Decidability, Arithmetic Subsequences and Eigenvalues of Morphic Subshifts

Fabien Durand   Valérie Goyheneche

Abstract

We prove decidability results on the existence of constant subsequences of uniformly recurrent morphic sequences along arithmetic progressions. We use spectral properties of the subshifts they generate to give a first algorithm deciding whether, given $p \in \mathbb{N}$, there exists such a constant subsequence along an arithmetic progression of common difference p. In the special case of uniformly recurrent automatic sequences we explicitly describe the sets of such p by means of automata.

1 Introduction

In this paper we are concerned with arithmetic subsequences $(x_{k+np})_{n\in\mathbb{Z}}$ of morphic sequences $x = (x_n)_{n\in\mathbb{Z}}$ and decision problems concerning constant such subsequences. Namely,

Input: Two finite alphabets $\mathcal{A}$ and $\mathcal{B}$, an endomorphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ and a morphism $\phi : \mathcal{A}^* \to \mathcal{B}^*$.

Question 1: Given $p \in \mathbb{N}$ and an admissible fixed point $x \in \mathcal{A}^{\mathbb{Z}}$ of σ, does there exist a constant subsequence $y = ((\phi(x))_{k+np})_{n\in\mathbb{Z}}$?

Question 2: Given an admissible fixed point $x \in \mathcal{A}^{\mathbb{Z}}$, does there exist a constant subsequence $y = ((\phi(x))_{k+np})_{n\in\mathbb{Z}}$?

Question 3: Does there exist an algorithm that computes the set of integers $p \in \mathbb{N}$ satisfying the requirement of Question 1 ?

Received by the editors in November 2018 - In revised form in April 2019.
Communicated by F. Point.
2010 *Mathematics Subject Classification* : 37B10, 54H20, 68R15, 68Q45.
Key words and phrases : constant arithmetic subsequence, eigenvalue, subshift, morphic.

Let us recall that Cobham showed [Cob72, Sec. 5] that arithmetic subsequences of l-automatic sequences are l-automatic. Durand extended this result [Dur96, Sec. I.4] to primitive substitutive sequences. The periodicity of such sequences is decidable and the length of the period can be found [HL86, Pan86, Hon86, Dur13a]. Therefore, Question 1 is decidable for l-automatic and primitive substitutive sequences.

We provide a different proof using dynamical systems, namely subshifts, and some of their spectral properties, for uniformly recurrent morphic sequences, and Presburger arithmetic for automatic sequences. For uniformly recurrent automatic sequences, we show Question 2 is algorithmically decidable and there exists an algorithm for Question 3. For the non uniformly recurrent automatic sequences, both questions are open.

In Section 2 the basic definitions on morphic sequences and on subshift dynamical systems are given. In Section 3 we show the relation between constant arithmetic subsequences of uniformly recurrent sequences and the spectral eigenvalues of the subshifts they generate. We recall some well-known results on dynamical eigenvalues of Sturmian [Kur03], constant length substitutions [Dek78] and Toeplitz subshifts [Wil84]. As a direct consequence we observe in Section 4 that intersections of languages coming from these subshifts should be finite. We prove the decidability of Question 1 for uniformly recurrent morphic sequences in Section 5. The decidability of Question 2 and Question 3 is proven in Section 6 but only for uniformly recurrent automatic sequences. The description of the set of integers $p \in \mathbb{N}$ satisfying the requirement of Question 1 is given by means of allowed paths in a finite automaton.

2 Definitions and background

2.1 Words and sequences

In all this article, $\mathcal{A}$ will stand for an *alphabet*, that is a finite set of elements called *letters*. A *word* is an element of the free monoid $\mathcal{A}^*$ generated by $\mathcal{A}$. The neutral element of $\mathcal{A}^*$ is called the *empty word* and is denoted by ϵ. We set $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\epsilon\}$. For $u = u_0u_1u_2u_3 \cdots u_{n-1}$ in $\mathcal{A}^n \subset \mathcal{A}^*$, n is called the *length* of u and is denoted by $|u|$. *One-sided sequences* are elements of $\mathcal{A}^{\mathbb{N}}$ and *sequences* are elements of $\mathcal{A}^{\mathbb{Z}}$. The sets $\mathcal{A}^{\mathbb{N}}$ and $\mathcal{A}^{\mathbb{Z}}$ are endowed with the product topology. For convenience, we use a dot to separate negative and positive indices: a sequence $x \in \mathcal{A}^{\mathbb{Z}}$ would be written $\cdots x_{-2}x_{-1}.x_0x_1x_2 \cdots$. We set $u_{[p,q]} = u_pu_{p+1} \cdots u_q$. The integer p is the *occurrence* of the *factor* $u_{[p,q]}$ in the word u. If a factor has several occurrences in a word u, we call the difference between two successive occurrences a *gap*. We denote by $\mathcal{L}(u)$ the set of all factors of u and we call it the *language of* u. The number of occurrences of a word v in the word u is denoted by $|u|_v$. These definitions also hold when u belongs to $\mathcal{A}^{\mathbb{Z}}$ or $\mathcal{A}^{\mathbb{N}}$.

Let u and v be two words in $\mathcal{A}^*$. The set $\{x \in \mathcal{A}^{\mathbb{Z}} : x_{[-|u|,|v|-1]} = uv\}$ is called a *cylinder* set and denoted $[u.v]$, or $[v]$ if u is the empty word. These sets generate the product topology of $\mathcal{A}^{\mathbb{Z}}$.

If $u = vw$ belongs to $\mathcal{A}^*$ or $\mathcal{A}^{\mathbb{N}}$, the word v is called a *prefix* of u. An *arithmetic subsequence* of x in $\mathcal{A}^{\mathbb{Z}}$ (resp. $\mathcal{A}^{\mathbb{N}}$) is a sequence of the form $(u_{k+np})_{n\in\mathbb{Z}}$ (resp. $(u_{k+np})_{n\in\mathbb{N}}$) for some k and p. The integer p is called the *common difference* of this arithmetic subsequence of x.

The sequence $x \in \mathcal{A}^{\mathbb{N}}$ (resp. in $\mathcal{A}^{\mathbb{Z}}$) is *periodic* with period p if there exists $p \in \mathbb{N}$ such that $x_i = x_{i+p}$ for all $i \in \mathbb{N}$ (resp. $\mathbb{Z}$). Otherwise, u is *non-periodic*.

A sequence x is *recurrent* if every factor of x occurs in x infinitely often. It is *uniformly recurrent* if every factor of x occurs in x with bounded gaps.

In the sequel $\mathcal{A}$ and $\mathcal{B}$ will stand for alphabets.

2.2 Morphisms and matrices

Let σ be a *morphism* from $\mathcal{A}^*$ to $\mathcal{B}^*$. When $\sigma(\mathcal{A}) = \mathcal{B}$, we say that σ is a *coding*. If $\sigma(\mathcal{A})$ is included in $\mathcal{B}^+$, it induces by concatenation a map from $\mathcal{A}^{\mathbb{N}}$ to $\mathcal{B}^{\mathbb{N}}$. These two maps are also called σ. To the morphism σ is naturally associated the matrix $M_\sigma = (m_{i,j})_{i\in\mathcal{B},j\in\mathcal{A}}$ where $m_{i,j}$ is the number of occurrences of i in the word $\sigma(j)$. We call it the *incidence matrix* of σ. We set $|\sigma| = \max_{a\in\mathcal{A}} |\sigma(a)|$ and $\langle\sigma\rangle = \min_{a\in\mathcal{A}} |\sigma(a)|$.

2.3 Substitutions

In the sequel we use the definition of substitution presented in [Que10] and the notion of substitutive sequence defined in [Dur98a].

The *language* of the endomorphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ is the set $\mathcal{L}(\sigma)$ of words appearing in some $\sigma^n(a)$, $a \in \mathcal{A}$.

If there exist a letter $a \in \mathcal{A}$ and a non-empty word $u \in \mathcal{A}^*$ such that $\sigma(a) = au$ and moreover, if $\lim_{n\to+\infty} |\sigma^n(a)| = +\infty$, then σ is said to be *right-prolongable on* a. Analogously, we define the notion to be *left-prolongable*. The endomorphism σ is *prolongable* on $b.a$ if it is left-prolongable on b, right-prolongable on a and if ba belongs to $\mathcal{L}(\sigma)$. The endomorphism σ is a *substitution* whenever it is prolongable and *growing* (that is, $\lim_n \langle\sigma^n\rangle = +\infty$). We say σ is *left-proper* if there exists $a \in A$ such that $\sigma(A)$ is included in aA^*. It is *right-proper* if there exists $b \in A$ such that $\sigma(A)$ is included in A^*b. It is *proper* whenever it is both left and right proper.

A letter c such that $\lim_{n\to+\infty} |\sigma^n(c)| = +\infty$ is called a *growing letter*.

It is an exercise to check that when σ is right-prolongable on a, the sequence $(\sigma^n(aa\cdots))_{n\geq 0}$ converges to a sequence denoted by $\sigma^\infty(a)$. The map $\sigma : \mathcal{A}^{\mathbb{N}} \to \mathcal{A}^{\mathbb{N}}$ being continuous, $\sigma^\infty(a)$ is a fixed point of σ: $\sigma(\sigma^\infty(a)) = \sigma^\infty(a)$. We say it is an *admissible one-sided fixed point*. In the same way $x = \sigma^\infty(b.a)$ can be defined and is a *two-sided* fixed point of σ. When ba belongs to $\mathcal{L}(\sigma)$ we say x is an *admissible* fixed point of the endomorphism σ. We also say that x is *purely morphic* (with respect to σ) or *purely substitutive* when σ is a substitution. If $x \in \mathcal{A}^{\mathbb{Z}}$ is purely morphic and $\phi : \mathcal{A}^* \to \mathcal{B}^*$ is a morphism then the sequence $y = \phi(x)$ is said to be a *morphic sequence* (with respect to ϕ and σ). We say y is *α-morphic* whenever the dominant eigenvalue of σ is α. When ϕ is a coding and σ a substitution, we say that y is *substitutive* (with respect to ϕ and σ).

Whenever the matrix associated to σ is primitive (that is, when it has a power with positive coefficients) we say that σ is a *primitive endomorphism*. In this situation we easily check that $\sigma^\infty(a)$ and $\sigma^\infty(b.a)$ are uniformly recurrent. Moreover, one has $\mathcal{L}(\sigma) = \mathcal{L}(\sigma^\infty(a)) = \mathcal{L}(\sigma^\infty(b.a))$.

A sequence is *primitive substitutive* if it is substitutive with respect to a primitive substitution. Such sequences are uniformly recurrent. As we will see later (Theorem 22), the set of uniformly recurrent morphic sequences is exactly the set of primitive substitutive sequences. If $|\sigma(a)| = p$ is constant for all $a \in \mathcal{A}$, we say σ is of *constant-length* p. In the other case, σ is of *non constant-length*.

2.4 Dynamical systems, subshifts and eigenvalues

For more details, we refer to Queffélec's book [Que10].

