Return Words and Bifix Codes in Eventually Dendric Sets

Francesco Dolce, Dominique Perrin

Return words and bifix codes in eventually dendric sets

Francesco Dolce[1] and Dominique Perrin[2]

[1] IRIF, Université Paris Diderot, France
[2] LIGM, Université Paris-Est Marne-la-Vallée, France

Abstract. A shift space (or its set of factors) is eventually dendric if the possible extensions of all long enough factors are described by a graph which is a tree. We prove two results on eventually dendric shifts. First, all sets of return words to long enough words have the same cardinality. Next, this class of shifts is closed under complete bifix decoding.

Keywords: Formal languages, Symbolic dynamics, Neutral words

1 Introduction

A shift space X can be defined as the set of two-sided infinite words with all their factors in a given extendable factorial set, called the *language* of the shift, and denoted $L(X)$. Thus shift spaces and extendable factorial sets are two aspects of the same notion. The traditional hierarchy of classes of languages translates into a hierarchy of shift spaces. In particular, a shift space X is called sofic when its language $L(X)$ is a regular language. It is called of finite type when its language is the complement of a finitely generated ideal.

The complexity of a shift space is the function $n \mapsto p(n)$ where $p(n)$ is the number of factors of length n of the shift. 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. Such shifts arise in many other contexts(see, e.g., [**?**] or [**?**]). A shift space X is recurrent if for every $u, v \in L(X)$ there exists some w such that $uwv \in L(X)$. It is uniformly recurrent if for every element $w \in L(X)$ there is an integer n_w such that w occurs as a factor in each elements of $L(X)$ longer than n_w. Thus, the notion of recurrence expresses the property that every factor has a second occurrence. Uniform recurrence correspond to the appearance of the second occurrence after bounded time. Is is known that all uniformly recurrent factorial extendable sets of at most linear complexity have a finite $\mathcal{S}$-adic representation (i.e., a generalization to several morphisms of a fixed point of a morphism). Conversely, it is an open problem, known as the *$\mathcal{S}$-adic conjecture*, to characterise the S-adic representations of uniformly recurrent sets of at most linear complexity. Note that all substitutive shifts defined by a primitive morphism are both uniformly recurrent and of at most linear complexity.

In this contribution we study a class of shift spaces of at most linear complexity, called eventually dendric, recently introduced in [**?**]. We also call eventually dendric the language of an eventually dendric shift. This class extends the class of dendric sets introduced in [**?**] (under the name of *tree sets*) which themselves extend naturally episturmian sets (also called Arnoux-Rauzy sets) and interval exchange sets. It is known that the class of eventually dendric shifts is closed under the natural equivalence on shifts called conjugacy (see [**?**]). We prove here 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 sets, giving a strong motivation for its introduction.

A dendric set S is defined by introducing the extension graph of a word and by requiring that this graph is a tree for every word in S. It has many interesting properties which involve free groups. In particular, in a dendric set S on an alphabet A, the group generated by the set of return words (see Section 4) to some word in S is the free group on the alphabet and, in particular, has $\operatorname{Card} A$ free generators. This generalizes a property known for Sturmian sets whose link with automorphisms of the free group was noted by Arnoux and Rauzy. The class of eventually dendric sets, studied in this paper, is defined by the property that the extension graph of every long enough word in the set is a tree (for short words the graphs may be arbitrary). These sets are contained in the class of eventually neutral sets, where a weaker hypothesis on the extensions is required (see Section 3) Our main results are that: all sets of return words to a (long enough) word in a recurrent eventually neutral set S have the same cardinality (Theorem 1); the class of eventually dendric sets is closed under complete bifix decoding (Theorem 4). An interesting consequence of Theorem 1 is the equivalence of the notions of recurrence and uniform recurrence for eventually neutral (and thus eventually dendric) sets (Theorem 2)

The paper is organized as follows. In Section 2, we introduce the definition of extension graphs and of eventually dendric sets. In Section 3, we recall some known properties on the complexity of a factorial extendable set of words and of special words. In Section 4 we focus on (uniformly) recurrent eventually neutral sets and on return words in such sets. In particular we prove that for every word w the set of return words on w is finite and that when w is long enough, all these sets have the same cardinality (Theorem 1). In the same section, and actually as a consequence of Theorem 1, we also prove that an eventually neutral set is recurrent if and only if it is uniformly recurrent (Theorem 2), a property already known for neutral sets ([**?**,**?**]). In Section 5 we introduce generalized extension graphs in which extension by words over particular sets replaces extension by letters. We prove that one obtains an equivalent definition of eventually dendric shifts using these generalized extension graphs (Theorem 3). In Section 6, we use generalized extesion graphs to prove that the class of recurrent eventually dendric sets is closed under complete bifix decoding (Theorem 4), a result already known for dendric sets. We conclude with some open questions.

