}$ is denoted $\mathcal{H}$. In general, $\mathcal{J}$ is not the smallest equivalence relation containing both $\mathcal{R}$ and $\mathcal{L}$, which is denoted $\mathcal{D}$, but it is so in every compact semigroup.

Every subgroup of a semigroup S (meaning a subsemigroup that is a group) is contained in a $\mathcal{J}$-class or, more precisely, in an $\mathcal{H}$-class. The $\mathcal{H}$-classes that are subgroups (and therefore maximal subgroups) are precisely the $\mathcal{H}$-classes that contain a (unique) idempotent. If S is compact, then all the maximal subgroups contained in the same $\mathcal{J}$-class are isomorphic as topological groups. Compact semigroups also satisfy the following stability condition: if $x \leq_{\mathcal{R}} y$ and $x \mathcal{J} y$, then $x \mathcal{R} y$, and dually for $\mathcal{L}$.

Lemma 2.3. *Let w be a uniformly recurrent pseudoword over a finite alphabet A.*

(1) *Every element of $X(w)$ is a factor of w.*
(2) *All elements of $X(w)$ lie in the same $\mathcal{J}$-class of $\widehat{A^*}$.*
(3) *Every element of $X(w)$ is regular.*

Proof.

(1) This is an immediate consequence of Lemma 2.1.

(2) Suppose that $u, v \in X(w)$. By Lemma 2.2, u and v have the same finite factors. Hence, by (1), u and v are factors of each other, i.e., they are $\mathcal{J}$-equivalent.

(3) Suppose that u is an infinite pseudoword that is the limit of a sequence $(u_n)_n$ of finite factors of w. Since w is recurrent and the u_n are finite factors of w, there are finite words v_n such that $u_n v_n u_n$ is a factor of w. If v is an accumulation point of the sequence $(v_n)_n$, then the infinite pseudoword uvu belongs to $X(w)$ and therefore, by (2), uvu is $\mathcal{J}$-equivalent to u. In a compact semigroup, this implies that u is regular. □

For a uniformly recurrent pseudoword w, denote by $J(w)$ the unique $\mathcal{J}$-class containing $X(w)$.

Lemma 2.4. *Let w be a uniformly recurrent pseudoword. Then every $\mathcal{H}$-class contained in $J(w)$ contains some element of $X(w)$.*

Proof. Let $u \in J(w)$. Let x_n and y_n denote respectively the prefix and the suffix of u of length n. Since u is uniformly recurrent by Lemma 2.2, y_n can be found as a factor of u within bounded distance from the left and there is a factor t_n of u, of length at least $2n$, such that x_n is a prefix of t_n and y_n is a suffix of t_n. Let $(n_k)_k$ be a strictly increasing sequence such that the sequences $(x_{n_k})_k$, $(y_{n_k})_k$, and $(t_{n_k})_k$ converge and let x, y, and t be their respective limits. By Lemma 2.3(2), we know that $x, y, t \in J(w)$. Since $x \geq_{\mathcal{R}} z \leq_{\mathcal{L}} y$ for $z \in \{u, t\}$, by stability it follows that u and t lie in the same $\mathcal{H}$-class. □

Lemma 2.5. *Let v be a uniformly recurrent pseudoword and suppose that a is a letter such that va is still uniformly recurrent. Then v and va are $\mathcal{R}$-equivalent.*

Proof. Let v_n be the suffix of v of length n. Since v is an infinite factor of the uniformly recurrent pseudoword va, by Lemma 2.2 they have the same finite factors. Hence, for every n there exists some m_n such that there is a factorization $v_{m_n} = x_n v_n a y_n$ for some words x_n and y_n. By compactness, there exists some strictly increasing sequence of indices $(n_k)_k$ such that each of the sequences $(v_{n_k})_k$, $(x_{n_k})_k$, and $(y_{n_k})_k$ converges, respectively, say, to v', x, and y. Then, by the continuity of multiplication in $\widehat{A^*}$, the sequence $(v_{m_{n_k}})_k$ converges to $xv'ay$. Since it is well known and easy to verify that the limits of two convergent sequences of suffixes of increasing length of the same pseudoword are $\mathcal{L}$-equivalent, it follows that $v' \leq_{\mathcal{J}} v'a$, and so $v' \mathcal{R} v'a$. Moreover, since v' is the limit of a sequence of suffixes of v, there is some factorization of the form $v = zv'$. Since $\mathcal{R}$-equivalence is a left congruence, we finally conclude that $va = zv'a \mathcal{R} zv' = v$. □

We are now ready for the main result of this section.

Theorem 2.6. *Let w be an infinite pseudoword over a finite alphabet. Then w is uniformly recurrent if and only if w is $\mathcal{J}$-maximal as an infinite pseudoword.*

Proof. Suppose first that $w \in \widehat{A^*}$ is uniformly recurrent and let $u \in J(w)$. By Lemma 2.3(1), $u \geq_{\mathcal{J}} w$, say $w = puq$, with $p, q \in \widehat{A^*}$. We claim that $u \mathcal{J} w$. Indeed, otherwise, by [2, Corollary 5.6.2(b)], there is a continuous homomorphism $\varphi\colon \widehat{A^*} \to M$ onto a finite monoid such that $\varphi(u) >_{\mathcal{J}} \varphi(w)$. We will show that this leads to a contradiction.

Let $(p_n)_n$ and $(q_n)_n$ be sequences of finite words converging to p and q, respectively, such that $\varphi(p_n) = \varphi(p)$ and $\varphi(q_n) = \varphi(q)$ for all n. For each n, considering p_n and q_n as being factorized into their letter factors, we may view $\varphi(w)$ as being obtained from $\varphi(u)$ by successively multiplying, on the left and then on the right, by the image under φ of those letters. Since $\varphi(u) >_{\mathcal{J}} \varphi(w)$, at some point in this sequence of multiplications there is a first step in which we leave the $\mathcal{J}$-class of $\varphi(u)$. In other words, either there is a factorization of the form $p_n = x_n a_n y_n$, where $a_n \in A$, such that $\varphi(u) \mathcal{L} \varphi(y_n u) >_{\mathcal{L}} \varphi(a_n y_n u) \geq_{\mathcal{J}} \varphi(w)$ or $\varphi(pu) = \varphi(p_n u) \mathcal{L} \varphi(u)$, in which case there is a factorization of the form $q_n = z_n b_n t_n$, where $b_n \in A$, such that $\varphi(pu) \mathcal{R} \varphi(puz_n) >_{\mathcal{R}} \varphi(puz_n b_n) \geq_{\mathcal{J}} \varphi(w)$. Since the alphabet is finite and $\widehat{A^*}$ is compact, one may extract subsequences such that the relevant letter sequences are constant, the factor sequences converge, and the same alternative holds for every n. Hence, there is

either some factorization of the form $p = xay$, with $a \in A$, such that $\varphi(u)\ \mathcal{L}\ \varphi(yu) >_{\mathcal{L}} \varphi(ayu) \geq_{\mathcal{J}} \varphi(w)$ or some factorization of the form $q = zbt$, with $b \in A$, such that $\varphi(pu)\ \mathcal{R}\ \varphi(puz) >_{\mathcal{R}} \varphi(puzb) \geq_{\mathcal{J}} \varphi(w)$. The two cases are essentially dual; so we consider only the second one. Since $\varphi(puz) >_{\mathcal{R}} \varphi(puzb)$, we cannot have $puz\ \mathcal{R}\ puzb$. On the other hand, both puz and $puzb$ are infinite factors of w and, therefore, by Lemma 2.2 they are both uniformly recurrent. By Lemma 2.5, we obtain $puz\ \mathcal{R}\ puzb$, which is a contradiction. This proves the claim.

Now, given an infinite factor v of w, let u be the infinite limit of a sequence of finite prefixes of v. Then $u \in X(w) \subseteq J(w)$ and, by the prefix version of Lemma 2.1, we obtain $u \geq_{\mathcal{R}} v \geq_{\mathcal{J}} w$. By the claim, $u\ \mathcal{J}\ w$, which implies that $v\ \mathcal{J}\ w$. Hence w is $\mathcal{J}$-equivalent to all of its infinite factors, i.e., w is $\mathcal{J}$-maximal as an infinite pseudoword.

Conversely, suppose that w is $\mathcal{J}$-maximal as an infinite pseudoword. If v is an infinite factor of w, then, by $\mathcal{J}$-maximality of w, v is $\mathcal{J}$-equivalent to w. Hence, v and w have the same factors and, in particular, they have the same finite factors. By Lemma 2.2, it follows that w is uniformly recurrent. □

Theorem 2.6 has a number of important consequences, which we proceed to state. The proofs are now all straightforward. The first corollary could also be derived directly from the definition of a uniformly recurrent pseudoword.

Corollary 2.7. *The $\mathcal{J}$-classes of uniformly recurrent pseudowords consist entirely of uniformly recurrent pseudowords.* □

Corollary 2.8. *The $\mathcal{J}$-class of a uniformly recurrent pseudoword w is completely determined by the finite factors of w as well as by the finite prefixes (respectively, suffixes) of w.* □

Corollary 2.9. *Every uniformly recurrent pseudoword is $\mathcal{H}$-equivalent to the limit of a sequence of its finite factors.* □

Corollary 2.10. *If u and v are two uniformly recurrent pseudowords and every finite factor of u is also a factor of v, then u and v are $\mathcal{J}$-equivalent.* □

We say that an infinite pseudoword is *periodic* if it is $\mathcal{J}$-equivalent to some pseudoword of the form u^ω for some $u \in A^+$. To end this section, we present a characterization of periodicity for uniformly recurrent pseudowords in terms of the combinatorial and topological properties of its set of factors. The basic idea in the proof is the pumping lemma of automata theory, as in the proof of [14, Corollary 6.1.11].

Theorem 2.11. *Let $w \in \widehat{A^*}$ be a uniformly recurrent pseudoword. Then the following conditions are equivalent*:

(1) *w is periodic*;
(2) *the language $F(w)$ of finite factors of w is rational*;
(3) *the set of all factors of w is clopen in $\widehat{A^*}$.*

Proof. Let F denote the set of all the factors of w in $\widehat{A^*}$.

(1) $\implies$ (3) Suppose that w is $\mathcal{J}$-equivalent to u^ω, where $u \in A^+$. Then the finite factors of w are the words that are factors of some power of u. In particular, there are at most n factors of w of length n. By [10, Theorem 6.3], the factors of w are the factors of u together with all words of the form $xu^\nu y$, where x is a suffix of u, y is a prefix of u, a^ν denotes an arbitrary element of the monogenic free profinite monoid $\widehat{\{a\}^*}$, and $u^\nu = \psi(a^\nu)$ for the unique continuous homomorphism $\psi\colon \widehat{\{a\}^*} \to \widehat{A^*}$ such that $\psi(a) = u$. It follows that F is the closure of the rational language

$$F(u) \cup \bigcup \{xu^*y\colon x \in (A^*)^{-1}u,\ y \in u(A^*)^{-1}\}$$

and, therefore, it is a clopen subset of $\widehat{A^*}$.

(3) $\implies$ (2) Suppose that F is clopen. Then $F(w) = F \cap A^*$ is a rational language by [2, Theorem 3.6.1].

(2) $\implies$ (1) Finally, suppose that $F(w)$ is a rational language. Then $F(w)$ is recognized by some finite deterministic automaton. If n is the number of states of such an automaton and $v \in F \cap A^*$ is a word of length at least n, then there is a factorization $v = xyz$, with $y \in A^+$, such that $xy^*z \subseteq F \cap A^*$. This is the pumping lemma and results from the fact that the path labeled v from the initial state has to repeat some state. Since F is closed under taking factors, it follows that $y^* \subseteq F$. Since F is closed, we deduce that the infinite pseudoword $y^\omega = \lim_{n\to\infty} y^{n!}$ belongs to F. By Theorem 2.6, we conclude that $y^\omega \mathcal{J} w$, which shows that w is periodic. □

3. Uniform Recurrence vs. Substitutions

A well-known way to produce uniform recurrence phenomena in symbolic dynamics is through the iteration of primitive substitutions. For pseudowords, we have an extension of the analogous result, which we proceed to present.