A *topological dynamical system* is a pair (X, T) where X is a compact metric space and $T : X \to X$ a homeomorphism. The topological dynamical system (X', T') is a *factor* of (X, T) whenever there exists a continuous and onto map $f : X \to X'$ such that $f \circ T = T' \circ f$. We say $f : (X, T) \to (X', T)$ is a factor map. If f is one-to-one and onto we say (X, T) and (X', T') are *isomorphic* and that f is an *isomorphism*. We call the set $\mathcal{O}(x) = \{T^n x : n \in \mathbb{Z}\}$ the *orbit of* $x \in X$. We say that (X, T) is *minimal* if every orbit is dense in X. Equivalently, the only closed T-invariant sets in X are X and $\emptyset$. We say it is *p-periodic* for some $p \in \mathbb{N}$ whenever there exists $x \in X$ such that $X = \{x, Tx, T^2x, \ldots, T^{p-1}x\}$ and $T^p x = x$. If such an integer p does not exist, (X, T) is *non-periodic*. If X is a Cantor set (i.e., a compact space without isolated points and having a countable base consisting of closed open sets, called *clopen* sets), we say (X, T) is a *Cantor system*.

We say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of (X, T) whenever there exists a continuous function $f : X \to \mathbb{C}$ satisfying $f \circ T = \lambda f$. Such a f is called an *eigenfunction* of (X, T).

Suppose (X, T) is minimal. Then, by compactness, f is a constant function and $|\lambda| = 1$. It follows that there exists $\alpha \in \mathbb{R}$ such that $\lambda = \exp(2i\pi\alpha)$. Such α are called *additive eigenvalues*. They form an additive subgroup of $\mathbb{R}/\mathbb{Z}$

Let $\mathcal{A}$ be an alphabet. The shift map on $\mathcal{A}^\mathbb{Z}$ is the map $S : \mathcal{A}^\mathbb{Z} \to \mathcal{A}^\mathbb{Z}$ defined by $(Sx)_n = x_{n+1}$ for all $n \in \mathbb{Z}$. A *subshift* is a topological dynamical system $(X, S_{/X})$ where X is a subset of $\mathcal{A}^\mathbb{Z}$. In what follows, S will stand for the shift map independently of the alphabet we consider and S will be used instead of $S_{/X}$.

If σ is a primitive substitution and x one of its fixed points, we call the subshift $(\overline{\mathcal{O}(x)}, S)$, also denoted (X_σ, S), the *primitive substitution subshift* generated by σ. It does not depend on the choice of the fixed point x [Que10, Prop. 5.3].

3 Relation between eigenvalues of subshifts and constant arithmetic subsequences

In this section, given a minimal Cantor system, we recall some well-known facts concerning the relations between its periodic behaviours and its eigenvalues. This is part of the folklore of ergodic theory of such dynamical systems. First we need some definitions.

Let us introduce some terminology already used to study Toeplitz subshifts [Wil84]. Let (X, T) be a minimal Cantor system, $x \in X$ and U be a clopen subset of X. We call the set

$$\mathrm{PS}_p(x, U) = \{k \in \mathbb{Z} : T^{k+np}x \in U,\ \forall n \in \mathbb{Z}\}$$

the *period-p skeleton* of x relatively to U. We say p is an *essential period* of x for U if:

1. $\mathrm{PS}_p(x, U) \neq \emptyset$ and
2. p divides q for every q satisfying $\mathrm{PS}_p(x, U) = \mathrm{PS}_p(x, U) - q$.

The proofs of the following lemmas are left to the reader.

Lemma 1. *Let $x \in X$ and U be a clopen subset of X. Then, for all n,*

$$\mathrm{PS}_p(T^n x, U) = \mathrm{PS}_p(x, U) - n.$$

Lemma 2. *For all $x, y \in X$ and all clopen sets U one has*

$$\mathrm{PS}_p(x, U) = \emptyset \Longleftrightarrow \mathrm{PS}_p(y, U) = \emptyset.$$

We call the following set

$$\mathbb{P}(X, T) = \{p \geq 2 : p \text{ is an essential period for some clopen set } U\}$$

the *periodic spectrum of* (X, T)

The following proposition will be useful for our study. It states in the two first properties that the set $\mathbb{P}(X, T)$ can be interpreted as the set of denominators of additive rational eigenvalues associated to (X, T).

Together with the third property, it gives a necessary condition to have constant arithmetic subsequences, which will be detailed in the next lemma.

Proposition 3. *Let (X, T) be a minimal dynamical system and $p \in \mathbb{N}$. Then, the following are equivalent:*

1. *$\lambda = \exp(2i\pi/p)$ is an eigenvalue of (X, T);*
2. *p belongs to $\mathbb{P}(X, T)$;*
3. *there exists a closed subset V of X such that $\{V, T^{-1}V, \ldots, T^{-p+1}V\}$ is a partition of X;*
4. *X admits a periodic factor with essential period p.*

Remark 4. If these properties are satisfied, the set of eigenvalues associated with (X,T) that are roots of unity is $\{\exp(2ik\pi/p) : p \in \mathbb{P}(X,T), k \in \mathbb{Z}\}$.

It is observed, and easy to prove, in [Dek78, Lemma II.2] that the partition of Property (3) is unique up to cyclic permutation.

Proof. (1) $\Rightarrow$ (2) Suppose that $\lambda = \exp(2i\pi/p)$ is an eigenvalue of (X,T) and $f : X \to \mathbb{C}$ a continuous eigenfunction associated to λ. Let $z \in X$. Replacing f by $f/f(z)$ if needed, one can suppose $f(z) = 1$. Let us prove that p is an essential period for the clopen set $U = f^{-1}(\{1\})$. By minimality $\mathrm{PS}_p(x,U)$ is non-empty for all $x \in X$. Suppose $\mathrm{PS}_p(x,U) = \mathrm{PS}_p(x,U) - q$ and let $n \in \mathrm{PS}_p(x,U)$. Then, $\lambda^{n-q} f(x) = \lambda^n f(x)$. Thus $\lambda^q = 1$ and p divides q.

(2) $\Rightarrow$ (3) There exists some $x \in X$ and a clopen set U such that $\mathrm{PS}_p(x,U)$ is not empty. From Lemma 2, the set $\mathrm{PS}_p(y,U)$ is not empty for all $y \in X$.

Let V be the closed subset $\{y \in X : \mathrm{PS}_p(y,U) = \mathrm{PS}_p(x,U)\}$. It suffices to show that $\{V, T^{-1}V, \dots, T^{-p+1}V\}$ is a partition of X. Observe that $T^{-p}V = V$. Thus $\cup_{i=0}^{p-1} T^{-i}V$ is a non-empty closed T-invariant set and by minimality it is X. Suppose there exists y belonging to $V \cap T^{-i}V$ with $0 \leq i < p$. Then, using Lemma 1, $\mathrm{PS}_p(y,U) = \mathrm{PS}_p(T^i y,U) = \mathrm{PS}_p(y,U) - i$. Moreover, p being an essential period it should divide i. Contradiction.

(3) $\Rightarrow$ (4) Let $\{V, T^{-1}V, \cdots T^{-p+1}V\}$ be a partition of X. Define the map $\pi : X \to \{0,\dots,p-1\}$ such that $\pi(x) = i$ if $x \in T^{-i}V$. Then, the system $(\mathbb{Z}/p\mathbb{Z}, R)$, where R is the addition of 1 modulo p, is a p-periodic factor of (X,T).

(4) $\Rightarrow$ (1) We can suppose the periodic factor is $(\mathbb{Z}/p\mathbb{Z}, R)$ where R is the addition of 1 modulo p. Choose some factor map f from (X,T) onto $(\mathbb{Z}/p\mathbb{Z}, R)$. Then, it suffices to consider the map $\phi : X \to \mathbb{C}$ defined by $\phi(x) = \exp(2i\pi f(x)/p)$. ■

Corollary 5. *Let (X',T') be a factor of (X,T). Then $\mathbb{P}(X',T')$ is included in $\mathbb{P}(X,T)$.*

Let us illustrate these notions and results in the framework of minimal subshifts. Let (X,S) be a minimal subshift defined on the alphabet $\mathcal{A}$ and $x = (x_n)_{n\in\mathbb{Z}} \in X$. It is clear that x has a constant arithmetic subsequence $(x_{k+np})_{n\in\mathbb{Z}}$ if and only if k belongs to $\mathrm{PS}_p(x,[a])$ for some $a \in \mathcal{A}$. We set

$$\begin{aligned}\mathrm{Per}(x) &= \{p \geq 1 : \exists a \in \mathcal{A}, \mathrm{PS}_p(x,[a]) \neq \emptyset\} \\ &= \{p \geq 1 : \exists a \in \mathcal{A}, \exists k\, \forall n, x_{k+np} = a\}.\end{aligned}$$

Let $\mathbb{P}'(X,S)$ be the set

$$\{p \in \mathbb{P}(X,S) : p \text{ is an essential period for some clopen set } [a], a \in \mathcal{A}\}.$$

and $\mathbb{Z}\mathbb{P}'(X,S)$ be the set of multiples of the elements of $\mathbb{P}'(X,S)$.

Lemma 6. *Let (X,S) be a minimal subshift and $x \in X$. Then, one has*

$$\mathrm{Per}(x) = \mathbb{Z}\mathbb{P}'(X,S). \tag{3.1}$$

Proof. It is clear that $\mathbb{P}'(X,S)$ is included in $\mathrm{Per}(x)$. Hence $\mathbb{Z}\mathbb{P}'(X,S)$ is also included in $\mathrm{Per}(x)$. The converse inclusion is clear from the definitions. ■

To answer Question 2, it is enough to answer positively Question 3. For this purpose, we should algorithmically determine the set $\mathbb{P}'(X,S)$. This, together with Proposition 3, provides a strategy to find the constant arithmetic subsequences: one has to find the essential periods, and thus the eigenvalues that are roots of unity, and then to check whether this provides such a sequence.

Let us enlighten what is above by considering some well-known families of minimal subshifts.

Toeplitz subshifts

Let $x = (x_n)_{n\in\mathbb{Z}} \in \mathcal{A}^{\mathbb{Z}}$. We say that x is a *Toeplitz sequence* if for all $k \in \mathbb{Z}$ there exists $p > 1$ such that $x_{k+np} = x_k$ for all $n \in \mathbb{Z}$. That is, for every k, we can find a constant arithmetic subsequence $(x_{k+np})_{n\in\mathbb{Z}}$. Observe that this does not force x to be periodic as p is depending on k.

Some of the Toeplitz sequences are substitutive [CK97, sec. 3]. For example, consider the substitution σ defined by $\sigma(0) = 1000$ and $\sigma(1) = 1010$. It has a fixed point $\sigma^{\omega}(0.1) = \cdots 0010001000.10101000101010\cdots$. Every index belongs to a constant arithmetic progression with common difference 2^p for some integer $p \in \mathbb{N}$. Thus this fixed point is a Toeplitz sequence.

The following proposition makes Proposition 3 more precise in the context of Toeplitz sequences. It asserts that eigenvalues are not only characterized by essential periods for clopen sets but for cylinder sets.

Proposition 7. *[Wil84] Let x be a Toeplitz sequence and (X,S) the subshift it generates. The following properties are equivalent:*

1. *λ is an eigenvalue of (X,S);*
2. *there exist p and an essential period q for $[a]$, for some letter a, such that $\lambda = \exp(2i\pi/p)$ and p divides q.*

Sturmian subshifts

Given a sequence $x \in \mathcal{A}^{\mathbb{Z}}$, we define its complexity function as $p_x(n) = \mathrm{card}\{x_{[i;i+n-1]} : i \in \mathbb{Z}\}$, which gives the number of factors of length n that occur in x. The Morse-Hedlund theorem [MH38] asserts that x is non-periodic if and only if it satisfies $p_x(n) \geq n+1$ for all $n \in \mathbb{N}$. We say that x is *sturmian* whenever x is uniformly recurrent and $p_x(n) = n+1$ for all $n \in \mathbb{N}$. A subshift is called *sturmian* if it is of the form $(\overline{\mathcal{O}(x)}, S)$, where x is a sturmian sequence.

