HD0L ω-EQUIVALENCE AND PERIODICITY PROBLEMS IN THE PRIMITIVE CASE

Fabien Durand

ABSTRACT. In this paper I would like to witness the mathematical inventiveness of G. Rauzy through a personnal exchanges I had with him. The objects that will emerge will be used to treat the decidability of the HD0L ω-equivalence and periodicity problems in the primitive case.

Communicated by Robert Tichy

Dedicated to the memory of G. Rauzy

1. Introduction

1.1. Some words about Gérard Rauzy

I started my PhD in 1992 under the direction of G. Rauzy. The topic of my PhD was *Substitutions and Bratteli diagrams*. G. Rauzy was a worldwide recognized expert on substitutions (and many other topics) but Bratteli diagram representations of substitution subshifts was not one of these subjects. Thus I had an unformal second advisor, B. Host. Later, as a common agreement, he became my advisor.

Some days after I begun my PhD he told me that he noticed something amazing about substitutions. He wrote on the blackboard the Fibonacci fixed point on the alphabet $\{0, 1\}$ starting with 0. He considered all occurrences of the letter 0 and coded the words between two consecutive such occurrences. It is not surprising (see [Morse and Hedlund 1940]) that only two words appear: 01 and 0. If you respectively code them by 0 and 1 then you again obtain the same sequence. This is not surprising too because this sequence is the fixed point of the morphism defined by $0 \mapsto 01$ and $1 \mapsto 0$. Then he wrote a long prefix of the fixed

point of the Morse sequence that we call x. Applying the same coding process (see below) to x it clearly appeared that the new sequence called the *derived sequence of x*, and denoted by $\mathcal{D}(x)$, is not the same for the obvious reason that 3 different letters occur. He repeated the process to this new sequence. Let's do it. The result was a sequence on a 4 letter alphabet. Once again he repeated the process. Amazingly we obtained the same sequence (the same long prefix). Repeating the process will clearly produce the same sequence. The process is periodic.

x	$= 011$	01	0	011	0	01	011	01	0	01	011	0	011	01	$0\dots$
$\mathcal{D}(x)$	$= 0$	1	2	0	2	1	0	1	2	1	0	2	0	1	$2\dots$

$\mathcal{D}(x)$	$= 012$	021	0121	02	012	021	02	0121	012	021	0121	$02\dots$
$\mathcal{D}^2(x)$	$= 0$	1	2	3	0	1	3	2	0	1	2	$3\dots$

$\mathcal{D}^2(x)$	$= 0123$	0132	01232	013	0123	0132	013	01232	0123	$0132\dots$
$\mathcal{D}^3(x)$	$= 0$	1	2	3	0	1	3	2	0	$1\dots$

Then he showed me another example where the period was 2. He was wondering if such a period always exists for fixed points of primitive substitutions. He knew that when this process is eventually periodic then the sequence is primitive substitutive and he explained me the proof on the blackboard.

This proof was very useful for my PhD and for the present paper: this is the construction exposed in Proposition 17. It became a second subject in my PhD that I succeeded to solve in [Durand 1998a] and that provided me many arguments to obtain an extension of Cobham's theorem for primitive substitutions ([Durand 1998b]).

When he showed me this process I found curious to look at it. Why was it interesting ? In fact, at this time, I did not know much about the work of G. Rauzy and almost nothing about interval exchange transformations (iets) and what he did with. I realized quickly that this process was in fact the induction process defined on dynamical systems by the Poincaré first return map: the derived sequence $\mathcal{D}(x)$ generates a subshift that is conjugate to the induced dynamical system (X, T) on the cylinder set $[0]$, where (X, T) is the subshift generated by x. Thus, this corresponds to the induction process he defined [Rauzy 1979] to study iets. He induced on the left most interval and showed that the induced tranformation is again an iet with the same permutation. This is also the case for primitive substitutions : derived sequences of purely primitive substitutive sequences are purely primitive substitutive sequences ([Durand 1998a]). His goal

was to tackle the Keane's conjecture saying that almost all iets are uniquely ergodic. This was solved independently in [Masur 1982] and [Veech 1982]. Later M. Boshernitzan and C. R. Carroll [Boshernitzan and Carroll 1997] proved that, for an iet defined on a quadratic field, "every consistent method of induction is eventually periodic" (up to rescaling). This is what I obtained in [Durand 1998a] in the context of primitive substitutions and that was expected by G. Rauzy.

Thus I imagine the intuition of G. Rauzy was certainly led by iet considerations.

I am very grateful to Gérard Rauzy for the wonderful subject he gave me and for the opportunity he offered me to do research.

1.2. HD0L decidability problems

In this paper we apply what the author developped in [Durand 1998a] (to answer the G. Rauzy's problem on derived sequences presented above) and in [Durand 1998b] to solve the following decidability problems in the primitive case. To describe these problems we need some notations. Let $\sigma : A^* \to A^*$ and $\tau : B^* \to B^*$ be morphisms, and, u and v be words such that $\lim_{n \to \infty} \sigma^n(u)$ and $\lim_{n \to \infty} \tau^n(v)$ converge to respectively $x \in A^{\mathbb{N}}$ and $y \in B^{\mathbb{N}}$. We also need two morphisms $\phi : A^* \to C^*$ and $\psi : B^* \to D^*$.