Given an element x of a profinite monoid, the sequence $(x^{n!})_n$ must converge to an idempotent, and this is the only idempotent that is the limit of a sequence of positive powers of x, simply because this is the case in every finite monoid and profinite monoids are residually finite as topological monoids. This special idempotent associated with x is denoted by x^ω.

We have shown elsewhere that, provided a profinite monoid M is finitely generated, the monoid of continuous endomorphisms of M is a profinite monoid with respect to the pointwise convergence topology [5, Theorem 4.14]. In particular, for a finite alphabet A, given any continuous endomorphism φ of $\widehat{A^*}$, there is a (unique) idempotent "infinite iterate" φ^ω.

The following is a trivial but crucial observation.

Lemma 3.1. *If u is a factor of $\varphi^\omega(v)$, then so is $\varphi^\omega(u)$.*

Proof. By the hypothesis, u is a factor of $\varphi^\omega(v)$, say $\varphi^\omega(v) = xuy$, where $x, y \in \widehat{A^*}$. Since φ^ω is an idempotent homomorphism, it follows that

$$\varphi^\omega(v) = \varphi^\omega(\varphi^\omega(v)) = \varphi^\omega(x)\varphi^\omega(u)\varphi^\omega(y),$$

which shows that $\varphi^\omega(u)$ is a factor of $\varphi^\omega(v)$. □

Since $\widehat{A^*}$ is the free profinite monoid on the set A of free generators, every homomorphism $\varphi\colon A^* \to B^*$ induces a continuous homomorphism $\hat{\varphi}\colon \widehat{A^*} \to \widehat{B^*}$. A continuous homomorphism $\psi\colon \widehat{A^*} \to \widehat{B^*}$ is said to be *finite* if ψ transforms finite words into finite words, i.e., ψ is induced by some homomorphism $A^* \to B^*$. Both a homomorphism $A^* \to B^*$ and the unique finite continuous homomorphism $\widehat{A^*} \to \widehat{B^*}$ that it induces are called *substitutions from A to B*, or simply *over A* in the case where $B = A$.

For a pseudoword $w \in \widehat{A^*}$, $c_{\leq n}(w)$ denotes the set of all nonempty factors of w of length at most n. An element of $c_{\leq n}(w)$ is also said to *occur in w*. It is well known that the *content* function defined by $c(w) = c_{\leq 1}(w) \setminus \{1\}$ is a continuous homomorphism with values in the semilattice of all subsets of A with respect to union and the discrete topology. For $\varphi \in \operatorname{End} \widehat{A^*}$, let $c_{\leq n}(\varphi) = \bigcup_{a\in A} c_{\leq n}(\varphi(a))$ and let $c(\varphi) = c_{\leq 1}(\varphi) \setminus \{1\}$.

Lemma 3.2. *Let $\varphi \in \operatorname{End} \widehat{A^*}$. Then $\big(c(\varphi^n)\big)_n$ is a sequence of subsets of A that strictly decreases until it stabilizes. In particular, $c(\varphi^\omega) = c(\varphi^{|A|})$.*

Proof. If the letter a occurs in $\varphi^{n+1}(b) = \varphi^n(\varphi(b))$, then there is some letter $d \in c(\varphi(b))$ such that a occurs in the factor $\varphi^n(d)$. Hence the sequence $(c(\varphi^n))_n$ is decreasing. Suppose next that $c(\varphi^n) = c(\varphi^{n+1})$. Given $a \in c(\varphi^n)$, there is some letter b such that a occurs in $\varphi^{n+1}(b) = \varphi(\varphi^n(b))$ and so there is some letter $d \in c(\varphi^n(b))$ such that a occurs in $\varphi(d)$. Since $c(\varphi^n) = c(\varphi^{n+1})$, there is some letter e such that $d \in c(\varphi^{n+1}(e))$. Hence $a \in c(\varphi^{n+2})$, since a occurs in $\varphi^{n+2}(e)$. □

For $B \subseteq A$, denote by $B^{\leq n}$ the set of all words in the letters of B of length at most n. We say that a mapping $\varphi\colon \widehat{A^*} \to \widehat{B^*}$ *erases* a letter a if $\varphi(a) = 1$.

Lemma 3.3. *Let $\varphi \in \operatorname{End}\widehat{A^*}$ and let $B = c(\varphi^\omega)$. If φ does not erase letters of B, then $\big(c_{\le r}(\varphi^n|_B)\big)_n$ is a sequence of subsets of $B^{\le n}$ that strictly increases until it stabilizes.*

Proof. If u is a factor of $\varphi^n(a)$ and $a \in B$, then $a \in c(\varphi(b))$ for some letter $b \in B$, whence u is a factor of $\varphi^{n+1}(b)$. This shows that the sequence is increasing. Suppose next that $c_{\le r}(\varphi^n|_B) = c_{\le r}(\varphi^{n+1}|_B)$ and let u be a factor of $\varphi^{n+2}(a)$ for some $a \in B$. Since $\varphi^{n+2}(a) = \varphi(\varphi^{n+1}(a))$, there is some factor v of $\varphi^{n+1}(a)$ such that u is a factor of $\varphi(v)$. If we choose v to be of minimal length, since φ does not erase letters of B, we must have $|v| \le r$. Hence v belongs to $c_{\le r}(\varphi^{n+1}|_B)$ and so also to $c_{\le r}(\varphi^n|_B)$. This shows that $u \in c_{\le r}(\varphi^{n+1}|_B)$. By induction, it follows that the sequence $\big(c_{\le r}(\varphi^n|_B)\big)_n$ stabilizes as soon as it repeats some term. □

A continuous endomorphism φ of $\widehat{A^*}$ is said to be *weakly primitive* if there exists n such that the set $c_{\le 2}(\varphi^n(a))$ is the same for every letter $a \in A$ and it is not contained in A, i.e., $\varphi^n(a)$ has at least one factor of length 2. We also say that φ is *primitive* if there exists n such that $c(\varphi^n(a)) = A$ for every letter $a \in A$. An endomorphism of A^* is also said to be *weakly primitive* (respectively, *primitive*) if its unique extension to a continuous endomorphism of $\widehat{A^*}$ has the same property.

Lemma 3.4. *Let $\varphi \in \operatorname{End}\widehat{A^*}$ and suppose that $c_{\le 2}(\varphi^n(a))$ is the same set for all $a \in A$ and a fixed n. Then, for every $m \ge n$, the set $c_{\le 2}(\varphi^m(a))$ is also independent of $a \in A$.*

Proof. Proceeding by induction, it suffices to show that if $a, b \in A$ and u is a factor of $\varphi^{n+1}(a)$ of length at most 2, then u is also a factor of $\varphi^{n+1}(b)$. Indeed, since $\varphi^{n+1}(a) = \varphi(\varphi^n(a))$ and $|u| \le 2$, there is $v \in c_{\le 2}(\varphi^n(a))$ such that u is a factor of $\varphi(v)$. Hence $v \in c_{\le 2}(\varphi^n(b))$ and u is also a factor of $\varphi^{n+1}(b)$. □

The set of elements $w \in \widehat{A^*}$ with a given $c_{\le 2}(w)$ is the intersection of a finite set of clopen subsets of $\widehat{A^*}$, each stipulating the presence or absence of a certain factor of length at most 2. Hence the function $c_{\le 2}\colon \widehat{A^*} \to \mathcal{P}(A^{\le 2})$ is continuous for the discrete topology of the power set $\mathcal{P}(A^{\le 2})$. In view of Lemma 3.4, we conclude that $\varphi \in \operatorname{End}\widehat{A^*}$ is weakly primitive if and only if $c_{\le 2}(\varphi^\omega(a))$ is the same for every letter $a \in A$, in which case the common value is $c_{\le 2}(\varphi^\omega)$. Since $c(w) = c_{\le 2}(w) \cap A$ for every $w \in \widehat{A^*}$, if φ is weakly primitive then we also have $c(\varphi^\omega(a)) = c(\varphi^\omega)$ for every letter $a \in A$.

The following result is useful to perform computations in concrete examples. It follows from a related result in the Perron–Frobenius theory presented without proof by Wielandt [24], where an example shows that the bound is sharp. A proof can be found in [20].

Lemma 3.5. *Let φ be a weakly primitive substitution, let $B = c(\varphi^\omega)$, and let $r = |B|$ and $N = r^2 - 2r + 2$. Then $c(\varphi^N(b)) = B$ for every $b \in B$.*

Part of the following simple observation has already appeared in [10], but its proof is also included here for the sake of completeness.

Lemma 3.6. *Let $\varphi \in \operatorname{End}\widehat{A^*}$.*

1. *If all $c(\varphi^\omega(a)) = B$ for every letter $a \in A$, then all $\varphi^\omega(a)$, with $a \in B$, lie in the same $\mathcal{J}$-class.*
2. *If all $\varphi^\omega(a)$, with $a \in A$, lie in the same $\mathcal{J}$-class, then, for each $n \ge 1$, $c_{\le n}(\varphi^\omega(a))$ is the same for every letter $a \in A$.*
3. *If all $\varphi^\omega(a)$, with $a \in A$, have the same factors of length at most 2, then they have the same finite factors.*

Proof.

(1) Let $a, b \in B$. By the hypothesis, a is a factor of $\varphi^\omega(b)$ and so, by Lemma 3.1, $\varphi^\omega(a)$ is a factor of $\varphi^\omega(b)$. By symmetry, it follows that $\varphi^\omega(a)$ and $\varphi^\omega(b)$ are $\mathcal{J}$-equivalent.

(2) Given $a, b \in A$, the pseudowords $\varphi^\omega(a)$ and $\varphi^\omega(b)$ have the same factors and, in particular, they have the same finite factors.

(3) It suffices to show that, given $a, b \in A$, every finite factor of $\varphi^\omega(a)$ is also a factor of $\varphi^\omega(b)$. Now, by [10, Lemma 7.2] every finite factor u of $\varphi^\omega(a) = \varphi^\omega(\varphi^\omega(a))$ is also a factor of $\varphi^\omega(x)$ for some factor x of $\varphi^\omega(a)$ of length at most 2. Then, by the hypothesis, x is a factor of $\varphi^\omega(b)$ and, therefore, by Lemma 3.1 so is $\varphi^\omega(x)$. Hence u is a factor of $\varphi^\omega(b)$. □

Lemma 3.6 implies that primitive substitutions are weakly primitive. An example of a weakly primitive substitution that is not primitive is given by the substitution φ over the alphabet $\{a, b, c\}$ defined by $\varphi(a) = ab$, $\varphi(b) = ba$, and $\varphi(c) = a^3b^3$: for every letter x, the factors of $\varphi^3(x)$ of length at most 2 are the words of such length in the proper subalphabet $\{a, b\}$.

In view of Lemma 3.6, for a weakly primitive substitution $\varphi \in \operatorname{End}\widehat{A^*}$ with $B = c(\varphi^\omega)$, we must have $c_{\le n}(\varphi^\omega) = c_{\le n}(\varphi^\omega|_B)$, which, together with Lemma 3.3, provides a simple algorithm to compute $c_{\le n}(\varphi^\omega)$. A rough upper bound of how many iterations are needed is provided by the equality $c_{\le n}(\varphi^\omega) = c_{\le n}((\varphi|_B)^M)$, where $M = \sum_{i=1}^{n} |B|^i = O(|B|^n)$, which now follows from Lemma 3.3.

Theorem 3.7. *Let φ be a substitution over a finite alphabet A. Then the following conditions are equivalent*:

(1) *the pseudowords $\varphi^\omega(a)$, with $a \in A$, are all $\mathcal{J}$-equivalent*;
(2) *the pseudowords $\varphi^\omega(a)$, with $a \in A$, are all uniformly recurrent and have the same content*;
(3) *φ is weakly primitive.*

Proof. The implication (1) $\implies$ (3) follows from Lemma 3.6(2).

To prove (2) $\implies$ (1), consider a letter $b \in c(\varphi^\omega)$. By the hypothesis, b is a factor of $\varphi^\omega(a)$ for every $a \in A$, and therefore so is $\varphi^\omega(b)$ by Lemma 3.1. Since $\varphi^\omega(b)$ is itself uniformly recurrent, it must be infinite. By Theorem 2.6, we conclude that all $\varphi^\omega(a)$ lie in the same $\mathcal{J}$-class as $\varphi^\omega(b)$, since uniformly recurrent pseudowords do not have infinite pseudowords strictly $\mathcal{J}$-above them.