Proposition 8. *Let x be a sturmian sequence. Then it admits no constant arithmetic subsequence.*

Proof. It is well-known that the set of eigenvalues of Sturmian subshifts is $\{\exp(2i\pi n\alpha) : n \in \mathbb{Z}\}$ for some $\alpha \notin \mathbb{Q}$ [Kur03, Section 4.5.3]. We conclude using Proposition 3. ∎

Substitutive subshifts

In this section we recall some results concerning eigenvalues associated to primitive substitutions that will be exploited in the next section in order to solve Question 1 of the introduction. A complete algebraic characterization of the eigenvalues of minimal substitution subshifts can be found in [FMN96].

We fix an integer $d \in \mathbb{N}$ and consider the alphabet $\mathcal{A} = \{0, 1, \ldots, d-1\}$.

We begin with two preliminary results where $\mathbf{1} \in \mathbb{Z}^d$ denotes the vector $(1, 1, \ldots, 1)$.

Lemma 9. *Let M be a square $d \times d$-matrix and p be a prime integer. The following properties are equivalent:*

1. *there exists $m \in \mathbb{N}$ such that $\mathbf{1}M^m \in p\mathbb{Z}^d$;*
2. *$\mathbf{1}M^d \in p\mathbb{Z}^d$.*

Proof. We will prove that (1) $\Rightarrow$ (2), the converse is obvious.

If m is less or equal to d then the conclusion is clear. Suppose m greater than d. One can suppose $m = \min\{i \in \mathbb{N} : \mathbf{1}M^i \in p\mathbb{Z}^d\}$. Applying Cayley-Hamilton theorem to M ensures that there exist some integers $a_0, a_1, \cdots, a_{d-1}$ such that $\mathbf{1}M^d = a_0\mathbf{1} + a_1\mathbf{1}M + \cdots + a_{d-1}\mathbf{1}M^{d-1}$. Then $\mathbf{1}M^{m+d-1} = a_0\mathbf{1}M^{m-1} + a_1\mathbf{1}M^m + \cdots + a_{d-1}\mathbf{1}M^{m-1+d-1}$, where every term $\mathbf{1}M^i$ except $\mathbf{1}M^{m-1}$ belongs to $p\mathbb{Z}^d$. It follows that $a_0\mathbf{1}M^{m-1}$ belongs to $p\mathbb{Z}^d$. As p is prime and $\mathbf{1}M^{m-1} \notin p\mathbb{Z}^d$, p divides a_0. With the same method and multiplying successively $\mathbf{1}M^d$ by $M^{m-2}, M^{m-3}, \ldots, M^{m-d}$, we establish that p divides $a_1, a_2, \ldots, a_{d-1}$. As a consequence, p divides $\mathbf{1}M^d$. ∎

Lemma 10. *Let (X, S) be a non-periodic subshift generated by a left-proper primitive substitution $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ with incidence matrix M and let p be an integer. The following properties are equivalent:*

1. *$\exp(2i\pi/p)$ is an eigenvalue of (X, S) ;*
2. *there exists $m \in \mathbb{N}$ such that p divides $|\sigma^m(a)|$ for all $a \in \mathcal{A}$;*

Moreover, when p is a prime number, this is equivalent to:

3. *$\mathbf{1}M^d = (|\sigma^d(a)|)_{a\in\mathcal{A}}$ belongs to $p\mathbb{Z}^d$, where $d = |\mathcal{A}|$.*

Proof. The equivalence between (1) and (2) follows from the proof of Lemma 27 in Durand's article [Dur00]. The equivalence between (2) and (3) comes from the fact that M^m is the incidence matrix of σ^m, for every $m \in \mathbb{N}$. Therefore, $\mathbf{1}M^m \in p\mathbb{Z}^d$ is the vector $(|\sigma^m(a)|)_{a\in\mathcal{A}}$. We conclude using Lemma 9. ∎

The previous lemma is stated for left-proper substitutions, but it is clear that it also holds for right-proper substitutions.

Given a minimal subshift (X, S) and $p \in \mathbb{P}(X, S)$ two situations can occur, $\sup\{n : p^n \in \mathbb{P}(X, S)\}$ is bounded or not. We will need this information to describe precisely $\mathbb{P}(X, S)$.

To this end, we denote $\mathbb{PP}(X,S)$ the set of prime numbers belonging to $\mathbb{P}(X,S)$ and $\mathbb{PP}^\infty(X,S)$ the set of $p \in \mathbb{PP}(X,S)$ such that p^n belongs to $\mathbb{P}(X,S)$ for all n.

Lemmas 10 and 11 respectively determine algorithmically $\mathbb{PP}(X,S)$ and $\mathbb{PP}^\infty(X,S)$. For $p \in \mathbb{PP}(X,S) \setminus \mathbb{PP}^\infty(X,S)$, Lemmas 12 and 13 provide an algorithm to compute the integer $n_{\max}$ such that $p^{n_{\max}}$ belongs to $\mathbb{P}(X,S)$ but $p^{n_{\max}+1}$ does not. We will need this result for our decidability problem in the next section.

Lemma 11. *[Dur00, Lemma 29] Let M be a $d \times d$ matrix and p a prime number. The following properties are equivalent:*

1. $\forall n \in \mathbb{N}, \exists k \in \mathbb{N} : \mathbf{1}M^k \in p^n\mathbb{Z}^d$;
2. *p divides $\gcd(a_0,\dots,a_r)$, where*

$$r = \max\{i \in \mathbb{N} : \{\mathbf{1}, \mathbf{1}M, \dots, \mathbf{1}M^i\} \text{ is free}\}$$

and $Q(X) = \sum_{i=0}^{r+1} a_i X^i \in \mathbb{Z}[X]$ is the characteristic polynomial of the restriction of M to the vector subspace spanned by $\mathbf{1}$, $\mathbf{1}M, \dots, \mathbf{1}M^r$.

Lemma 12. *Let M be a $d \times d$ matrix and p a prime number. If there exist $n \in \mathbb{N}$ and $i \in \mathbb{N}$ such that each of $\mathbf{1}M^i, \dots, \mathbf{1}M^{i+d}$ belong to $p^n\mathbb{Z}^d$ and does not belongs to $p^{n+1}\mathbb{Z}^d$, then for all $j \geq i$, $\mathbf{1}M^j$ does not belong to $p^{n+1}\mathbb{Z}^d$.*

As a consequence, if there exist two integers n and k satisfying $\mathbf{1}M^k \in p^n\mathbb{Z}^d$, then, in particular, the vector $\mathbf{1}M^{nd}$ belongs to $p^n\mathbb{Z}^d$.

Proof. We only sketch the proof, the arguments used are those of Lemma 9. By contradiction, let $j > i$ be the smallest integer such that $\mathbf{1}M^{j+d}$ belongs to $p^{n+1}\mathbb{Z}^d$. Let $Q(X) = \sum_{l=0}^{r+1} a_l X^l$ be the characteristic polynomial of M. Then, by Cayley-Hamilton theorem, every coefficient a_l, $0 \leq l \leq r$, is a multiple of p, which can be seen by multiplying the equality by successively $M^{j+d-1}, M^{j+d-2}, \dots, M^j$. This leads to a contradiction. ∎

Lemma 13. *Let M be a $d \times d$ matrix and p a prime number such that*

- $\mathbf{1}M^d \in p\mathbb{Z}^d$,
- $\exists n \in \mathbb{N} : \forall k \in \mathbb{N}, \mathbf{1}M^k \notin p^n\mathbb{Z}^d$.

Let $n_{\max} = \max\{n \in \mathbb{N} : \exists k \in \mathbb{N}, \mathbf{1}M^k \in p^n\mathbb{Z}^d\}$ and $k_{\min} = \min\{k \in \mathbb{N} : \mathbf{1}M^k \in p^{n_{\max}}\mathbb{Z}^d\}$. Let $K = \max\{k \in \mathbb{N} : p^{k-1} \leq \max_j \sum_i M_{i,j}\}$. Then

$$k_{\min} \leq p^d \text{ and } n_{\max} \leq Kp^d.$$

Proof. The inequality $k_{\min} \leq dn_{\max}$ is a consequence of the previous lemma. We detail the proof of the second inequality.

We define a graph $\mathcal{G}$ whose vertices belong to $[0, p-1]^d \setminus \{(0,\dots,0)\}$ and are given by the decomposition in base p of the vectors $\mathbf{1}M^k$ for $k \in \mathbb{N}$. More precisely, $\mathcal{G}$ is the unique oriented graph defined as follows:

- the vector $\mathbf{1}$ belongs to $\mathcal{G}$;
- there is an edge from V to W with label i if

$$VM = p^k V_k + p^{k-1} V_{k-1} + \cdots + pV_1 + V_0,$$

with $V_0, \ldots, V_k$ in $[0, p-1]^d \setminus \{(0, \ldots, 0)\}$, and $W = V_i$ for some i.

Notice that the labels of the edges are bounded by K.

Let $N(k) = \max\{n \in \mathbb{N} : \mathbf{1}M^k \in p^n \mathbb{Z}^d\}$. The sequence $k \mapsto N(k)$ is non-decreasing and is eventually constant equal to $n_{\max}$.

The aim of this graph is to give a bound for $k_{\min}$, that is, the first k such that $N(k) = n_{\max}$.

Observe that the growth of $N(k)$ is controlled by the paths of length k starting from $\mathbf{1}$. Indeed, for $k \in \mathbb{N}$, let $P_1, \ldots, P_m$ be these paths. To each path P_i, we associate $s(i)$, the sum of its labels. Then, $N(k) = \min_{1 \le i \le m} s(i)$. Let P_i be a path realizing this minimum. Notice that, as $k \mapsto N(k)$ is bounded by $n_{\max}$, any cycle in P_i should be labelled by 0. Hence, deleting the cycles in P_i, whose weight is 0 in $s(i)$, the remaining path has length at most p^d and thus $s(i) \le Kp^d$. In other words, one has $k_{\min} \le p^d$ and $n_{\max} \le Kp^d$. ∎

Constant-length substitutions

Let us now consider some particular minimal substitution subshifts where the situation is easier to handle.

The group of eigenvalues for minimal constant-length substitution subshifts has been determined by Dekking [Dek78]. To this end he introduced the notion of height of a substitution. Let σ be a primitive substitution of constant-length and $x = (x_n)_{n \in \mathbb{Z}}$ one of its fixed points. The *height* of σ is

$$h(\sigma) = \max\{n \ge 1 : (n, |\sigma|) = 1, n \text{ divides } g_0\},$$

where $g_0 = \gcd\{n \ge 1, x_n = x_0\}$.

Theorem 14 ([Dek78]). *Let σ be a primitive constant-length substitution and (X, S) the subshift it generates. Suppose (X, S) is non-periodic. Then,*

$$\mathbb{P}(X, S) = \{p \in \mathbb{N} : p \text{ divides } h(\sigma)|\sigma|^n \text{ for some } n \in \mathbb{N}\}.$$

Moreover, the group of eigenvalues of (X, S) is

$$\{\exp(2i\pi q/p) : q \in \mathbb{Z}, p \in \mathbb{P}(X, S)\}.$$

Remark 15. As a consequence, let us recall the well-known fact that subshifts arising from primitive constant-length substitutions only admit eigenvalues that are roots of unity.