1.2.1. The HD0L ω-equivalence problem

This problem is open for more than 30 years and asks whether it is decidable to know if two morphic sequences are equal. More precisely,

Is the equality "$\phi(x) = \psi(y)$" decidable?

In this paper we answer positively to this problem for primitive morphisms. The equality problem "$x = y$" (also called the D0L ω-equivalence problem) was solved in 1984 by K. Culik II and T. Harju [Culik and Harju 1984]. An other proof was given by J. Honkala in [Honkala 2007] (see also [Honkala 2009a]).

1.2.2. More on the equality of sequences generated by morphisms

Once the equality "$\phi(x) = \psi(y)$" or "$x = y$" is satisfied, it is natural to look for relations between the involved morphisms.

In [Séébold 1998] it is shown than when two invertible substitutions, σ and τ, on a two letter alphabet, $\{0, 1\}$, share the same fixed point ($\sigma^\infty(0) = \tau^\infty(0)$), then these substitutions are powers of some other morphism γ (for different proofs see also [Berthé, Frettlöh and Sirvent:preprint] and [Rao and Wen 2010]): there exist k and l such that $\sigma = \gamma^l$ and $\tau = \gamma^k$. For more general results on this topic see [Krieger 2008, Diekert and Krieger 2009].

We will give a proof of a weaker result but that can be extended to bigger alphabets: if two primitive substitutions share the same fixed point and satisfy some conditions on their return words then they have a common power.

1.2.3. The decidability of the periodicity of HD0L infinite sequences

This is also an open problem for more than 30 years. It ask:

Is it decidable to know whether $\phi(x)$ is ultimately periodic?

J.-J. Pansiot proved in [Pansiot 1986] that the periodicity of x is decidable. We answer positively to this problem in the primitive case and in a forthcoming paper the author should complete this work to close the problem.

In the sequel the vocabulary of D0L and HD0L systems will be replace by the vocabulary of substitutions.

1.3. Some definitions

1.3.1. Words and sequences

An *alphabet* A is a finite set of elements called *letters*. A *word* over A is an element of the free monoid generated by A, denoted by A^*. Let $x = x_0 x_1 \cdots x_{n-1}$ (with $x_i \in A$, $0 \le i \le n-1$) be a word, its *length* is n and is denoted by $|x|$. The *empty word* is denoted by ϵ, $|\epsilon| = 0$. The set of non-empty words over A is denoted by A^+. The elements of $A^{\mathbb{N}}$ are called *sequences*. If $x = x_0 x_1 \cdots$ is a sequence (with $x_i \in A$, $i \in \mathbb{N}$) and $I = [k, l]$ an interval of $\mathbb{N}$ we set $x_I = x_k x_{k+1} \cdots x_l$ and we say that x_I is a *factor* of x. If $k = 0$, we say that x_I is a *prefix* of x. The set of factors of length n of x is written $\mathcal{L}_n(x)$ and the set of factors of x, or the *language* of x, is denoted $\mathcal{L}(x)$. The *occurrences* in x of a word u are the integers i such that $x_{[i, i+|u|-1]} = u$. If u has an occurrence in x, we also say that u *appears* in x. When x is a word, we use the same terminology with similar definitions.

The sequence x is *ultimately periodic* if there exist a word u and a non-empty word v such that $x = uv^{\omega}$, where $v^{\omega} = vvv \cdots$. In this case $|v|$ is called a *length period* of x. It is *periodic* if u is the empty word. A word u is *recurrent* in x if it appears in x infinitely many times. A sequence x is *uniformly recurrent* if every factor u of x appears infinitely often in x and the greatest difference of two successive occurrences of u is bounded.

1.3.2. Morphisms and matrices

Let A and B be two alphabets. Let σ be a *morphism* from A^* to B^*. When $\sigma(A) = B$, we say σ is a *coding*. Thus, codings are onto. If $\sigma(A)$ is included in B^+,

it induces by concatenation a map from $A^{\mathbb{N}}$ to $B^{\mathbb{N}}$. These two maps are also called σ. To the morphism σ is naturally associated the matrix $M_\sigma = (m_{i,j})_{i \in B, j \in A}$ where $m_{i,j}$ is the number of occurrences of i in the word $\sigma(j)$.

1.3.3. Substitutions and substitutive sequences

A *substitution* is a morphism $\sigma : A^* \to A^*$. If there exists a letter $a \in A$ and a word $u \in A^+$ such that $\sigma(a) = au$ and moreover, if $\lim_{n \to +\infty} |\sigma^n(a)| = +\infty$, then σ is said to be *prolongable on a*.

Since for all $n \in \mathbb{N}$, $\sigma^n(a)$ is a prefix of $\sigma^{n+1}(a)$ and because $|\sigma^n(a)|$ tends to infinity with n, the sequence $(\sigma^n(aaa\cdots))_{n \geq 0}$ converges (for the usual product topology on $A^{\mathbb{N}}$) to a sequence denoted by $\sigma^\omega(a)$. The morphism σ being continuous for the product topology, $\sigma^\omega(a)$ is a fixed point of σ: $\sigma(\sigma^\omega(a)) = \sigma^\omega(a)$. A sequence obtained in this way by iterating a prolongable substitution is said to be *purely substitutive* (w.r.t. σ). If $x \in A^{\mathbb{N}}$ is purely substitutive and $\phi : A^* \to B^*$ is a morphism then the sequence $y = \phi(x)$ is said to be a *morphic sequence* (w.r.t. σ). When ϕ is a coding, we say y is *substitutive* (w.r.t. σ).

