Eventually dendric shifts

Francesco Dolce, Dominique Perrin

Eventually dendric shift spaces

Francesco Dolce[*]and Dominique Perrin[†]

April 2, 2020

Abstract

We define a new class of shift spaces which contains a number of classes of interest, like Sturmian shifts used in discrete geometry. We show that this class is closed under two natural transformations. The first one is called conjugacy and is obtained by sliding block coding. The second one is called the complete bifix decoding, and typically includes codings by non overlapping blocks of fixed length.

1   Introduction

Shift spaces are the sets of two-sided infinite words avoiding the words of a given language F denoted X_F. In this way the traditional hierarchy of classes of languages translates into a hierarchy of shift spaces. The shift space X_F is called of finite type when one starts with a finite language F and sofic when one starts with a regular language F.

There is a natural equivalence between shift spaces called conjugacy. Two shift spaces are conjugate if there is a sliding block coding sending bijectively one upon the other (in this case the inverse map has the same form). Many basic questions are still open concerning conjugacy. For example, it is surprisingly not known whether the conjugacy of shifts of finite type is decidable.

The complexity of a shift space X is the function $n \mapsto p(n)$ where $p(n)$ is the number of admissible blocks of length n in X. The complexities of conjugate shifts of linear complexity have the same growth rate (see [19, Corollary 5.1.15]).

In this paper, we are interested in shift spaces of at most linear complexity. This class is important for many reasons and includes the class of Sturmian shifts which are by definition those of complexity $n + 1$, which play a role as binary codings of discrete lines. Several books are devoted to the study of such shifts (see [19] or [21] for example). We define a new class of shifts of at most linear complexity, called eventually dendric, extends the class of dendric shifts introduced in [5] (under the name of *tree sets* given to their language) which

---

[*]FNSPE, Czech Technical University in Prague - Trojanova 13, 120 00 Praha 2, Czech Republic

[†]LIGM, Université Gustave Eiffel, France

themselves extend naturally strict episturmian shifts (also called Arnoux-Rauzy shifts [1]) and interval exchange shifts.

Our first main result is that this class is closed under conjugacy. We also prove that it is closed under a second transformation, namely complete bifix decoding, which is important because it includes coding by non overlapping blocks of fixed length. These two results show the robustness of the class of eventually dendric shifts, giving a strong motivation for its introduction.

Recently, Michael Damron and Jon Fickenscher have studied a class of recurrent shifts satisfying a condition called the regular bispecial condition. They proved that the number of ergodic measures on such shifts is bounded by $(K + 1)/2$, where K is the limiting value of the differences $p(n + 1) - p(n)$ and p is the complexity [11]. We will see that these shifts are precisely the recurrent eventually dendric shifts.

The class of dendric shifts (defined below) is known to be closed under complete bifix decoding (see [7]) but it is not closed under conjugacy. This fact was the initial motivation for introducing eventually dendric shifts, following a suggestion of Fabien Durand.

We now describe the results in some more detail.

A dendric shift X is defined by introducing the extension graph of a word in the language $\mathcal{L}(X)$ of X and by requiring that this graph is a tree for every word in $\mathcal{L}(X)$. It has many interesting properties which involve free groups. In particular, in a dendric shift X on the alphabet A, the group generated by the set of return words to some word in $\mathcal{L}(X)$ is the free group on the alphabet and, in particular, has Card A free generators. This generalizes a property known for Sturmian shifts whose link with automorphisms of the free group was noted by Arnoux and Rauzy.

The class of eventually dendric shifts, introduced in this paper, is defined by the property that the extension graph of every word w in the language of the shift is a tree for every long enough word w.

Our main results are that the class of eventually dendric shifts is closed under

- conjugacy (Theorem 6.1), and

- complete bifix decoding (Theorem 9.2).

The paper is organized as follows. In the first section, we introduce the definition of the extension graph and of an eventually dendric shift. In Section 3, we recall some mostly known properties on the complexity of a shift and of special words. We prove a result which characterizes eventually dendric shifts by the extension properties of special words (Proposition 3.5). In Section 4, we use the classical notion of asymptotic equivalence to give a second characterization of eventually dendric shifts (Theorem 4.6). In Section 5, we introduce the notion of a simple tree and we prove that for an eventually dendric shift, the extension graph of every long enough word is a simple tree (Proposition 5.1), a property which holds trivially for every word in a Sturmian shift. In Section 6 we prove the first of our main results (Theorem 6.1).

2

In the next section (Section 7), we prove additional properties of eventually dendric shifts. We first prove that the cardinality of sets of return words is eventually constant (Theorem 7.3). Next, we prove that eventually dendric shifts are minimal as soon as they are irreducible (Theorem 7.6), a property already known for dendric shifts [13]..

In Section 8 we introduce generalized extension graphs in which extension by words of fixed length replace extension by letters. We prove that one obtains an equivalent definition of eventually dendric shifts using these generalized extension graphs (Theorem 8.4).

Finally, in Section 9, we prove that the class of eventually dendric shifts is closed under complete bifix decoding, a result already known for dendric shifts.

Preliminary versions of the results of this paper have been presented at the conferences CSR 2019 [14] and WORDS 2019 [15].

2   Eventually dendric shifts

Let A be a finite alphabet. We consider the set $A^{\mathbb{Z}}$ of bi-infinite words on A as a topological space for the product topology. The *shift map* $\sigma_A : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is defined by $y = \sigma_A(x)$ if $y_i = x_{i+1}$ for every $i \in \mathbb{Z}$. It is a one-to-one continuous map.

We also consider the topological space $A^{\mathbb{N}}$ of one-sided infinite words. We still denote by σ_A the map from $A^{\mathbb{N}}$ to $A^{\mathbb{N}}$ defined by $\sigma_A(x) = y$ if $y_i = x_{i+1}$ for all $i \in \mathbb{N}$. Note that σ_A is not one-to-one as soon as $\mathrm{Card}(A) \geq 2$.

A *shift space* on the alphabet A is a subset X of the set $A^{\mathbb{Z}}$ which is closed and invariant under the shift, that is such that $\sigma_A(X) = X$ (for more on shift spaces see, for instance, [19]).

We denote by A^* the set of (finite) words on the alphabet A. A subset of A^* is called a *language* on A. A word $w \in A^*$ is a *factor* of a bi-infinite word $x \in A^{\mathbb{Z}}$ if $w = x_i \cdots x_{i+n-1}$ for some $i \in \mathbb{Z}$.

We denote by $\mathcal{L}(X)$ the language of X, which is the set of finite factors of the elements of X. A language $\mathcal{L}$ on the alphabet A is the language of a shift if and only if it is *factorial* (that is contains the factors of its elements) and *extendable* (that is for any $w \in \mathcal{L}$ there are letters $a, b \in A$ such that $awb \in \mathcal{L}$).

For $n \geq 0$ we denote

$$\begin{aligned} \mathcal{L}_n(X) &= \mathcal{L}(X) \cap A^n \\ \mathcal{L}_{\geq n}(X) &= \cup_{m \geq n} \mathcal{L}_m(X). \end{aligned}$$

For $w \in \mathcal{L}(X)$ and $n \geq 1$, we denote

$$
\begin{aligned}
L_n(w, X) &= \{u \in \mathcal{L}_n(X) \mid uw \in \mathcal{L}(X)\} \\
R_n(w, X) &= \{v \in \mathcal{L}_n(X) \mid wv \in \mathcal{L}(X)\} \\
E_n(w, X) &= \{(u, v) \in L_n(w, X) \times R_n(w, X) \mid uwv \in \mathcal{L}(X)\}
\end{aligned}
$$

The *extension graph* of order n of w, denoted $\mathcal{E}_n(w, X)$, is the undirected bipartite graph whose set of vertices is the disjoint union of $L_n(w, X)$ and $R_n(w, X)$ and whose edges are the elements of $E_n(w, X)$.

When the context is clear, we denote $L_n(w), R_n(w), E_n(w)$ and $\mathcal{E}_n(w)$ instead of $L_n(w, X), R_n(w, X), E_n(w, X)$ and $\mathcal{E}_n(w, X)$.

A path in an undirected graph is *reduced* if it does not contain successive equal edges (such a path is also called simple). For any $w \in \mathcal{L}(X)$, since any vertex of $L_n(w)$ is connected to at least one vertex of $R_n(w)$, the bipartite graph $\mathcal{E}_n(w)$ is a tree if and only if there is a unique reduced path between every pair of vertices of $L_n(w)$ (resp. $R_n(w)$).

The shift X is said to be *eventually dendric* with *threshold* $m \geq 0$ if $\mathcal{E}_1(w)$ is a tree for every word $w \in \mathcal{L}_{\geq m}(X)$. It is said to be *dendric* if we can choose $m = 0$. Thus, a shift X is dendric if and only if $\mathcal{E}_1(w)$ is a tree for every word $w \in \mathcal{L}(X)$.

The languages of dendric shifts were introduced in [5] under the name of tree sets. An important example of dendric shifts is formed by *strict episturmian shifts* (also called Arnoux-Rauzy shifts), which are by definition such that $\mathcal{L}(X)$ is closed by reversal and such that for every n there exists a unique $w_n \in \mathcal{L}_n(X)$ such that $\mathrm{Card}(R_1(w_n)) = \mathrm{Card}(A)$ and such that for every $w \in \mathcal{L}_n(X) \setminus \{w_n\}$ one has $\mathrm{Card}(R_1(w)) = 1$ (see [5]).

Example 2.1 Let X be the *Fibonacci shift*, which is generated by the morphism $a \mapsto ab, b \mapsto a$. It is well known that it is a Sturmian shift (see [19]). The graph $\mathcal{E}_1(a)$ is shown in Figure 1 on the left. The graph $\mathcal{E}_3(a)$ is shown on the right.
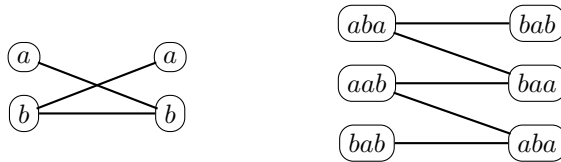
Figure 1: The graphs $\mathcal{E}_1(a)$ and $\mathcal{E}_3(a)$.

The tree sets of characteristic $c \geq 1$ introduced in [4, 13] give an example of eventually dendric shifts. The language $\mathcal{L}(X)$ of a shift space X is said to be a *tree sets of characteristic c* if for any $w \in \mathcal{L}_{\geq 1}(X)$, the extension graph $\mathcal{E}_1(w)$ is a tree and if $\mathcal{E}_1(\varepsilon)$ is a disjoint union of c trees.

Example 2.2 Let X be the shift generated by the morphism $a \mapsto ab, b \mapsto cda, c \mapsto cd, d \mapsto abc$. Its language is a tree set of characteristic 2 ([4, Example 4.2]) and it is actually a specular set. The extension graph $\mathcal{E}_1(\varepsilon)$ is shown in Figure 2.
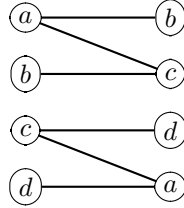

Figure 2: The extension graph $\mathcal{E}_1(\varepsilon)$.

Since the extension graphs of all nonempty words are trees, the shift space is eventually dendric with threshold 1.

Example 2.3 Let S be the *Tribonacci set*, which is the set of factors of the fixed point of the morphism $\psi : a \mapsto ab, b \mapsto ac, c \mapsto a$. S is an Arnoux-Rauzy set and a dendric set (see [5]).