This does not hold for non-constant length substitution subshifts. For example, consider the subshift arising from the Fibonacci substitution $\sigma : 0 \to 01$, $1 \to 0$. The set of its eigenvalues is $\{\exp(2i\pi n\alpha) : n \in \mathbb{N}\}$ with $\alpha = (\sqrt{5} - 1)/2$ [Kur03, Sec. 4.3.1]. The only eigenvalue that is a root of unity is 1.

Example. Let us consider the substitution σ defined by $0 \to 0213; 1 \to 1341; 2 \to 4104; 3 \to 0413; 4 \to 2134$ [Que10, sec. 6.1.1]. Its length is $|\sigma| = 4$. Following the algorithm of Dekking [Dek78, Remark II.9], we can compute $h(\sigma) = 3$. We conclude that the periodic spectrum of the underlying dynamical system (X, S) is $\mathbb{P}(X, S) = \{3^\delta \times 2^n : \delta \in \{0, 1\}, n \in \mathbb{N}\}$.

Automatic sequences

Let $\sigma : A^* \to \mathcal{A}^*$ be a substitution of constant-length l and $\phi : \mathcal{A}^* \to \mathcal{B}^*$ be a coding. Let x be a fixed point (in $\mathcal{A}^{\mathbb{N}}$ or $\mathcal{A}^{\mathbb{Z}}$) of the σ. The sequence $y = \phi(x)$ is called *l-automatic*. It is a substitutive sequence with respect to a constant-length substitution. A huge literature exists on this family of sequences [AS03].

4 An application to intersections of languages

Below we say that a language is Toeplitz, Sturmian, morphic, automatic, ... whenever it is the language of a sequence of the same type.

In a recent paper [RS18] it is observed that k-automatic sequences and Sturmian sequences cannot have arbitrary large factors in common. It is not surprising as it is a straightforward consequence of the well-known results recalled in the previous section and basics on dynamical systems. In the same spirit the following proposition can be established. Before we need the following lemma. We recall a language $\mathcal{L}$ is *factorial* if for all $u \in \mathcal{L}$, all words occurring in u belongs to $\mathcal{L}$.

Lemma 16. *Let $\mathcal{L}_1$ be a p-automatic language and $\mathcal{L}_2$ be a factorial language such that $\mathcal{L}_1 \cap \mathcal{L}_2$ is infinite. Then there exists a primitive p-automatic language $\mathcal{L}_1'$ included in $\mathcal{L}_1$ such that $\mathcal{L}_1' \cap \mathcal{L}_2$ is infinite.*

Proof. The proof is left to the reader. Hint: decomposition into sub-substitutions as in Proposition 15 in [Dur98b] may be used. ∎

Lemma 17. *Let $\mathcal{L}_1$ be a morphic language and $\mathcal{L}_2$ be an infinite uniformly recurrent factorial language such that $\mathcal{L}_2$ is included in $\mathcal{L}_1$. Then $\mathcal{L}_2$ is morphic.*

Proof. We left the proof to the reader. It uses the decomposition into sub-substitutions. ∎

Proposition 18. *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two languages. Then, $\mathcal{L}_1 \cap \mathcal{L}_2$ is finite in all the following situations:*

1. *$\mathcal{L}_1$ is automatic and $\mathcal{L}_2$ is sturmian;*
2. *$\mathcal{L}_1$ is sturmian and $\mathcal{L}_2$ is Toeplitz;*
3. *$\mathcal{L}_1$ is morphic and $\mathcal{L}_2$ is sturmian associated to a non quadratic rotation number;*

Proof. We give some hints. The details are left to the reader. We proceed by contradiction. We suppose $\mathcal{L}_1 \cap \mathcal{L}_2$ is infinite.

For Assertion (1) we first have to apply Lemma 16 in order to deal with primitive morphic languages. Assertions (1) and (2) can be deduced from the previous section applied to the subshifts generated by the languages. For Assertion (3), as sturmian sequences are uniformly recurrent one necessarily has that $\mathcal{L}_2$ is included in $\mathcal{L}_1$. Then we apply Lemma 17 and conclude with the fact that morphic subshifts that are also sturmian are α-morphic for some quadratic number α [DDM00]. ■

When $\mathcal{L}_1$ is α-morphic primitive with $\alpha \notin \mathbb{Z}$ and $\mathcal{L}_2$ is Toeplitz we would also like to conclude that $\mathcal{L}_1 \cap \mathcal{L}_2$ is finite but we did not find obvious arguments. One can show that when the underlying substitution has determinant ± 1 or has all its (matrix) eigenvalues with modulus greater or equal to 1 then $\mathcal{L}_1 \cap \mathcal{L}_2$ is finite but we leave the whole case as a question.

5 Constant arithmetic subsequences and eigenvalues for minimal morphic subshifts

In this section we prove the following theorem. Notice that its second statement corresponds to a positive answer to Question 1 for uniformly recurrent morphic sequences.

Theorem 19. *Let $y \in \mathcal{B}^{\mathbb{Z}}$ be a uniformly recurrent morphic sequence with respect to the endomorphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ and the morphism $\phi : \mathcal{A}^* \to \mathcal{B}^*$. Let (Y, S) be the minimal subshift it generates. Then the following properties hold.*

1. *The periodic spectrum of (Y, S) is algorithmically computable.*
2. *Given an integer $p \in \mathbb{N}$, it is algorithmically decidable whether y contains a letter in arithmetic progression with common difference p.*

Let us describe the way we proceed. We will consider two cases. We will begin considering purely substitutive sequences. Our algorithms rely on properties concerning non-periodic sequences. Thus we first consider the periodic case. It is easy to treat and provides all the letters appearing in x in arithmetic progression. We can then process, in the non-periodic case, with Algorithm 1 to compute the periodic spectrum. It is a requirement to apply Algorithm 2, which proves the second statement of the above theorem.

The general case of uniformly recurrent morphic sequences will come as a consequence.

Remark 20. Let $y \in \mathcal{A}^{\mathbb{Z}}$ be a uniformly recurrent sequence and p be some positive integer. If there exists n_0 such that $(y_{k+np})_{n>n_0}$ is constant, then, due to uniform recurrence, for any z in the subshift generated by y, there exists $i \in [0, p)$ such that $(z_{i+np})_{n\in\mathbb{Z}}$ is constant. As we are considering uniformly recurrent sequences in this section, it is enough to consider any admissible one-sided fixed point of σ to check whether x contains constant arithmetic subsequences.

Below, the inputs are an endomorphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, a morphism $\phi : \mathcal{A}^* \to \mathcal{B}^*$, $x \in \mathcal{A}^{\mathbb{Z}}$ an admissible fixed point of σ and $y = \phi(x)$ a uniformly recurrent sequence belonging to $\mathcal{B}^{\mathbb{Z}}$. We recall it is decidable to check whether y is uniformly recurrent or not, and uniformly recurrent morphic sequences are substitutive sequences with respect to primitive substitutions [Dur13b]. Thus, in Sections 5.1, 5.2 and 5.3, we suppose the endomorphism σ is primitive.

Let (X, S) be the minimal subshift generated by x and (Y, S) be the minimal subshift generated by y. We set $M = M_\sigma$. Algorithms 1 and 2 below answer positively to the decidability of Question 1 for purely substitutive sequences with respect to primitive substitutions.

We treat the general case in Section 5.4.

5.1 The periodic case

In this section, we consider the particular case of a left-proper primitive substitution $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ right-prolongable on $a \in \mathcal{A}$. Let $x = \sigma^\omega(a)$, (X, S) be the minimal substitution subshift generated by σ and M its incidence matrix.

It is decidable to check whether x is periodic [Dur12]. Moreover, if the answer is positive, this algorithm gives a period q for the sequence x. As a consequence of the Fine and Wilf Theorem [Lot02, Theorem 8.1.4], the essential period of x is a divisor of q. Thus, to find this essential period, it suffices to consider the factor $x_{[0,q-1]}$ of x and check if it has period p for every divisor p of q. The smallest such period p is the essential period of x.

Then the periodic subshift generated by x has the following periodic spectrum:

$$\mathbb{P}(X, S) = \{p' \in \mathbb{N} : p' \text{ divides } p\}.$$

Hence, for each $i \in \mathbb{N}$, there exists a constant arithmetic subsequence starting at index i, with an essential period p_i that divides p. All the p_i can be determined by studying the occurrences of letter x_i in the factor $x_{[0,p-1]}$ of x.

For the sequel, we suppose that the subshift (X, S) is non-periodic.

5.2 Algorithm 1: determining $\mathbb{P}(X, S)$ when σ is left-proper

In order to determine $\mathbb{P}(X, S)$ we first determine the set $\mathbb{PP}(X, S)$ of prime numbers belonging to $\mathbb{P}(X, S)$. Observe that from Proposition 3 and Lemma 10 the set $\mathbb{PP}(X, S)$ is finite. This set is composed of two types of primes. As in Section 3, let $\mathbb{PP}^\infty(X, S)$ be the set of $p \in \mathbb{PP}(X, S)$ such that p^n belongs to $\mathbb{P}(X, S)$ for all n. For $p \in \mathbb{PP}(X, S) \setminus \mathbb{PP}^\infty(X, S)$ we set

$$n_{\max}(p) = \max\{n : p^n \in \mathbb{P}(X, S)\}.$$

Let

$$\mathbb{PP}^\infty(X, S) = \{p_1, \ldots, p_k\} \text{ and } \mathbb{PP}(X, S) \setminus \mathbb{PP}^\infty(X, S) = \{q_1, \ldots, q_l\}.$$

Observe that the set of additive eigenvalues of (X,S) being a subgroup of $\mathbb{R}/\mathbb{Z}$, $\mathbb{P}(X,S)$ is the set of integers

$$\prod_{1\le i\le k} p_i^{r_i} \prod_{0\le j\le l} q_j^{s_j} \tag{5.1}$$

with r_i and s_j in $\mathbb{N}$ such that $s_j \in [0, n_{\max}(q_j)]$ for $1 \le i \le k$, $0 \le j \le l$.

Step 1. Determine $\mathbb{PP}(X,S)$

According to Lemma 10, denoting by $\mathbf{P}$ the set of prime numbers,

$$\mathbb{PP}(X,S) = \{p \in \mathbf{P} : p \text{ divides} \gcd((\mathbf{1}M^d)_i, 1 \le i \le d)\},$$

which is clearly computable.

Step 2. Determine $\mathbb{PP}^\infty(X,S)$

We use Lemmas 10 and 11.

- Compute $r = \max\{i \in \mathbb{N} : \{\mathbf{1}, \mathbf{1}M, \ldots \mathbf{1}, M^i\} \text{ is free}\}$.
- Compute $\tilde{M}$ the restriction of M to the vector subspace spanned by $\mathbf{1}, \mathbf{1}M, \ldots, \mathbf{1}M^r$.
- Determine its characteristic polynomial $Q(X) = \sum_{i=0}^{r+1} a_i X^i \in \mathbb{Z}[X]$.
- Compute $\gcd(a_0, \ldots, a_r)$. Its prime divisors form the set $\mathbb{PP}^\infty(X,S)$. We set:

  $$\mathbb{PP}^\infty(X,S) = \{p_1, p_2, \ldots, p_k\}.$$

  where the p_i's are pairwise distinct.