Whenever the matrix associated to σ is primitive (i.e., when it has a power with strictly positive coefficients) we say that σ is a *primitive substitution*. We say σ is *erasing* if there exists $b \in A$ such that $\sigma(b)$ is the empty word.

2. Substitutions and return words

The main results of the paper are consequences of results in [Durand 1998a] (see also [Durand 2000]) taking care to be explicit on some bounds (Section 2.1). Since we are dealing with decidability problems, it is important to pay attention that these bounds are algorithmically computable because they are involved in the decidability problems we are dealing with.

2.1. Useful results on return words of morphic sequences

Let x be a uniformly recurrent sequence on the alphabet A and u a non-empty prefix of x. We call *return word to u* every word $x_{[i,j-1]} = x_i x_{i+1} \cdots x_{j-1}$, where i and j are two successive occurrences of u in x. The reader can check that a word v is a return word to u of x if and only if vu belongs to $\mathcal{L}(x) = \{x_{[i,j]}; 0 \leq i \leq j\}$, u is a prefix of vu and u has exactly two occurrences in vu. The set of return words to u is finite, because x is uniformly recurrent, and is denoted by $\mathcal{R}_{x,u}$. The sequence x can be written as a concatenation

$$x = m_0 m_1 m_2 \cdots \ , m_i \in \mathcal{R}_{x,u}, i \in \mathbb{N},$$

of return words to u, and this decomposition is unique. We enumerate the elements of $\mathcal{R}_{x,u}$ in the order of their first appearance in $(m_n)_{n\in\mathbb{N}}$. This defines a bijective map

$$\Theta_{x,u} : R_{x,u} \to \mathcal{R}_{x,u} \subset A^*$$

where $R_{x,u} = \{1, \cdots, \#(\mathcal{R}_{x,u})\}$. It defines a morphism and the set $\Theta_{x,u}(R^*_{x,u})$ consists of all concatenations of return words to u. We should point out that the order used to define this morphism will be of particular importance in the sequel. In fact, you can change the enumeration and this will change the morphism. What is important in the sequel is that once you fixed an algorithm to define $\Theta_{x,u}$ then you should use the same for all x and u.

We denote by $\mathcal{D}_u(x)$ the unique sequence on the alphabet $R_{x,u}$ characterized by

$$\Theta_{x,u}(\mathcal{D}_u(x)) = x.$$

We call it the *derived sequence of x on u*. It is clearly uniformly recurrent. The following proposition points out the basic properties of return words.

PROPOSITION 1 ([Durand 1998a]). *Let x be a uniformly recurrent sequence and u a non-empty prefix of x.*

(1) *The set $\mathcal{R}_{x,u}$ is a code, i.e., $\Theta_{x,u} : R^*_{x,u} \to \Theta_{x,u}(R^*_{x,u})$ is one-to-one.*

(2) *If u and v are two prefixes of x such that u is a prefix of v then each return word to v belongs to $\Theta_{x,u}(R^*_{x,u})$, i.e., it is a concatenation of return words to u.*

(3) *Let v be a non-empty prefix of $\mathcal{D}_u(x)$ and $w = \Theta_{x,u}(v)u$. Then*
- *w is a prefix x,*
- *$\Theta_{x,u}\Theta_{\mathcal{D}_u(x),v} = \Theta_{x,w}$ and*
- *$\mathcal{D}_v(\mathcal{D}_u(x)) = \mathcal{D}_w(x).$*

COROLLARY 2. *Let x be a uniformly recurrent sequence and u a non-empty prefix of x. If u is a prefix of v and v is a prefix of x then there is a unique morphism Θ satisfying $\Theta_{x,u}\Theta = \Theta_{x,v}$.*

The next lemma asserts that the derived sequences of purely primitive substitutive sequences are purely primitive substitutive sequences.

PROPOSITION 3 ([Durand 1998a]). *Let $x = \sigma^\omega(a)$ where $\sigma : A^* \to A^*$ is a primitive substitution, and, u be a non-empty prefix of x. The derived sequence $\mathcal{D}_u(x)$ is the unique fixed point of the unique substitution σ_u satisfying*

$$\Theta_{x,u}\sigma_u = \sigma\Theta_{x,u}.$$

The substitution σ_u is called the *return substitution* on u.

Remark that it is easy to compute σ_u and $\Theta_{x,u}$. Take w the first return word to u having an occurrence in x: $\Theta_{x,u}(1) = w$. Then, compute $\sigma(w)$. It is necessarily a (computable) concatenation of return words to u and a prefix of x: $\sigma(w) = \Theta_{x,u}(v_1)$ for some $v_1 \in R_{x,u}^*$. This defines $\sigma_u(1) = v_1$. If no return words different from w appears then we stop here and x is periodic: $x = www\ldots$. If not, we do the same with $\Theta_{x,u}(2)$ and we continue until no new return word appears. This will stop in finite time because x is ultimately recurrent. Remark that this procedure also defines the map $\Theta_{x,u}$. We proved the following lemma.

Lemma 4. *Let $x = \sigma^\omega(a)$ where $\sigma : A^* \to A^*$ is a primitive substitution. Let u be a prefix of x. Then, σ_u and $\Theta_{x,u}$ are algorithmically computable.*

Bounds will be given with the help of the next 3 lemmas. For any morphism $\sigma : A^* \to B^*$, we set $|\sigma| = \max_{a \in A} |\sigma(a)|$.