Acknowledgements We thank Valérie Berthé, Paulina Cecchi, Fabien Durand and Samuel Petite for useful conversations on this subject.

2 Eventually dendric sets

Let A be a finite alphabet. We denote by A^* the set of all words on A and by A^n the set of words of n letters, with $n \geq 0$. We denote by ε the empty word. A *factor* of a word x is a word v such that $x = uvw$. If u (resp. w) is the empty word, we say that v is a *suffix* (resp. *prefix*) of x. These definitions can be extended in a natural way to infinite words. A set of words S on the alphabet A is *factorial* if it contains the factors of its elements as well as the alphabet A. It is called *extendable* if for any $w \in S$ there are letters $a, b \in A$ such that $awb \in S$. An example of factorial and extendable set is given by the *language* of an infinite word $\mathbf{w}$, i.e., the set $\mathcal{L}(\mathbf{w})$ containing all finite factors of $\mathbf{w}$.

Given a factorial set S and an integer $n \geq 0$ we denote $S_n = S \cap A^n$ and $S_{\geq n} = \bigcup_{m \geq n} S_m$. For $w \in S$ and $n \geq 1$, we denote $L_n(w, S) = \{u \in S_n \mid uw \in S\}$, $R_n(w, S) = \{v \in S_n \mid wv \in S\}$ and $E_n(w, S) = \{(u, v) \in L_n(w, S) \times R_n(w, S) \mid uwv \in S\}$. The *extension graph* of order n of w, denoted $\mathcal{E}_n(w, S)$, is the undirected bipartite graph whose set of vertices the disjoint union of $L_n(w, S)$ and $R_n(w, S)$ and with edges the elements of $E_n(w, S)$. When the context is clear, we denote $L_n(w), R_n(w), E_n(w)$ and $\mathcal{E}_n(w)$ instead of $L_n(w, S)$, $R_n(w, S)$, $E_n(w, S)$ and $\mathcal{E}_n(w, S)$. A path in an undirected graph is *reduced* if it does not contain successive equal edges. For any $w \in S$, 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 in $\mathcal{E}_n(w)$ between every pair of vertices of $L_n(w)$ (resp. $R_n(w)$). A factorial and extendable set S is said to be *eventually dendric* with *threshold* $m \geq 0$ if $\mathcal{E}_1(w)$ is a tree for every word $w \in S_{\geq m}$. It is said to be (purely) *dendric* if we can choose $m = 0$. Dendric sets were introduced in [**?**] under the name of tree sets. An important example of dendric sets is formed by *episturmian sets* (also called Arnoux-Rauzy sets), which are by definition factorial extendable sets closed by reversal and such that for every n there exists a unique $w_n \in S_n(X)$ such that $\mathrm{Card}(R_1(w_n)) = \mathrm{Card}(A)$ and such that for every $w \in S_n \setminus \{w_n\}$ one has $\mathrm{Card}(R_1(w)) = 1$ (see [**?**,**?**]).

Example 1. Let F be the *Fibonacci set*, which is the set of factors of the words $\varphi^n(a)$, where φ is the morphism $a \mapsto ab, b \mapsto a$. It is also the set of factors of the one-sided infinite word $\mathbf{x}$ having all $\varphi^n(a)$ as prefixes, called a *fixed point* of φ, since $\varphi(\mathbf{x}) = \mathbf{x}$. It is well known that F is a Sturmian set (see [**?**]). 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.

The tree sets of *characteristic* $c \geq 1$ introduced in [**?**,**?**] give an example of eventually dendric sets of threshold 1 (while $\mathcal{E}_1(\varepsilon)$ is a forest of c trees).