3 Complexity of shift spaces

Let X be a shift space. For a word $w \in \mathcal{L}(X)$ and $k \geq 1$, we denote

$$\ell_k(w) = \mathrm{Card}(L_k(w)), \; r_k(w) = \mathrm{Card}(R_k(w)), \; e_k(w) = \mathrm{Card}(E_k(w)).$$

For any $w \in \mathcal{L}(X)$, we have $1 \leq \ell_k(w), r_k(w) \leq e_k(w)$. The word w is *left-k-special* if $\ell_k(w) > 1$, *right-k-special* if $r_k(w) > 1$ and *k-bispecial* if it is both left-k-special and right-k-special. For $k = 1$, we use ℓ, e, r instead of ℓ_1, e_1, r_1 and we simply say special instead of k-special.

We define the *multiplicity* of w as (see [9])

$$m(w) = e(w) - \ell(w) - r(w) + 1.$$

We say that w is *strong* if $m(w) > 0$, *weak* if $m(w) < 0$ and *neutral* if $m(w) = 0$.

It is clear that

1. if $\mathcal{E}_1(w)$ is acyclic, then w is weak or neutral,

2. if $\mathcal{E}_1(w)$ is connected, then w is strong or neutral,

3. if $\mathcal{E}_1(w)$ is a tree, then w is neutral.

Proposition 3.1 *Let X be a shift space and let $w \in \mathcal{L}(X)$. If w is neutral, then*

$$\ell(w) - 1 = \sum_{b \in R_1(w)} (\ell(wb) - 1) \tag{3.1}$$

5

Proof. Since w is neutral, we have $e(w) = \ell(w) + r(w) - 1$. Thus

$$\sum_{b \in R_1(w)} (\ell(wb) - 1) \quad = \quad e(w) - r(w)$$
$$= \quad \ell(w) - 1.$$

■

Note that the symmetrical of Proposition 3.1 also holds: if $w \in \mathcal{L}(X)$ is neutral then
$$r(w) - 1 = \sum_{b \in L_1(w)} (r(bw) - 1).$$

Set further

$$p_n(X) \quad = \quad \mathrm{Card}(\mathcal{L}_n(X)),$$
$$s_n(X) \quad = \quad p_{n+1}(X) - p_n(X),$$
$$b_n(X) \quad = \quad s_{n+1}(X) - s_n(X).$$

The sequence $p_n(X)$ is called the *complexity* of the shift space X.

The following result is from [8] (see also [5, Lemma 2.12] and [9, Theorem 4.5.4]). We include a proof for convenience.

Proposition 3.2 *We have for all $n \geq 0$,*

$$s_n(X) = \sum_{w \in \mathcal{L}_n(X)} (\ell(w) - 1) = \sum_{w \in \mathcal{L}_n(X)} (r(w) - 1) \tag{3.2}$$

and

$$b_n(X) = \sum_{w \in \mathcal{L}_n(X)} m(w). \tag{3.3}$$

In particular, the number of left-special (resp. right-special) words of length n is bounded by $s_n(X)$.

Proof. We have

$$\sum_{w \in \mathcal{L}_n(X)} (\ell(w) - 1) \quad = \quad \sum_{w \in \mathcal{L}_n(X)} \ell(w) - \mathrm{Card}(\mathcal{L}_n(X))$$
$$= \quad \mathrm{Card}(\mathcal{L}_{n+1}(X)) - \mathrm{Card}(\mathcal{L}_n(X)) = p_{n+1} - p_n$$
$$= \quad s_n(X)$$

with the same result for $\sum_{w \in \mathcal{L}_n(X)} (r(w) - 1)$. N

$$\sum_{w \in \mathcal{L}_n(X)} m(w) \quad = \quad \sum_{w \in \mathcal{L}_n(X)} (e(w) - \ell(w) - r(w) + 1)$$
$$= \quad p_{n+2}(X) - 2p_{n+1}(X) + p_n(X) = s_{n+1}(X) - s_n(X) = b_n(X).$$

■

We will use the following easy consequence of Proposition 3.2.

Proposition 3.3 *Let X be a shift space. If X is eventually dendric, then the sequence $s_n(X)$ is eventually constant.*

Proof. Let $n \geq 1$ be such that the extension graph of every word in $\mathcal{L}_{\geq n}(X)$ is a tree. Then $b_m(X) = 0$ for every $m \geq n$. Thus $s_m(X) = s_{m+1}(X)$ for every $m \geq n$, whence our conclusion. ∎

The previous result implies that eventually dendric sets have eventual linear complexity.

The converse of Proposition 3.3 is not true, as shown by the following example.

Example 3.4 Let X be the *Chacon ternary shift*, which is the substitutive shift space generated by the morphism $\varphi : a \mapsto aabc, b \mapsto bc, c \mapsto abc$. It is well known that the complexity of X is $p_n(X) = 2n + 1$ and thus that $s_n = 2$ for all $n \geq 0$ (see [19, Section 5.5.2]). The extension graphs of abc and bca are shown in Figure 3.

Figure 3: The extension graphs of abc and bca.

Thus $m(abc) = 1$ and $m(bca) = -1$. Let now α be the map on words defined by $\alpha(x) = abc\varphi(x)$. Let us verify that if the extension graph of x is the graph of Figure 3 on the left, the same holds for the extension graph of $y = \alpha(x)$. Indeed, since $axa \in \mathcal{L}(X)$, the word $\varphi(axa) = aabc\varphi(x)aabc = ayaabc$ is also in $\mathcal{L}(X)$ and thus $(a, a) \in \mathcal{E}_1(y)$. Since $cxa \in \mathcal{L}(X)$ and since a letter c is always preceded by a letter b, we have $bcxa \in \mathcal{L}(X)$. Thus $\varphi(bcxa) = bcyaabc \in \mathcal{L}(X)$ and thus $(c, a) \in \mathcal{E}_1(y)$. The proof of the other cases is similar. The same property holds for a word x with the extension graph on the right of Figure 3. This shows that there is an infinity of words whose extension graph is not a tree and thus the Chacon set is not eventually dendric.

Let X be a shift space. We define $LS_n(X)$ (resp. $LS_{\geq n}(X)$) as the set of left-special words of $\mathcal{L}(X)$ of length n (resp. at least n). We denote $LS(X) = \cup_{n \geq 1} LS_n(X)$.

The following result expresses the fact that eventually dendric shift spaces are characterized by an asymptotic property of left-special words which is a local version of the property defining Sturmian shift spaces.

Proposition 3.5 *A shift space X is eventually dendric if and only if there is an integer $n \geq 0$ such that any word w of $LS_{\geq n}(X)$ has exactly one right extension $wb \in LS_{\geq n+1}(X)$ with $b \in A$. Moreover, in this case, one has $\ell(wb) = \ell(w)$.*

Proof. Assume first that X is eventually dendric with threshold m. Then any word w in $LS_{\geq m}(X)$ has at least one right extension in $LS(X)$. Indeed, since $R_1(w)$ has at least two elements and since the graph $\mathcal{E}_1(w)$ is connected, there is at least one element of $R_1(w)$ which is connected by an edge to more than one element of $R_1(w)$.

Next, Equation (3.1) shows that for any $w \in LS_{\geq m}(X)$ which has more than one right extension in $LS(X)$, one has $\ell(wb) < \ell(w)$ for each such extension. Thus the number of words in $LS_{\geq m}(X)$ which are prefix of one another, and which have more than one right extension, is bounded by $\mathrm{Card}(A)$. This proves that there exists an $n \geq m$ such that for any $w \in \mathcal{L}_{\geq n}(X)$ there is exactly one $b \in A$ for which $wb \in LS(X)$. Moreover, one has then $\ell(wb) = \ell(w)$ by Equation (3.1).

Conversely, assume that the condition is satisfied for some integer n. For any word w in $\mathcal{L}_{\geq n}(X)$, the graph $\mathcal{E}_1(w)$ is acyclic since all vertices in $R_1(w)$ except at most one have degree 1. Thus w is weak. Let N be the length of w. Then for every word u of length N and every $b \in R_1(u)$, one has $\ell(ub) = 1$ except for one letter b such that $\ell(ub) = \ell(u)$. Thus, by Proposition 3.2,

$$s_N(X) = \sum_{u \in \mathcal{L}_N(X)} (\ell(u) - 1) = \sum_{v \in \mathcal{L}_{N+1}(X)} (\ell(v) - 1) = s_{N+1}(X).$$

This shows that $b_N = 0$ for every $N \geq n$ and thus, by Proposition 3.2 again, all words in $\mathcal{L}_{\geq n}(X)$ are neutral. Since all graphs $\mathcal{E}_1(w)$ are moreover acyclic, this forces that these graphs are trees and thus that X is eventually dendric with threshold n. ∎

A symmetric result on left extensions of right -pecial words also holds.

In [11], a bispecial word is called *regular* if it has only one left extension wich is right-special and only one right extension which is left-special. A shift space is said to satisfy the *regular bispecial condition* (RBC) if every long enough bispecial word is regular. The following statement is a direct consequence of Proposition 3.5.

Corollary 3.6 *A shift space satisfies the regular bispecial condition if and only if it is eventually dendric.*

We give below an example of a shift space which is shown to be eventually dendric using Proposition 3.5.

Example 3.7 Let X be the *Tribonacci shift*, which is the strict episturmian shift space generated by the substitution $\varphi : a \mapsto ab, b \mapsto ac, c \mapsto a$ and let α be the morphism $\alpha : a \mapsto a, b \mapsto a, c \mapsto c$. Let $\varphi^\omega(a)$ be the right infinite word having all $\varphi^n(a)$ for $n \geq 1$ as prefixes. The left-special words for X are the prefixes of $\varphi^\omega(a)$. Indeed, it is easy to verify that if w is left-special, then $\varphi(w)$ is also left-special.

Note that the set $\mathcal{L}(X) \cap c\{a, b\}^* c$ is

$$\{cabac, cabaabac, cababac\}.$$

8

Since these three words are of distinct lengths, it follows that the restriction of α to the set $\mathcal{L}(X) \cap c\{a,b\}^*c$ is injective.

Next we claim that the left-special words for $\alpha(X)$ containing a letter c are the prefixes of $\alpha(\varphi^\omega(a))$ or $aa\alpha(\varphi^\omega(a))$ containing a letter c. Indeed, if w is a prefix of $\varphi^\omega(a)$, we have $\mathcal{L}_1(w, X) = \{a, b, c\}$ and thus $\mathcal{L}_1(\alpha(w), \alpha(X)) = \{a, c\}$ showing that $\alpha(w)$ is left-special. Next, $\mathcal{L}_3(w, X) = \{aba, bac, cab\}$ and thus $\mathcal{L}_1(aa\alpha(w), \alpha(X)) = \{a, c\}$ showing t hat $aa\alpha(w)$ is left-special. Conversely, assume that u is left-special for $\alpha(X)$ and contains a c. Since u is a prefix of a word ending with c, we may assume that u ends with c. Set $u = a^j cvc$ with $j \geq 0$. By a previous remark, there is a unique word $s \in \mathcal{L}(X)$ such that $csc \in \mathcal{L}(X)$ and $\alpha(csc) = cvc$. Since every word in $\mathcal{L}(X)$ of length at least 7 contains a c, we have $j \leq 6$. It is easy to verify by inspection of the possible left extensions of c in $\mathcal{L}(X)$ that u is left-special only when $j = 3$ or $j = 5$ (see Figure 4).
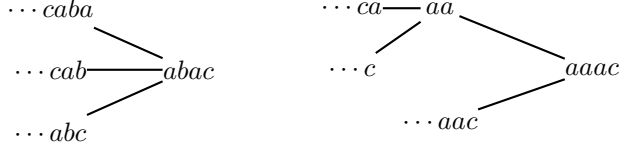

Figure 4: The possible left extensions of c in $\mathcal{L}(X)$ and in $\alpha(\mathcal{L}(X))$.