To conclude the proof, it remains to show that (3) $\implies$ (2), i.e., weak primitivity of φ implies uniform recurrence of the pseudowords $\varphi^\omega(a)$, since the content condition in (2) is an immediate consequence of weak primitivity. By Lemma 3.6(3), all $\varphi^\omega(a)$, with $a \in A$, have the same finite factors. Fix $a \in A$ and let v be a finite factor of $\varphi^\omega(a)$. Since $\varphi^\omega(a) = \lim_{n\to\infty} \varphi^{n!}(a)$ and all the pseudowords $\varphi^\omega(b)$ have the same finite factors, we see that for every $b \in A$ there exists $k \ge 1$ such that v is a factor of $\varphi^k(b)$. Since b is a factor of $\varphi^p(a)$ for all $a \in A$ and all sufficiently large p, it follows that there exists k such that v is a factor of $\varphi^j(b)$ for all $j \ge k$. Following [21, Sec. 5.2], let

$$K = \max_{b\in A} \min\{k \ge 1 \colon j \ge k \implies v \in F(\varphi^j(b))\}.$$

Let l be the maximum of the lengths $|\varphi^K(b)|$ with $b \in A$. Then, for a factor z of $\varphi^\omega(a)$ of length $2l - 1$, z must be a factor of $\varphi^r(a)$ for some $r \ge K$. Note that $\varphi^r(a)$ is a product of words of the form $\varphi^K(b)$ ($b \in A$), each of which has v as a factor. But z is a too long factor of $\varphi^r(a)$ to overlap these words without completely containing one of them as a factor. Hence v is a factor of z, which shows that $\varphi^\omega(a)$ is uniformly recurrent. □

We examine next what happens when we apply a finite continuous endomorphism to a uniformly recurrent pseudoword.

Theorem 3.8. *Let $\varphi\colon \widehat{A^*} \to \widehat{B^*}$ be a finite continuous homomorphism and let $w \in \widehat{A^*}$ be such that φ does not erase all the letters from $c(w)$. If w is uniformly recurrent, then so is $\varphi(w)$.*

Proof. Let w be uniformly recurrent. Note that $\varphi(w)$ is an infinite pseudoword, since φ does not erase every letter from w and no letter can occur only a finite number of times in a uniformly recurrent pseudoword.

Suppose v is a finite factor of $\varphi(w)$. By Corollary 2.9, there exists a sequence $(w_n)_n$ of finite factors of w converging to a pseudoword in the $\mathcal{H}$-class of w. Since φ is a continuous homomorphism, the sequence

$(\varphi(w_n))_n$ of factors of $\varphi(w)$ converges to a pseudoword that is $\mathcal{H}$-equivalent to $\varphi(w)$ and, therefore, has the same finite factors as $\varphi(w)$ does. Hence v is a factor of some $\varphi(w_n)$. Let N be such that every factor of w of length N admits w_n as a factor. Let $K = (N+1)M + 1$, where $M = \max\{|\varphi(a)| : a \in A\}$, and let z be a factor of $\varphi(w)$ of length K. Then z is a factor of some $\varphi(w_m)$, which in turn is a product of words of the form $\varphi(a)$. Now, z was chosen sufficiently long to contain a factor of the form $\varphi(y)$ with y a factor of w_m of length N. Hence w_n is a factor of y and so v is a factor of $\varphi(y)$, which in turn is a factor of z. This shows that $\varphi(w)$ is uniformly recurrent. □

We say that a subgroup of $\widehat{A^*}$ is *$\mathcal{J}$-maximal* if it consists of infinite pseudowords and there is no other such subgroup of $\widehat{A^*}$ that lies strictly $\mathcal{J}$-above it. By Theorem 2.6, the $\mathcal{J}$-maximal subgroups are the subgroups that consist of uniformly recurrent pseudowords. By Lemma 2.3(3), the $\mathcal{J}$-class of every uniformly recurrent pseudoword contains $\mathcal{J}$-maximal subgroups and, in fact, all of its subgroups are $\mathcal{J}$-maximal.

Example 3.9. Let $A = \{a_1, \ldots, a_m\}$ and let $w_i = \varphi^\omega(a_i)$, where $\varphi(a_i) = a_1 \cdots a_{i-1} a_i^2 a_{i+1} \cdots a_m$ for $i = 1, \ldots, m-1$ and $\varphi(a_m) = a_1 \cdots a_m$. It is shown in [10] that the w_i freely generate a free profinite subgroup H of $\widehat{A^*}$. By Theorems 3.7 and 2.6, this is a $\mathcal{J}$-maximal subgroup. It follows from the results of Secs. 4 and 5 below that H is an $\mathcal{H}$-class of $\widehat{A^*}$ and, therefore, a maximal subgroup.

We say that an equality $u_1 \cdots u_m = v_1 \cdots v_n$ with $u_i, v_j \in \widehat{A^*}$ is *reducible* if there exist indices r and s with $2 < r + s \leq m + n$ and $u_r \cdots u_m = v_s \cdots v_n$.

Let φ be a continuous homomorphism $\widehat{A^*} \to \widehat{B^*}$. We say that φ is an *encoding* if it is injective. By [19, Proposition 2.1], in the case where φ is finite, φ is an encoding if and only if its restriction to A^* is injective.

We say that $C \subseteq A^*$ is *of bounded delay with respect to* a given $w \in \widehat{A^*}$ if there exists an integer N such that any equality between factors of w of one of the forms

$$uc_1 \cdots c_m v = c'_1 \cdots c'_n \quad \text{or} \quad uc_1 \cdots c_m = c'_1 \cdots c'_n v,$$

where $c_i, c'_j \in C$, $u, v \in A^*$, $A^*u \cap C^* \neq \varnothing$, $vA^* \cap C^* \neq \varnothing$, and $m + n > N$, is reducible. In this case, we also say that C has *delay at most N with respect to w*. If there is an integer N such that C has delay at most N with respect to every $w \in \widehat{A^*}$, then we also say that C *has bounded delay* and C *has delay at most N*. See [1, 22] for effective procedures to test this stronger property in the case where C is finite. From those procedures it is not hard to derive algorithms to test whether C has bounded delay with respect to a given $w \in \widehat{A^*}$ provided one can effectively test whether any of a certain finite number of words (depending only on C) is a factor of w.

For a finite continuous homomorphism $\varphi\colon \widehat{A^*} \to \widehat{B^*}$, we say that φ is *of bounded delay with respect to* a given $w \in \widehat{B^*}$ if $\varphi(A)$ is of bounded delay with respect to w. We also say that φ has *delay at most N with respect to w* if $\varphi(A)$ does.

The following result is a partial converse for Theorem 3.8, which will play a key role in Sec. 4.

Theorem 3.10. *Let $\varphi\colon \widehat{A^*} \to \widehat{B^*}$ be a finite continuous homomorphism and let $w \in \widehat{A^*}$. If $\varphi(w)$ is uniformly recurrent and φ is an encoding of bounded delay with respect to $\varphi(w)$, then w is also uniformly recurrent.*

Proof. Let N be an integer such that φ has delay at most N with respect to $\varphi(w)$. Let $u \in F(w)$. Let $n = \lceil |u|/N \rceil + 2$ and let m and M be respectively the minimum and maximum of the lengths of the words in $\varphi(A)$. Since $\varphi(w)$ is uniformly recurrent and A is finite, there exists an integer K such that every finite factor of $\varphi(w)$ of length at least K contains as a factor every factor of $\varphi(w)$ of length at most MnN.

Let v be a factor of w of length $\lceil K/m \rceil$. Then $\varphi(v)$ is a factor of $\varphi(w)$ of length at least K, which, therefore, contains as a factor some word of the form $\varphi(xuy)$ with $x, y \in A^N$, since the length of such a word is at most MnN. Let $v = v_1 \cdots v_p$, $x = x_1 \cdots x_N$, $u = u_1 \cdots u_q$, and $y = y_1 \cdots y_N$, where

$v_i, x_j, u_k, y_j \in A$, and denote, for each $a \in A$, $\varphi(a)$ by $\bar{a}$. Then we have an equality between factors of $\varphi(w)$ of the form

$$\bar{v}_1 \cdots \bar{v}_p = z\,\bar{x}_1 \cdots \bar{x}_N\,\bar{u}_1 \cdots \bar{u}_q\,\bar{y}_1 \cdots \bar{y}_N\,t$$

for some $z, t \in A^*$. Since φ has delay at most N with respect to $\varphi(w)$, there exist indices i, j, k, and l such that

$$\bar{v}_i \cdots \bar{v}_j = \bar{x}_k \cdots \bar{x}_N \bar{u}_1 \cdots \bar{u}_q \bar{y}_1 \cdots \bar{y}_l.$$

Since φ is an encoding, it follows that the preceding equality still holds if we remove the bars and so u is a factor of v. Hence w is uniformly recurrent. □

Note that one cannot expect that w be uniformly recurrent just under the hypothesis that $\varphi(w)$ is uniformly recurrent for a nonerasing continuous homomorphism $\varphi\colon \widehat{A^*} \to \widehat{B^*}$. Indeed, simply by changing one occurrence of one letter in a uniformly recurrent pseudoword v into a new letter we obtain a pseudoword w that is not uniformly recurrent. But, letting φ be the identity on all the old letters and mapping the new letter to the replaced letter, we have $\varphi(w) = v$. In this example, φ is not an encoding. We do not know whether there are any examples in which φ is an encoding that fails the bounded-delay hypothesis of Theorem 3.10 and does not preserve uniform recurrence.

4. Weakly Primitive Substitutions

Throughout this section, let φ be a finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$, where A is a finite alphabet. By Theorem 3.7, all the pseudowords $\varphi^\omega(a)$ $(a \in A)$ lie in the same $\mathcal{J}$-class of $\widehat{A^*}$. We denote this $\mathcal{J}$-class by J_φ. The factors of members of J_φ will be called simply *factors of* J_φ. By Theorem 2.6 and Lemma 2.3(3), J_φ is a $\mathcal{J}$-maximal regular $\mathcal{J}$-class of $\widehat{A^+}$ and it consists of uniformly recurrent pseudowords. Hence the infinite factors of J_φ are the members of J_φ. By compactness and continuity of multiplication, the set of factors of J_φ is closed.

Lemma 4.1.

(1) *If u is a finite factor of J_φ, then so is $\varphi(u)$.*
(2) *If u is a nonempty factor of J_φ, then $\varphi^\omega(u) \in J_\varphi$.*

Proof.

(1) Let a be any letter in $c(\varphi^\omega)$ and let $w = \varphi^\omega(a)$. Since $w = \lim_{n\to\infty} \varphi^{n!}(a)$, u is a finite factor of w, and the set $\widehat{A^*}u\widehat{A^*}$ is an open subset of $\widehat{A^*}$, we see that u must be a factor of $\varphi^{n!}(a)$ for all sufficiently large n. Since φ is weakly primitive, for all sufficiently large m the letter a occurs in $\varphi^{m-n!}(a)$, whence the word $\varphi^{n!}(a)$ is a factor of $\varphi^m(a)$. Hence, for all sufficiently large m, u is a factor of $\varphi^m(a)$ and therefore it is also a factor of $\varphi^{\omega-1}(a) = \lim_{n\to\infty} \varphi^{n!-1}(a)$. Since φ is a homomorphism, $\varphi(u)$ is a factor of $\varphi^\omega(a) = w$.