Example 2. Let S be the language of the infinite word obtained as fixed point of the morphism $\psi : a \mapsto ab, b \mapsto cda, c \mapsto cd, d \mapsto abc$. Its language is a tree set of characteristic 2 and it is actually a specular set ([**?**, Example 4.2]). The

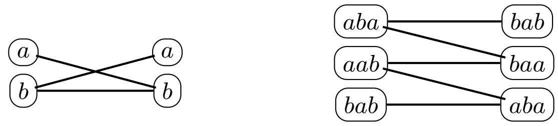

Fig. 1. The graphs $\mathcal{E}_1(a)$ and $\mathcal{E}_3(a)$.

extension graph $\mathcal{E}_1(\varepsilon)$ is shown in Figure 2. Since the extension graphs of all nonempty words are trees, the set is eventually dendric with threshold 1.

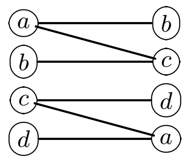

Fig. 2. The extension graph $\mathcal{E}_1(\varepsilon)$.

Example 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 [**?**]). Let α be the morphism $\alpha : a \mapsto a, b \mapsto a, c \mapsto c$. The set $\alpha(S)$ is eventually dendric with threshold 4 (see [**?**]).

3 Complexity of eventually dendric sets

Let S be a factorial extendable set. For a word $w \in S$, we denote $\ell_k(w) = \operatorname{Card} L_k(w), e_k(w) = \operatorname{Card} E_k(w)$, and $r_k(w) = \operatorname{Card} R_k(w)$. For any $w \in S$, we have $1 \le \ell_k(w), r_k(w) \le e_k(w)$. 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, r, e and we simply say special instead of k-special. We define the *multiplicity* of w as $m(w) = e(w) - \ell(w) - r(w) + 1$. We say that w is *strong* if $m(w) \ge 0$, *weak* if $m(w) \le 0$ and *neutral* if $m(w) = 0$ (see [**?**]). It is clear that if $\mathcal{E}_1(w)$ is acyclic (resp., connected, a tree), then w is weak (resp., strong, neutral). The following proposition is easily verified.

Proposition 1. *Let S be a factorial extendable set and let $w \in S$. If w is neutral, then*

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

A factorial and extendable set S is said to be *eventually neutral* with *threshold* $m \ge 0$ if w is neutral for every word $w \in S_{\ge m}$. It is said to be (purely) *neutral* if we can choose $m = 0$. Set further $p_n(S) = \operatorname{Card} S_n$, $s_n(S) = p_{n+1}(S) - p_n(S)$ and $b_n(S) = s_{n+1}(S) - s_n(S)$. The sequence $p_n(S)$ is called the *complexity* of S.

The following result is from [**?**] (see also [**?**] and [**?**, Theorem 4.5.4]).

Proposition 2. *We have for all $n \geq 0$,*

$$s_n(S) = \sum_{w \in S_n} (\ell(w) - 1) = \sum_{w \in S_n} (r(w) - 1) \quad \text{and} \quad b_n(S) = \sum_{w \in S_n} m(w).$$

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

We will use the following easy consequence of Proposition 2.

Proposition 3. *Let S be a factorial extendable set. If S is eventually dendric, then the sequence $s_n(S)$ is eventually constant.*

The previous result implies that eventually dendric sets have eventual linear complexity. The converse of Proposition 3 is not true.

Example 4. Let C be the *Chacon ternary set*, which is the set of factors of the fixed point of the morphism $\varphi : a \mapsto aabc, b \mapsto bc, c \mapsto abc$. It is well known that the complexity of C is $p_n(C) = 2n + 1$ and thus that $s_n(C) = 2$ for all $n \geq 0$ (see [**?**, Section 5.5.2]). The extension graphs of abc and bca are shown in Figure 3. 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 $axa \in C$, the word $\varphi(axa) = aabc\varphi(x)aabc = ayaabc$ is also in C and thus $(a, a) \in \mathcal{E}_1(y)$. Since $cxa \in C$ and since a letter c is always preceded by a letter b, we have $bcxa \in C$. Thus $\varphi(bcxa) = bcyaabc \in C$ 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.

Fig. 3. The extension graphs of abc and bca.