If $j = 3$, then $u = \alpha(w)$ where $w = abacsc$ is left-special in $\mathcal{L}(X)$ and thus is a prefix of $\varphi^\omega(a)$. If $j = 5$, then u is the common image by α of $ababacsc$ and $baabacsc$. Then $w = abacsc$ is left-special in X and thus is a prefix of $\varphi^\omega(a)$. Since $u = aa\alpha(w)$, the claim is proved.

It follows from the claim that the shift space $\alpha(X)$ satisfies the condition of Proposition 3.5 with $n = 4$. Thus we conclude that $\alpha(X)$ is dendric with threshold at most 4. The threshold is actually 4 since a^3 has multiplicity 1 in $\alpha(X)$.

4 Asymptotic equivalence

The *orbit* of $x \in A^{\mathbb{Z}}$ is the equivalence class of x under the action of the shift transformation. Thus y is in the orbit of x if there is an $n \in \mathbb{Z}$ such that $x = \sigma_A^n(y)$. We say that x is a shift of y if they belong to the same orbit.

For $x \in A^{\mathbb{Z}}$, denote

$$x^- = \cdots x_{-2} x_{-1} \text{ and } x^+ = x_0 x_1 \cdots$$

and $x = x^- \cdot x^+$. When X is a shift space, we denote X^+ the set of right infinite words u such that $u = x^+$ for some $x \in X$.

A right infinite word $u \in A^{\mathbb{N}}$ is a *tail* of the two-sided infinite word $x \in A^{\mathbb{Z}}$ if $u = y^+$ for some shift y of x, that is $u = x_n x_{n+1} \cdots$ for some $n \in \mathbb{Z}$.

Let X be a shift space on the alphabet A. The *right asymptotic equivalence* is the equivalence on X defined as follows. Two elements x, y of X are right asymptotically equivalent if there exists two shifts x', y' of x, y such that $x'^{+} = y'^{+}$. In other words, x, y are right asymptotic equivalent if they have a common tail (see Figure 5 where we assume $x = x'$ and $y = y'$).

$$x^{-}$$
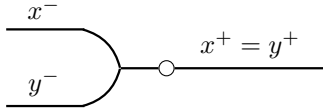$$x^{+} = y^{+}$$
$$y^{-}$$

Figure 5: Two right asymptotic sequences x, y.

The classes of the right asymptotic equivalence not coinciding with only one orbit are called *right asymptotic classes* (they are called in [16] *asymptotic components*).

Example 4.1 The Fibonacci shift X has one only right asymptotic class. It is formed of the shifts of the two elements $x, y \in X$ such that $x^{+} = y^{+} = \varphi^{\omega}(a)$ where $\varphi^{\omega}(a)$ is the Fibonacci word, that is the right infinite word having all $\varphi^{n}(a)$ for $n \geq 1$ as prefixes. Indeed, let $x, y \in X$ be such that $x^{+} = y^{+}$ with $x \neq y$. Then all finite prefixes of $x^{+} = y^{+}$ are left-special and thus are prefixes of $\varphi^{\omega}(a)$ (see, for instance, [19]). Thus $x^{+} = y^{+} = \varphi^{\omega}(a)$.

If C is a right asymptotic class, it is, by the definition of asymptotic equivalence, a union of orbits. The following result is proved in [16, Lemma 3.2] under a weaker hypothesis that we shall not need here. We give a proof for the sake of completeness.

Proposition 4.2 *Let X be a shift space such that the sequence $s_n(X)$ is bounded by k. Then the number of right asymptotic classes is finite and at most equal to k.*

Proof. Let $(x_1, y_1), \ldots, (x_\ell, y_\ell)$ be ℓ pairs of distinct elements of X belonging to asymptotic classes $C_1, \ldots, C_\ell$ such that for all $1 \leq i \leq \ell$, one has $x_i^{+} = y_i^{+}$ and $(x_i)_{-1} \neq (y_i)_{-1}$. For n large enough the prefixes of length n of the x_i^{+} are ℓ distinct left-special words and thus $\ell \leq s_n(X)$ since by Proposition 3.2 the number of left-special words is bounded by $s_n(X)$. This shows that the number of right asymptotic classes is finite and bounded by k. ∎

Let X be a shift space. For an asymptotic class C of X, we denote $\omega(C) = \mathrm{Card}(o(C)) - 1$ where $o(C)$ is the set of orbits contained in C. For a right infinite word $u \in X^{+}$, let

$$\ell_C(u) = \mathrm{Card}\{a \in A \mid x^{+} = au \text{ for some } x \in C\}.$$

We denote by $LS_{\omega}(C)$ the set of right infinite words u such that $\ell_C(u) \geq 2$.

The following statement can be seen as an infinite counterpart of Proposition 3.2.

Proposition 4.3 *Let X be a shift space and let C be a right asymptotic class. Then*

$$\omega(C) = \sum_{u \in LS_\omega(C)} (\ell_C(u) - 1) \tag{4.1}$$

where both sides are simultaneously finite.

In order to prove Proposition 4.3, we use the notion of a *cluster of trees* that we now define.

A *cluster of trees* is a directed graph which is the union of a (non-trivial) cycle Γ and a family of disjoint trees (oriented from child to father) T_v with root v indexed by the vertices v on Γ (see Figure 6). It is easy to verify that a finite connected graph is a cluster of trees if and only if every vertex has outdegree 1 and there is a unique strongly connected component.
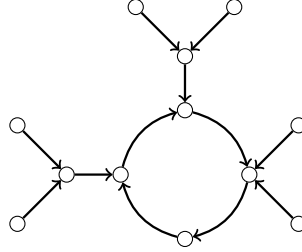
Figure 6: A cluster of trees.

In a cluster of trees, the number of leaves (that is, the leaves of the trees T_v not reduced to their root) is equal to $\sum_u (d^-(u) - 1)$, where d^- stands for the indegree function and the sum runs over the set of internal nodes. Indeed, this is true for one cycle alone since there are no leaves and every internal node has indegree 1. The formula remains valid when suppressing a leaf in one of the trees not reduced to its root.

Proof of Proposition 4.3. We first suppose that C does not contain periodic points which implies that $LS_\omega(C)$ does not contain periodic points either.

It is easy to verify that if $u, v \in LS_\omega(C)$, there exist $n, m \geq 0$ such that $\sigma^n(u) = \sigma^m(v)$.

We build a graph $T(C)$ as follows. The set of vertices of $T(C)$ is $o(C) \cup LS_\omega(C)$. There will be for each vertex u of $T(C)$ at most one edge going out of u, called its father.

Let first $x \in C$ and let u be the orbit of x. There is, up to a shift of x, at least one $y \in C$ with $x \neq y$ such that $y^+ = x^+$. Let $n \geq 0$ be the minimal integer such that $x_{-n} \neq y_{-n}$. Then $v = \sigma^{-n+1}(x)^+$ is in $LS_\omega(C)$ and depends only on the orbit u of x. We choose the vertex v as the father of u.

Next, for every $u \in LS_\omega(C)$, we consider the minimal integer, if it exists, such that $v = \sigma^n(u)$ is in $LS_\omega(C)$. Then we choose v as the father of u.

Assume now that $\omega(C)$ is finite. Then $LS_\omega(C)$ is also finite and $T(C)$ is a finite tree. Indeed, if $u \in LS_\omega(C)$, there is at least one $x \in C$ such that

$x^+ = u$ and thus such that u is an ancestor of the orbit of x. By the claim made above, any two elements of $LS_\omega(C)$ have a common ancester. Since C does not contain periodic points, two vertices cannot be ancestors of one another. Thus there is a unique element of $LS_\omega(C)$ which has no father, namely the unique $u \in SL(C)$ with a maximal number of elements of $o(C)$ as descendants. Since it is an ancestor of all vertices of $T(C)$, this shows that $T(C)$ is a finite tree.

Formula (4.1) now follows from the fact that in any finite tree with n leaves and and a set V of internal vertices, one has $n - 1 = \sum_{v \in V}(d^-(v) - 1)$.

Assume next that the right hand side of Equation (4.1) is finite. Then the set $LS_\omega(C)$ is finite and thus $T(C)$ is again a tree with a finite number of internal nodes. Since the degree of each node is finite, it implies that it has also a finite number of leaves. Thus $\omega(C)$ is finite and Equation (4.1) also holds.

Finally, assume that C contains a periodic point. It follows from the definition of an asymptotic class that there is exactly one such periodic orbit, since two periodic points having a common tail are in the same orbit.

The proof follows the same lines as in the first case, but this time $T(C)$ will be a cluster of trees instead of a tree.

The set of leaves of $T(C)$ is, as above, the set $o(C)$ of non periodic orbits and the the other vertices are the elements of $LS_\omega(C)$. The unique father of a vertex is defined in the same way as above. The fact that there is a unique strongly connected component is a consequence of the fact that there is a unique periodic orbit in C. Finally, Formula (4.1) holds with since the number of leaves is equal to $\sum(d^-(u) - 1) - 1$, where the sum runs over the set of internal nodes and the -1 corresponds to the unique periodic orbit. ■

Example 4.4 Consider again the image $\alpha(X)$ of the Tribonacci shift by the morphism $\alpha : a \mapsto a, b \mapsto a, c \mapsto c$ (Example 3.7).
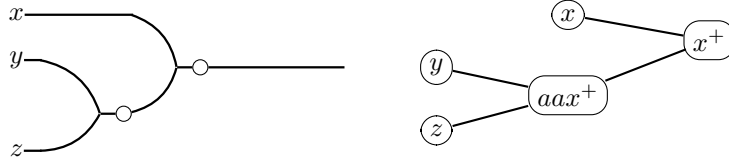
Figure 7: The asymptotic class C and the tree $T(C)$.

There is one asymptotic class C made of three orbits represented in Figure 7 on the left. The class is formed of the orbits of x, y, z where $x^+ = \alpha(\varphi^\omega(a))$ and $y^+ = z^+ = aax^+$. The tree $T(C)$ is shown on the right.

In the next example we use the notation u^ω for the right infinite word $uuu \cdots$ and symmetrically $^\omega u$ for the left-infinite word $\cdots uuu$.

Example 4.5 Let X be the shift space which is the closure under the shift of the set $\{^\omega c.(ab)^\omega \ \cup \ ^\omega d.(ab)^\omega \ \cup \ ^\omega(ab) \cdot (ab)^\omega\}$. The shift has just one right asymptotic class C, the one associated to the tail $(ab)^\omega$, containing three orbits.

Since the tail $(ab)^\omega$ can be prolonged on the left by either c, d or b, we have that Formula 4.1 is verified. The cluster of trees $T(C)$ is represented in Figure 8 where we denote by xy the orbit of $x \cdot y$.
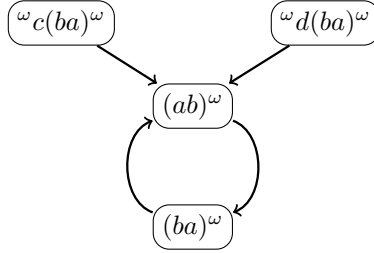

Figure 8: The cluster $T(C)$.