In the sequel R_σ will denote the maximal difference between two successive occurrences of a word of length 2 in any fixed point of the primitive substitution σ. In order to bound R_σ we need the following lemma.

Lemma 5. [Horn and Johnson 1990] *If M is a primitive $n \times n$ matrix then there exists $k \leq (n-1)n^n$ such that M^k has strictly positive entries.*

Lemma 6. *If σ is a primitive substitution defined on an alphabet with d letters, then R_σ is bounded by $2|\sigma|^{(d-1)d^d}$.*

P r o o f. The proof is left to the reader. $\qquad\square$

Lemma 7. *For each primitive substitution σ defined on an alphabet A with d letters and all n the following holds:*

$$|\sigma^n| \leq Q_\sigma \min_{a \in A} |\sigma^n(a)|.$$

where $Q_\sigma = \max \left(\max_{0 \leq n \leq (d-1)d^d + 1} \frac{|\sigma^n|}{\min_a |\sigma^n(a)|}, |\sigma| \right)$.

P r o o f. The proof is left to the reader. $\qquad\square$

Now we can give upper and lower bounds for the length of the return words and for the cardinality of the sets of return words.

Theorem 8 ([Durand 1998a, Durand, Host and Skau 1999]). *Let $x = \sigma^\omega(a)$, where σ is a primitive substitution, is non-periodic. For all non-empty prefixes u of x,*

(1) for all words v belonging to $\mathcal{R}_{x,u}$, $\frac{1}{K_\sigma}|u| \leq |v| \leq K_\sigma|u|$, and

(2) $\#(\mathcal{R}_{x,u}) \leq 4K_\sigma^3$,

where $K_\sigma = Q_\sigma R_\sigma |\sigma|$.

P r o o f. The upper bound in (1) is easily deduced from [Durand 1998a]. The lower bound can also be deduced from [Durand 1998a] but it is easier to obtain from [Durand, Host and Skau 1999]. The last bound can be deduced from (1).
□

Note that the right inequality of Statement (1) is sufficient to imply the two other inequalities (see [Durand, Host and Skau 1999]). From this theorem and the definition of σ_u we deduce the following corollary.

COROLLARY 9. *For each primitive substitution σ and all u such that u is a prefix of $\sigma(u)$ we have*

$$|\sigma_u| \leq |\sigma| K_\sigma^2.$$

The result we stated above were used to prove the two following results that will be involved in the decidability problems. The first will be important in the purely substitutive (or D0L) context and the second in the morphic (or HD0L) case.

PROPOSITION 10 ([Durand 1998a]). *Let σ be a primitive substitution. The set of the return substitutions of σ is finite and bounded by $\left(4K_\sigma^3\right)^{|\sigma| K_\sigma^2 + 1}$.*

PROPOSITION 11. *Let $x = \sigma^\omega(a)$, where σ is a primitive substitution, and $\phi : A^* \to B^*$ be a coding. For all prefixes u of x there exists a unique morphism λ_u satisfying*

$$\phi\Theta_{x,u} = \Theta_{\phi(x),\phi(u)}\lambda_u.$$

Moreover,

(1) $\lambda_u(\mathcal{D}_u(x)) = \mathcal{D}_{\phi(u)}(\phi(x))$,

(2) $\#\{\lambda_u; u \text{ prefix of } x\} \leq (K_\sigma + 1)^{K_\sigma^2}$.

PROPOSITION 12. *With the settings of Proposition 11, let u be a prefix of x. Then, $\Theta_{\phi(x),\phi(u)}$ and λ_u can be algorithmically defined.*

3. HD0L ω-equivalence problem

3.1. Decidability results for purely substitutive sequences in a primitive context

In this section we propose a strategy to decide whether two purely substitutive sequences are equal when the underlying substitutions are primitive.

The following lemma is easy to establish (see [Durand, Host and Skau 1999]). In the sequel the constants K_σ and K_τ are those given by Theorem 8.

LEMMA 13. *Let x be a uniformly recurrent sequence such that there exists K satisfying: for all $u \in \mathcal{L}(x)$ and all $w \in \mathcal{R}_{x,u}$, $|w| \leq K|u|$. Then, for all n, all words of length $(K+1)n$ of $\mathcal{L}(x)$ have an occurrence of all words of length n.*

THEOREM 14. *Let $\sigma : A^* \to A^*$ and $\tau : B^* \to B^*$ be primitive substitutions, $x = \sigma^\omega(a) \in A^{\mathbb{N}}$ and $y = \sigma^\omega(b) \in B^{\mathbb{N}}$. Let $K = 1 + \left((4K_\sigma)^3\right)^{|\sigma|K_\sigma^2+1} \left((4K_\tau)^3\right)^{|\tau|K_\tau^2+1}$, $u_n = x[0, (K_\sigma + 1)^n]$, $v_n = y[0, (K_\sigma + 1)^n]$ and $T_n = (\sigma_{u_n}, \tau_{v_n})$. Then, the following are equivalent:*

(1) $x = y$;

(2) there exist n and m less than K such that

$$u_n = v_n, u_m = v_m, \Theta_{x,u_n} = \Theta_{y,v_n}, \Theta_{x,u_m} = \Theta_{y,v_m} \text{ and } T_n = T_m.$$

P r o o f. The necessary condition is obtained applying Proposition 10 and the pigeon hole principle.