4 Recurrent eventually dendric sets

A factorial set S is recurrent if for any $u, v \in S$ there is a word w such that $uwv \in S$. A set is *uniformly recurrent* whenever for all $w \in S$ there exists an $n \geq 0$ such that w is a factor of any word in S_n. This last property is called *minimality* in the context of dynamical systems. If S is uniformly recurrent and infinite, then either there exists for every $w \in S$ an integer $n \geq 1$ such that $w^n \notin S$ or S is equal to the set of factors of ane infinite periodic word $uuu \cdots$. A recurrent set is uniformly recurrent but the converse is false, since for example the A^* is recurrent but not uniformly recurrent as soon as A has at least two elements.

Let S be a factorial extendable set. The set of *complete return words* to a word $w \in S$ is the set $\mathcal{CR}_S(w)$ 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 S is uniformly recurrent if and only if it is recurrent 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 $\mathcal{R}_S(w)$ the set of return words to w. Clearly $\mathrm{Card}(\mathcal{CR}_S(w)) = \mathrm{Card}(\mathcal{R}_S(w))$.

Example 5. Let S be the Tribonacci set (see Example 3). Then $\mathcal{R}_S(a) = \{a, ba, ca\}$ and $\mathcal{R}_S(c) = \{abac, ababac, abaabac\}$.

By a result of [**?**], if S is uniformly recurrent and neutral (a fortiori, if S is dendric) the set $\mathcal{R}_S(w)$ has $\mathrm{Card}(A)$ elements for every $w \in S$. This is not true anymore for eventually dendric sets, as shown in the following example.

Example 6. Let S be the Tribonacci set and let $Y = \alpha(X)$ be, as in Example 3 the image of X under the morphism $\alpha : a, b \to a, c \to c$. Then, using Example 5, 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 $w \in S$, set $\rho_S(w) = r_1(w) - 1$ and for a set $W \subset S$, set $\rho_S(W) = \sum_{w \in W} \rho_S(w)$. By the symmetric of Proposition 1, for every neutral word $w \in S$, we have

$$\rho_S(w) = \sum_{a \in L_1(w)} \rho_S(aw). \tag{2}$$

Theorem 1. *Let S be a uniformly recurrent set which is eventually neutral with threshold m. For every $w \in S$, the set $\mathcal{R}_S(w)$ is finite. Moreover, for every $w \in S_{\geq m}$, we have*

$$\mathrm{Card}(\mathcal{R}_S(w)) = 1 + \rho(S_m). \tag{3}$$

Note that for $m = 0$ we have $\mathrm{Card}(\mathcal{R}_S(w)) = \mathrm{Card}(A)$ since $\rho_S(\varepsilon) = \mathrm{Card}(A) - 1$.

A *prefix code* (resp. a *suffix code*) is a set X of words such that none of them is a prefix (resp. a suffix) of another one. A prefix code (resp. a suffix code) $U \subset S$ is called S-*maximal* if it is not properly contained in a prefix code (resp. suffix code) $Y \subset S$ (see, for instance, [**?**]).

Proposition 4. *Let S be an eventually neutral set with threshold m. Then $\rho_S(U)$ is finite for every suffix code $U \subset S$. If U is a finite S-maximal suffix code with $U \subset S_{\geq m}$, then*

$$\rho_S(U) \ = \rho_S(S_m). \tag{4}$$

Proof. For any suffix code $U \subset S$, let us set $U_m = (U \cap S_{<m}) \cup (V \cap S_m(X))$, where V 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 $S_m(X)$ if U is S-maximal and contained in $S_{\geq m}(X)$. Assume first that $U \subset S$ is a finite suffix code. We prove the result by induction on the sum $\ell(U)$ of the lengths of the words of U that

$$\rho_S(U) \leq \rho_S(U_m) \text{ with equality if } U \text{ is } S\text{-maximal and } U \subset S_{\geq m}(X). \tag{5}$$

If all words of U are of length at most m, then $U \subset U_m$ with equality if U is S-maximal and $U \subset S_{\geq m}$, since in this case $U_m = U = S_m$. Thus Equation (5) holds. Otherwise, let $u \in U$ be of maximal length. Set $u = av$ with $a \in A$. Then $Av \cap S \subset U$. Set $U' = (U \setminus Av) \cup \{v\}$. Thus U' is a suffix code with $\ell(U') < \ell(U)$ which is S-maximal if U is S-maximal. We have the inclusion $U \subset (U' \setminus v) \cup L_1(v)v$ with equality if U is S-maximal. Since v is neutral (its extension graph is a tree), we have, by Equation (2), $\rho_S(U) \leq \rho_S(U') - \rho_S(v) + \sum_{a \in L_1(v)} \rho_S(av) = \rho_S(U')$, with equality if U is S-maximal. By induction hypothesis, Equation (5) holds for U'. Thus $\rho_S(U) \leq \rho_S(U_m)$. If U is X-maximal and $U \subset S_{\geq m}$, then $\rho_S(U) = \rho_S(S_m(X))$ since $U'_m = U_m = S_m(X)$. Thus Equation (5) is proved.