Let us now deduce from Proposition 3.5 a characterization of eventually dendric shift spaces in terms of asymptotic classes. For a shift space X, denote

$$\omega(X) = \sum \omega(C)$$

where the sum is over the asymptotic classes C of X.

Theorem 4.6 *A shift space X is eventually dendric if and only if:*

1. *The sequence $s_n(X)$ is eventually constant, and*

2. *We have $\lim s_n(X) = \omega(X)$.*

Proof. Assume first that X is eventually dendric. Then b 1 holds by Proposition 3.3. To prove assertion 2, consider an integer n large enough so that the condition of Proposition 3.5 holds (it implies that $s_m(X)$ is constant for $m \geq n$). Let us consider an asymptotic class C.

Let π be the map assigning to $u \in A^\mathbb{N}$ its prefix of length n. Then π maps $LS_\omega(C)$ into $LS_n(X)$. The map π is injective since otherwise some word in $LS_{\geq}n(X)$ would have more than one extension on the right, contrary to Proposition 3.5. Next the sets $\pi(LS_\omega(C))$ for all asymptotic classes C form a partition of $LS_n(X)$.

Thus, by Equation (3.2),

$$s_n(X) \quad = \quad \sum_{w \in LS_n(X)} (\ell(w) - 1) = \sum_C \sum_{u \in LS_\omega(C)} (\ell_C(u) - 1)$$

$$= \quad \sum_C \omega(C)$$

where the last equality follows from Equation (4.1).

Conversely, if the two conditions are satisfied, let n be large enough so that $s_m(X) = s_n(X)$ for all $m \geq n$. We may also assume that n is large enough so

13

that the prefixes of length n of the words of $LS_\omega(C)$ for every asymptotic class C are distinct. Then, every word w of $LS_n(X)$ has exactly one right extension wb in $LS_{n+1}(X)$. It is moreover such that $\ell(w) = \ell(wb)$ and thus X is eventually dendric by Proposition 3.5. ∎

For example, the Tribonacci shift is such that $s_n(X) = 2$ for every $n \geq 0$ and $\omega(X) = 2$ since there is only one asymptotic class made of 3 orbits. Note that the Chacon shift X (Example 3.4) satisfies condition 1 of Theorem 4.6 but not condition 2. Indeed, one can verify that $s_n(X) = 2$ for all $n \geq 0$ but $\omega(X) = 1$.

5 Simple trees

The *diameter* of a tree is the maximal length of simple paths. We call a tree *simple* if its diameter is at most 3. Note that if the simple tree is the extension graph $\mathcal{E}_n(w)$ in some shift space X of a bispecial word w, then the diameter of $\mathcal{E}_n(w)$ is equal to 3 and this happens if and only if any two vertices of $\mathcal{E}_n(w)$ on the same side (that is, both in $L_n(x)$ or both in $R_n(w)$) are connected to a common vertex on the opposite side.

Note that Corollary 3.6 expresses the fact that a shift space X is eventually dendric if and only if for any long enough word $w \in \mathcal{L}(X)$, the extension graph $\mathcal{E}_1(w)$ is a simple tree. However, $\mathcal{E}_1(w)$ can be a simple tree while $\mathcal{E}_k(w)$, for a $k > 1$ is not. For example, if X is the Fibonacci shift, then $\mathcal{E}_1(a)$ is simple while $\mathcal{E}_3(a)$ is not (see Example 2.1).

We prove the following additional property of the graphs $\mathcal{E}_k(w)$.

Proposition 5.1 *Let X be an eventually dendric shift space. For any $k \geq 1$ there exists an $n \geq 1$ such that $\mathcal{E}_k(w)$ is a simple tree for every $w \in \mathcal{L}_{\geq n}(X)$.*

We first prove the following lemma.

Lemma 5.2 *Let X be an eventually dendric shift space. For every $k \geq 1$ there is an $n \geq 1$ such that if $p, w \in \mathcal{L}(X)$ with $|p| \leq k$ and $|w| \geq n$ are such that $pw, w \in LS(X)$, then pw, w have a unique right extension in $LS(X)$ by some letter $b \in A$ which is moreover such that $\ell(pwb) = \ell(pw)$ and $\ell(wb) = \ell(w)$.*

Proof. Consider two asymptotic classes C, D and let $u \in LS_\omega(C)$, $v \in LS_\omega(D)$. If C, D are distinct, we cannot have $pu = v$ for some word p. Thus there is an integer n such that if w is the prefix of length n of u, then pw is not a prefix of v. Since there is a finite number of words p of length at most k, a finite number of asymptotic classes (by Proposition 4.2) and since for each such class the set $LS_\omega(C)$ is finite, we infer that for every k there exists an n such that for every pair of asymptotic classes C, D and any $u \in S(C), v \in LS(D)$, if w is a prefix of u and pw a prefix of v, with $|p| \leq k$ and $|w| = n$, then $C = D$.

Next, assume that w is a prefix of u and pw a prefix of v with $u, v \in S(C)$ for some asymptotic class C. If $v \neq pu$, then there is a right extension w' of w

such that pw' is not a prefix of v. By contraposition, if n is large enough, we have $v = pu$.

We thus choose n large enough so that:

1. All elements of $LS_\omega(C)$ for all asymptotic components C have distinct prefixes of length n;

2. For every pair of asymptotic classes C, D and any $u \in LS_\omega(C), v \in LS(D)$, if w is prefix of u and pw is prefix of v with $|p| \leq k$ and $|w| = n$ then $C = D$ and $pu = v$.

We moreover assume that n is large enough so that the condition of Proposition 3.5 holds.

Consider p, w with $|p| = k$ and $|w| = n$ such that pw, w are left-special. By condition 1, there are asymptotic components C, D and elements $u \in LS_\omega(C)$ and $v \in LS(D)$ such that w is a prefix of u and pw a prefix of v. Because of condition 2, we must have $\sigma^k(v) = u$ (and in particular $C = D$). Thus there is a unique letter $b \in A$ such that $wb, pwb \in LS(X)$ which is moreover such that $\ell(wb) = \ell(w)$ and $\ell(pwb) = \ell(pw)$ by Proposition 3.5. ∎

Proof of Proposition 5.1. We choose n such that Proposition 3.5 and Lemma 5.2 hold.

We prove by induction on ℓ with $1 \leq \ell \leq k$ that $\mathcal{E}_\ell(w)$ is a simple tree and thus that for any $p, q \in L_\ell(w)$ there is an $r \in R_k(w)$ such that $pwr, qwr \in \mathcal{L}(X)$.

The property is true for $\ell = 1$. Indeed, set $p = a$ and $q = b$. Apply iteratively Proposition 3.5 to obtain letters $c_1, \ldots, c_k$ such that $\ell(wc_1 \cdots c_i) = \ell(wc_1 \cdots c_i c_{i+1})$ and set $r = c_1 \cdots c_k$. Then $awr, bwr \in \mathcal{L}(X)$.

Assume next that the property is true for $\ell - 1$ and consider $ap, bq \in L_\ell(w)$ with $a, b \in A$. Replacing if necessary w by some longer word, we may assume that p, q end with different letters and thus that w is left-special. By the induction hypothesis, there is a word $r \in R_k(w)$ such that $pwr, qwr \in \mathcal{L}(X)$. By Lemma 5.2, the first letter of r is the unique letter c such that $\ell(pwc) = \ell(pw)$ and $\ell(qwc) = \ell(qw)$. Thus $apwc, bqwc \in \mathcal{L}(X)$. Applying Lemma 5.2 iteratively in this way, we obtain that $apwr, bqwr \in \mathcal{L}(X)$. ∎

6 Conjugacy

Let A, B be two alphabets, and $X \subset A^{\mathbb{Z}}$ and $Y \subset B^{\mathbb{Z}}$ be two shift spaces. A map $\phi : X \to Y$ is called a *sliding block code* if there exists $m, n \in \mathbb{N}$ and a map $f : \mathcal{L}_{m+n+1}(X) \to B$ such that $\phi(x)_i = f(x_{i-m} \cdots x_{i+n})$ for all $i \in \mathbb{Z}$ and $x = (x_i) \in X$. It can be shown that a map $\phi : X \to Y$ is a sliding block code if and only if it is continuous and commutes with the shift, that is $\phi \circ \sigma_A = \sigma_B \circ \phi$ (see, for instance, [20]).

Two shift spaces X, Y are said to be *conjugate* when there is a bijective sliding block code $\phi : X \to Y$. The following result shows that the property of being eventually dendric is a dynamical property, in the sense that it only depends on the conjugacy class of a shift.

Theorem 6.1 *The class of eventually dendric shift spaces is closed under conjugacy.*

We first treat the following particular case of conjugacy. Let X be a shift space on the alphabet A and let $k \geq 1$. Let $f : \mathcal{L}_k(X) \to A_k$ be a bijection from the set $\mathcal{L}_k(X)$ of blocks of length k of X onto an alphabet A_k. The map $\gamma_k : X \to A_k^{\mathbb{Z}}$ defined for $x \in X$ by $y = \gamma_k(x)$ if for every $n \in \mathbb{Z}$

$$y_n = f(x_n \cdots x_{n+k-1})$$

is the k-th *higher block code* on X. The shift space $X^{(k)} = \gamma_k(X)$ is called the k-th *higher block shift space* of X. It is well known that the k-th higher block code is a conjugacy.

We extend the bijection $f : \mathcal{L}_k(X) \to A_k$ to a map still denoted f from $\mathcal{L}_{\geq k}(X)$ to $\mathcal{L}_{\geq 1}(X^{(k)})$ by $f(a_1 a_2 \cdots a_n) = f(a_1 \cdots a_k) \cdots f(a_{n-k+1} \cdots a_n)$. Note that all nonempty elements of $\mathcal{L}(X^{(k)})$ are images by f of elements of $\mathcal{L}(X)$, that is, $\mathcal{L}(X^{(k)}) = \{f(w) \mid w \in \mathcal{L}_{\geq k}(X)\} \cup \{\varepsilon\}$.

Example 6.2 Let X be the Fibonacci shift. We show that the 2-block extension $X^{(2)}$ of X is eventually dendric with threshold 1. Set $A_2 = \{u, v, w\}$ with $f : aa \mapsto u, ab \mapsto v, ba \mapsto w$. Since X is dendric, the graph $\mathcal{E}_1(w)$ is a tree for every word $w \in \mathcal{L}(X^{(2)})$ of length at least 1 (but not for $w = \varepsilon$). Thus $X^{(2)}$ is eventually dendric. It is actually a tree shift space of characteristic 2 since the graph $\mathcal{E}_1(\varepsilon)$ is the union of two trees (see Figure 9).
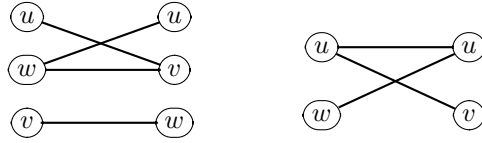

Figure 9: The extension graphs $\mathcal{E}_1(\varepsilon)$ and $\mathcal{E}_1(vw)$.

Lemma 6.3 *For every $k \geq 1$, the k-th higher block shift space $X^{(k)}$ is eventually dendric if and only if X is eventually dendric.*

Proof. We define for every $w \in \mathcal{L}_{\geq k}(X)$ a map from $\mathcal{E}_1(w)$ to $\mathcal{E}_1(f(w))$ as follows.