Step 3. Determine the maximal power q^n for $q \in \mathbb{PP}(X,S) \setminus \mathbb{PP}^\infty(X,S)$

We use Lemma 12. Let $Q = \max(\mathbb{PP}(X,S) \setminus \mathbb{PP}^\infty(X,S))$.

- Starting from $n = 1$:
  - Compute $g_n = \gcd((\mathbf{1}M^{nd})_1, \ldots, (\mathbf{1}M^{nd})_d)$.
  - Determine $\tilde{g}_n = \max\{g \in \mathbb{N} : g | g_n \text{ and } \gcd(g, p_m) = 1 \text{ for } 1 \le m \le j\}$.
- Carry on till $\tilde{g}_n = \tilde{g}_{n+1}$. According to Lemma 13, it will halt in a finite time bounded by KQ^d, where K is defined as in Lemma 13.

  Let $\tilde{g}_n = q_1^{n_1} \cdots q_l^{n_l}$. Then, $\mathbb{PP}(X,S) \setminus \mathbb{PP}^\infty(X,S) = \{q_1, \ldots, q_l\}$ and $n_i = n_{\max}(q_i)$, $1 \le i \le l$.

Step 4. Output of the algorithm

One obtains $\mathbb{P}(X,S)$ as described in (5.1). With this output and Proposition 3 we are able to describe the whole group of rational eigenvalues associated to the system (X,S): they are the complex numbers $\exp(2i\pi q/p)$ with $q \in \mathbb{Z}$ and $p \in \mathbb{P}(X,S)$.

Remark 21. In the case of a 2-letter alphabet, Host [Hos86, sec. 2.3] established an algorithm that computes $\mathbb{P}(X,S)$ for non constant-length substitution subshifts. We recall it below.

Let σ be a non constant-length substitution. Let M be its incidence matrix, with determinant d and trace T. Then

$$\mathbb{P}(X_\sigma, S) = \{p \in \mathbb{N} : p \text{ divides some } w \times r^n \text{ for some } n \in \mathbb{N}\},$$

where:

- $r = \gcd(d,T)$;
- the prime divisors of w are those of $|\sigma(0)|$ and $|\sigma(1)|$ that do not divide r, with the same exponents as in $|\sigma(0)| - |\sigma(1)|$.

5.3 Algorithm 2: for a given p, checking if x has a constant arithmetic subsequence with common difference p, when σ is left-proper

This corresponds to the decidability of Question 1 for fixed points of primitive substitutions. Recall that we consider a non-periodic substitution subshift (X,S) and that the inputs are given by Theorem 19.

Let p be an positive integer.

Step 1. Determine $\mathbb{P}(X,S)$ as described in Algorithm 1.

We start by determining the greatest divisor $\tilde{p}$ of p that belongs to $\mathbb{P}(X,S)$. Following the notations given in Algorithm 1, $\tilde{p}$ is of the form $\tilde{p} = p_1^{r_1}\cdots p_k^{r_k} q_1^{s_1}\cdots q_l^{s_l}$ with $r_i \in \mathbb{N}$, with $1 \leq i \leq k$, and, $0 \leq \beta_j \leq n_{\max}(q_j)$, with $1 \leq j \leq l$. We can easily observe that the sequence x contains a letter in arithmetic progression with common difference p if and only if it contains a letter in arithmetic progression with common difference $\tilde{p}$ (see Lemma 6).

Step 2. Check if $\tilde{p}$ corresponds to a letter in arithmetic progression in x.

There exists, according to Lemma 10, an integer m_p such that $\tilde{p}$ divides $|\sigma^{m_p}(a)|$ for all $a \in \mathcal{A}$. Due to Lemma 12, we can take $m_p = d\max\{r_1,\ldots,r_k,s_1,\ldots,s_l\}$. For every integer i such that $0 \leq i \leq \tilde{p}-1$, check if $(n,a) \mapsto (\sigma^{m_p}(a))_{i+n\tilde{p}}$ is constant on $\{(n,a) : a \in \mathcal{A}, 0 \leq i+n\tilde{p} \leq |\sigma^{m_p}(a)|-1\}$. If such an integer i exists, then x has a letter in arithmetic progression with common difference $\tilde{p}$ (and thus, also with common difference p) starting at index i. If not, there does not exist any letter in arithmetic progression with common difference $\tilde{p}$ (and thus, none of common difference p).

Example. Let σ be the left-proper primitive substitution defined on the alphabet $\mathcal{A} = \{0,1\}$ by $0 \mapsto 01$ and $1 \mapsto 0110$. Its incidence matrix is

$$M = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}.$$

We can check that the unique fixed point $x \in \mathcal{A}^{\mathbb{Z}}$ of σ is not periodic. As σ is left-proper and $|\mathcal{A}| = 2$, it suffices to compute $M^2 = \begin{pmatrix} 3 & 6 \\ 3 & 6 \end{pmatrix}$, which gives $\mathbf{1}M^2 = (6,12)$. Since $\gcd(6,12) = 6$, the prime integers in $\mathbb{P}(X_\sigma, S)$ are 2 and 3. Following Algorithm 1, we establish that $\mathbb{P}(X_\sigma, S) = \{3^m, 2 \times 3^m : m \in \mathbb{N}\}$.

Remark that $\sigma(1)$ has no arithmetic progression with common difference 2. As $|\sigma(0)|$ and $|\sigma(1)|$ are divisible by 2, it follows that x does not have any letter in arithmetic progression with common period 2.

Using the same method (Step 2 of Algorithm 2) with

$$\sigma^2 : \begin{cases} 0 \to 010110 \\ 1 \to 010110011001 \end{cases}$$

we see that x does not contains any letter in arithmetic progression with common difference 3, but it admits arithmetic progressions with common difference 6: $x_{6n} = 0$ and $x_{6n+1} = 1$ for all $n \in \mathbb{Z}$.

5.4 Algorithm 3: for a given p, checking if y has a constant arithmetic subsequence with common difference p, general case

We recall that the general case is that y is a uniformly recurrent morphic sequence. Thus the inputs are an endomorphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, a morphism $\phi : \mathcal{A}^* \to \mathcal{B}^*$, $x \in \mathcal{A}^{\mathbb{Z}}$ an admissible fixed point of σ with $y = \phi(x)$. Let (X, S) and (Y, S) be the respective subshifts these sequences generate.

Observe that even if y is uniformly recurrent, σ is not necessarily primitive, it can even have erasing letters, and ϕ is not necessarily a coding. Nevertheless one can find a primitive substitution σ, with an admissible fixed point x', and a coding ϕ' such that $y = \phi'(x')$. Moreover, this can be done algorithmically.

Theorem 22. *[Dur13b] Uniformly recurrent morphic sequences are primitive substitutive sequences.*

The fact that this can be done algorithmically can be easily deduced from the proof of this theorem [Dur13b, Section 5.1] with the additional property that the substitution can be chosen left-proper.

Consequently, we suppose, without loss of generality, that σ is a left-proper primitive substitution and ϕ is a coding. In addition, and again without loss of generality, we suppose $\phi : (X, S) \to (Y, S)$ is an isomorphism [DHS99, Dur00, Prop. 31], and thus that $\mathbb{P}(Y, S) = \mathbb{P}(X, S)$.

Then, to check whether, for a given $p > 1$, there exists a constant arithmetic subsequence $(y_{k+np})_n$ equal to some letter a, it is necessary and sufficient to check whether $(x_{k+np})_n$ is a sequence on the alphabet $\phi^{-1}(\{a\})$.

Let us translate this into an algorithm. Let $p \geq 1$.

Step 1.

Following the algorithms in [Dur13b] and [Dur00], we compute

- a left-proper primitive substitution $\zeta : \mathcal{B}^* \to \mathcal{B}^*$ with fixed point z,
- a coding $\psi : \mathcal{B}^* \to \mathcal{A}^*$ such that $\psi(z) = y$.

Let (Z, S) be the subshift generated by ζ. The factor map $\psi : (Z, S) \to (Y, S)$ is an isomorphism [DHS99, Dur00].

Step 2.

Apply Algorithm 1 to ζ to determine $\mathbb{P}(Y, S) = \mathbb{P}(Z, S)$. Choose an integer $\tilde{p} \in \mathbb{P}(Y, S)$ as in Step 1 of Algorithm 2.

Step 3. Check if there exists a subsequence $(z_{k+n\tilde{p}})_n$ defined on an alphabet $\phi^{-1}(\{a\})$, $a \in \mathcal{A}$.

We proceed as in the Step 2 of Algorithm 2 except that one has to check whether one of the maps $(n, b) \mapsto (\zeta^{m_p}(b))_{k+n\tilde{p}}$ has all its images in some $\phi^{-1}(\{a\})$, $a \in \mathcal{A}$.

If it is the case for some i and $a \in \mathcal{A}$, then $\tilde{p}$, and thus p, corresponds to a constant arithmetic subsequence of y. Otherwise, it is not.

Example. Let σ be the substitution defined on the alphabet $\mathcal{A} = \{0, 1, 2\}$ by $0 \to 02$, $1 \to 2$ and $2 \to 10$. Following the algorithm of Durand [Dur00, proof of Prop. 31], we find that (X_σ, S) is isomorphic to (X_τ, S) where τ is the left-proper substitution defined by $1 \to 6134242, 2 \mapsto 61342426134242, 3 \mapsto 6134261356135$, $4 \mapsto 613426135, 5 \mapsto 6134261356135, 6 \mapsto 613426135$, whose incidence matrix is

$$M_\tau = \begin{pmatrix} 1 & 2 & 3 & 2 & 3 & 2 \\ 2 & 4 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 2 & 3 & 2 \\ 2 & 4 & 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 1 & 2 & 1 \\ 1 & 2 & 3 & 2 & 3 & 2 \end{pmatrix}.$$

If we compute the sum of each column of M_τ^6 (to apply algorithm 1), we obtain the vector $(930072, 1860144, 1675961, 1159797, 1675961, 1159797)$ whose entries have greatest common divisor equal to 1. Thus, the substitution σ does not admit any non-trivial rational eigenvalue. As a consequence, no sequence of X_σ has a constant arithmetic subsequence.

6 The case of uniformly recurrent automatic sequences

In Section 3 we defined automatic sequences. We recall below that these sequences correspond exactly to definable sets in well-chosen extensions of Presburger arithmetic, Section 6.1. Then, we aim to show that for such a framework Question 1 is decidable even in the non-primitive case. The proof is immediate once we recall classical results on Presburger arithmetic. Then using the characterization of the eigenvalues of minimal subshifts generated by constant-length substitutions given by Dekking [Dek78], we answer positively to Question 2 and Question 3 for uniformly recurrent automatic sequences.

6.1 Automatic sequences and Presburger arithmetic

The *Presburger arithmetic* [Pre29, Pre91] on $\mathbb{N} = \{0, 1, \dots\}$ is the first order logical structure $\langle \mathbb{N}, + \rangle$, on $\mathbb{Z}$ it is $\langle \mathbb{Z}, \geq, + \rangle$. That is the set of formulas without free variables composed of elements of $\mathbb{K} = \mathbb{N}$ or $\mathbb{Z}$, variables taking values in $\mathbb{K}$, addition, equality,

- the connectives $\vee$ (or), $\wedge$ (and), $\neg$ (not), $\rightarrow$ (then), $\leftrightarrow$ (iff), and,
- the quantifiers: $\forall$ (for all), $\exists$ (there exists),

We refer the reader to Rigo's book [Rig14] for more details. Observe that the order relation $\geq$ is definable in $\langle \mathbb{N}, + \rangle$ noticing that $n \geq m$ if and only if there exists $k \in \mathbb{N}$ such that $n = m + k$. Thus, $\langle \mathbb{N}, + \rangle = \langle \mathbb{N}, \geq, + \rangle$.