If U is infinite, then $\rho_S(U)$ is the supremum of the values of $\rho_S(W)$ on the finite subsets W of U and thus it is bounded by Equation (5).

Proof (of Theorem 1). Consider a word $w \in S$ and let P be the set of proper prefixes of $\mathcal{CR}_S(w)$. For $p \in P$, denote $\alpha(p) = \mathrm{Card}\{a \in A \mid pa \in P \cup \mathcal{CR}_S(w)\} - 1$. Then $\mathcal{CR}_S(w)$ is finite if and only if P is finite. Moreover in this case, since $\mathcal{CR}_S(w)$ is a prefix code, we have by a well known property of trees (one can see $\mathcal{CR}_S(w)$ as set of leaves and P as set of internal nodes)

$$\mathrm{Card}(\mathcal{CR}_S(w)) = \alpha(P) + 1, \tag{6}$$

where $\alpha(P) = \sum_{p \in P} \alpha(p)$. Let U be the set of words in P which are not proper prefixes of w. We claim that U is an S-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 $r \in S$. Either r has a suffix in U or r is a suffix of a word in u. Indeed, let us suppose that r has no suffixes in U. Then, since S is recurrent, there is some $s \in S$ such that $wsr \in S$. Let u be the shortest prefix of wsr which has a proper suffix equal to w . Then $u \in U$. This shows that U is an S-maximal suffix code. We have $\alpha(p) = 0$ for any proper prefix p of w since any word in $\mathcal{CR}_S(w)$ has w as a proper prefix. Next we have $\alpha(p) = \rho_S(p)$ for any $p \in U$. Indeed, if $ua \in S$ for $u \in P$ and $a \in A$, then $ua \in \mathcal{CR}_S(w) \cup P$ since S is recurrent. Thus we have $\alpha(P) = \rho_S(U)$. By Proposition 4, $\rho_S(U)$ is finite. Therefore, Equation 6 shows that $\mathrm{Card}(\mathcal{CR}_S(w)) = \mathrm{Card}(\mathcal{R}_S(w))$ is finite.

Assume finally that $|w| \geq m$. Then $U \subset S_{\geq m}(X)$ and thus, by Proposition 4, we have $\rho_S(U) = \rho_S(S_m)$. Thus we have $\alpha(P) = \rho_S(S_m)$. By Equation (6), this implies Equation (3).

It is known that for neutral set recurrence is enough to guarantee uniform recurrence [**?**]. We obtain as a direct corollary of Theorem 1 the following:

Theorem 2. *An eventually neutral set is recurrent if and only if it is uniformly recurrent.*

Proof. Let S be a recurrent eventually neutral set. By Theorem 1, the set $\mathcal{R}_S(w)$ is finite for every $w \in S$. Thus S is uniformly recurrent.

TTheorem 2 shows also that in a recurrent eventually neutral set the cardinality of complete return words is bounded. There exist (uniformly) recurrent sets which do not have this property (see [**?**, Example 3.17]).

5 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.

Proposition 5. *For every $n \geq 1$ and $m \geq 0$, the graph $\mathcal{E}_n(w)$ is a tree for all $w \in S_{\geq m}$ if and only if $\mathcal{E}_{n+1}(w)$ is a tree for all words $w \in S_{\geq m}$.*

To prove Proposition 5 we need some preliminary result as well as the following notions. Let S be a factorial extendable set of words over an alphabet A. It is not difficult to show that given an S-maximal prefix code (resp. a S-maximal suffix code) $U \subset S$, every word of S either has a prefix in U or is a prefix of a word of U.

For $U, V \subset A^*$ and $w \in S$, let $L_U(w) = \{u \in U \mid uw \in S\}$ and $R_V(w) = \{v \in V \mid wv \in S\}$.