To every $a \in L_1(w)$, we associate the first letter $\lambda(a)$ of $f(aw)$ and to every $b \in R_1(w)$, we associate the last letter $\rho(b)$ of $f(wb)$. Then, since $f(awb) = \lambda(a)f(w)\rho(b)$, the pair (a, b) is in $E_1(w)$ if and only if $(\lambda(a), \rho(b))$ is in $E_1(f(w))$. Thus, the maps λ, ρ define an isomorphism from $\mathcal{E}_1(w)$ onto $\mathcal{E}_1(f(w))$.

Thus we conclude that X is eventually dendric with threshold m if and only if $X^{(k)}$ is eventually dendric with threshold M with $0 \leq M \leq \max(1, m-k+1)$. ∎

Example 6.4 Let X be the Fibonacci shift. For all $k \geq 2$, $X^{(k)}$ is an eventually dendric shift space with threshold 1.

Example 6.5 Let X be the shift space associated to the two-sided infinite word $\cdots abab \cdot abab \cdots$. X is an eventually dendric shift space with threshold 1 (the empty word has 2 connected components). For every $k \geq 1$, the shift space $X^{(k)}$ is eventually dendric with threshold 1.

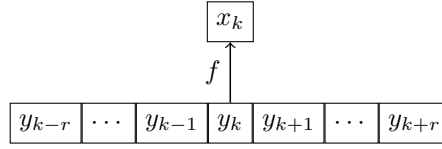A morphism $\alpha : A^* \to B^*$ is called *alphabetic* if $\alpha(A) \subset B$.

Figure 10: The sliding block code.

Lemma 6.6 *Let X be an eventually dendric shift space on the alphabet A and let $\alpha : A^* \to B^*$ be an alphabetic morphism which induces a conjugacy from X onto a shift space Y. Then Y is eventually dendric.*

Proof. Since α is invertible, there exists an integer $r \geq 0$ and a map $f : B^{2r+1} \to A$ such that for $x = (x_k)_{k \in \mathbb{Z}}$ and $y = (y_k)_{k \in \mathbb{Z}}$, one has $y = \alpha(x)$ if and only if for every $k \in \mathbb{Z}$, one has (see Figure 10)

$$x_k = f(y_{k-r} \cdots y_{k-1} y_k y_{k+1} \cdots y_{k+r}).$$

We extend the definition of f to a map from $\mathcal{L}_{\geq 2r+1}(X)$ to A as follows. For $w = b_{1-r} \cdots b_{n+r} \in \mathcal{L}_{\geq 2r+1}(Y)$, set $f(w) = a_1 \cdots a_n$ where $a_i = f(b_{i-r} \cdots b_i \cdots b_{i+r})$. Note that if $u = f(w)$ and $w = svt$ with $s, t \in \mathcal{L}_r(Y)$, then $v = \alpha(u)$ (see Figure 11).

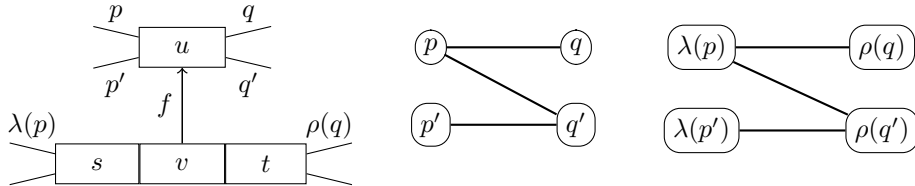Figure 11: The map f (on the left), the graph $\mathcal{E}_k(u)$ (on the center) and the graph $\mathcal{E}_1(w)$ (on the right).

Let n be the integer given by Proposition 5.1 for $k = r + 1$. We claim that every graph $\mathcal{E}_1(w)$ for $|w| \geq n + 2r$ is a tree. Let indeed $s, t \in \mathcal{L}_r(Y)$ and $v \in \mathcal{L}_{\geq n}(Y)$ be such that $w = svt$. Let $u = f(svt)$ (see Figure 11).

Let $E'_k(u) = \{(p,q) \in L_k(u) \times R_k(u) \mid \alpha(puq) \in BwB\}$ and let $L'_k(u)$ (resp. $R'_k(u)$) be the set of $p \in L_k(u)$ (resp. $q \in R_k(u)$) which are connected to $L_k(u)$ (resp. $R_k(u)$) by an edge in $E'_k(u)$. Let $\mathcal{E}'_k(u)$ be the subgraph of $\mathcal{E}_k(u)$ obtained by restriction to the set of vertices which is the disjoint union of $L'_k(u)$ and $R'_k(u)$ (and that has, thus, $E'_k(u)$ as set of edges).

Claim 1. The graph $\mathcal{E}'_k(u)$ is a simple tree. Indeed, by Proposition 5.1, the graph $\mathcal{E}_k(u)$ is a simple tree. We may assume that u is k-bispecial (otherwise, the property is obviously true). Let (p,q) be an edge of $\mathcal{E}'_k(u)$. Then (p,q) is an edge of $\mathcal{E}_k(u)$ and since the latter is a simple tree either p is the unique vertex in $L_k(u)$ such that pu is right-special or q is the unique vertex in $R_k(u)$ such that uq is left-special (both cases can occur simultaneously). Assume the first case, the other being proved in a symmetric way. If (p',q') is another edge of $\mathcal{E}'(u)$, then (p,q') is an edge of $\mathcal{E}_k(u)$. Since $\alpha(p) \in Bs$ and $\alpha(q) \in tB$, we have actually $(p,q') \in E'_k(u)$. Thus $\mathcal{E}'_k(u)$ contains the two vertices of $\mathcal{E}_k(u)$ connected to more than one other vertex and this implies that $\mathcal{E}'_k(u)$ is a simple tree.

For $p \in L'_k(u)$, let $\lambda(p)$ be the first letter of $\alpha(p)$ and for $q \in R'_k(u)$, let $\rho(q)$ be the last letter of $\alpha(q)$.

Claim 2. The graph $\mathcal{E}_1(w)$ is the image by the maps λ, ρ of the graph $\mathcal{E}'_k(u)$. Indeed, one has $(a,b) \in E_1(w)$ if and only if there exist $(p,q) \in E'_k(u)$ such that $\lambda(p) = a$ and $\rho(q) = b$.

Let us consider a graph homomorphism ϕ preserving bipartiteness and such such that left vertices are sent to left vertices and right vertices to right ones: Then, it is easy to verify that the image of a simple tree by ϕ is again a simple tree. Thus $\mathcal{E}_1(w)$ is a simple tree, which concludes the proof. ∎

We are now ready to prove the theorem.

Proof of Theorem 6.1. Every conjugacy is a composition of a higher block code and an alphabetic morphism (see [20, Proposition 1.5.12]). Thus Theorem 6.1 is a direct consequence of Lemmas 6.3 and 6.6. ∎

Example 6.7 We have seen in Example 3.7 that the image $Y = \alpha(X)$ of the Tribonacci shift X by the morphism $\alpha : a \mapsto a, b \mapsto a, c \mapsto c$ is eventually dendric. This is actually a consequence of Theorem 6.1. Indeed α is a injective by Example 3.7, and thus a conjugacy from X to $\alpha(x) = Y$. The images of a Sturmian shift space by a non trivial alphabetic morphism have been investigated in [22].

7   Minimal eventually dendric shifts

A shift space X is *irreducible* if for any $u, v \in \mathcal{L}(X)$ there is a word w such that $uwv \in \mathcal{L}(X)$ (equivalently $\mathcal{L}(X)$ is called *recurrent*).

A nonempty shift space is *minimal* if it does not contain properly another nonempty shift space. As well known, X is minimal if and only if it is *uniformly recurrent*, that is for any $w \in \mathcal{L}(X)$ there exists an $n \geq 0$ such that w is a factor

of any word in $\mathcal{L}_n(X)$. If X is minimal and infinite, then there exists for every $w \in \mathcal{L}(X)$ an integer $n \geq 1$ such that $w^n \notin \mathcal{L}(X)$. Indeed, otherwise, $\mathcal{L}(X)$ contains the periodic word with period w and thus X is equal to the finite shift space formed by the shifts of $\cdots ww \cdot ww \cdots$.

A minimal shift space is irreducible but the converse is false, since for example the full shift $A^{\mathbb{Z}}$ is irreducible but not minimal as soon as A has at least two elements.

Let X be a shift space. The set of *complete return words* to a word $w \in \mathcal{L}(X)$ is the set $\mathcal{CR}_X(w)$ of words having exactly two factors equal to w, one as a proper prefix and the other one as a proper suffix. It is clear that X is minimal if and only if it is irreducible and if for every word w the set of complete return words to w is finite.

If wu is a complete return word to w, then u is called a (right) *return word* to w. We denote by $\mathcal{R}_X(w)$ the set of return words to w. Clearly $\mathrm{Card}(\mathcal{CR}_X(w)) = \mathrm{Card}(\mathcal{R}_X(w))$.

Example 7.1 Let X be the Tribonacci shift (see Example 3.7). The image of X under the morphism $\alpha : a, b \to a, c \to c$. Then $\mathcal{R}_X(a) = \{a, ba, ca\}$ and $\mathcal{R}_X(c) = \{abac, ababac, abaabac\}$.

By a result of [2], if X is minimal and neutral (a fortiori if X is dendric) the set $\mathcal{R}(w)$ has $\mathrm{Card}(A)$ elements for every $w \in \mathcal{L}(X)$. This is not true anymore for eventually dendric shift spaces, as shown in the following example.

Example 7.2 Let X be the Tribonacci shift and let $Y = \alpha(X)$ be, as in Example 6.7 the image of X under the morphism $\alpha : a, b \to a, c \to c$. Then, using Example 7.1, we find $\mathcal{R}_Y(a) = \{a, ca\}$ while $\mathcal{R}_Y(c) = \{aaac, aaaaac, aaaaaac\}$.

We will prove that for eventually dendric sets, a weaker property is true. It implies that the cardinality of sets of return words is eventually constant.

For $w \in \mathcal{L}(X)$, set $\rho_X(w) = r_1(w) - 1$ and for a set $W \subset \mathcal{L}$, set $\rho_X(W) = \sum_{w \in W} \rho_X(w)$ (if W is infinite, $\rho_X(W)$ is the supremum of the values of $\rho_X(U)$ over the finite subsets U of W). By the symmetric of Proposition 3.1, for every neutral word $w \in \mathcal{L}(X)$, we have

$$\rho_X(w) = \sum_{a \in L_1(w)} \rho_X(aw). \tag{7.1}$$

Theorem 7.3 *Let X be an irreducible shift space which is eventually dendric with threshold m. For every $w \in \mathcal{L}(X)$, the set $\mathcal{R}_X(w)$ is finite. Moreover, for every $w \in \mathcal{L}_{\geq m}(X)$, we have*

$$\mathrm{Card}(\mathcal{R}_X(w)) = 1 + \rho_X(\mathcal{L}_m(X)). \tag{7.2}$$

Note that for $m = 0$, we obtain $\mathrm{Card}(\mathcal{R}_X(w)) = \mathrm{Card}(A)$ since $\rho_X(\varepsilon) = \mathrm{Card}(A) - 1$.

To prove Theorem 7.3 we need to introduce some definitions.

Let X be a shift space on an alphabet A. A *prefix code* (resp. a *suffix code*) is a set of non-empty words such that none of them is a prefix (resp. a suffix) of another one. A prefix code (resp. a suffix code) $U \subset \mathcal{L}(X)$ is called X-*maximal* if it is not properly contained in a prefix code (resp. suffix code) $V \subset \mathcal{L}(X)$ (see, for instance, [3]).