(2) By Lemma 3.1, $\varphi^\omega(u)$ is an infinite factor of J_φ. Hence $\varphi^\omega(u) \in J_\varphi$, since J_φ is a $\mathcal{J}$-maximal $\mathcal{J}$-class of infinite pseudowords. □

Proposition 4.2. *The action of φ on $\widehat{A^*}$ induces an action on J_φ.*

Proof. Take $a \in A$ and let $w = \varphi^\omega(a)$. Since w is uniformly recurrent by Theorem 3.7, by Theorem 3.8 $\varphi(w)$ is also uniformly recurrent. Hence, by Corollary 2.10, it suffices to show that every finite factor of $\varphi(w)$ is also a factor of w. Now, by [10, Lemma 7.2], every finite factor u of $\varphi(w)$ is a factor of $\varphi(v)$ for some finite factor v of w. Moreover, by Lemma 4.1(1), whenever v is a finite factor of w, so is $\varphi(v)$. Hence u is a factor of w. □

Lemma 4.3. *Let v be a pseudoword of J_φ that belongs to* $\operatorname{Im}\varphi^\omega$ *and let a be its first letter. Then a is also the first letter of $\varphi^\omega(a)$ and there is some $k \geq 1$ such that a is the first letter of $\varphi^k(a)$ and $\varphi^{\omega-k}(a) \in J_\varphi$.*

Proof. By the hypothesis, there is a factorization of the form $v = av'$ for some $v' \in \widehat{A^+}$ and so the first letter, a, of $v = \varphi^\omega(v) = \varphi^\omega(a)\varphi^\omega(v')$ is also the first letter of $\varphi^\omega(a)$. Since the alphabet A is finite so that the set $a\widehat{A^*}$ is open, there is some $k \geq 1$ such that $\varphi^k(a)$ starts with the letter a. By Proposition 4.2 and since J_φ is a closed set,

$$\varphi^{\omega-k}(a) = \varphi^{\omega-k}(\varphi^\omega(a)) = \lim_{n\to\infty} \varphi^{n!-k}(\varphi^\omega(a))$$

also lies in J_φ. □

Lemma 4.4. *Let H be a maximal subgroup of J_φ containing an element of the form $\varphi^\omega(v)$ for some pseudoword v. Then $\varphi^\omega(H) \subseteq H$.*

Proof. Let $K = \varphi^\omega(H)$. Then K is a continuous homomorphic image of a profinite group. Since K is a closed subsemigroup of a profinite semigroup, it is itself a profinite semigroup [4, Proposition 4.3]. Hence K is a profinite group by the same result. On the other hand, $H \cap K$ is nonempty, since both contain the pseudoword $\varphi^\omega(v)$. Since H is a maximal subgroup of $\widehat{A^*}$, it follows that $K \subseteq H$. □

Let a and b be letters such that ba is a factor of J_φ. Denote by $X_\varphi(a, b)$ the set of all finite words u such that bua is a factor of J_φ and u starts with a, ends with b, and cannot be properly factorized into such words, i.e., it does not contain ba as a factor. There is another related set, which is, in general, larger and which will also play a role here. It is the set $Y_\varphi(a, b)$ consisting of all finite factors of J_φ that start with a, end with b, and do not contain the factor ba. Since the elements of J_φ are uniformly recurrent, the set $Y_\varphi(a, b)$ is finite, and therefore so is its subset $X_\varphi(a, b)$. More precisely, we have the following result, which shows that $X_\varphi(a, b)$ may be effectively computed.

Lemma 4.5. *Let ba be a 2-letter factor of J_φ, $B = c(\varphi^\omega)$, and $r = |B|$. Let M be the smallest integer such that $c_{\leq 2}(\varphi^M) = c_{\leq 2}(\varphi^{M+1})$ and let N be the smallest integer such that $c(\varphi^N(b)) = c(\varphi^\omega)$ for every $b \in B$. Then $X_\varphi(a, b)$ consists of factors of words of the form $\varphi^{M+N}(u)$, where $u \in c_{\leq 2}(\varphi^\omega|_B)$. Hence $X_\varphi(a, b)$ may be effectively computed.*

Proof. Let $w \in X_\varphi(a, b)$. Then w is a factor of J_φ and, therefore, it is a factor of $\varphi^{M+N}(u)$ for some $u \in B^*$. By Lemma 3.3, ba is a factor of $\varphi^M(d)$ for some $d \in B$. On the other hand, by assumption, d appears in every word of the form $\varphi^N(e)$ with $e \in B$. Hence ba is a factor of every word of the form $\varphi^{M+N}(e)$ with $e \in B$. Since w does not admit ba as a factor, it follows that w cannot contain any factor of the form $\varphi^{M+N}(e)$, where $e \in B$, and hence it must be a factor of $\varphi^{M+N}(u_0)$ for some $u_0 \in c_{\leq 2}(\varphi^\omega|_B)$, which establishes the lemma. □

Note that the numbers M and N of Lemma 4.5 satisfy the following inequalities: by Lemma 3.3, $M \leq r^2 + r$ and, by Lemma 3.5, $N \leq r^2 - 2r + 2$. The upper bound for N is optimal by [24], but the upper bound for M is probably not optimal.

Example 4.6. For the substitution $\varphi \in \mathrm{End}\,\widehat{\{a, b\}^*}$ that sends the letter a to ab^2 and b to a, where $r = 2$, we see that $N = 2$, $M = 3$, and every 2-letter word is a factor of J_φ. Hence, to compute the set $X_\varphi(a, a)$, it suffices to compute the words

$$\varphi^5(a) = abbaaabbabbabbaaabbaaabbaaabbabbabbaaabbabb,$$
$$\varphi^5(b) = abbaaabbabbabbaaabbaa,$$

concatenate the two in any order, and find the factors between consecutive occurrences of the factor aa. Performing this routine calculation, we conclude that $X_\varphi(a, a) = \{a, ab^2a, (ab^2)^3a\}$. Since every word in $Y_\varphi(a, a)$ is a factor of some word in $X_\varphi(a, a)$, it is now immediate to deduce that $Y_\varphi(a, a) = \{a, ab^2a, (ab^2)^2a, (ab^2)^3a\}$.

For a subset X of $\widehat{A^+}$, denote by X^+ the subsemigroup of $\widehat{A^+}$ generated by X.

Lemma 4.7. *Suppose that ba is a factor of J_φ and that there is some pseudoword of the form $w = \varphi^\omega(u)$ with $u \in X_\varphi(a,b)$ such that w starts with a and ends with b. Let M be the maximum length of elements of $X_\varphi(a,b)$. Let v be a finite word that has the same prefix of length $M+1$, the same suffix of length $M+1$, and also the same factors of length $M+2$ as w does. Then v belongs to $X_\varphi(a,b)^+$.*

Proof. Note that, by the choice of M, every factor of length $M+1$ of the uniformly recurrent pseudoword w must contain ba as a factor. Since every factor of w of length $M+1$ is a factor of v, the word v must therefore contain ba as a factor. Since v starts with a and ends with b, it follows that v admits a factorization of the form $u_1u_2\cdots u_r$, with each u_i starting with a, ending with b, not having ba as a factor, and of length at most M. Hence u_1a is a prefix, bu_r is a suffix, and each bu_ia $(1 < i < r)$ is a factor of w. Moreover, as ba is a factor of J_φ, so is $\varphi^\omega(ba)$. Since $\varphi^\omega(u)$ starts with a and ends with b, it follows that $\varphi^\omega(bua)$ belongs to J_φ. Hence bu_1a and bu_ra are also factors of w by Lemma 4.3. By the definition of $X_\varphi(a,b)$, this implies that all the factors u_j belong to $X_\varphi(a,b)$. Hence $v \in X_\varphi(a,b)^+$. □

Proposition 4.8. *Let a and b be letters such that ba is a factor of J_φ.*

1. *All pseudowords of the form $\varphi^\omega(u)$ with $u \in X_\varphi(a,b)$ belong to the same $\mathcal{H}$-class H of J_φ, which is a group.*
2. *If $\varphi^\omega(a)$ starts with a and $\varphi^\omega(b)$ ends with b, then $\varphi^\omega(H)$ is generated by the set $\varphi^\omega(X_\varphi(a,b))$ as a closed subgroup.*

Proof. The pseudowords $\varphi^\omega(a)$ and $\varphi^\omega(b)$, as well as, by Lemma 4.1(2), all the elements of the set $\varphi^\omega(X_\varphi(a,b))$, belong to J_φ. Moreover, $\varphi^\omega(a)$ is a prefix and $\varphi^\omega(b)$ is a suffix of every element of $\varphi^\omega(X_\varphi(a,b))$. Hence $\varphi^\omega(X_\varphi(a,b))$ is contained in the $\mathcal{H}$-class H that is the intersection of the $\mathcal{R}$-class of $\varphi^\omega(a)$ with the $\mathcal{L}$-class of $\varphi^\omega(b)$.

Since ba is a factor of J_φ and the pseudowords of J_φ are uniformly recurrent, there is some finite factor of J_φ of the form $baubavba$ and therefore also one of the form xy with $x, y \in X_\varphi(a,b)$. Again by Lemma 4.1(2), the pseudoword $\varphi^\omega(xy) = \varphi^\omega(x)\varphi^\omega(y)$ belongs to J_φ, which implies that H is a group. This proves (1). Moreover, by Lemma 4.4, $\varphi^\omega(H)$ is a profinite subgroup of H.

Next assume that $\varphi^\omega(a)$ starts with a and $\varphi^\omega(b)$ ends with b. Then every element of H starts with a and ends with b.

Let M be as in Lemma 4.7 and let w be an arbitrary element of $\varphi^\omega(H)$. Let $(w_n)_n$ be a sequence of finite words converging to w, which we may choose so that all w_n have the same prefix and the same suffix of length $M+1$ as w, as well as the same factors of length $M+2$ as w. By Lemma 4.7, it follows that each w_n belongs to $X_\varphi(a,b)^+$. This shows that $w \in \overline{X_\varphi(a,b)^+}$. By the continuity of φ^ω, we deduce that $\varphi^\omega(w)$ belongs to the closure of $\varphi^\omega(X_\varphi(a,b))^+$. Finally, by (1), the closure of the subsemigroup $\varphi^\omega(X_\varphi(a,b))^+$ is the closed subgroup of H generated by $\varphi^\omega(X_\varphi(a,b))$, which proves (2). □

We say that a finite weakly primitive substitution φ on an alphabet A is of *relative bounded delay* if φ is of bounded delay with respect to elements of J_φ, and we say that φ is *special* if φ is of relative bounded delay and its restriction to $\widehat{c(\varphi^\omega)^*}$ is an encoding.

Lemma 4.9. *If φ is a finite, special, weakly primitive, continuous endomorphism of $\widehat{A^*}$, then so is φ^n for every $n \geq 1$.*

Proof. Let $B = c(\varphi^\omega)$. Since φ is assumed to be an injective continuous endomorphism of $\widehat{B^*}$, so are its powers φ^n. Since $\varphi^\omega = (\varphi^n)^\omega$, and so $J_\varphi = J_{\varphi^n}$, it remains to show that φ^n is of bounded delay with respect to elements of J_φ. In order to prove this for $n > 1$, we assume inductively that φ^{n-1} is of bounded delay with respect to elements of J_φ.