Let $U \subset A^*$ (resp. $V \subset A^*$) be a suffix code (resp. prefix code) and $w \in S$ be such that $L_U(w)$ is an S-maximal suffix code (resp. $R_V(w)$ is an S-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 ices 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 $uwv \in S$. In particular $\mathcal{E}_n(w) = \mathcal{E}_{S_n,S_n}(w)$. The only if part is [**?**, Lemmas 3.8 and 3.10].

Lemma 1. *Let S be a factorial extendable set and let $w \in S$. Let $U \subset S$ be a finite S-maximal suffix code and let $V \subset S$ be finite S-maximal prefix code. Let $\ell \in S$ be such that $A\ell \cap S \subset U$ and such that $\mathcal{E}_{A,V}(\ell w)$ is a tree. Set $U' = (U \setminus A\ell) \cup \{\ell\}$. 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 by π' a path passing by the same vertices than π. 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, 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^* pass 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 S 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 holds for $r \in S$ such that $rA \cap S \subset V$ and $\mathcal{E}_{U,A}(wr)$ is a tree, with $V' = (V \setminus rA) \cup \{r\}$: the graph $\mathcal{E}_{U,V}(w)$ is a tree if and only if $\mathcal{E}_{U,V'}(w)$ is a tree.

Lemma 2. *Let $n \geq 1$, let $m \geq 0$ and let V be a finite S-maximal prefix code. If $\mathcal{E}_{S_n,V}(w)$ is a tree for every $w \in S_{\geq m}$ then for each word $u \in S_{\geq n+m-1}$, the graph $\mathcal{E}_{A,V}(u)$ is a tree.*

Proof. The graph $\mathcal{E}_{A,V}(u)$ is obtained from $\mathcal{E}_{S_n,V}(u)$ by identifying the vertices of $L_n(u)$ ending with the same letter. Since $\mathcal{E}_{S_n,V}(u)$ is connected, $\mathcal{E}_{A,V}(u)$ is also connected. Set $u = \ell u'$ with $|\ell| = n-1$. The graph $\mathcal{E}_{A,V}(u)$ is isomorphic to $\mathcal{E}_{A\ell,V}(u')$ which is a subgraph of $\mathcal{E}_n(u')$ and thus it is acyclic.

A symmetric statement holds for $n \geq 1$ and U a finite S-maximal suffix code: $\mathcal{E}_{U,S_n}(w)$ is a tree for every $w \in S_{\geq m}$ if and only if $\mathcal{E}_{U,A}(wr)$ is a tree for every $r \in S_{\geq n-1}$ and $w \in S_{\geq m}$.

Proof (of Proposition 5). Assume first that $\mathcal{E}_n(w)$ is tree for every word $w \in S_{\geq m}$. We fix some $w \in S_{\geq m}$. We claim that for any finite S-maximal suffix code U formed of words of length n or $n+1$, the graph $\mathcal{E}_{U,S_n}(w)$ is a tree. The proof is done 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 $\mathcal{E}_{U,S_n}(w) = \mathcal{E}_n(w)$. 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 S-maximal suffix code with words of length n or $n+1$, we have $A\ell \cap S \subset U$. Let us consider $U' = (U \setminus A\ell) \cup \{\ell\}$. Since $\gamma_{n+1}(U') < \gamma_{n+1}(U)$, by induction hypothesis the graph $\mathcal{E}_{U',S_n}(w)$ is a tree. Moreover, by Lemma 2, the graph $\mathcal{E}_{A,S_n}(\ell w)$ is a tree. Thus, by assertion 1 of Lemma 1, the graph $\mathcal{E}_{U,S_n}(w)$ is a tree. This proves the claim. We now claim that for any finite X-maximal prefix code V formed of words of length n or $n+1$, the graph $\mathcal{E}_{S_{n+1},V}(w)$ is a tree by 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 $\mathcal{E}_{S_{n+1},V}(w) = \mathcal{E}_{S_{n+1},S_n}(w)$, 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 $rA \cap S \subset U$. Let us consider $V' = (V \setminus rA) \cup \{r\}$. Since $\delta_{n+1}(V') < \delta_{n+1}(V)$, by induction hypothesis the graph $\mathcal{E}_{S_{n+1},V'}(w)$ is a tree. Moreover, by the symmetric version of Lemma 2, the graph $\mathcal{E}_{S_{n+1},A}(wr)$ is a tree. This proves the claim. Since $\mathcal{E}_{n+1}(w) = \mathcal{E}_{S_{n+1},S_{n+1}}(w)$, we conclude that $\mathcal{E}_{n+1}(w)$ is a tree.