Let $U \subset \mathcal{L}(X)$ be a finite X-maximal suffix code, let v be a word such that $Av \cap \mathcal{L}(X) \subset U$. We denote

$$\Delta_u(U) = (U \setminus Au) \cup \{u\}. \tag{7.3}$$

Conversely, given a suffix code $V \subset \mathcal{L}(X)$ and a word $v \in V$, we denote

$$\Delta'_v(V) = (V \setminus \{v\}) \cup (Av \cap \mathcal{L}(X)). \tag{7.4}$$

For a finite set U we denote $\ell(U)$ as the sum of the lengths of the words of U. The following elementary lemma will be used repeatedly.

Lemma 7.4 *Let U, V, u, v be as above.*

1. *The set $\Delta_u(U)$ is either reduced to the empty word or it is an X-maximal suffix code. Moreover, $\ell(\Delta_u(U)) < \ell(U)$.*

2. *The set $\Delta'_v(V)$ is a suffix code.*

3. *One has $U = \Delta'_u(\Delta_u(U))$ and $V = \Delta_v(\Delta'_v(V))$.*
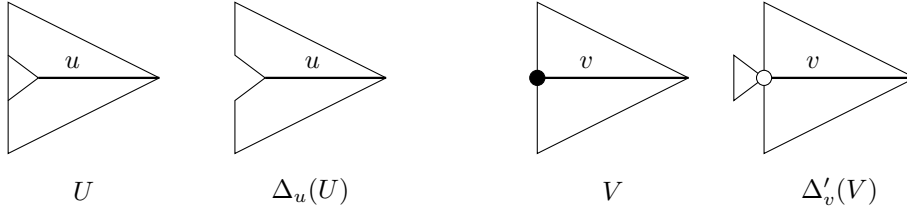
Figure 12: The sets U and $\Delta_u(u)$ (on the left) and V and $\Delta'_v(V)$ (on the right).

Proposition 7.5 *Let X be a shift space which is eventually dendric with threshold m. Then $\rho_X(U)$ is finite for every suffix code $U \subset \mathcal{L}(X)$. If U is a finite X-maximal suffix code with $U \subset \mathcal{L}_{\geq m}(X)$, then*

$$\rho_X(U) = \rho_X(\mathcal{L}_m(X)). \tag{7.5}$$

Proof. For any suffix code $U \subset \mathcal{L}(X)$, let U_m be the union

$$U_m = (U \cap \mathcal{L}_{<m}(X)) \cup (\mathcal{L}_m(X) \cap S),$$

where S is the set of words which are suffixes of some words of U. Note that U_m is a finite suffix code. It is equal to $\mathcal{L}_m(X)$ if U is X-maximal and contained in $\mathcal{L}_{\geq m}(X)$.

Assume first that $U \subset \mathcal{L}(X)$ is a finite X-maximal suffix code. We prove by induction on the sum $\ell(U)$ of the lengths of the words of U that

$$\rho_X(U) = \rho_X(U_m). \tag{7.6}$$

If all words of U are of length at most m, then $U = U_m$ and thus Equation (7.6) holds. Otherwise, let $w \in U$ be of maximal length. Set $w = au$ with $a \in A$. Then $Au \cap \mathcal{L}(X) \subset U$. Set $U' = \Delta_u(U)$. Since $U = \Delta'_u(U')$ by Lemma 7.4, we have by Equation (7.4)

$$\rho_X(U) = \rho_X(U') - \rho_X(u) + \sum_{a \in L_1(v)} \rho_X(au).$$

Since u is neutral (because $|u| \geq m$), we have, by Equation (7.1),

$$\rho_X(U) = \rho_X(U').$$

By induction hypothesis, Equation (7.6) holds for U'. Since $U_m = U'_m$, we have $\rho_X(U) = \rho_X(U') = \rho_X(U'_m) = \rho_X(U_m)$ and Equation (7.6) is proved.

If $U \subset \mathcal{L}_{\geq m}(X)$, then $U_m = \mathcal{L}_m(X)$ and thus Equation (7.5) is proved.

If U is infinite, then $\rho_X(U)$ is the supremum of the values of $\rho_X(V)$ on the finite subsets V of U. Any finite suffix code $V \subset \mathcal{L}(X)$ is contained in a finite X-maximal suffix code W and $\rho_X(V) \leq \rho_X(W)$. By Equation (7.6), this implies $\rho_X(V) \leq \rho_X(W_m)$. There is a finite number of possible W_m and thus $\rho_X(V)$ is bounded. We conclude that $\rho_X(U)$ is finite. ∎

Proof of Theorem 7.3. Consider a word $w \in \mathcal{L}(X)$ and let P be the set of proper prefixes of $\mathcal{CR}(w)$. For $p \in P$, denote $\alpha(p) = \mathrm{Card}\{a \in A \mid pa \in P \cup \mathcal{CR}(w)\} - 1$. Then $\mathcal{CR}(w)$ is finite if and only if P is finite. Set $\alpha(P) = \sum_{p \in P} \alpha(p)$. Since $\mathcal{CR}(w)$ is a prefix code, it is the set of leaves of a tree with $f = \mathrm{Card}(\mathcal{CR}(w))$ leaves, i internal nodes and $\alpha(P) + i$ edges. Since $f = e - i + 1$, we have

$$\mathrm{Card}(\mathcal{CR}(w)) = \alpha(P) + 1, \tag{7.7}$$

Let U be the set of words in P which are not proper prefixes of w. We claim that U is an X-maximal suffix code.

Indeed, if $u, vu \in U$, then w is a proper prefix of u and thus is an internal factor of vu, a contradiction unless $v = \varepsilon$. Thus U is suffix.

Consider $r \in \mathcal{L}(X)$. Then, since $\mathcal{L}(X)$ is recurrent, there is some $s \in \mathcal{L}(X)$ such that $wsr \in \mathcal{L}(X)$. Let u be the shortest suffix of wsr which has a proper prefix equal to w. Then $u \in U$. This shows that U is an X-maximal suffix code.

We have $\alpha(p) = 0$ for any proper prefix p of w since any word in $\mathcal{CR}(w)$ has w as a proper prefix. Next we have $\alpha(p) = \rho_X(p)$ for any $p \in U$. Indeed, if $ua \in \mathcal{L}(X)$ for $u \in P$ and $a \in A$, then $ua \in \mathcal{CR}(w) \cup P$ since $\mathcal{L}(X)$ is recurrent. Thus we have $\alpha(P) = \rho_X(U)$.

By Proposition 7.5, $\rho_X(U)$ is finite. Therefore, Equation 7.7 shows that $\mathrm{Card}(\mathcal{CR}(w)) = \mathrm{Card}(\mathcal{R}(w))$ is finite.

Assume finally that $|w| \geq m$. Then $U \subset \mathcal{L}_{\geq m}(X)$ and thus, by Proposition 7.5, we have $\rho_X(U) = \rho_X(\mathcal{L}_m(X))$. Thus we have

$$\alpha(P) = \rho_X(\mathcal{L}_m(X)).$$

By Equation (7.7), this implies Equation (7.2). ∎

It is known that for dendric shift spaces, irreducibility is enough to guarantee minimality [13]. We obtain as a direct corollary of Theorem 7.3 that this still holds for eventually dendric shifts.

Corollary 7.6 *An eventually dendric shift space is minimal if and only if it is irreducible.*

Proof. Let X be an irreducible shift space. By Theorem 7.3, the set $\mathcal{R}(w)$ is finite for every $w \in \mathcal{L}(X)$. Thus X is minimal. ∎

Note that the proof shows that Corollary 7.6 holds for the more general class of shift spaces which are *eventually neutral with threshold m*, in the sense that every word of length at least m is neutral.

The class of eventually neutral shift spaces includes the shift spaces X such that $\mathcal{L}(X)$ is neutral with characteristic c introduced in [13]. By definition, $\mathcal{L}(X)$ is *neutral with characteristic c* if every non-empty word $w \in \mathcal{L}(X)$ is neutral and if $m(\varepsilon) = 1 - c$. Such a set is eventually neutral with threshold at most 1. Corollary 7.6 is proved in [13] for neutral sets of characteristic c with a similar proof.

As observed in [11], one may use the result on neutral sets with characteristic c to obtain a proof of Corollary 7.6. Indeed, if X is eventually dendric with threshold m, then the m-th higher block shift space $X^{(m)}$ is eventually neutral with threshold at most 1 (a priori, we know nothing about the behavior of the empty word).

Note also that Theorem 7.3 shows that in a minimal eventually dendric shift space the cardinality of sets of complete return words is bounded. There exist minimal shift spaces which do not have this property (see [18, Example 3.17]).

8 Generalized extension graphs

We will now see how the conditions on extension graphs can be generalized to graphs expressing the extension by words having different length. We will prove the following statement.

Proposition 8.1 *For every $n \geq 1$ and $m \geq 0$, the graph $\mathcal{E}_n(w)$ is a tree for all $w \in \mathcal{L}_{\geq m}(X)$ if and only if $\mathcal{E}_{n+1}(w)$ is a tree for all words $w \in \mathcal{L}_{\geq m}(X)$.*

We will need the following notions which allow to replace in the graphs $E_n(X)$ the set of all words of length n by a set of words of different lengths.

Let X be a shift space on an alphabet A. A set $U \subset \mathcal{L}(X)$ is said to be *right X-complete* (resp. *left X-complete*) if every word of $\mathcal{L}(X)$ either has a prefix (resp. suffix) in U or is a prefix (resp. suffix) of a word of U.

It is not difficult to show that a prefix code (resp. a suffix code) $U \subset \mathcal{L}(X)$ is X-maximal if and only if it is right X-complete (resp. left X-complete).

For $U, V \subset A^*$ and $w \in \mathcal{L}(X)$, let

$$L_U(w) = \{u \in U \mid uw \in \mathcal{L}(X)\} \quad \text{and} \quad R_V(w) = \{v \in V \mid wv \in \mathcal{L}(X)\}.$$

Let $U \subset A^*$ (resp. $V \subset A^*$) be an suffix code (resp. prefix code) and $w \in \mathcal{L}(X)$ be such that $L_U(w)$ is an X-maximal suffix code (resp. $R_V(w)$ is an X-maximal prefix code). The *generalized extension graph* of w relative to U, V is the following undirected bipartite graph $\mathcal{E}_{U,V}(w)$. The set of vertices is the disjoint union of $L_U(w)$ and $R_V(w)$. The edges are the pairs $(u, v) \in L_U(w) \times R_V(w)$ such that $uwv \in \mathcal{L}(X)$. In particular $\mathcal{E}_n(w) = \mathcal{E}_{\mathcal{L}_n(X), \mathcal{L}_n(X)}(w)$.

The proof uses the following statement. The only if part of the next result is [5, Lemmas 3.8 and 3.10].

Lemma 8.2 *Let X be a shift space and let $w \in \mathcal{L}(X)$. Let $U \subset \mathcal{L}(X)$ be a finite X-maximal suffix code and let $V \subset \mathcal{L}(X)$ be finite X-maximal prefix code. Let $\ell \in \mathcal{L}(X)$ be such that $A\ell \cap \mathcal{L}(X) \subset U$ and such that $\mathcal{E}_{A,V}(\ell w)$ is a tree. Set $U' = \Delta_\ell(U)$. The graph $\mathcal{E}_{U',V}(w)$ is a tree if and only if the graph $\mathcal{E}_{U,V}(w)$ is a tree.*

Proof. We need only to prove the if part.