Let us show it is sufficient: suppose there exist n, m satisfying $n < m \leq K$ such that $u_n = v_n$, $u_m = v_m$, $\Theta_{x,u_n} = \Theta_{y,v_n}$, $\Theta_{x,u_m} = \Theta_{y,v_m}$ and $T_n = T_m$.

From Proposition 3 we deduce that $\mathcal{D}_{u_n}(x) = \mathcal{D}_{u_m}(x)$ and $\mathcal{D}_{u_n}(y) = \mathcal{D}_{u_m}(y)$. Moreover, from Corollary 2 there exists a unique morphism $\Theta : R^* \to R^*$ such that $\Theta_{x,u_n}\Theta = \Theta_{x,u_m}$ where $R = R_{x,u_n} = R_{x,u_m} = R_{y,u_n} = R_{y,u_m}$. Then,

$$\Theta_{x,u_n}\Theta(\mathcal{D}_{u_m}(x)) = \Theta_{x,u_m}(\mathcal{D}_{u_m}(x)) = x \text{ and}$$
$$\Theta_{x,u_n}\Theta(\mathcal{D}_{u_m}(y)) = \Theta_{x,u_m}(\mathcal{D}_{u_m}(y)) = y.$$

Thus, from the definition of $\mathcal{D}_{u_n}(x)$ and $\mathcal{D}_{u_n}(y)$ we deduce:

$$\Theta(\mathcal{D}_{u_m}(x)) = \mathcal{D}_{u_n}(x) = \mathcal{D}_{u_m}(x)$$
$$\Theta(\mathcal{D}_{u_m}(y)) = \mathcal{D}_{u_n}(y) = \mathcal{D}_{u_m}(y)$$

and $\mathcal{D}_{u_n}(x)$ and $\mathcal{D}_{u_n}(y)$ are fixed points of Θ. But remark that from Lemma 13 and the choice of the u_n, in all images of Θ there are all the letters of R

and all images of Θ start with the same letter. Hence Θ is a primitive substitution and Θ has a unique fixed point. Consequently, $\mathcal{D}_{u_n}(x) = \mathcal{D}_{u_n}(y)$ and $x = \Theta_{x,u_n}(\mathcal{D}_{u_n}(x)) = \Theta_n(\mathcal{D}_{u_n}(y)) = y$. $\qquad\square$

COROLLARY 15. *The D0L ω-equivalence problem is decidable in the primitive case.*

P r o o f. This is a direct consequence of Theorem 14. Let us keep the notation of this theorem. The bounds K_σ, K_τ and K are algorithmically computable. So are the u_n and v_n, $n \le K$. We gave the algorithm to compute return substitutions after their definition, hence T_n is algorithmically computable and the D0L ω-equivalence problem is decidable. $\qquad\square$

3.2. Decidability results for morphic sequences in the primitive context

The strategy to tackle the decidability of the HD0L ω-equivalence problem in the primitive case is the same as for the D0L case.

THEOREM 16. *Let $\sigma : A^* \to A^*$ and $\tau : B^* \to B^*$ be two primitive substitutions, $x = \sigma^\omega(a) \in A^\mathbb{N}$ and $y = \sigma^\omega(b) \in B^\mathbb{N}$. Let $\phi : A^* \to C^*$ and $\psi : B^* \to D^*$ be two codings. Let $K = 1 + \left((4K_\sigma)^3\right)^{|\sigma|K_\sigma^2+1} (K_\sigma+1)^{K_\sigma^2} \left((4K_\tau)^3\right)^{|\tau|K_\tau^2+1} (K_\tau+1)^{K_\tau^2}$, $u_n = \phi(x[0,(K_\sigma+1)^n])$, $v_n = \psi(y[0,(K_\sigma+1)^n])$ and $T_n = (\lambda_{u_n}, \sigma_{u_n}, \lambda_{v_n}, \tau_{v_n})$. Then, the following are equivalent:*

(1) $x = y$;

(2) there exist n and m less than K such that

$$u_n = v_n, u_m = v_m, \Theta_{x,u_n} = \Theta_{y,u_n}, \Theta_{x,u_m} = \Theta_{y,u_m} \text{ and } T_n = T_m.$$

P r o o f. The proof proceed as the proof of Theorem 14. $\qquad\square$

The following proposition will allow us to reduce the HD0L ω-equivalence problem to substitutive sequences. This result can also be found in several papers: [Cobham 1968, Pansiot 1983, Cassaigne and Nicolas 2003, Durand 1998a, Honkala 2009b]. Below I present the proof G. Rauzy showed me.

PROPOSITION 17. *Let x be a morphic sequence w.r.t. a primitive substitution. Then, x is substitutive w.r.t. a primitive substitution.*

P r o o f. Let y be a purely substitutive sequence generated by the primitive substitution $\sigma : A^* \to A^*$ and $\rho : A^* \to B^*$ be a morphism such that $x = \rho(y)$. Suppose $y = \sigma^\omega(a)$.

From Lemma 5, there exists $k \leq (m-1)^m$, where $m = \#A$, such that all images of σ^k have an occurrence of all letters of A. Thus, the morphism $\phi = \rho\sigma^k$ does not have the empty word as an image of letter. Moreover, $x = \phi(y)$.