Assume now that $\mathcal{E}_{n+1}(w)$ is a tree for every $w \in S_{\geq m}$. Fix some $w \in S_{\geq m}$. We first claim that $\mathcal{E}_{U,S_{n+1}}(w)$ 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 $\mathcal{E}_{U,S_{n+1}}(w) = \mathcal{E}_{n+1}(w)$. Assume next that $\gamma_n(U) > 0$. Let $\ell \in L_U(w) \cap A^n$. Set $W = (U \setminus \{\ell\}) \cup A\ell$ or equivalently $U = (W \setminus A\ell) \cup \{\ell\}$. Then $\delta_n(W) < \delta_n(U)$ and consequently $\mathcal{E}_{W,S_{n+1}}(w)$ is a tree by induction hypothesis. On the other hand, by Lemma 2, the graph $\mathcal{E}_{A,S_{n+1}}(\ell w)$ is also a tree. By Assertion 2 of Lemma 1, the graph $\mathcal{E}_{U,S_{n+1}}(w)$ is a tree and thus the claim is proved. We now claim that $\mathcal{E}_{S_n,V}(w)$ is a tree for every S-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 $T = (V \setminus \{r\}) \cup rA$ or equivalently $V = (T \setminus rA) \cup \{r\}$. Then $\delta_n(T) < \delta_n(V)$ and thus $\mathcal{E}_{S_n,T}(w)$ is a tree by induction hypothesis. On the other hand, by the symmetric version of Lemma 2, the graph $\mathcal{E}_{S_n,A}(wr)$ is also a tree. By Assertion 2 of Lemma 1, the graph $\mathcal{E}_{S_n,T}(w)$ is a tree and thus the claim is proved. Since $\mathcal{E}_n(w) = \mathcal{E}_{U,V}(w)$ for $U = V = S_n$, it follows from the claim that $\mathcal{E}_n(w)$ is a tree.

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

Theorem 3. *Let S be a factorial extendable set. For every $m \geq 1$, the following conditions are equivalent.*

(i) *S is eventually dendric with threshold m,*
(ii) *the graph $\mathcal{E}_n(w)$ is a tree for every $n \geq 1$ and every word $w \in S_{\geq m}$,*
(iii) *there is an integer $n \geq 1$ such that $\mathcal{E}_n(w)$ is a tree for every word $w \in S_{\geq m}$.*

Proof. (i) $\Rightarrow$ (ii) and (iii) $\Rightarrow$ (i) follows from Proposition 5 using respectively ascending and descending induction on n. Finally, (ii) clearly implies $\Rightarrow$ (iii).

6 Complete bifix decoding

Let S be a factorial extendable set of words over an alphabet A. A set $U \subset S$ is said to be *right S-complete* (resp. *left S-complete*) if any long enough word of S has a prefix (resp. suffix) in U. It is *two-sided S-complete* if it is both left and right S-complete. A *bifix code* is a set of words that is both a prefix code and a suffix code. Similary to what seen in Section 5, we say that a bifix code $U \subset S$ is S-maximal if it is not properly contained in a bifix code $V \subset S$. If a bifix code $U \subset S$ is right S-complete (resp. left S-complete), it is an S-maximal bifix code since it is already an S-maximal prefix code (resp. suffix code). It can be proved conversely that if S is recurrent, a finite bifix code is S-maximal if and only if it is two-sided S-complete (see [**?**, Theorem 4.2.2]). This is not true in general, as shown by the following example.

Example 7. Let $S = a^*b^*$. The set $U = \{aa, b\}$ is an S-maximal bifix code. Indeed, it is a bifix code and it is left S-complete as one may verify. However it is not right S-complete since no word in ab^* has a prefix in U.