A subset $E \subset \mathbb{K}$ is *definable* if there exists a map ϕ from $\mathbb{K}$ to the set of formulas over $\langle \mathbb{K}, \geq, + \rangle$ such that E is the set of numbers $n \in \mathbb{K}$ such that $\phi(n)$ is true. It is well-known that the definable sets are the finite unions of arithmetic progressions and that this logical structure is decidable. This means that given a formula there exists an algorithm answering in finite time whether this sentence is true or false.

Fix $l \geq 2$ an integer. We define the function V_l on $\mathbb{K}$ as $V_l(0) = 1$ and $V_l(n) = l^i$ where l^i is the largest power of l dividing n (e.g. $V_2(12) = 4$). We consider the first-order logical structure $\langle \mathbb{K}, \geq, +, V_l \rangle$. It is an extension of the Presburger arithmetic. We define formulas and l-definable sets as we did before. This structure is again decidable [BHMV94, Theorem 4.1 and Corollary 6.2]. We say $x \in \mathcal{A}^{\mathbb{K}}$ is l-definable if for all $a \in \mathcal{A}$ there exists a formula ϕ_a defining the set $\{n \in \mathbb{K} : x_n = a\}$ in the theory $\langle \mathbb{K}, \geq, +, V_l \rangle$. We say the formulas $(\phi_a)_{a \in \mathcal{A}}$ define x.

Theorem 23. *Let $l \in \mathbb{N}$ and $x \in \mathcal{A}^{\mathbb{K}}$. Then the following properties are equivalent:*

1. *x is $l-$automatic ;*
2. *x is $l-$definable.*

In the theory $\langle \mathbb{K}, \geq, +, V_l \rangle$, the property "$x$ admits a letter a in arithmetic progression with common difference $p \geq 1$" is given by the sentence

$$\exists a, \exists i : \forall k \in \mathbb{K}, \phi_a(i + pk)$$

and therefore is decidable, for any given $p \in \mathbb{K}$. Observe that multiplication by a constant, here p, is definable in the Presburger arithmetic, that is, once some $p \in \mathbb{N}$ is fixed, the set $S = \{n : \exists k \in \mathbb{Z}, n = kp\}$ is definable where kp is the abbreviation for $k + \cdots + k$ (p times). Whereas the multiplication of natural numbers is not definable in $\langle \mathbb{Z}, \geq, +, V_l \rangle$ [Bès01]. That is, the set $\{(x, y, n) \in \mathbb{K}^3 : xy = n\}$ is not definable in $\langle \mathbb{K}, \geq, +, V_l \rangle$. If it was, then $\langle \mathbb{K}, \geq, +, V_l \rangle$ would include the Peano arithmetic which is known to be undecidable.

Now we suppose x belongs to $\mathcal{A}^{\mathbb{K}}$ and is l-automatic. Then, given some $p \geq 1$, the property "x admits some letter a in arithmetic progression with common difference $p \geq 1$" is given by the following formulas indexed by a:

$$\exists i : \forall k \in \mathbb{K}, \phi_a(i + pk).$$

We proved the following theorem.

Proposition 24. *Question 1 is decidable for automatic sequences.*

Observe that Question 2 can be described by the following formula:

$$\exists p \geq 1, \exists i : \forall k \in \mathbb{K}, \phi_a(i + pk).$$

But here this formula uses the multiplication of two variables and thus it is not a formula from $\langle \mathbb{K}, \geq, +, V_l \rangle$. Hence we cannot conclude directly the decidability of this question. It is possible that this statement could be rewritten into a formula from $\langle \mathbb{K}, \geq, +, V_l \rangle$ but we did not find it.

6.2 Heights and eigenvalues of minimal constant length substitutions subshifts

We recall some well-known results on minimal constant length substitution subshifts.

Let us consider a primitive constant-length substitution σ defined on the alphabet $\mathcal{A}$. We set $d = \operatorname{card} \mathcal{A}$. We denote l its length and h its height (possibly equal to 1). Let $x \in \mathcal{A}^{\mathbb{Z}}$ be one of its admissible fixed points and (X, S) the minimal subshift generated by σ. It is decidable to know whether x is periodic or not [Hon86, ARS09]. As the periodic case has been studied in Section 5.1, we suppose for the sequel that (X, S) is non-periodic.

In that case, we recall (Theorem 14) that the periodic spectrum is

$$\mathbb{P}(X, S) = \{\text{divisors of some } h \times l^m, \text{ with } m \in \mathbb{N}\}. \tag{6.1}$$

Let us recall some well-known properties of the height. Let

$$g_k = \gcd\{n \geq 1, x_{k+n} = x_k\}.$$

Then one has the following property [Dek78, Rk. II.9 (ii)]

$$\begin{aligned} h &= \max\{n \geq 1 : (n, l) = 1, n \text{ divides } g_k\} \\ &= \max\{n \geq 1 : (n, l) = 1, n \text{ divides } g_0\}. \end{aligned}$$

For all $i \in \mathbb{N}$ we set $\mathcal{A}_i = \mathcal{A}_i(x) = \{x_{i+nh} : n \in \mathbb{Z}\}$.

Proposition 25. *[Dek78, Rk. II.9 (ii)] The height h of σ is algorithmically computable. Moreover $(\mathcal{A}_i)_{0\le i<h}$ is a partition of $\mathcal{A}$ and it is algorithmically computable.*

Of course, $i \equiv j \mod h$ if and only if $\mathcal{A}_i = \mathcal{A}_j$. Observe that if $h = 1$ then this partition is reduced to the whole alphabet $\mathcal{A}$. For the other extremal situation, $h = d$ implies x is periodic.

6.3 Periods in primitive automatic subshifts

Let $y \in \mathcal{B}^{\mathbb{Z}}$ be a primitive automatic sequence, that is $y = \phi(x)$ where σ is a primitive constant-length substitution, $x \in \mathcal{A}^{\mathbb{Z}}$ one of its admissible fixed points and ϕ a coding.

Consider (X, S) and (Y, S) the minimal subshifts defined by x and y respectively. The subshift (Y, S) is clearly a factor of (X, S) as ϕ defines a factor map from (X, S) onto (Y, S). Consequently the set $\mathbb{P}(Y, S)$ is included in $\mathbb{P}(X, S)$.

To answer Question 2, it is sufficient to answer positively Question 3. For this purpose, we should algorithmically determine the set $\mathrm{Per}(y)$. We recall it is equal to $\mathbb{Z}\mathbb{P}'(Y, S)$, Section 3. Thus, to answer Question 3 it is sufficient to determine $\mathbb{P}'(Y, S)$ which is included in $\mathbb{P}(Y, S)$ and thus in $\mathbb{P}(X, S)$.

We proceed as follows. We determine the alphabet consisting of the letters occurring in every possible arithmetic subsequence with common difference of the form $h \times l^m, m \in \mathbb{N}$. Due to the regularity of the construction (inherited by the constant-length of the substitutions we are dealing with) described below, all these alphabets can be represented as vertices of a directed edge-labelled graph $G = (V, E)$ and the sequence y admits a constant arithmetic subsequence if and only if one of those alphabets equals $\{b\}$, $b \in \mathcal{B}$.

6.4 A graph to describe the sets $\mathbb{P}'(X, S)$ and $\mathbb{P}'(Y, S)$

We keep the assumptions and notations of Section 6.2 and Section 6.3, in particular the partition of $\mathcal{A}$ into the subsets $\mathcal{A}_i$ given by Proposition 25 for σ.

We define a graph, $G(\sigma)$, that will characterize $\mathbb{P}'(X, S)$ and $\mathbb{P}'(Y, S)$.

Let $G' = (V', E')$ be the directed graph where V' is the family of subsets of $\mathcal{A}$ and where $(\mathcal{C}, \mathcal{D})$ is an edge of E' whenever there exists some integer i, $0 \le i < l$, such that

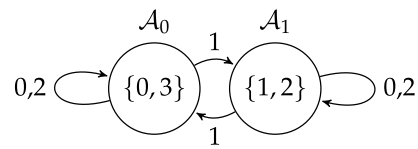
$$\mathcal{D} = \{\sigma(b)_i : b \in \mathcal{C}\}.$$

Moreover we will consider the edges-labelling function $f : E' \to \{0, \ldots, l-1\}$ defined by $f(\mathcal{C}, \mathcal{D}) = i$. We say the vertex $\mathcal{D}$ is *reachable* from $\mathcal{C}$ whenever there exists a finite sequence of edges $(\mathcal{C}, \mathcal{C}_1)(\mathcal{C}_1, \mathcal{C}_2)\ldots(\mathcal{C}_{i-1}, \mathcal{C}_i)(\mathcal{C}_i, \mathcal{D})$.

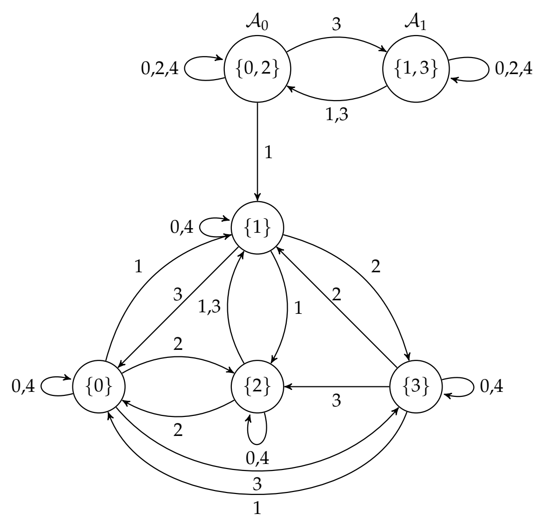
Let $G(\sigma) = (V, E)$ be the subgraph of G' where V is the set of vertices that are reachable from some vertices $\mathcal{A}_i$, $i \in \{0, \ldots, h-1\}$. A *walk* of $G(\sigma)$ of *length* i is a finite sequence of edges of the type $(\mathcal{C}_1, \mathcal{C}_2)(\mathcal{C}_2, \mathcal{C}_3)\ldots(\mathcal{C}_{i-1}, \mathcal{C}_i)$. The vertex $\mathcal{C}_1$ is called the *starting* vertex of this walk and $\mathcal{C}_i$ the *terminal* vertex. The *label* of this path is the finite sequence $(f(\mathcal{C}_1, \mathcal{C}_2), f(\mathcal{C}_2, \mathcal{C}_3), \ldots, f(\mathcal{C}_{i-1}, \mathcal{C}_i))$. A walk of $G(\sigma)$ is called *admissible* if it starts from one of the vertices $\mathcal{A}_i$, $0 \le i \le h-1$.

Example 1. Let us consider the substitution defined on the alphabet $\mathcal{A} = \{0,1,2,3\}$ by $\sigma(0) = 013$, $\sigma(1) = 102$, $\sigma(2) = 231$, $\sigma(3) = 320$. It has height $h = 2$, which leads to the following partition of $\mathcal{A}$: $\mathcal{A}_0 = \{0,3\}$, $\mathcal{A}_1 = \{1,2\}$.