First, note that the hypothesis that $\mathcal{E}_{A,V}(\ell w)$ is a tree guarantees that the left vertices $A\ell$ in $\mathcal{E}_{U,V}(w)$ are clusterized: for any pair of vertices $a\ell, b\ell$ there exists a unique reduced path from $a\ell$ to $b\ell$ in $\mathcal{E}_{U,V}(w)$ using as left vertices only elements of $A\ell$. Indeed, such a path exists since the subgraph $\mathcal{E}_{A\ell,V}(w)$ of $\mathcal{E}_{U,V}(w)$ is isomorphic to $\mathcal{E}_{A,V}(\ell w)$ that is connected. Since $\mathcal{E}_{U,V}(w)$ is a tree, this path is unique.

Let $v, v' \in R_V(w)$ be two distinct vertices and let π be the unique reduced path from v to v' in $\mathcal{E}_{U,V}(w)$. We show that we can find a unique reduced path π' from v to v' in $\mathcal{E}_{U',V}(w)$.

If π does not pass by $A\ell$, we can simply define $\pi' = \pi$. Otherwise, we can decompose π in a unique way as a concatenation of a path π_1 from v to a vertex in $A\ell$ not passing by $A\ell$ before, followed by a path from $A\ell$ to $A\ell$ (using on the left only vertices from $A\ell$) and a path π_2 from $A\ell$ to v' without passing in $A\ell$ again. We consider in $\mathcal{E}_{U',V}(w)$ the unique path π_1' from v to ℓ obtained by replacing the last vertex of π_1 by ℓ and the unique reduced path π_2' from ℓ to v' obtained by replacing the first vertex of π_2 by ℓ. In this case we define π' as the concatenation of π_1' and π_2'.

The reduced path π' is unique. Indeed, let us suppose that we have a different path π^* from v to v' in $\mathcal{E}_{U',V}(w)$. If π^* does not pass (on the left) by ℓ then we would find a path having the same vertices in $\mathcal{E}_{U,V}(w)$ which is impossible since the graph is acyclic. Let us suppose that both π' and π^* passes by ℓ. Without loss of generality let us suppose that we have a cycle in $\mathcal{E}_{U',V}(w)$ passing by ℓ

and v (the case with v' being symmetric). Let us define by π_0' and π_0^* the two distinct subpaths of π' and π^* respectively going from v to ℓ. Since $\mathcal{L}(X)$ is extendable, we can find $a\ell, b\ell \in U$, with $a, b \in A$ not necessarily distinct, and two reduced paths π_1 from v to $a\ell$ and and π_2 from v to $b\ell$ in $\mathcal{E}_{U,V}(w)$ obtained from π_0' and π_0^* by replacing the vertex ℓ by $a\ell$ and $b\ell$ respectively. From the remark at the beginning of the proof we know that we can find a reduced path in $\mathcal{E}_{U,V}(w)$ from $a\ell$ to $b\ell$. Thus we can find a nontrivial cycle in $\mathcal{E}_{U,V}(w)$, which contradicts the acyclicity of the graph. ∎

A symmetric statement of Lemma 8.2 holds. To state it, we introduce the following complement to the Δ operator. For a finite X-maximal prefix code V and a word r such that $rA \cap \mathcal{L}(X) \subset V$, set

$$\Gamma_r(V) = (V \setminus rA) \cup \{r\}. \tag{8.1}$$

Conversely, for a finite X-maximal prefix code W and $r \in W$, we denote

$$\Gamma_r'(W) = (W \setminus \{r\}) \cup (rA \cap \mathcal{L}(X)). \tag{8.2}$$

A symmetric statement of Lemma 7.4 holds. It allows to state a symmetric statement of Lemma 8.2 for $r \in \mathcal{L}(X)$ such that $rA \cap \mathcal{L}(X) \subset V$ and $\mathcal{E}_{U,A}(wr)$ is a tree, with $V' = \Gamma_r(V)$: the graph $\mathcal{E}_{U,V}(w)$ is a tree if and only if $\mathcal{E}_{U,V'}(w)$ is a tree.

Lemma 8.3 *Let $n \geq 1$, let $m \geq 0$ and let V be a finite X-maximal prefix code. If $\mathcal{E}_{\mathcal{L}_n(X),V}(w)$ is a tree for every $w \in \mathcal{L}_{\geq m}(X)$, then for each word $u \in \mathcal{L}_{\geq m+n-1}(X)$, the graph $\mathcal{E}_{A,V}(u)$ is a tree.*

Proof. Set $u = \ell w$ with $|w| = m$. The graph $\mathcal{E}_{A,V}(u)$ is obtained from $\mathcal{E}_{\mathcal{L}_n(X),V}(u)$ by identifying the vertices of $L_n(u)$ ending with the same letter. Since $\mathcal{E}_{\mathcal{L}_n(X),V}(u)$ is connected, $\mathcal{E}_{A,V}(u)$ is also connected.

Set $\ell = \ell'\ell''$ with $|\ell'| = n - 1$. The graph $\mathcal{E}_{A,V}(\ell w)$ is isomorphic to $\mathcal{E}_{A\ell',V}(\ell''w)$ which is a subgraph of $\mathcal{E}_n(\ell''w)$ and thus it is acyclic.

Thus $\mathcal{E}_{A,V}(u)$ is a tree. ∎

A symmetric statement holds. For $n \geq 1$, $m \geq 0$ and U a finite X-maximal suffix code: If $\mathcal{E}_{U,\mathcal{L}_n(X)}(w)$ is a tree for every $w \in \mathcal{L}_{\geq m}(X)$, then $\mathcal{E}_{U,A}(v)$ is a tree for every $v \in \mathcal{L}_{\geq m+n-1}(X)$.

Proof of Proposition 8.1. We proceed in several steps.

Step 1. Assume first that $\mathcal{E}_n(w)$ is tree for every word $w \in \mathcal{L}_{\geq m}(X)$. We fix some $w \in \mathcal{L}_{\geq m}(X)$.

Step 1.1 We claim that for any finite X-maximal suffix code U formed of words of length n or $n+1$, the graph $\mathcal{E}_{U,\mathcal{L}_n(X)}(w)$ is a tree by induction on $\gamma_{n+1}(U) = \mathrm{Card}(L_U(w) \cap A^{n+1})$.

The property is true for $\gamma_{n+1}(U) = 0$, since then $\mathcal{E}_{U,\mathcal{L}_n(X)}(w) = \mathcal{E}_n(w)$. Assume now that $\gamma_{n+1}(U) > 0$. Let $a\ell$ with $a \in A$ be a word of length $n+1$ in $L_U(w)$. Since U is an X-maximal suffix code with words of length n or $n+1$,

we have $A\ell \cap \mathcal{L}(X) \subset U$. Let us consider $U' = \Delta_\ell(U)$. Since $\gamma_{n+1}(U') < \gamma_{n+1}(U)$, by induction hypothesis the graph $\mathcal{E}_{U',\mathcal{L}_n(X)}(w)$ is a tree. Moreover, by Lemma 8.3, the graph $\mathcal{E}_{A,\mathcal{L}_n(X)}(\ell w)$ is a tree.

Thus, by Lemma 8.2, the graph $\mathcal{E}_{U,\mathcal{L}_n(X)}(w)$ is a tree. This proves the claim. Step 1.2 We now claim that for any finite X-maximal prefix code V formed of words of length n or $n + 1$, the graph $\mathcal{E}_{\mathcal{L}_{n+1}(X),V}(w)$ is a tree. We use an induction on $\delta_{n+1}(V) = \mathrm{Card}(R_V(w) \cap A^{n+1})$.

The property is true for $\delta_{n+1}(V) = 0$, since the graph $\mathcal{E}_{\mathcal{L}_{n+1}(X),V}(w) = \mathcal{E}_{\mathcal{L}_{n+1}(X),\mathcal{L}_n(X)}(w)$, is a tree by Step 1.1. Assume now that $\delta_{n+1}(V) > 0$. Let ra with $a \in A$ be a word of length $n+1$ in $R_V(w)$. Since V is an X-maximal prefix code with words of length n or $n + 1$, we have $rA \cap \mathcal{L}(X) \subset U$. Let us consider $V' = \Gamma_r(V)$. Since $\delta_{n+1}(V') < \delta_{n+1}(V)$, by induction hypothesis the graph $\mathcal{E}_{\mathcal{L}_{n+1}(X),V'}(w)$ is a tree. Moreover, by the symmetric version of Lemma 8.3, the graph $\mathcal{E}_{\mathcal{L}_{n+1}(X),A}(wr)$ is a tree. This proves the claim.

Since $\mathcal{E}_{n+1}(w) = \mathcal{E}_{\mathcal{L}_{n+1}(X),\mathcal{L}_{n+1}(X)}(w)$, we conclude that $\mathcal{E}_{n+1}(w)$ is a tree. Step 2 Assume now that $\mathcal{E}_{n+1}(w)$ is a tree for every $w \in \mathcal{L}_{\geq m}(X)$. Fix some $w \in \mathcal{L}_{\geq m}(X)$.
Step 2.1 We first claim that $\mathcal{E}_{U,\mathcal{L}_{n+1}(X)}$ is a tree for every X-maximal suffix code U formed of words of length n or $n + 1$ by induction on $\gamma_n(U) = \mathrm{Card}(L_U(w) \cap A^n)$.

The property is true if $\gamma_n(U) = 0$, since then $\mathcal{E}_{U,\mathcal{L}_{n+1}(X)}(w) = \mathcal{E}_{n+1}(w)$.

Assume next that $\gamma_n(U) > 0$. Let $\ell \in L_U(w) \cap A^n$. Set $W = \Delta'_\ell(U)$ or equivalently $U = \Delta_\ell(W)$. Then $\delta_n(W) < \delta_n(U)$ and consequently $\mathcal{E}_{W,\mathcal{L}_{n+1}(X)}(w)$ is a tree by induction hypothesis. On the other hand, by Lemma 8.3, the graph $\mathcal{E}_{A,\mathcal{L}_{n+1}(X)}(\ell w)$ is also a tree. By Lemma 8.2, the graph $\mathcal{E}_{U,\mathcal{L}_{n+1}(X)}(w)$ is a tree and thus the claim is proved.
Step 2.2 We now claim that $\mathcal{E}_{\mathcal{L}_n(X),V}(w)$ is a tree for every X-maximal prefix code V formed of words of length n or $n + 1$ by induction on $\delta_n(V) = \mathrm{Card}(R_V(w) \cap A^n)$.

The property is true if $\delta_n(V) = 0$ by Step 2.1. Assume now that $\delta_n(V) > 0$. Let $r \in R_V(w) \cap A^n$ and let $T = \Gamma'_r(V)$ or equivalently $V = \Gamma_r(T)$. Then $\delta_n(T) < \delta_n(V)$ and thus $\mathcal{E}_{\mathcal{L}_n(X),T}(w)$ is a tree by induction hypothesis. On the other hand, by the symmetric version of Lemma 8.3, the graph $\mathcal{E}_{\mathcal{L}_n(X),A}(wr)$ is also a tree. By Lemma 8.2, the graph $\mathcal{E}_{\mathcal{L}_n(X),T}(w)$ is a tree and thus the claim is proved.

Since $\mathcal{E}_n(w) = \mathcal{E}_{U,V}(w)$ for $U = V = \mathcal{L}_n(X)$, it follows from the claim that $\mathcal{E}_n(w)$ is a tree. $\blacksquare$

The following result shows that in the definition of eventually dendric shift spaces, one can replace the graphs $\mathcal{E}_1(w)$ by $\mathcal{E}_n(w)$ with the same threshold.