Let $S \subset A^*$ be a factorial extendable set and let U be a two-sided S-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 $\varphi^{-1}(S)$ is factorial and, since U is two-sided complete, it is extendable. $\varphi^{-1}(S)$ is called the *complete bifix decoding* of S with respect to U. For example, for any $n \geq 1$, the set S_n is a two-sided complete bifix code and the corresponding complete bifix decoding is the decoding of S by n-blocks. In [**?**, Theorem 3.13] it is proved that the maximal bifix decoding of a recurrent dendric set is a dendric set. Actually, the hypothesis that S is recurrent is only used to guarantee that the S-maximal bifix code used for the decoding is also an S-maximal prefix code and an S-maximal suffix code. In the definition used here of complete bifix decoding, we do not need this hypothesis. Note, however, that when S is recurrent, the two notions of complete and maximal bifix decoding coincide.

Theorem 4. *Any complete bifix decoding of an eventually dendric set is an eventually dendric set having the same threshold.*

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

Lemma 3. *A set S is an eventually dendric set with threshold n if and only if for any $w \in S_{\geq n}$, for any S-maximal suffix code U and for any S-maximal prefix code V, the graph $\mathcal{E}_{U,V}(w)$ is a tree.*

Proof. The "if" part is trivial. To prove the other direction, 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 $2\,\mathrm{Card}(X \cap A)$. Indeed, for every $w \in S_{\geq n}$ one has $U = L(w)$ and $V = R(w)$ and thus $\mathcal{E}_{U,V}(w) = \mathcal{E}_1(w)$ is a tree. Otherwise, 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 S-maximal suffix code, we have $A\ell \cap S \subset U$. Set $U' = (U \setminus A\ell) \cup \{\ell\}$. By induction hypothesis, both $\mathcal{E}_{U',V}(w)$ and $\mathcal{E}_{A,V}(\ell w)$ are trees. Thus, by Lemma 1, $\mathcal{E}_{U,V}(w)$ is also a tree.

Proof (of Theorem 4). Assume that S is eventually dendric with threshold n. Let $\varphi : B \to U$ be a coding morphism for U and let T be the decoding of S corresponding to U. Consider a word w of T of length at least n. By Lemma 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 $bwc \in T$ if and only if $\varphi(bwc) \in S$, 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 T is eventually dendric with threshold n.

Example 8. Let S be the Fibonacci set. Then $U = \{aa, aba, b\}$ is an S-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 T of S with respect to U is a purely dendric set. It is actually the natural coding of an interval exchange transformation on three intervals (see [**?**]). The extension graphs $\mathcal{E}_1(\varepsilon, T)$ and $\mathcal{E}_1(v, T)$ are shown in Figure 4.

Fig. 4. The graphs $\mathcal{E}_1(\varepsilon, T)$ and $\mathcal{E}_1(w, T)$.

A particular case of complete bifix decoding is related to a notion which is well-known in topological dynamics, namely the skew product of two dynamical systems (see [**?**]). Indeed, assume that when we start with a sfhit X, a permutation group G on a set Q and a morphism $f : A^* \to G$. We denote $q \mapsto q \cdot w$ the result of the action of the permutation $f(w)$ on the point $q \in Q$. 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 (this follows from [**?**, Theorem 4.2.11]). The corresponding decoding is a shift space which is related to the skew product of S and (G, Q). It is the shift space Y on the alphabet $A \times Q$ formed by the labels of the two-sided infinite paths on the graph with vertices Q and edges (p, q) labeled (a, p) for $a \in A$ such that $p \cdot f(a) = q$. The decoding of X corresponding to U is the dynamical system induced by Y on the set of $y \in Y$ such that $y_0 = (a, i)$ for some $a \in A$.

Example 9. 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 8.

7 Conclusion

We have seen that the class of eventually dendric shifts is closed under complete bifix decoding. It is also known to be closed under conjugacy (see [**?**]), and thus it has strong closure properties. It would be interesting to know how properties which are known to hold for dendric sets (or language of dendric shifts) extend to this more general class. For instance, to which extent the properties of return words proved for recurrent dendric sets extend to eventually dendric ones? More precisely, what can we say about the subgroup of the free group generated by return words to a given word? In [**?**] it is proved that for recurrent dendric sets, every set of return words to a fixed word is a basis of the free group, while in the case of specular sets, the set of return words to a fixed word is a basis of a particular subgroup called the even subgroup (see [**?**]). Also, is there a finite S-adic representation for all recurrent eventually dendric sets? There is one for recurrent dendric sets [**?**].