Consider an equality between factors of J_φ of the following form:

$$u\varphi^n(a_1)\cdots\varphi^n(a_r)v = \varphi^n(b_1)\cdots\varphi^n(b_s), \tag{4.1}$$

where $a_i, b_j \in B$. Since φ^{n-1} is assumed to be of bounded delay with respect to factors of J_φ, provided $r+s$ is sufficiently large, there are indices i_1 and j_1 and factorizations $\varphi(a_{i_1}) = x_{i_1}x'_{i_1}$ and $\varphi(b_{j_1}) = y_{j_1}y'_{j_1}$

such that

$$u\varphi^{n-1}(\varphi(a_1\cdots a_{i_1-1})x_{i_1}) = \varphi^{n-1}(\varphi(b_1\cdots b_{j_1-1})y_{j_1}), \tag{4.2}$$

$$\varphi^{n-1}(x'_{i_1}\varphi(a_{i_1+1}\cdots a_r))v = \varphi^{n-1}(y'_{j_1}\varphi(b_{j_1+1}\cdots b_s)). \tag{4.3}$$

Provided $r+s$ is sufficiently large, one of the numbers i_1+j_1 or $r+s-i_1-j_1$ will be still sufficiently large to guarantee that at least one of the equalities (4.2) and (4.3) will reduce similarly, say producing equalities

$$\begin{aligned} u\varphi^{n-1}(\varphi(a_1\cdots a_{i_2-1})x_{i_2}) &= \varphi^{n-1}(\varphi(b_1\cdots b_{j_2-1})y_{j_2}),\\ \varphi^{n-1}(x'_{i_2}\varphi(a_{i_2+1}\cdots a_{i_1-1})x_{i_1}) &= \varphi^{n-1}(y'_{j_2}\varphi(b_{j_2+1}\cdots b_{j_1-1})y_{j_1}), \end{aligned} \tag{4.4}$$

where $\varphi(a_{i_2}) = x_{i_2}x'_{i_2}$ and $\varphi(b_{j_2}) = y_{j_2}y'_{j_2}$. Again assuming that $r+s$ is sufficiently large, we may proceed in this manner until, by concatenating the equalities of type (4.4), we obtain an equality of the form

$$\varphi^{n-1}(x'_l\varphi(a_{l+1}\cdots a_{m-1})x_m) = \varphi^{n-1}(y'_p\varphi(b_{p+1}\cdots b_{q-1})y_q),$$

where $\varphi(a_t) = x_tx'_t$, $\varphi(b_t) = y_ty'_t$, and $m-l+q-p$ is as large as desired. Since φ^{n-1} is injective on B^*, $B = c(\varphi^\omega)$, and φ is weakly primitive, it follows that

$$x'_l\varphi(a_{l+1})\cdots\varphi(a_{m-1})x_m = y'_p\varphi(b_{p+1})\cdots\varphi(b_{q-1})y_q. \tag{4.5}$$

Moreover, since φ^{n-1} sends J_φ into itself by Proposition 4.2 and φ^{n-1} is injective on $\widehat{B^*}$, the common value of the two sides of (4.5) is still a factor of J_φ. Now, since $m-l+q-p$ may be taken to be as large as required and φ is of bounded delay with respect to factors of J_φ, we conclude that Eq. (4.5) is reducible, say

$$x'_l\varphi(a_{l+1}\cdots a_f) = y'_p\varphi(b_{p+1}\cdots b_g).$$

Taking into account how the equality (4.5) was obtained, this implies that

$$u\varphi^n(a_1\cdots a_f) = \varphi^n(b_1\cdots b_g),$$

which shows that the equality (4.1) is reducible. The other type of equality needed to show that φ^n is of bounded delay with respect to J_φ is treated similarly. □

Under suitable hypotheses, we have described in Proposition 4.8(2) a set of generators for the closed subgroup $\varphi^\omega(H)$ corresponding to a maximal subgroup H of J_φ. The extra hypothesis of bounded delay with respect to elements of J_φ will now allow us to show that, if H and $\varphi^\omega(H)$ have some point in common, then they are equal. The proof of this result turns out to be somewhat technical and long.

Proposition 4.10. *Let φ be a special, finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$ and let $w \in J_\varphi$. If w lies in the same $\mathcal{H}$-class as some element of $\operatorname{Im}\varphi^\omega$, then $w \in \operatorname{Im}\varphi^\omega$.*

Proof. Let v be an element of the $\mathcal{H}$-class of w that lies in $\operatorname{Im}\varphi^\omega$. Since φ^ω is an idempotent homomorphism, we have $\varphi^\omega(v) = v$. Note that v and w have the same factors, as well as the same finite prefixes and the same finite suffixes. Since $v = \lim_{n\to\infty} v_n$ for some sequence $(v_n)_n$ of finite words and so $v = \varphi^\omega(v) = \lim_{n\to\infty}\varphi^\omega(v_n)$, every finite factor of w is therefore a factor of some word in $\varphi^k(A^+)$ for every positive integer k.

Let a be the first letter of v. By Lemma 4.3, there is some $l > 0$ such that $\varphi^l(a)$ starts with a. Now, $(\varphi^l)^\omega = \varphi^\omega$ (and so $J_{\varphi^l} = J_\varphi$) so that, taking also into account Lemma 4.9, by replacing φ by φ^l if necessary, we may assume that a is the first letter of $\varphi(a)$. This implies that $\varphi^n(a)$ is a prefix of $\varphi^{n+1}(a)$ for all $n \geq 0$, whence $\varphi^n(a)$ is a prefix of v for all $n \geq 0$. In particular, w has arbitrarily long prefixes of the form $\varphi^k(u)$ with $u \in A^+$ and $k > 0$. Similarly, w has arbitrarily long suffixes in $\varphi^k(A^+)$ for every $k > 0$.

Let k be an arbitrary positive integer and let $\psi = \varphi^k$. By Lemma 4.9, there exists N such that ψ has delay at most N with respect to w. Let m and M be respectively the minimum and maximum of the lengths of the words in $\psi(A)$, and let $K = \big(N + \lceil\frac{M}{m}\rceil + 2\big)M$.

Suppose next that x is a finite word that has the same factors, the same prefixes, and the same suffixes of length at most K as w does. We claim that $x \in \operatorname{Im}\psi$. Since w has arbitrarily long prefixes in $\psi(A^+)$, x has a prefix y_0 and a suffix z in $\psi(A^{N+\lceil \frac{M}{m} \rceil+2})$.

On the other hand, given any factor u of w of length K, it has already been observed that u is a factor of some word in $\psi(A^+)$. Hence u must have the form $u = u_1u_2u_3u_4u_5u_6u_7$ with each of the u_1 and u_7 of length less than M, $|u_1u_2| = |u_6u_7| = M$, $u_2u_3, u_5u_6 \in \psi(A^+)$, and $u_4 = \psi(\tilde{u}_4)$ for some $\tilde{u}_4 \in A^+$ of length at least $N-2$. Note that some of the factors u_i, particularly u_3 and u_5, may be empty. The factorization is depicted in Fig. 1, where the waves are meant to represent elements of $\psi(A)$.

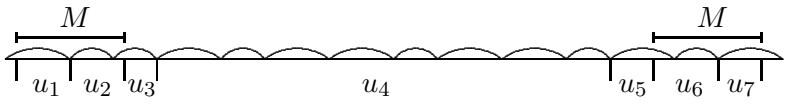


Fig. 1. Factorization of a factor of length K.

Now suppose that we have already identified a prefix y_n of x that belongs to $\psi(A^+)$ and that $K \leq |y_n| < |x| - M$. Let u be the suffix of length K of the prefix of x of length $|y_n| + M$ and consider its associated factorization $u = u_1u_2u_3u_4u_5u_6u_7$ as described in the preceding paragraph. Then for the word $u_2u_3u_4u_5u_6$, which is a product of at least N factors from $\psi(A)$, the prefix $u_2u_3u_4u_5$ is also a suffix of the word y_n. Since ψ has delay at most N with respect to w, it follows that the equality that expresses the overlapping factor must split, since it expresses two factorizations of a factor of u, and therefore also of a factor of w. Hence we can further extend y_n by jumping from the factorization within y_n to that of u and thus find a word y_{n+1} such that $|y_{n+1}| > |y_n|$ and $y_{n+1} \in \psi(A^+)$. The extension of the factorization of a prefix of x in terms of elements of $\psi(A)$ is depicted in Fig. 2.

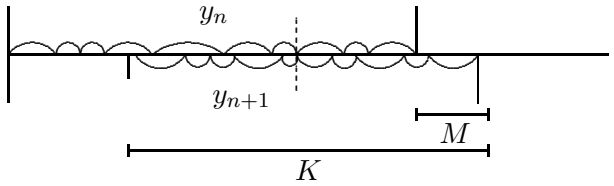


Fig. 2. Extension of the prefix of x in $\psi(A^+)$.

Suppose that y is the longest prefix of x that belongs to $\psi(A^+)$. By the preceding argument and the assumptions on x, we have $|y| \geq K$ and $|y| > |x| - M$. We claim that $y = x$. Otherwise, we consider the overlap of y with the suffix z of x introduced above. By length considerations, there are factorizations $y = y't$ and $z = tz'$ with $1 \leq |z'| < M$. See Fig. 3. Recall that z is the product of $N + \lceil \frac{M}{m} \rceil + 2$ factors

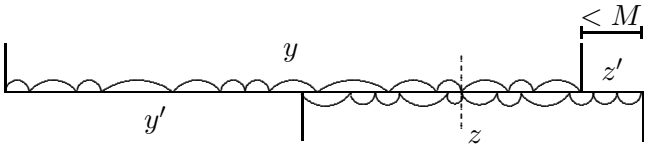


Fig. 3. Final extension of the prefix of x in $\psi(A^+)$.

from $\psi(A)$. Reading from right to left, excluding those factors that lie completely within z', certainly there will remain at least N factors. Since ψ has delay at most N with respect to w, we find a factorization (of a factor of w) involving those remaining factors that must split, which allows us to extend y to a longer prefix that still lies in $\psi(A^+)$, in contradiction with the assumption at the beginning of this paragraph. This completes the proof of the claim that $x \in \psi(A^+)$.

Now, since prescribing a finite number of conditions on finite factors determines a clopen subset of $\widehat{A^+}$, w is the limit of a sequence $(z_n)_n$ of finite words all of which have the same factors, the same prefixes, and the same suffixes of length K as w does. By the above, each z_n belongs to $\psi(A^+)$. Since $\psi = \varphi^k$ is continuous, it follows that $w \in \operatorname{Im}\varphi^k$, say $w = \varphi^k(w_k)$.

Since $\widehat{A^*}$ is a compact metric space, there is a strictly increasing sequence $(k_r)_r$ such that $(w_{k_r!})_r$ converges to some pseudoword w_∞. By the continuity of the evaluation mapping of continuous endomorphisms of $\widehat{A^*}$ [5, Theorem 4.14], we conclude that

$$w = \lim_{r\to\infty} \varphi^{k_r!}(w_{k_r!}) = \varphi^\omega(w_\infty),$$

whence $w \in \operatorname{Im}\varphi^\omega$. □

For a subset X of a group G, denote by $\langle X\rangle$ the subgroup generated by X.

A 2-letter factor ba of J_φ such that $\varphi^\omega(a)$ starts with a and $\varphi^\omega(b)$ ends with b will be called a *connection* for φ. The name is intended to suggest that the factor ba establishes precisely the needed connection between elements of an $\mathcal{H}$-class of J_φ that makes their product remain in J_φ, i.e., that forces the $\mathcal{H}$-class to be a maximal subgroup. This is justified by Theorem 4.13 below, which summarizes the main conclusions of this section.

Lemma 4.11. *Let φ be a finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$ and suppose that H is a maximal subgroup of J_φ such that H contains some element of $\operatorname{Im}\varphi^\omega$. Then there is a unique connection ba for φ such that $\varphi^\omega(X_\varphi(a,b)) \subseteq H$.*

Proof. Let a be the first letter and b the last letter of the elements of H. Then, by Lemma 4.3 and its left-right dual, a is the first letter of $\varphi^\omega(a)$ and b is the last letter of $\varphi^\omega(b)$. Since H is a group, ba is a factor of J_φ, which shows that ba is a connection for φ. Since H contains some element of the form $\varphi^\omega(u)$ that starts with a and ends with b, we see that H is the intersection of the $\mathcal{R}$-class of $\varphi^\omega(a)$ with the $\mathcal{L}$-class of $\varphi^\omega(b)$. Hence every element of the form $\varphi^\omega(v)$, with $v \in X_\varphi(a,b)$, belongs to H. This proves the existence of a connection as stated. Uniqueness follows from the fact that, ba being a connection such that $\varphi^\omega(X_\varphi(a,b)) \subseteq H$, a must be the first letter and b the last letter of the elements of H. □

Since there is at least one maximal subgroup of J_φ that meets nontrivially $\operatorname{Im}\varphi^\omega$ by Proposition 4.8(1), we obtain the following result.

Corollary 4.12. *Every finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$ has at least one connection.* □

Assuming further that φ is special, we obtain more precise conclusions.