Let $D = \{(c,k); c \in C, 0 \leq k \leq |\phi(c)| - 1\}$ and $\psi : C^* \to D^*$ the morphism defined by:

$$\psi(c) = (c,0)\ldots(c,|\phi(c)| - 1).$$

There is an integer n $(\leq \max_{a \in A} |\phi(a)|)$ such that $|\sigma^n(c)| \geq |\phi(c)|$ for all c in C.

Let τ be the morphism from D to D^* defined by:

$$\tau((c,k)) = \psi(\sigma^n(c)_{[k,k]}) \ \text{ if } 0 \leq k < |\phi(c)| - 1,$$
$$\text{and} \quad \tau((c,|\phi(c)| - 1)) = \psi(\sigma^n(c)_{[|\phi(c)|-1,|\sigma^n(c)|-1]}) \ \text{ otherwise.}$$

For all c in C we have
$$\tau(\psi(c)) \quad = \tau((c,0)\cdots(c,|\phi(c)| - 1))$$

$$= \psi(\sigma^n(c)_{[0,0]})\cdots\psi(\sigma^n(c)_{[|\sigma^n(c)|-1,|\sigma^n(c)|-1]}) = \psi(\sigma^n(c)),$$

hence $\tau(\psi(y)) = \psi(\sigma^n(y)) = \psi(y)$. Hence $\psi(y)$ is a fixed point of the substitution τ which begins with some letter $(e,0)$ and $\tau\psi = \psi\sigma^n$.

Let χ be the coding from D^* to B^* defined by $\chi((c,k)) = \phi(c)_{[k,k]}$ for all (c,k) in D. For all c in C we obtain

$$\chi(\psi(c)) = \chi((c,0)\cdots(c,|\phi(c)| - 1)) = \phi(c),$$

and then $\chi(\psi(y)) = \phi(y)$. Moreover, the relation $\tau\psi = \psi\sigma^n$ implies that for all $k \in \mathbb{N}$ we have $\tau^k\psi = \psi\sigma^{kn}$. Hence if σ is primitive then τ is primitive too. $\square$

COROLLARY 18. *The HD0L-ω-equivalence problem is decidable in the primitive case.*

P r o o f. It is a consequence of Proposition 17 and Theorem 16. $\square$

4. More on substitutions sharing the same fixed point

A substitution on a finite alphabet is *invertible* if when extended to a free group, it is an automorphism. Let x be a sequence on the finite alphabet A and suppose all letters of A have an occurrence in x. The *stabilizer* of x, denoted by $\text{Stab}(x)$, is the set of morphisms $\sigma : A^* \to A^*$ such that $\sigma(x) = x$. We say that x is *rigid* whenever the monoid $\text{Stab}(x)$ is generated by a single element

(see [Krieger 2008] for more informations on this topic). It has been shown in [Séébold 1998] that all invertible substitutions generate rigid sequences (Theorem 19). There was in fact a gap in the proof. In [Rao and Wen 2010] the authors gave a geometrical proof (see also [Berthé, Frettlöh and Sirvent:preprint] for a different proof). The gap in the proof in [Séébold 1998] was recently corrected in [Richomme and Séébold:preprint]. The statement of this result is the following.

THEOREM 19. *Let σ and τ be primitive invertible substitutions defined on the alphabet $\{a, b\}$. They have powers sharing the same fixed point if and only if there exist a substitution ζ and positive integers i and j such that $\sigma = \zeta^i$ and $\zeta^j = \tau$.*

As pointed out in [Krieger 2008], the invertibility assumption is crucial in this theorem. The following example due to Séébold (see [Krieger 2008]) gives a counterexample:

$$\sigma : a \mapsto ab, \; b \mapsto baabba, \; \tau : a \mapsto abbaab, \; b \mapsto ba.$$

Indeed, $\sigma^\omega(a) = \tau^\omega(a)$ but σ and τ are not powers of the same substitution.

For a 3-letter alphabet, it has been shown in [Wen and Zhang 1999] (see also [Richomme 2003]) that the monoid of invertible substitutions is not finitely generated. This led D. Krieger [Krieger 2008] to find a non rigid purely substitutive sequence x w.r.t. an invertible substitution.

Below we propose a weaker version of Theorem 19 (Corollary 22) that applies to any kind of alphabets. Before we need a lemma, a proposition and the following notation: for any word w, $V(w)$ is the vector whose entry of index a is the number of a in w. It is sometimes called the *Parikh vector of w* or the *abelianized* of w.

LEMMA 20. *Let $x = \sigma^\omega(a)$ and $y = \tau^\omega(a)$ where σ and τ are two primitive substitutions (on an alphabet with d letters). Suppose x and y have a common prefix u satisfying:*

(1) $\sigma(w) = \tau(w)$ for all $w \in \mathcal{R}_{x,u}$;

(2) $\{V(w); w \in \mathcal{R}_{x,u}\}$ generates $\mathbb{Z}^d$.

Then $\tau = \sigma$.

P r o o f. The proof is left to the reader. $\qquad\qquad\square$

PROPOSITION 21. [Durand 1998b] *Let τ and σ be two primitive substitutions. Then, they share the same non-periodic fixed point $\sigma^\omega(a) = \tau^\omega(a)$ if and only if $\sigma^\omega(a)$ and $\tau^\omega(a)$ share a common prefix u such that there exist two integers i and j verifying*

$$\tau_u^i = \sigma_u^j. \tag{1}$$

Remark that Equality (1) means that τ^i and σ^j coincide on the set $\mathcal{R}_{x,u}$. Thus, using Lemma 20 we obtain the announced corollary.