We obtain the following graph:



Example 2. Now, consider the substitution σ defined on the alphabet $\mathcal{A} = \{0,1,2,3\}$ by $\sigma(0) = 01230$, $\sigma(1) = 12301$, $\sigma(2) = 21012$, $\sigma(3) = 30123$. It has height $h = 2$, which leads to the following partition of $\mathcal{A}$: $\mathcal{A}_0 = \{0,2\}$, $\mathcal{A}_1 = \{1,3\}$. We obtain the following graph.



Let us point out some properties of these graphs that will allow us to conclude with Question 3.

Lemma 26. *The graph $G(\sigma)$ is algorithmically computable.*

Proof. It is clear from Proposition 25 and the fact that the number of vertices is bounded by 2^d. ■

For a sequence z in $\mathcal{A}^{\mathbb{K}}$ we set $\mathcal{A}(z) = \{z_n : n \in \mathbb{K}\}$, where $\mathbb{K} = \mathbb{N}$ or $\mathbb{Z}$. We call it the *alphabet* of z.

Let $m \geq 0$ and k be integers such that $0 \leq k < hl^m$. We will denote by $(k_m, k_{m-1}, \ldots, k_1, k_0)$ the expansion of k in base $(l^m, l^{m-1}, \ldots, l, 1)$, i.e. $k = \sum_{i=0}^{m} k_i l^i$ with $0 \leq k_m < h$ and $0 \leq k_i < l$ for $0 \leq i < m$.

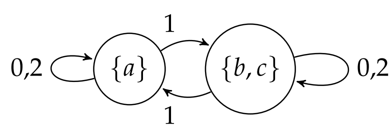
Lemma 27. *Let $m \geq 0$ and k be integers such that $0 \leq k < hl^m$. Let $(k_m, k_{m-1}, \ldots, k_1, k_0)$ be the expansion of k in base $(l^m, l^{m-1}, \ldots, l, 1)$. Then, the set $\mathcal{A}((y_{k+nhl^m})_n)$ is the image under ϕ of the terminal edge of the admissible walk in the graph $G(\sigma)$ starting from the vertex $\mathcal{A}_{k_m}$ with label $(k_{m-1}, \ldots, k_1, k_0)$, read from left to right.*

Proof. We proceed by induction on m. If $m = 0$, the statement is clear from the definition of the $\mathcal{A}_i$'s.

Suppose that the property holds for some integer $m \geq 0$.

Let k be an integer such that $0 \leq k < hl^{m+1}$ and $(k_{m+1}, \dots, k_1, k_0)$ be its representation in base $(l^{m+1}, l^m, \dots, l, 1)$. Then, $k + nhl^{m+1} = k_0 + l(k_{m+1}l^m + k_m l^{m-1} + \cdots + k_1 + nhl^m)$. Therefore, for all n, $y_{k+nhl^{m+1}} = \phi(x_{k+nhl^{m+1}})$ is the k_0-th letter of the word $\phi(\sigma(x_{k_{m+1}l^m+\cdots+k_1+nhl^m}))$. By induction hypothesis, the alphabet $\mathcal{A}((y_{k_{m+1}l^m+\cdots+k_1+nhl^m})_n)$ is the image under ϕ of the terminal vertex of the admissible walk starting from $\mathcal{A}_{k_{m+1}}$ with label $(k_m, \dots, k_1)$. By construction of the graph $G(\sigma)$, the alphabet $\mathcal{A}((y_{k+nhl^{m+1}})_n)$ is the image under ϕ of the terminal vertex of the admissible walk starting from $\mathcal{A}_{k_{m+1}}$ with label $(k_m, \dots k_1, k_0)$, which achieves the proof of the claim. ■

Example 3. We consider the substitution σ defined in Example 1 above. Let ϕ be the morphism defined by $0 \to a, 1 \to b, 2 \to c$ and $3 \to a$, and y the image of x under ϕ. We apply the morphism ϕ to each vertex of the graph $G(\sigma)$, which could be represented by the following labelling of the graph $G(\sigma)$.



Remark 28. As a consequence, the alphabet of any arithmetic subsequence of y with common difference hl^m, $m \in \mathbb{N}$, is the image under ϕ of a vertex of the graph $G(\sigma)$. In particular, there exists a constant arithmetic subsequence $(y_{k+nhl^m})_n$ equal to a if and only if, the image under ϕ of the terminal edge of the unique walk starting from vertex $\mathcal{A}_m$ with label $(k_{m-1}, \dots, k_1, k_0)$ is $\{a\}$.

Example 1 (continued). From Lemma 27, none of its four (one-sided) fixed points admit a letter in arithmetic progression. In fact, the substitution being primitive, it is sufficient to check this for just one of them.

Example 2 (continued). We observe that every letter of $\mathcal{A}$ appears in x in arithmetic progression. In fact, the substitution σ being primitive and of constant length, if there exists a letter that appears in arithmetic progression, then every letter of $\mathcal{A}$ would occur in x in arithmetic progression. From Lemma 27 a period for the letter 1 is $2 \times 5 = 10$. Thus, the essential period should be among 2, 5 or 10. It cannot be 2, else either $\mathcal{A}_0$ or $\mathcal{A}_1$ would be $\{1\}$. Moreover, from (6.1) and (3.1) it is a multiple of $h = 2$, so the essential period for the letter 1 is equal to 10: more precisely, $x_{1+10n} = 1$ for all n and this is the only constant arithmetic subsequence with common difference 10 in x.

Example 3 (continued). The letter a occurs in y in arithmetic progression with common difference 2. Notice that the letters b and c occur in y at the same indices as 1 and 2 in x, thus they do not appear in arithmetic progression. As a consequence, the sequence y has only one constant arithmetic subsequence with an essential period: $y_{2n} = 0$ for all $n \in \mathbb{N}$.

Thus from Lemma 27 and the properties of the graph $G(\sigma)$, we are able to answer positively to Question 2 and Question 3 with the following proposition.

Proposition 29. *Let $y \in \mathcal{B}^{\mathbb{Z}}$ be a primitive automatic sequence, that is $y = \phi(x)$ where σ is a primitive constant-length substitution, $x \in \mathcal{A}^{\mathbb{Z}}$ one of its admissible fixed points and ϕ a coding. Then, $\mathbb{Z}\mathbb{P}'(Y,S) = \mathrm{Per}(y)$ is the set of integers hl^m such that $\phi(\mathcal{A}((x_{k+nhl^m})_n))$ is a singleton for some k with $0 \leq k < hl^m$. Moreover, this set is algorithmically computable.*

The following proposition makes precise the positive answer to Question 2.

Proposition 30. *The graph $G(\sigma)$ satisfies the following properties.*

1. *The sequence y admits a constant arithmetic subsequence if, and only if, there exists a vertex $\mathcal{C}$ of $G(\sigma)$ whose image under ϕ is a singleton.*

2. *The sequence y is periodic if, and only if, every long enough walk in the graph $G(\sigma)$ ends in a vertex whose image under ϕ is a singleton.*

Remark 31. The quantity $b(\sigma) = \min_{\mathcal{B}\in V} |\mathcal{B}|$ is called the *branching number* [Kam72]. The sequence x admits a constant arithmetic subsequence if and only if $b(\sigma) = 1$.

6.5 More properties of the graph $G(\sigma)$

We continue with the notations and assumptions of the two previous sections. We will now use the graph $G(\sigma)$ to characterize the set of essential periods of the letters (i.e. of $[a]$ with $a \in \mathcal{A}$).

For the sequel, we will call any vertex reachable from a given vertex a *successor* of this vertex.

We need the following proposition.

Proposition 32. *The graph $G(\sigma)$ satisfies exactly one of the following properties.*

1. *It doesn't contain any singleton.*

2. *Every long enough walk ends in a singleton.*

3. *There exists a cycle joining vertices of cardinality ≥ 2, with one having a singleton in his successors.*

Proof. Suppose that we are neither in Case (1) nor in Case (2). We will show that Case (3) is satisfied.

By hypothesis, the graph $G(\sigma)$ contains at least one singleton. Due to Remark 28, to this singleton corresponds a constant arithmetic subsequence of x, with common difference hl^m for some integer m. Moreover, there exist arbitrarily long walks with vertices of cardinality greater or equal to 2. As the number of edges is finite, there exists a cycle in these vertices.

Suppose by contradiction that the vertices of this cycle do not have any singleton in their successors. Pick $\mathcal{C}$ among these vertices. According to Lemma 27, there exist two integers k and p such that $\mathcal{C}$ is the alphabet of the subsequence $(x_{k+nhl^p})_{n\in\mathbb{N}}$. Let $(\mathcal{C},\mathcal{C}_i)$, $0 \leq i < l$, be the l edges starting in $\mathcal{C}$. Each $\mathcal{C}_i$ is the alphabet of the subsequence $(x_{i+kl+nhl^{p+1}})_n$ for $0 \leq i < l$. By a direct induction, the j-th successors of $\mathcal{C}$ contains the alphabets of each subsequence $x_{i+kl^j+nhl^{p+j}}$ for $0 \leq i < l^j$. None of them are constant because $\mathcal{C}$ has no singleton in its successors. Then, the sequence x contain arbitrarily long subwords with no letter in arithmetic progression. This contradicts the fact that there exist a constant arithmetic subsequence with common difference hl^m. Therefore, $\mathcal{C}$ has a singleton in its successors and Property (3) holds. ∎

For each of these cases, we now detail the consequences for arithmetic progressions. Property (1) holds if, and only if, there is no constant arithmetic subsequence in the sequence y (Proposition 30).

Remark 33. In the graph $G(\sigma)$, each alphabet has a cardinality greater or equal to each of its successors. In particular, every successor of a singleton is also a singleton.

Proposition 34. *Every long enough walk in the graph $G(\sigma)$ ends in a singleton if, and only if, the sequence x is periodic.*

Proof. Suppose every long enough walk in the graph $G(\sigma)$ ends in a singleton The graph $G(\sigma)$ contains at most 2^d vertices (recall that $d = \operatorname{card} \mathcal{A}$), thus every path with length 2^d leads to a singleton. As a consequence, each arithmetic subsequence with common difference hl^{2^d} is constant. Then x is periodic and its period is a divisor of hl^{2^d}.

If x is a periodic sequence, its period is 1 or is a divisor of hl^m for some m (Lemma 6 and Theorem 14). Then each subsequence with period hl^m is constant and every walk of length greater or equal than m ends in a singleton. ∎

From $G(\sigma)$ we define a forest $F(\sigma)$, that is a finite union of (infinite) trees $T_1, \ldots, T_h$. Let $i \in \{0, 1, \ldots, h-1\}$. The root of the tree T_i is $(\mathcal{A}_i, 0)$. The vertices of T_i are divided into *floors* (0th floor, 1st floor, ...). The 0th floor is $\{(\mathcal{A}_i, 0)\}$. The nth floor consists of a finite collection of elements $(\mathcal{B}, n)$ where $(\mathcal{B}', \mathcal{B})$ is an edge of $G(\sigma)$ with $(\mathcal{B}', n-1)$ belonging to the $(n-1)$th floor, and, edges are the pairs $((\mathcal{B}', n-1), (\mathcal{B}, n))$. Roughly speaking in $F(\sigma)$ each vertex has the same successors as in graph $G(\sigma)$. Admissible walks in $F(\sigma)$ start from some $\mathcal{A}_i$. We denote the number of vertices $(\mathcal{C}, m)$, where $\mathcal{C}$ is a singleton, by s_m .