Theorem 4.13. *Let φ be a special, finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$.*

1. *Let H be a maximal subgroup of J_φ such that $H \cap \operatorname{Im}\varphi^\omega \neq \varnothing$. Then there is a connection ba for φ such that $H = \overline{\langle \varphi^\omega(X_\varphi(a,b))\rangle}$.*
2. *Let ba be a connection for φ. Then $\overline{\varphi^\omega(X_\varphi(a,b))^+}$ is a maximal subgroup of J_φ.*

Proof. Given a connection ba for φ, by Proposition 4.8(2), $\overline{\varphi^\omega(X_\varphi(a,b))^+}$ is contained in a maximal subgroup of J_φ and therefore it coincides with $\overline{\langle \varphi^\omega(X_\varphi(a,b))\rangle}$, since $x^{\omega-1} = \lim_{n\to\infty} x^{n!-1}$. If we take H to be this maximal subgroup and prove (1), then (2) will follow by the uniqueness part of Lemma 4.11. So we proceed with a maximal subgroup as in (1). By the existence part of Lemma 4.11, we already have the inclusion $\overline{\langle \varphi^\omega(X_\varphi(a,b))\rangle} \subseteq H$ for a connection ba for φ. On the other hand, by Proposition 4.8(2), $\overline{\langle \varphi^\omega(X_\varphi(a,b))\rangle} = \varphi^\omega(H)$. Finally, by Proposition 4.10, we have $H = \varphi^\omega(H)$, which establishes (1). □

Let ba be a connection for a substitution φ satisfying the hypothesis of Theorem 4.13. The maximal subgroup of J_φ generated, as a topological group, by the set $\varphi^\omega(X_\varphi(a,b))$ is called the *maximal subgroup associated with ba*.

5. Ultimately G-Invertible Substitutions

We continue to assume here the general assumptions of Sec. 4, namely that φ is a finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$, where A is a finite alphabet.

We say that a weakly primitive substitution φ is *ultimately* G*-invertible* if there exists $\psi \in \operatorname{End} \widehat{FG_A}$ such that $\psi \circ p_{\mathsf{G}} \circ \varphi$ sends each letter $a \in c(\varphi^\omega)$ to the generator $p_{\mathsf{G}}(a)$. If we let $B = c(\varphi^\omega)$, then φ induces a continuous endomorphism φ' of $\widehat{FG_B}$ by $\varphi'(b) = p_{\mathsf{G}} \circ \varphi(b)$ for $b \in B$. Note that φ is ultimately G-invertible if and only if $(\varphi')^\omega$ is the identity transformation of $\widehat{FG_B}$ or, in other words, φ' has an inverse in the profinite monoid $\operatorname{End} \widehat{FG_B}$. In the case where $B = A$, we shall also say that φ is G*-invertible*. Note that, if φ and ψ are finite G-invertible continuous endomorphisms of $\widehat{A^*}$, then so is their composite $\varphi\psi$.

Without further reference, we shall view $\widehat{FG_B}$ as the closed subgroup of $\widehat{FG_A}$ generated by B. We shall also view A^* as being naturally embedded in $\widehat{FG_A}$, namely as the submonoid generated by A.

Proposition 5.1. *Let φ be a special, finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$ and let $B = c(\varphi^\omega)$. Suppose φ is ultimately G-invertible and let ba be a connection for φ with an associated maximal subgroup H. Then the image of the restriction to H of the natural projection $p_{\mathsf{G}} \colon \widehat{A^*} \to \widehat{FG_A}$ is the closed subgroup generated by $Y_\varphi(a, b)$.*

Proof. Since $X_\varphi(a, b) \subseteq Y_\varphi(a, b)$ and $\varphi^\omega(Y_\varphi(a, b)) \subseteq H$, we see that H is also generated, as a topological group, by $\varphi^\omega(Y_\varphi(a, b))$. Denote by ψ the unique continuous endomorphism of $\widehat{FG_A}$ that sends each $a \in A$ to the positive word $\varphi(a)$. Note that $p_{\mathsf{G}} \circ \varphi = \psi \circ p_{\mathsf{G}}$, from which it follows that the restriction of $p_{\mathsf{G}} \circ \varphi^\omega = \psi^\omega \circ p_{\mathsf{G}}$ to $\widehat{B^*}$ is the same as that of p_{G}, since φ is ultimately G-invertible. Hence the closed subgroup of $\widehat{FG_A}$ given by $p_{\mathsf{G}}(H)$ is that generated by $Y_\varphi(a, b)$. □

Let X be a finite nonempty subset of A^+. An *elementary splitting* performed over X consists of finding in X a pair of distinct words of one of the forms x, xy or x, yx and producing the set X', which is obtained from X by replacing the element xy, respectively yx, by y. Then we obviously have $\langle X \rangle = \langle X' \rangle$ and $X^* \subseteq (X')^*$. Note that $\sum_{x \in X} |x| > \sum_{x' \in X'} |x'|$, and so any sequence of elementary splittings performed on X must eventually lead to a finite subset Y of A^+ upon which no elementary splitting is possible, i.e., Y is a finite biprefix code. Moreover, it is easy to see that the elementary splitting transformation is locally confluent in the sense that, when two distinct elementary splitting transformations are applied to a set X to produce sets X' and X'', then there is a set Z that can be obtained from both X' and X'' by applying elementary splittings. Hence there is a unique biprefix code $\tilde{X}$ that can be obtained from X by applying a sequence of elementary splittings. We call it the *split code* of X. Taking also into account that if $w(a_1, \ldots, a_n)$ is a nontrivial reduced group word and we substitute for each a_i a distinct element u_i of a biprefix code, then the group word $w(u_1, \ldots, u_n)$ cannot be reduced to the empty word, we obtain the following result, which amounts to a simple and probably folkloric exercise in combinatorial group theory.

Proposition 5.2. *Let X be a finite nonempty subset of A^+. Then the split code $\tilde{X}$ is a set of free generators of the subgroup $\langle X \rangle$ of the free group FG_A and $X^* \subseteq \tilde{X}^*$.* □

We are ready for one of the main results of this paper.

Theorem 5.3. *Let φ be a special, finite, weakly primitive, continuous endomorphism of $\widehat{A^*}$. Suppose φ is ultimately G-invertible and let ba be a connection for φ with associated maximal subgroup H. Then the mapping $\chi \colon H \to \widehat{FG_A}$ obtained by restriction of the natural projection $p_{\mathsf{G}} \colon \widehat{A^*} \to \widehat{FG_A}$ is an isomorphism from H onto the closed subgroup generated by $Y_\varphi(a, b)$, which is a finitely generated, free profinite group with set of free generators $\widetilde{Y_\varphi(a, b)}$.*

Proof. By Proposition 5.1 and its proof, $\mathsf{G}(\varphi^\omega(u)) = u$ for $u \in A^*$. Let $K = \langle Y_\varphi(a, b) \rangle$. By Proposition 5.2, K is a free group on the set $\widetilde{Y_\varphi(a, b)}$. By a result of Coulbois, Sapir, and Weil [15, Theorem 1.1] applied

to the pseudovariety G of all finite groups, the closure $\overline{K}$ of K in $\widehat{FG_A}$ is a free profinite group on $\widetilde{Y_\varphi(a,b)}$. This already establishes that $\operatorname{Im}\chi$ is a finitely generated, free profinite group. It remains to show that χ is injective. For this purpose, it suffices to show that H is generated as a closed subgroup by elements that map into $\widetilde{Y_\varphi(a,b)}$ under χ.

The slight difficulty in the proof at this point is that $\varphi^\omega(v)$ may not belong to H for $v \in \widetilde{Y_\varphi(a,b)}$. To overcome this difficulty, we have to exhibit a modified pseudoword v' such that $\varphi^\omega(v') \in H$ and $\chi(\varphi^\omega(v')) = v$. This is simply done by mimicking in the group H the sequence of cancellations leading from the set $Y_\varphi(a,b)$ to $\widetilde{Y_\varphi(a,b)}$ by using the $(\omega-1)$-power operation. More precisely, suppose that X is a set of generators of H, as a topological group, and that x and y are two distinct elements of X. Then, by replacing y by $x^{\omega-1}y$ or by $yx^{\omega-1}$, we obtain another subset of H that is still a set of generators of H, as a topological group. Since χ is a group homomorphism and the $(\omega-1)$-power in $\widehat{FG_A}$ coincides with inversion, the result follows. □

Recall from [13] the notion of a circular code. A subset C of A^+ is a *circular code* if, whenever $p \in A^*$, $s \in A^+$, $c_1, \ldots, c_m, d_1, \ldots, d_n \in C$, $sc_2 \cdots c_m p = d_1 \cdots d_n$, and $c_1 = ps$, we have $m = n$, $p = 1$, and $c_i = d_i$ $(1 \le i \le n)$. Equivalently, the submonoid C^* of the free monoid A^* is *very pure*, in the sense that $uv, vu \in C^*$ implies $u, v \in C^*$, and C is a minimal set of generators of C^*. We say that $\varphi \in \operatorname{End}\widehat{A^*}$ is a *circular encoding* if $\varphi|_B$ is injective and $\varphi(B)$ is a circular code, where $B = c(\varphi^\omega)$.

Proposition 5.4. *Let φ be a finite, continuous, ultimately G-invertible endomorphism of $\widehat{A^*}$. Then φ is a circular encoding.*

Proof. Let $B = c(\varphi^\omega)$. It follows from Proposition 5.2 that $\widetilde{\varphi(B)} = B$ and so there is a sequence of elementary splittings starting from $\varphi(B)$ that ends in the set B, which is certainly a circular code. The idea of the proof is to trace back through elementary splittings and show that, at each stage, a circular code is obtained. To prove this fact we use a well-known result from the theory of codes, namely that the composition of two circular codes is again a circular code [13, Proposition 1.9]. Indeed, if $X = \{x_1, x_2, \ldots, x_n\}$ is a circular code, then $\{x_1x_2, x_2, \ldots, x_n\}$ is obtained by composing the code $\{d_1d_2, d_2, \ldots, d_n\}$, over the alphabet $D = \{d_1, d_2, \ldots, d_n\}$ with X and similarly for the dual construction $\{x_2x_1, x_2, \ldots, x_n\}$. Thus, it suffices to show that $Z = \{d_1d_2, d_2, \ldots, d_n\}$ (along with its dual) is a circular code, which corresponds to the very first step of the announced trace-back procedure.

Now, Z is even a prefix code, so it suffices to show that Z^* is a very pure submonoid of D^*. Indeed, if $u, v \in D^*$ and $uv, vu \in Z^*$, then $u, v \in Z^*$, since, for instance, if in u not every occurrence of the letter d_1 is followed by d_2, then the same holds for vu, which contradicts $vu \in Z^*$. □

The following is a variation of several similar results, which can be found in the literature on the algebraic theory of codes with various definitions of delay [13, Sec. VII.2]. Since none seems to involve precisely the definition of language of bounded delay that has been adopted in this paper, we provide a proof for the sake of completeness.

Lemma 5.5. *Every finite circular code has bounded delay.*

Proof. Let C be a finite circular code over the alphabet A. Then there is an upper bound on the number of different overlaps between words in C, meaning factorizations of the form $xc = c'y$ with $c, c' \in C$, $x \in A^*$, $y \in A^+$, and $|x| < |c'|$ as is depicted in Fig. 4. Hence, provided $m + n$ is sufficiently large, if we

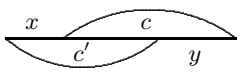


Fig. 4. An overlap between two code words.

have two factorizations

$$uc_1 \cdots c_m v = c'_1 \cdots c'_n \tag{5.1}$$

with $c_i, c'_j \in C$, then, considering the overlaps between factors c_i and c'_j, there will be at least two equal overlaps, which leads to equalities of the following forms: $sc_ic_{i+1}\cdots c_{i+r-1}p = c'_jc'_{j+1}\cdots c'_{j+t}$ and $ps = c_{i+r}$, where $s \neq 1$. Since C is a circular code, it follows that $p = 1$, which shows that the equality (5.1) is reducible. The other type of equality involved in the definition of language of bounded delay is handled similarly. □

Corollary 5.6. *Let φ be a finite, ultimately G-invertible, weakly primitive, continuous endomorphism of $\widehat{A^*}$. Then φ is special.*

Proof. This follows directly from Proposition 5.4 and Lemma 5.5. □

This leads to a simplified formulation of Theorem 5.3 without explicit reference to the technical hypothesis that the endomorphism be special.