Theorem 8.4 *Let X be a shift space. For every $m \geq 1$, the following conditions are equivalent.*

(i) *X is eventually dendric with threshold m,*

(ii) *the graph $\mathcal{E}_n(w)$ is a tree for every $n \geq 1$ and every word $w \in \mathcal{L}_{\geq m}(X)$,*

(iii) *there is an integer $n \geq 1$ such that $\mathcal{E}_n(w)$ is a tree for every word $w \in \mathcal{L}_{\geq m}(X)$.*

Proof. (i) $\Rightarrow$ (ii). It is proved by ascending induction on n using iteratively Proposition 8.1.

(ii) $\Rightarrow$ (iii). It is obvious.

(iii) $\Rightarrow$ (i). It is proved by descending induction on n using Proposition 8.1.
∎

9 Complete bifix decoding

Let X be a shift space on an alphabet A. A subset of $\mathcal{L}(X)$ is *two-sided X-complete* if it is both left and right X-complete.

A *bifix code* is both a prefix code and a suffix code. A bifix code $U \subset \mathcal{L}(X)$ is X-maximal if it is not properly contained in a bifix code $V \subset \mathcal{L}(X)$. If a bifix code $U \subset \mathcal{L}(X)$ is right X-complete (resp. left X-complete), it is an X-maximal bifix code since it is already an X-maximal prefix code (resp. suffix code). It can be proved conversely that if X is irreducible, a finite bifix code is X-maximal if and only if it is two-sided X-complete (see [3, Theorem 4.2.2]). This is not true in general, as shown by the following example.

Example 9.1 Let X be the shift space such that $\mathcal{L}(X) = a^*b^*$. The set $U = \{aa, b\}$ is an X-maximal bifix code. Indeed, it is a bifix code and it is left X-complete as one may verify. However it is not right X-complete since no word in ab^* has a prefix in U.

Let X be a shift space and let U be a two-sided X-complete finite bifix code. Let $\varphi : B \to U$ be a coding morphism for U, that is, a bijection from an alphabet B onto U extended to a morphism from B^* into A^*. Then $\varphi^{-1}(\mathcal{L}(X))$ is factorial and, since U is two-sided complete, it is extendable. Thus it is the language of a shift space called the *complete bifix decoding* of X with respect to U.

For example, for any $n \geq 1$, the set $\mathcal{L}_n(X)$ is a two-sided complete bifix code and the corresponding complete bifix decoding is the decoding of X by non-overlapping n-blocks. It can be identified with the dynamical system (X, σ^n).

In [5, Theorem 3.13] it is proved that the maximal bifix decoding of an irreducible dendric shift space is a dendric shift space. Actually, the hypothesis that X is irreducible is only used to guarantee that the X-maximal bifix code used for the decoding is also an X-maximal prefix code and an X-maximal suffix code. In the definitions used here of a maximal bifix decoding, we do not need this hypothesis.

Theorem 9.2 *Any complete bifix decoding of an eventually dendric shift space is an eventually dendric shift space having the same threshold.*

Note that any X-maximal suffix code U one has $\mathrm{Card}(U) \geq \mathrm{Card}(\mathcal{L}_1(X))$. Indeed, every $a \in \mathcal{L}_1(X)$ appears as a suffix of (at least) an element of X.

Lemma 9.3 *Let X be an eventually dendric shift space with threshold n. For any $w \in \mathcal{L}_{\geq n}(X)$, any X-maximal suffix code U and any X-maximal prefix code V, the graph $\mathcal{E}_{U,V}(w)$ is a tree.*

Proof. We use an induction on the sum of the lengths of the words in U, V. The property is true if the sum is equal to $2\,\mathrm{Card}(\mathcal{L}_1(X))$. Indeed, for every $w \in \mathcal{L}_{\geq n}(X)$ one has $U = L(w)$ and $V = R(w)$ and thus $\mathcal{E}_{U,V}(w) = \mathcal{E}_1(w)$ is a tree. Otherwise, we may assume that U contains words of length at least 2 (the case with V being symmetrical). Let $u \in U$ be of maximal length. Set $u = a\ell$ with $a \in A$. Since U is an X-maximal suffix code, we have $A\ell \cap \mathcal{L}(X) \subset U$. Set $U' = \Delta_\ell(U)$. By induction hypothesis, the graphs $\mathcal{E}_{U',V}(w)$ and $\mathcal{E}_{A,V}(\ell w)$ are trees. Thus, by Lemma 8.2, $\mathcal{E}_{U,V}(w)$ is also a tree. ∎

Proof of Theorem 9.2. Assume that X is eventually dendric with threshold n. Let $\varphi : B \to U$ be a coding morphism for U and let Y be the decoding of X corresponding to U. Consider a word w of $\mathcal{L}(Y)$ of length at least n. By Lemma 9.3, and since $|\varphi(w)| \geq n$, the graph $\mathcal{E}_{U,U}(\varphi(w))$ is a tree. But for $b, c \in B$, one has $bwc \in \mathcal{L}(Y)$ if and only if $\varphi(bwc) \in \mathcal{L}(X)$, that is, if and only if $(\varphi(b), \varphi(c)) \in E_1(\varphi(w))$. Thus $\mathcal{E}_1(w)$ is isomorphic to $\mathcal{E}_{U,U}(\varphi(w))$ and thus $\mathcal{E}_1(w)$ is a tree. This shows that Y is eventually dendric with threshold n. ∎

Example 9.4 Let X be the Fibonacci shift. Then $U = \{aa, aba, b\}$ is an X-maximal bifix code. Let $\varphi : \{u, v, w\} \to U$ be the coding morphism for U defined by $\varphi : u \mapsto aa, v \mapsto aba, w \mapsto b$. The complete bifix decoding of X with respect to U is an eventually dendric shift space with threshold 0. It is actually the natural coding of an interval exchange transformation on three intervals (see [6]). The extension graphs $\mathcal{E}_1(\varepsilon, Y)$ and $\mathcal{E}_1(v, Y)$ are shown in Figure 13.

Figure 13: The graphs $\mathcal{E}_1(\varepsilon, Y)$ and $\mathcal{E}_1(w, Y)$.

A particular case of complete bifix decoding is related to the skew product of two dynamical systems, a notion which is well-known in topological dynamics (see [10]). Indeed, assume that we start with a shift space X, a transitive permutation group G on a set Q and a morphism $f : A^* \to G$. We denote by $q \mapsto q \cdot w$ the result of the action of the permutation $f(w)$ on the point $q \in Q$. The skew product of X and (G, Q) is the shift space Y on the alphabet $A \times Q$ formed by the bi-infinite words (a_i, q_i) such that $(a_i) \in X$ and $p_{i+1} = p_i \cdot f(a_i)$ for all $i \in \mathbb{Z}$. Fix a point $i \in Q$. The set of words w such that $i \cdot w = i$ is a submonoid generated by a bifix code U which is two-sided complete. The decoding of $S = \mathcal{L}(X)$ with respect to $U \cap S$ is the language of the dynamical

system induced by Y on the set of $y \in Y$ such that $y_0 = (a, i)$ for some $a \in A$ (see [6] for more details).

Example 9.5 Let X be the Fibonacci shift, i.e., the shift whose language is the Fibonacci set. Let $Q = \{1, 2\}$, $G = \mathbb{Z}/2\mathbb{Z}$ and $f : A^* \to G$ defined by $a \mapsto (12), b \mapsto (1)$. Choosing $i = 1$, the bifix code U built as above is $U = \{aa, aba, b\}$ as in Example 9.4.

10 Conclusion

The class of eventually dendric shifts is shown in this paper to have strong closure properties. It leaves open the question of whether it is closed under taking *factors*, that is, images by a sliding block code not necessarily bijective.

It would be interesting to know how other properties which are known to hold for dendric shifts extend to this more general class. This includes the following:

1. To which extent the properties of return words proved for minimal dendric shifts extend to eventually dendric ones. For example, what can we say about the subgroup of the free group generated by return words to a given word? In [5] it is proved that for minimal dendric sets, every set of return words is a basis of the free group, while in the case of specular sets, the set of return word to a fixed word is a basis of a particular subgroup called the even subgroup (see [4]).

2. Is there a finite S-adic representation for all minimal eventually dendric shifts? There is one for minimal dendric shifts [7].

3. Is the property of being eventually dendric decidable for a substitutive shift, as it is for dendric ones [12]?

It would also be interesting to know whether the conjugacy of effectively given eventually dendric shifts is decidable (the conjugacy of substitutive shifts was recently shown to be decidable [17]).

References

[1] Pierre Arnoux and Gérard Rauzy. Représentation géométrique de suites de complexité $2n + 1$. *Bull. Soc. Math. France*, 119(2):199–215, 1991.

[2] Ľubomíra Balková, Edita Pelantová, and Wolfgang Steiner. Sequences with constant number of return words. *Monatsh. Math.*, 155(3-4):251–263, 2008.

[3] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.

[4] Valérie Berthé, Clelia De Felice, Vincent Delecroix, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Specular sets. *Theoret. Comput. Sci.*, 684:3–28, 2017.

[5] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176(4):521–550, 2015.

[6] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and interval exchanges. *J. Pure Appl. Algebra*, 219(7):2781–2798, 2015.

[7] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Dicrete Math.*, 338:725–742, 2015.

[8] Julien Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1):67–88, 1997. Journées Montoises (Mons, 1994).

[9] Julien Cassaigne and François Nicolas. Factor complexity. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 163–247. Cambridge Univ. Press, Cambridge, 2010.

[10] Isaac P. Cornfeld, Sergei V. Fomin, and Yakov G. Sinai. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskii.

[11] Michael Damron and Jon Fickenscher. The number of ergodic measures for transitive subshifts under the regular bispecial condition. 2019. arXiv:1902.04619.

[12] Francesco Dolce, Revekka Kyriakoglou, and Julien Leroy. Decidable properties of extension graphs for substitutive languages. 2016. 15èmes Journées Montoises d'informatique théorique, Liège (Belgique).

[13] Francesco Dolce and Dominique Perrin. Neutral and tree sets of arbitrary characteristic. *Theoret. Comput. Sci.*, 658(part A):159–174, 2017.

[14] Francesco Dolce and Dominique Perrin. Eventually dendric shifts. In René van Bevern and Gregory Kucherov, editors, *Computer Science - Theory and Applications - 14th International Computer Science Symposium in Russia, CSR 2019, Novosibirsk, Russia, July 1-5, 2019, Proceedings*, volume 11532 of *Lecture Notes in Computer Science*, pages 106–118. Springer, 2019.

[15] Francesco Dolce and Dominique Perrin. Return words and bifix codes in eventually dendric sets. In *Words 2019, Proceedings*. Springer Verlag, 2019.

[16] Sebastian Donoso, Fabien Durand, Alejandro Maass, and Samuel Petite. On automorphism groups of low complexity subshifts. *Ergod. Th. Dynam. Sys.*, 36:64–95, 2016.

[17] Fabien Durand and Julien Leroy. Decidability of the isomorphism and the factorization between minimal substitution subshifts. 2018. `https://arxiv.org/abs/1806.04891`.

[18] Fabien Durand, Julien Leroy, and Gwenaël Richomme. Do the properties of an S-adic representation determine factor complexity? *J. Integer Seq.*, 16(2):Article 13.2.6, 30, 2013.

[19] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[20] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.

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

[22] Vojtch Vesely. Properties of morphic images of S-adic sequences. Master's thesis, Czech Technical University, 2018.