Let us make an observation we will use in the proof of the next proposition. Each such vertex $(\mathcal{C}, m)$ corresponds to exactly one constant arithmetic subsequences of x with common difference hl^m (Proposition 29) and produces l distinct constant arithmetic subsequences of x with common difference hl^{m+1}. As a consequence, one has $s_{m+1} > ls_m$ if, and only if, the sequence x has an essential period, for some letter, greater than hl^m and dividing hl^{m+1}. In the converse case, we have $s_{m+1} = ls_m$.

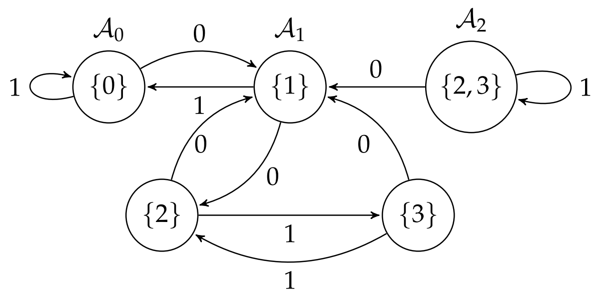
Proposition 35. *There exists in the graph $G(\sigma)$ a cycle joining vertices of cardinality greater or equal to 2, with one having a singleton in his successors, if, and only if, the essential periods for letters of x are unbounded.*

Proof. Suppose there exists in $G(\sigma)$ a cycle joining vertices of cardinality ≥ 2 with one having a singleton in his successors. Let $\mathcal{B}$ be a vertex, i.e. an alphabet, of this cycle. We will prove that, infinitely often, $s_{m+1} > ls_m$. In fact only one such m would be sufficient from the observation made above.

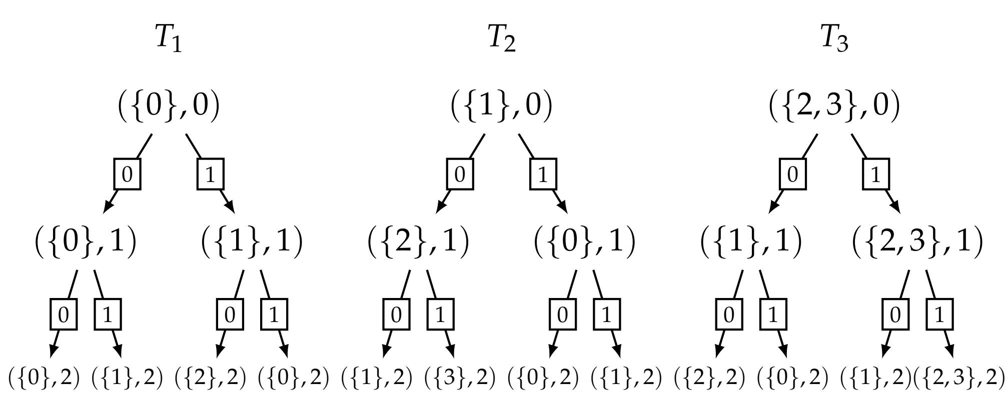
Let N be the length of a shortest path joining $\mathcal{B}$ to a singleton $\{a\}$. As $\mathcal{B}$ belongs to a cycle, it will appear infinitely often in the forest $F(\sigma)$. We denote $(b_k)_{k\in\mathbb{N}}$ the levels (in increasing order) containing $\mathcal{B}$. Then, each level $b_k + N$ contains the singleton $\{a\}$ as a direct successor of an alphabet with cardinality ≥ 2. The number of singletons at level $b_k + N$ is $s_{b_k+N} > s_{b_k+N-1}l$ (we have l successors of the singletons of level $b_k + N - 1$ and at least the singleton $\{a\}$). Therefore, for each $k \in \mathbb{N}$, x has a constant arithmetic subsequence with common difference greater than hl^{b_k+N-1} and dividing hl^{b_k+N}.

Suppose that the set of essential periods for letters is unbounded. For each essential period p_k (numbered in increasing order), let n_k be the smallest integer such that p_k divides hl^{n_k}. Of course, the sequence $(n_k)_k$ goes to infinity with $(p_k)_k$. Thus, this essential period corresponds to a singleton in the level n_k that is the direct descendant of some $(\mathcal{B}_k, k)$ with $\mathcal{B}_k$ having a cardinality greater or equal to 2. As the number of distinct alphabets in the forest is finite, there exists some k such that $\mathcal{B}_k$ appears in this forest an infinite number of times. Thus, it belongs to a cycle of $G(\sigma)$ or is the descendant of such a cycle, whose elements have cardinality greater or equal than the cardinality of $\mathcal{B}_k$. This ends the proof. ∎

Example 4. Let us consider the substitution: $0 \mapsto 01, 1 \mapsto 20, 2 \mapsto 13, 3 \mapsto 12$. It has length 2 and height 3, which lead to the alphabets: $\mathcal{A}_0 = \{0\}$, $\mathcal{A}_1 = \{1\}$ and $\mathcal{A}_2 = \{2, 3\}$. We obtain the following graph.



We are in Case (3) of Proposition 32, thus due to Proposition 35, the essential periods are unbounded. This can also be seen on the following forest.



We easily see that a new constant arithmetic progression appears at each level as a successor of vertex $(\{2,3\},2)$. The length of σ is a prime number, thus the set of essential periods is exactly $\{3 \times 2^n : n \in \mathbb{N}\}$.

6.6 Comments for words occurring in arithmetic progressions

In this work we concentrated on letters occurring periodically in substitutive sequences. The same questions could be asked for words. Using the substitutions on the words of length n (see [Que10]) it is clear that the main results we obtained, i.e. Theorem 19 and Propositions 29 and 30, can be applied to words. This is left as an exercise.

7 Questions

We leave open our three questions for the substitutive sequences (that are not uniformly recurrent). For example, consider the subshift (X,S) generated by the primitive substitution $0 \mapsto 0120$, $1 \mapsto 121$, $2 \mapsto 200$. Let x be an admissible fixed point. One has $\mathbb{P}(X,S) = \{2^m : m \geq 0\}$. By computer checking we did not find constant arithmetic subsequence for x with a period less than 2^{20} but we do not know whether it exists for greater periods.

Acknowledgements. We would like to thank Anna Frid for helpful discussions that were at the origin of the definition of the graph $G(\sigma)$ above. This work was partially supported by the MathAmSud project Dynamics of Cantor systems: computability, combinatorial and geometric aspects (17-MATH-01).

References

[ARS09] J.-P. Allouche, N. Rampersad, and J. Shallit. Periodicity, repetitions, and orbits of an automatic sequence. *Theoret. Comput. Sci.*, 410:2795–2803, 2009.

[AS03] J.-P. Allouche and J. O. Shallit. *Automatic Sequences, Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[Bès01] A. Bès. A survey of arithmetical definability. *Bull. Belg. Math. Soc. Simon Stevin*, suppl.:1–54, 2001.

[BHMV94] V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire. Logic and p-recognizable sets of integers. *Bull. Belg. Math. Soc.*, 1:191–238, 1994.

[CK97] J. Cassaigne and J. Karhumäki. Toeplitz words, generalized periodicity and periodically iterated morphisms. *European J. Combin.*, 18:497–510, 1997.

[Cob72] A. Cobham. Uniform tag sequences. *Math. Systems Theory*, 6:164–192, 1972.

[DDM00] P. Dartnell, F. Durand, and A. Maass. Orbit equivalence and Kakutani equivalence with Sturmian subshifts. *Studia Math.*, 142:25–45, 2000.

[Dek78] F. M. Dekking. The spectrum of dynamical systems arising from substitutions of constant length. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 41:221–239, 1977/78.

[DHS99] F. Durand, B. Host, and C. Skau. Substitutional dynamical systems, Bratteli diagrams and dimension groups. *Ergod. Th. & Dynam. Sys.*, 19:953–993, 1999.

[Dur96] F. Durand. *Contributions à l'étude des suites et systèmes dynamiques substitutifs*. PhD thesis, Université de la Méditerranée (Aix-Marseille II), 1996.

[Dur98a] F. Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179:89–101, 1998.

[Dur98b] F. Durand. Sur les ensembles d'entiers reconnaissables. *J. Théor. Nombres Bordeaux*, 10:65–84, 1998.

[Dur00] F. Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergod. Th. & Dynam. Sys.*, 20:1061–1078, 2000.

[Dur12] F. Durand. HD0L ω-equivalence and periodicity problems in the primitive case (To the memory of G. Rauzy). *J. Unif. Distrib. Theory*, 7:199–215, 2012.

[Dur13a] F. Durand. Decidability of the HD0L ultimate periodicity problem. *RAIRO Theor. Inform. Appl.*, 47:201–214, 2013.

[Dur13b] F. Durand. Decidability of uniform recurrence of morphic sequences. *Internat. J. Found. Comput. Sci.*, 24:123–146, 2013.

[FMN96] S. Ferenczi, C. Mauduit, and A. Nogueira. Substitution dynamical systems: algebraic characterization of eigenvalues. *Ann. Sci. École Norm. Sup.*, 29:519–533, 1996.

[HL86] T. Harju and M. Linna. On the periodicity of morphisms on free monoids. *RAIRO Theor. Inform. Appl.*, 20:47–54, 1986.

[Hon86] J. Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Theor. Inform. Appl.*, 20:395–403, 1986.

[Hos86] B. Host. Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable. *Ergod. Th. & Dynam. Sys.*, 6:529–540, 1986.

[Kam72] T. Kamae. A topological invariant of substitution minimal sets. *J. Math. Soc. Japan*, 24:285–306, 1972.

[Kur03] P. Kurka. *Topological and symbolic dynamics*, volume 11 of *Cours Spécialisés [Specialized Courses]*. Société Mathématique de France, Paris, 2003.

[Lot02] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.

[MH38] M. Morse and G. A. Hedlund. Symbolic Dynamics. *Amer. J. Math.*, 60:815–866, 1938.

[Pan86] J.-J. Pansiot. Decidability of periodicity for infinite words. *RAIRO Theor. Inform. Appl.*, 20:43–46, 1986.

[Pre29] M. Presburger. Über die volständigkeit eines gewissen systems der arithmetik ganzer zahlen, in welchem die addition als einzige operation hervortritt. *C. R. Premier congrès des Mathématiciens des pays slaves, Varsovie*, pages 92–101, 1929.

[Pre91] M. Presburger. On the completeness of a certain system of arithmetic of whole numbers in which addition occurs as the only operation. *Hist. Philos. Logic, Translated from the German*, 12:225–233, 1991.

[Que10] M. Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2010.

[Rig14] M. Rigo. *Formal languages, automata and numeration systems. 2*. Networks and Telecommunications Series. ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2014.

[RS18] N. Rampersad and J. Shallit. Common factors in automatic and sturmian sequences. *Preprint Arxiv:1802.00325*, 2018.

[Wil84] S. Williams. Toeplitz minimal flows which are not uniquely ergodic. *Z. Wahrsch. Verw. Gebiete*, 67:95–107, 1984.

Laboratoire Amiénois de Mathématiques Fondamentales et Appliquées,
CNRS-UMR 7352, Université de Picardie Jules Verne,
33 rue Saint Leu, 80000 Amiens, France.
email : {fabien.durand,valerie.goyheneche}@u-picardie.fr