Corollary 5.7. *Let φ be a finite, ultimately G-invertible, weakly primitive, continuous endomorphism of $\widehat{A^*}$. Then the maximal subgroups of J_φ are finitely generated, free profinite groups.*

Proof. By Corollary 5.6, φ is special. By Theorem 5.3, there is a maximal subgroup H of J_φ such that the restriction to H of the natural projection $p_{\mathsf{G}}\colon \widehat{A^*} \to \widehat{FG_A}$ is an embedding that sends H onto a finitely generated, free profinite group. □

Another consequence of Theorem 5.3 is the following result, for which we have found no direct proof.

Corollary 5.8. *Let φ be a finite, ultimately G-invertible, weakly primitive, continuous endomorphism of $\widehat{A^*}$ and let ba be a connection for φ. Then we have the following equality of subgroups of FG_A: $\langle X_\varphi(a,b)\rangle = \langle Y_\varphi(a,b)\rangle$ or, equivalently, $\widetilde{X_\varphi(a,b)} = \widetilde{Y_\varphi(a,b)}$.*

Proof. Let H be the maximal subgroup containing $\varphi^\omega(X_\varphi(a,b))$. By Theorem 4.13(2), H is generated, as a topological group, by $\varphi^\omega(X_\varphi(a,b))$. Let $\chi\colon H \to \widehat{FG_A}$ be the restriction of the natural projection $p_{\mathsf{G}}\colon \widehat{A^*} \to \widehat{FG_A}$ to H. By Theorem 5.3, we obtain the equality $\chi(H) = \overline{\langle Y_\varphi(a,b)\rangle}$. On the other hand, since χ is a continuous homomorphism and $H = \overline{\varphi^\omega(X_\varphi(a,b))^+}$, we also have the equality $\chi(H) = \overline{\langle X_\varphi(a,b)\rangle}$. This shows that

$$\overline{\langle X_\varphi(a,b)\rangle} = \overline{\langle Y_\varphi(a,b)\rangle}. \tag{5.2}$$

Finally, the result follows by M. Hall's theorem [16], since the finitely generated subgroups $\langle X_\varphi(a,b)\rangle$ and $\langle Y_\varphi(a,b)\rangle$ of the group FG_A are closed in the profinite topology and they have the same closure by (5.2). □

We conclude this section with a couple of examples. The first shows that the conclusion of Corollary 5.8 may fail without the assumption that φ be ultimately G-invertible. The second example illustrates the computation of a maximal subgroup. In both cases, we take $A = \{a, b, c\}$.

Example 5.9. Let φ be the substitution defined by $\varphi(a) = bac$, $\varphi(b) = cba$, and $\varphi(c) = acb$. Note that this is a primitive substitution and that ab is a connection for φ. Using Lemma 4.5, one computes

$$X_\varphi(b,a) = \{bacacbacbcba, bacacbbacacbcba, bacacbbacacbcbaacbcba, bacacbbacacbcbacba,$$
$$bacacbcba, bacacbcbaacbcba, bacbacacbcbaacbcba, baccba\},$$

from which it follows that $\widetilde{Y_\varphi(b,a)}$ contains the elements ba, bba, and $bcba$, which generate the free group FG_A. On the other hand, $\widetilde{X_\varphi(b,a)} = \{acb, bac, cba\}$, which coincides with $\varphi(A)$.

Example 5.10. Consider the continuous endomorphism of $\widehat{A^*}$ defined by $\varphi(a) = bcac$, $\varphi(b) = bcacbc$, and $\varphi(c) = cbcbcac$. A little computation using Lemma 4.5 shows that cb is a connection for φ and that $X_\varphi(b,c) = \{bc, bcc, bcac, bcacc\}$. Hence $\widetilde{X_\varphi(b,c)} = \{a,b,c\}$. By Theorem 5.3 and its proof, the maximal subgroup containing $\varphi^\omega(bc)$ is a free profinite group that is freely generated by the set

$$\Big\{\varphi^\omega\Big(((bc)^\omega c)^{\omega-1}\cdot\big(bc((bc)^\omega c)^{\omega-1}\big)^{\omega-1}\cdot bcac\cdot((bc)^\omega c)^{\omega-1}\Big), \varphi^\omega\big(bc\cdot((bc)^\omega c)^{\omega-1}\big), \varphi^\omega\big((bc)^{\omega-1}\cdot bcc\big)\Big\}.$$

Example 5.10 illustrates the computation of a specific maximal subgroup of the $\mathcal{J}$-class associated with a finite weakly primitive substitution. The computation is effective in the sense that a connection ba and the associated finite set of words $Y_\varphi(a,b)$ can be effectively computed. In the case where φ is ultimately G-invertible, the computation of the split code $\widetilde{Y_\varphi(a,b)}$ can be turned into the description of a set of free generators, as a topological group, of the maximal subgroup H of J_φ containing $\varphi^\omega(Y_\varphi(a,b))$. Moreover, taking into account that, in a profinite monoid, $u^{\omega-1} = \lim_{n\to\infty} u^{n!-1}$, we conclude that the computed free generators of H determine effectively computable "implicit operations." See [10], particularly its Proposition 4.5 and the subsequent remarks, for related computability and complexity questions.

6. Sturmian and Arnoux–Rauzy $\mathcal{J}$-Classes Generated by Substitutions

This section is dedicated to some important special cases of application of Theorem 5.3 in which it is possible to be more precise about the number of free generators of the maximal subgroups of the $\mathcal{J}$-class associated with a finite, ultimately G-invertible, weakly primitive, continuous endomorphism φ of $\widehat{A^*}$ without the need for a case-by-case computation. The results in this section have been previously announced, without proof, in [7].

The first simple application is obtained by considering Sturmian substitution subshifts.[1] Such subshifts are obtained by iterating primitive endomorphisms φ of the free monoid $\{a,b\}^*$ that are G-invertible, which are also known as *Sturmian substitutions*. Since the finite factors of J_φ must then be *balanced*, in the sense that factors of the same length cannot differ by more than 1 in the number of occurrences of a given letter and the number of factors of length n is $n+1$ [18, Chap. 2], it follows that the word ba is a factor and $X_\varphi(a,b)$ contains exactly two elements: if aa is not a factor, then the elements of $X_\varphi(a,b)$ are ab^n and ab^{n+1} for some $n \geq 1$; if bb is not a factor, then the elements are a^nb and $a^{n+1}b$ for some $n \geq 1$. In either case, $X_\varphi(a,b)$ generates the free group $FG_{\{a,b\}}$, and so Theorem 5.3 applies in the case where ba is a connection for φ. The case in which ab is a connection is dual. The case in which one of the words aa or bb is a connection is even easier. Indeed, say in the case where aa is a connection, $Y_\varphi(a,a)$ contains both a and the word aba (since, otherwise, the set of finite factors of J_φ would not be balanced), which shows that $Y_\varphi(a,a)$ generates the free group $FG_{\{a,b\}}$.

Corollary 6.1. *Let φ be a Sturmian substitution. Then the maximal subgroups of J_φ are free profinite groups on two free generators.* □

We have also announced in [7] that it follows that, for the maximal regular $\mathcal{J}$-class of $\widehat{\{a,b\}^*}$ associated with an arbitrary a Sturmian subshift over the alphabet $\{a,b\}$, the maximal subgroups are also free profinite groups on two free generators.

A generalization of Sturmian subshift proposed by Arnoux and Rauzy (see [14]) may be defined as follows. We first consider the *Arnoux–Rauzy* homomorphism

$$\rho\colon \widehat{A^*} \to \operatorname{End}\widehat{A^*},$$
$$w \mapsto \rho_w,$$

[1] By a *subshift* we mean a symbolic dynamic system over a finite alphabet A, i.e., a closed subset of $A^{\mathbb{Z}}$ that is stable under all shifts of the origin. A *substitution* subshift is one that is generated by iteration of a substitution in a sense that is described in detail in [14].

which is defined by the following formula for $a, b \in A$:

$$\rho_a(b) = \begin{cases} a & \text{if } a = b, \\ ab & \text{otherwise.} \end{cases}$$

We say that a word $u \in A^*$ has *full content* if $c(u) = A$.

Lemma 6.2. *Let $u \in A^*$ be a finite word with full content.*

(1) *The mapping ρ_u is a finite, G-invertible, primitive, continuous endomorphism of $\widehat{A^*}$.*
(2) *There is a connection ba for ρ_u such that the set $Y_{\rho_u}(a, b)$ generates the free group FG_A.*

Proof.

(1) Since each ρ_a, with $a \in A$, is G-invertible, so is ρ_u. To prove (1) it remains to show that ρ_u is primitive, which follows from the observation that, for $u, v \in A^*$,

$$c(\rho_u(v)) = c(u) \cup c(v), \tag{6.1}$$

which in turn is easily established by induction on the length of u.

(2) Let a be the first letter of the word u. Then $\operatorname{Im} \rho_{u^n} \subseteq \operatorname{Im} \rho_a$ and, since every finite factor of $\rho_u^\omega(a) = \rho_{u^\omega}(a)$ is a factor of all words of the form $\rho_{u^{n!}}(a)$ for sufficiently large n, such a factor must be a factor of $\rho_a(v)$ for some word v. Now, $\operatorname{Im}(\rho_a|_{A^*})$ is the submonoid of A^* generated by the words of the form ab, where $b \in A \setminus \{a\}$, together with the letter a. Moreover, by the content formula (6.1), if $u = au'$, then a occurs in $\rho_{u'u^n}(a)$ followed by some other letter for all $n \geq 1$, and so aa is a factor of $\rho_{u^n}(a)$ for $n > 1$. Note also that $\rho_{u^n}(a)$ starts and ends with the letter a. Hence the factor aa is a connection for ρ_u.

It remains to show that the set $Y_{\rho_u}(a, a)$ generates the free group FG_A. By definition of the set $Y_{\rho_u}(a, a)$, since its elements must be factors of words that are products of a and 2-letter words ab ($b \in A \setminus \{a\}$), $Y_{\rho_u}(a, a)$ contains both the letter a and all words of the form aba with $b \in A \setminus \{a\}$. From such words, by canceling the letter a, we obtain all other letters from A, and hence $Y_{\rho_u}(a, a)$ generates FG_A. □

Applying Theorem 5.3, we thus obtain the following result, of which Corollary 6.1 may be shown to be a consequence [7].

Corollary 6.3. *Let $u \in A^*$ be a word with full content. Then the maximal subgroups of the maximal regular $\mathcal{J}$-class J_{ρ_u} are free profinite groups on $|A|$ free generators.*

The subshifts corresponding to the $\mathcal{J}$-classes appearing in Corollary 6.3 are known as *Arnoux–Rauzy subshifts* (generated by substitutions) [7, 14, 18] and constitute a generalization of subshifts generated by Sturmian substitutions. More generally, a subshift (or a pseudoword) over the finite alphabet A is said to be Arnoux–Rauzy if, for every $n \geq 1$, it has precisely one right special and one left special factor of length n, each of degree $|A|$. Here, a factor u is said to be *right special of degree d* if there are precisely $d > 1$ letters $a \in A$ such that ua is still a factor, and a *left special factor of degree d* is defined dually. As has been argued in [7], it follows from results on right infinite words, i.e., words in $A^{\mathbb{N}}$, that every Arnoux–Rauzy pseudoword is $\mathcal{J}$-equivalent to some pseudoword of the form $\rho_v(a)$, where $v \in \widehat{A^*}$, in which every letter appears an unbounded number of times in the finite prefixes of v [12,14]. In fact, the sequence of the finite prefixes of v, i.e., a right infinite word, suffices to determine the $\mathcal{J}$-class. The Arnoux–Rauzy subshifts generated by substitutions correspond to the case where the right infinite word in question is periodic.

How to extend Corollary 6.3 to maximal subgroups of $\mathcal{J}$-classes associated with arbitrary Arnoux–Rauzy subshifts, not necessarily generated by the infinite iteration of a finite endomorphism is sketched in [7].

7. Some Examples

In this section, we present some examples to illustrate what happens beyond the nicer cases of Secs. 5 and 6, while the full picture for the calculation of $\mathcal{J}$-maximal subgroups of $\widehat{A^*}$ remains open.