COROLLARY 22. *Let τ and σ be two primitive substitutions (on an alphabet with d letters) sharing the same non-periodic fixed point $x = \sigma^\omega(a) = \tau^\omega(a)$. Let i, j be the positive integers and u the prefix of x given in Proposition 21. Suppose the set $\{V(w); w \in \mathcal{R}_{x,u}\}$ generates $\mathbb{Z}^d$. Then, there exist two integers i and j such that*

$$\tau^i = \sigma^j.$$

Applied to invertible substitutions on a 2-letter alphabet this corollary implies that when two invertible primitive substitutions share the same fixed point then they have non trivial common powers because it can be checked that for all u the set $\{V(w); w \in \mathcal{R}_{x,u}\}$ generates $\mathbb{Z}^2$. For the counterexample of Séébold recalled above this means that all u satisfying Equality (1) is such that all vectors in $\{V(w); w \in \mathcal{R}_{x,u}\}$ are $\mathbb{Q}$-colinear.

Proposition 21 provides a way to study rigity problems: with a weaker assumption than the one in Corollary 22 but with specific families of substitutions it is clear that the conclusion of this corollary would also hold.

To finish with this section we would like to mention the following complete substitutive extension of Cobham's theorem obtained in [Durand 2011]. This gives a strong restriction (without the primitivity assumption) to generate the same morphic sequence.

THEOREM 23. *Let α and β be two multiplicatively independent Perron numbers. Let A be a finite alphabet and x be a sequence of $A^{\mathbb{N}}$. Then, x is both α-substitutive and β-substitutive if and only if x is ultimately periodic,*

where "x is α-substitutive" means that x is substitutive w.r.t. a substitution whose dominant eigenvalue of its incidence matrix is α. This result also holds true for morphic sequences whenever the morphisms (instead of the codings) are non-erasing. For purely substitutive substitutions it (the necessary condition) was known from [Durand 2002].

5. Periodicity of HD0L infinite sequences

In the sequel $\sigma : A^* \to A^*$ is a primitive morphism, $\phi : A^* \to B^*$ is a morphism, $y = \sigma^\omega(a)$ is a sequence of $A^{\mathbb{N}}$ and $x = \phi(y)$. We have to find an

algorithm telling if x is ultimately periodic or not. Remark that as the morphism is primitive, the sequence x is uniformly recurrent, and, ultimate periodicity of x is equivalent to its periodicity. The following lemma is easy to prove.

LEMMA 24. *For any periodic sequence z with length period L, if u is a word of $\mathcal{L}(z)$ whose length is greater than L then $\#\mathcal{R}_u = 1$.*

We recall (Proposition 1) that for all n there exists a prefix u_n of y such that $\mathcal{D}^n(x) = \lambda_n(\mathcal{D}_{u_n}(y))$ and $u_1 = x_0$ where λ_n is the unique morphism from $R^*_{u_n}$ to $R^*_{\phi(u_n)}$ defined by

$$\phi\Theta_{x,u_n} = \Theta_{\phi(x),\phi(u_n)}\lambda_n.$$

We set $\sigma_n = \sigma_{u_n}$.

PROPOSITION 25. *Suppose ϕ is a coding and the incidence matrix of σ has strictly positive entries. Let $K = \left(4K_\sigma^3\right)^{|\sigma|K_\sigma^2+1}(K_\sigma+1)^{K_\sigma^2}$. Then, x is periodic if and only if there exists i less than K such that $\#\mathcal{R}_{\phi(u_i)} = 1$.*

P r o o f. If $\#\mathcal{R}_{\phi(u_i)} = 1$, then it is clear that x is periodic.

Now suppose x is periodic. From the choice of K and the pigeon hole principle, there exist i and j, $i < j$, such that $(\lambda_i, \sigma_i) = (\sigma_j, \lambda_j)$. Thus, $R_{u_i} = R_{u_j}$ and $R_{\phi(u_i)} = R_{\phi(u_j)}$. Moreover, remark that we have $\mathcal{D}^i(x) = \mathcal{D}^j(x)$. Thus, $\mathcal{D}^i(x) = \mathcal{D}^{i+k(j-i)}(x)$ for all $k \geq 1$. From Proposition 1, $\lim_{k\to+\infty} |u_{i+k(j-i)}| = +\infty$. We conclude with Lemma 24. □

THEOREM 26. *The HD0L periodicity problem for infinite sequences is decidable in the context of primitive morphisms.*

P r o o f. First, let us make some reductions of the problem. From Lemma 5, we can suppose the incidence matrix of σ has strictly positive entries. Suppose ϕ is erasing. It cannot erase all the letters. Hence $\psi = \phi\sigma$ is a non erasing morphism and remark that $x = \psi(y)$. Consequently, we can suppose ϕ non-erasing. To end with the reductions, from Proposition 17 we can suppose ϕ is a coding.

Now, with the notations introduced in this section, it suffices to compute $R_{\phi(u_n)}$ from $n = 1$ to $n = K$. If for some i less than K, $\#R_{\phi(u_n)} = 1$ then x is periodic, otherwise x is not periodic. It remains to explain how to compute $R_{\phi(u_n)}$. We recall the maps $\lambda_{u_n} : R_{y,u_n} \to R_{x,\phi(u_n)}$ are algorithmically definable (see Proposition 12). Hence it can be easily check that $\#R_{\phi(u_n)} = 1$. □

Remark that when x is periodic then a word period can be computed, that is a word u such that $x = u^\infty$. Indeed, once, in the proof of Theorem 26, we find n such that $\#R_{\phi(u_n)} = 1$ then $x = u^\infty$ where $\mathcal{R}_{\phi(u_n)} = \{u\}$. Moreover, $\mathcal{R}_{\phi(u_n)}$ is algorithmically computable.