Example 7.1. Let $A = \{a, b, c\}$ and let $\varphi \in \operatorname{End}\widehat{A^*}$ be defined by $\varphi(a) = bac$, $\varphi(b) = cbac$, and $\varphi(c) = bacb$. Then bc is a factor of the word $\varphi^2(b) = bacbcbacbacbacb$ and therefore also of φ^ω applied to any letter, since φ is primitive. A little calculation using Lemma 4.5 shows that bc is a connection for φ and that $X_\varphi(c, b) = \{cbacbacb, cbacbacbacb\}$, whence $Y_\varphi(c, b) = \{(cba)^n cb\colon n = 0, 1, 2, 3\}$. Hence the subgroup of FG_A generated by $Y_\varphi(c, b)$ is also generated by $\{a, cb\}$. Note that φ is G-invertible. Hence, by Theorem 5.3, the maximal subgroups of J_φ are free profinite groups on two generators.

Example 7.2. Let $A = \{a, b\}$ and consider the continuous endomorphism of $\widehat{A^*}$ defined by $\varphi(a) = ab$ and $\varphi(b) = a^3b$. Note that φ is a finite primitive substitution but that it is not G-invertible. We claim that the maximal subgroups of J_φ are not free profinite groups.

By [10, Proposition 4.3], the pseudowords $\varphi^\omega(a)$ and $\varphi^\omega(b)$ lie in the same maximal subgroup H of J_φ. Hence $\operatorname{Im}\varphi^\omega \subseteq H$. Note that φ is a prefix encoding with delay 1, whence φ is special. By Proposition 4.10, we conclude that $H = \operatorname{Im}\varphi^\omega$. On the other hand, by Theorem 4.13, the maximal subgroup H is generated, as a closed subgroup, by the set $\varphi^\omega(X_\varphi(a, b))$. Since $\varphi(A^*) = \{ab, a^3b\}$, it is easy to deduce that $X_\varphi(a, b) = \{ab, a^3b\}$. Hence H is generated, as a closed subgroup, by the set $\{\varphi^\omega(ab), \varphi^\omega(a^3b)\}$.

If the profinite group H were free on two generators, then it would also have to be free on any two generators. In particular, H would be freely generated, as a profinite group, by both the pair $\varphi^\omega(ab)$, $\varphi^\omega(a^3b)$ and the pair $\varphi^\omega(a)$, $\varphi^\omega(b)$. Hence, there is a continuous homomorphism $\psi\colon H \to \mathbb{Z}/2\mathbb{Z}$ that maps both $\varphi^\omega(a)$ and $\varphi^\omega(b)$ to 1 (which denotes here the generator of the additive group $\mathbb{Z}/2\mathbb{Z}$). Then ψ maps both $\varphi^\omega(ab)$ and $\varphi^\omega(a^3b)$ to $0 = 1 + 1$, which contradicts the already established fact that those two pseudowords generate a dense subgroup of H. Hence H is not a free group on two generators.

To conclude the proof that H is a not a free profinite group, it suffices to show that it is not a procyclic group. In order to establish this property, we consider the continuous homomorphism $\theta\colon \widehat{A^*} \to \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ that sends a to $(1, 0)$ and b to $(0, 1)$, where 1 denotes the usual generator of $\mathbb{Z}/3\mathbb{Z}$. Then $\theta(\varphi^n(a))$ and $\theta(\varphi^n(b))$ are easily seen to be, respectively, the first and second columns of the matrix M^n, computed over the field $\mathbb{Z}/3\mathbb{Z}$, where

$$M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Since M^3 is the identity matrix, it follows that $\theta(\varphi^\omega(a)) = (1, 0)$ and $\theta(\varphi^\omega(b)) = (0, 1)$. Hence H is not a procyclic group, since it has a noncyclic finite group as a continuous homomorphic image. This completes the proof of the claim that the maximal subgroup H of J_φ is not a free profinite group.

The previous example seems to be the first non-free-profinite maximal subgroup of a finitely generated free profinite monoid to be exhibited.

Example 7.3. Let $A = \{a, b\}$ and let φ be the continuous endomorphism of $\widehat{A^*}$ determined by $\varphi(a) = ab$ and $\varphi(b) = ba$. This is the well-known Prouhet–Thue–Morse substitution [14]. Consider the $\mathcal{J}$-class J_φ of $\alpha = \varphi^\omega(a)$. Note that a^2 is a connection for φ, whence α lies in a maximal subgroup H. A simple calculation shows that $X_\varphi(a, a) = \{a, aba, ab^2a\}$. While φ is not of bounded delay, it is of relative bounded delay and so it is special. Hence, by the general theory, H is the closure of the subgroup generated by $\alpha = \varphi^\omega(a)$, $\beta = \varphi^\omega(aba)$, and $\gamma = \varphi^\omega(ab^2a)$. We claim that α, β, and γ are not free generators of H, from which it follows that H is not a free profinite group on three generators, although it might still be a free profinite group on fewer generators. To prove the claim, consider the continuous endomorphism φ^2 of $\widehat{A^*}$, which is given by $\varphi^2(a) = abba$ and $\varphi^2(b) = baab$. Since members of H are of the form $\varphi^\omega(ava)$ for some pseudoword v, their images under φ^2 are of the form $\varphi^\omega(abbav'abba)$. Since the latter belongs

to J_φ by Proposition 4.2, it follows that it belongs to H. Hence $\varphi^2(H) \subseteq H$. On the other hand, we have

$$\varphi^\omega(ava) = \varphi^2(\varphi^{\omega-2}(\varphi^\omega(ava))) = \varphi^2\Big(\lim_{n\to\infty} \varphi^{n!-2}(\varphi^\omega(ava))\Big).$$

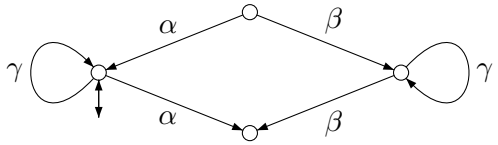
Since $\varphi^\omega(ava)$ is assumed to belong to H, $\varphi^{n!-2}$ is a power of φ^2 for $n \geq 3$, $\varphi^2(H) \subseteq H$, and H is closed, we deduce that $H = \varphi^2(H)$. Since φ^2 is injective by [19], it follows that φ^2 induces a continuous automorphism of H. Hence, if H were freely generated, as a profinite group, by α, β, and γ, then it would also be freely generated by their images under φ^2, which we now compute. Clearly $\varphi^2(\alpha) = \gamma$. On the other hand,

$$\varphi^2(\beta) = \varphi^\omega(abba\,baab\,abba) = \varphi^\omega(abba \cdot a^{\omega-1} \cdot aba \cdot aba \cdot a^{\omega-1} \cdot abba) = \gamma\alpha^{-1}\beta^2\alpha^{-1}\gamma$$

and, similarly,

$$\varphi^2(\gamma) = \varphi^\omega(abba\,baab\,baab\,abba) = \varphi^\omega(abba \cdot a^{\omega-1} \cdot aba \cdot abba \cdot aba \cdot a^{\omega-1} \cdot abba) = \gamma\alpha^{-1}\beta\gamma\beta\alpha^{-1}\gamma.$$

Thus, φ^2 sends the subgroup H_0 (discretely) generated by α, β, and γ to a subgroup of itself. Computing the minimal inverse automaton recognizing this subgroup, we obtain the following automaton:



Hence $\varphi^2\{\alpha, \beta, \gamma\}$ generates a proper subgroup K of H_0. Under the assumption that H is freely generated, as a profinite group, by α, β, and γ, H_0 would be freely generated, as a discrete group, by the same elements. Hence, by Hall's theorem [16], K is closed in the profinite topology of H_0, which is just the induced topology from H. This is a contradiction, since $\varphi^2\{\alpha, \beta, \gamma\}$ has been shown to generate a dense subgroup of H. Hence H is not freely generated by α, β, and γ.

Acknowledgments. The author wishes to acknowledge fruitful discussions on some of the ideas reflected in this work with Alfredo Costa, Kunitaka Shoji, and Benjamin Steinberg. Many thanks also to Alfredo Costa and Benjamin Steinberg for their comments on earlier versions of this paper and to Mikhail V. Volkov for the references to Wielandt's and Markowsky's papers.

This work was supported, in part, by *Fundação para a Ciência e a Tecnologia* (FCT) through the *Centro de Matemática da Universidade do Porto* and by the FCT and POCTI approved project POCTI/32817/MAT/2000, which is partly funded by the European Community Fund FEDER. It was done in part while the author was visiting the LIAFA at the University Denis Diderot (Paris 7), whose hospitality is gratefully acknowledged, benefiting from a sabbatical fellowship from FCT.

REFERENCES

1. J. Almeida, "Some algorithms on the star operation applied to finite languages," *Semigroup Forum*, **28**, 187–197 (1984).
2. J. Almeida, *Finite Semigroups and Universal Algebra*, World Scientific, Singapore (1995).
3. J. Almeida, "Dynamics of implicit operations and tameness of pseudovarieties of groups," *Trans. Amer. Math. Soc.*, **354**, 387–411 (2002).
4. J. Almeida, "Finite semigroups: An introduction to a unified theory of pseudovarieties," in: G. M. S. Gomes, J.-E. Pin, and P. V. Silva, eds., *Semigroups, Algorithms, Automata and Languages*, World Scientific, Singapore (2002), pp. 3–64.
5. J. Almeida, *Profinite Semigroups and Applications*, Tech. Rep. CMUP 2003-33, Univ. Porto, 2003.
6. J. Almeida, "Profinite structures and dynamics," *CIM Bulletin*, **14**, 8–18 (2003).
7. J. Almeida, "Symbolic dynamics in free profinite semigroups," in: *No. 1366 in RIMS Kokyuroku, Kyoto, Japan, April 2004*, pp. 1–12.

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, eds., *Algorithmic Problems in Groups and Semigroups*, Birkhäuser, 2000, pp. 1–23.
10. J. Almeida and M. V. Volkov, "Subword complexity of profinite words and subgroups of free profinite semigroups," *Internat. J. Algebra Comput.*, to appear.
11. J. Almeida and P. Weil, "Relatively free profinite monoids: An introduction and examples," in: J. B. Fountain, ed., *Semigroups, Formal Languages and Groups*, Vol. 466, Kluwer Academic, Dordrecht (1995), pp. 73–117.
12. J. Berstel, "Recent results on extensions of Sturmian words," *Internat. J. Algebra Comput.*, **12**, 371–385 (2002).
13. J. Berstel and D. Perrin, *Theory of Codes*, Academic Press, New York (1985).
14. V. Berthé, S. Ferenczi, C. Mauduit, and A. Siegel, eds., *Introduction to Finite Automata and Substitution Dynamical Systems* (2001), `http://iml.univ-mrs.fr/editions/preprint00/book/prebookdac.html`.
15. T. Coulbois, M. Sapir, and P. Weil, "A note on the continuous extensions of injective morphisms between free groups to relatively free profinite groups," *Publ. Mat.*, **47**, 477–487 (2003).
16. M. Hall, "A topology for free groups and related groups," *Ann. Math.*, **52**, 127–139 (1950).
17. G. Lallement, *Semigroups and Combinatorial Applications*, Wiley, New York (1979).
18. M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, Cambridge (2002).
19. S. Margolis, M. Sapir, and P. Weil, "Irreducibility of certain pseudovarieties," *Comm. Algebra*, **26**, 779–792 (1998).
20. G. Markowsky, "Bounds on the index and period of a binary relation on a finite set," *Semigroup Forum*, **13**, 253–259 (1976).
21. M. Queffélec, *Substitution Dynamical Systems—Spectral Analysis*, Lect. Notes Math., Vol. 1294, Springer, Berlin (1987).
22. J.-C. Spehner, "Quelques constructions et algorithmes relatifs aux sous-monoïdes d'un monoïde libre," *Semigroup Forum*, **9**, 334–353 (1975).
23. P. Weil, "Profinite methods in semigroup theory," *Internat. J. Algebra Comput.*, **12**, 137–178 (2002).
24. H. Wielandt, "Unzerlegbare, nicht negative Matrizen," *Math. Z.*, **52**, 642–648 (1950).

Jorge Almeida
Centro de Matemática da Universidade do Porto, Departamento de Matemática Pura, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
E-mail: `jalmeida@fc.up.pt`