REFERENCES

[Berthé, Frettlöh and Sirvent:preprint]
V. Berthé, D. Frettlöh, and V. Sirvent. Selfdual substitutions in dimension one. *Preprint* (2011).

[Boshernitzan and Carroll 1997]
M. D. Boshernitzan and C. R. Carroll. An extension of Lagrange's theorem to interval exchange transformations over quadratic fields. *J. Anal. Math.* **72** (1997), 21–44.

[Cassaigne and Nicolas 2003]
J. Cassaigne and F. Nicolas. Quelques propriétés des mots substitutifs. *Bull. Belg. Math. Soc. Simon Stevin* **10** (2003), 661–676.

[Cobham 1968]
A. Cobham. On the Hartmanis-Stearns problem for a class of tag machines. In *IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory*, pp. 51–60, 1968. Also appeared as IBM Research Technical Report RC-2178, August 23 1968.

[Culik and Harju 1984]
K. Culik II and T. Harju. The ω-sequence equivalence problem for D0L systems is decidable. *J. Assoc. Comput. Mach.* **31** (1984), 282–298.

[Diekert and Krieger 2009]
V. Diekert and D. Krieger. Some remarks about stabilizers. *Theoret. Comput. Sci.* **410** (2009), 2935–2946.

[Durand 1998a]
F. Durand. A characterization of substitutive sequences using return words. *Discrete Math.* **179** (1998), 89–101.

[Durand 1998b]
F. Durand. A generalization of Cobham's theorem. *Theory Comput. Syst.* **31** (1998), 169–185.

[Durand 2000]
F. Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergod. Th. & Dynam. Sys.* **20** (2000), 1061–1078.

[Durand 2002]
F. Durand. A theorem of Cobham for non primitive substitutions. *Acta Arith.* **104** (2002), 225–241.

[Durand 2011]
F. Durand. Cobham's theorem for substitutions. *J. Eur. Math. Soc* **13** (2011), 1797–1812.

[Durand, Host and Skau 1999]
F. Durand, B. Host, and C. Skau. Substitutive dynamical systems, Bratteli diagrams and dimension groups. *Ergod. Th. & Dynam. Sys.* **19** (1999), 953–993.

[Honkala 2007]
J. Honkala. The D0L ω-equivalence problem. *Int. J. Found. Comp. Sci.* **18** (2007), 181–194.

[Honkala 2009a]
J. Honkala. The equality problem for infinite words generated by primitive morphisms. *Inform. and Comput.* **207** (2009), 900–907.

[Honkala 2009b]
J. Honkala. On the simplification of infinite morphic words. *Theoret. Comput. Sci.* **410** (2009), 997–1000.

[Horn and Johnson 1990]
R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[Krieger 2008]
D. Krieger. On stabilizers of infinite words. *Theoret. Comput. Sci.* **400** (2008), 169–181.

[Masur 1982]
H. Masur. Interval exchange transformations and measured foliations. *Ann. of Math.* **115** (1982), 169–200.

[Morse and Hedlund 1940]
M. Morse and G. A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.* **62** (1940), 1–42.

[Pansiot 1983]
J.-J. Pansiot. Hiérarchie et fermeture de certaines classes de tag-systèmes. *Acta Inform.* **20** (1983), 179–196.

[Pansiot 1986]
J.-J. Pansiot. Decidability of periodicity for infinite words. *RAIRO Inform. Théor. App.* **20** (1986), 43–46.

[Rao and Wen 2010]
H. Rao and Z.-Y. Wen. Invertible substitutions with a common periodic point. In *Recent developments in fractals and related fields*, Appl. Numer. Harmon. Anal., pp. 401–409. Birkhäuser Boston Inc., Boston, MA, 2010.

[Rauzy 1979]
G. Rauzy. Échanges d'intervalles et transformations induites. *Acta Arith.* **34** (1979), 315–328.

[Richomme 2003]
G. Richomme. Conjugacy and episturmian morphisms. *Theoret. Comput. Sci.* **302** (2003), 1–34.

[Richomme and Séébold:preprint]
G. Richomme and P. Séébold. Completing a combinatorial proof of the rigidity of sturmian words generated by morphisms. *Preprint* (2011).

[Séébold 1998]
P. Séébold. On the conjugation of standard morphisms. *Theoret. Comput. Sci.* **195** (1998), 91–109.

[Veech 1982]
W. A. Veech. Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math.* **115** (1982), 201–242.

[Wen and Zhang 1999]
Z. Wen and Y. Zhang. Some remarks on invertible substitutions on three letter alphabet. *Chinese Sci. Bull.* **44** (1999), 1755–1760.

HD0L ω-EQUIVALENCE AND PERIODICITY PROBLEMS IN THE PRIMITIVE CASE

Université de Picardie Jules Verne
LAMFA CNRS-UMR 6140
33 rue Saint Leu
80039 Amiens Cedex 01
France.
E-mail: fabien.durand@u-picardie.fr