Maximal Bifix Decoding of Recurrent Sets

João Pedro Marques Pires
Mestrado em Matemática
Departamento de Matemática
2018

Orientador
Jorge Manuel Meneses Guimarães de Almeida,
Professor Catedrático,
Faculdade de Ciências da Universidade do Porto

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ______/______/________

MSc

Acknowledgments

.

I would like to express my deep gratitude and appreciation to my parents and brother for their constant love and support throughout my entire life and also to thank my parents for providing me the opportunity to further my education and my dreams. To all my friends for their encouragement, for the several moments that made my life much richer and for liking me as the person that I am.

I am also greatfull to my friends Raquel and Inês for the big help and support they gave me during the course. I also wish to thank my friend Margarida for helping and saving me in a peculiar moment that my TeXstudio and MiKTeX decided to betray me...

I express my sincere thanks to Professor Jorge Almeida for his availability in helping me in several moments, also by his patience and support, including extended discussions and valuable suggestions which have contributed greatly to the improvement of my work.

I also express my sincere thanks to Professor and Director of the Course Jorge Rocha for his attention and support given in several moments, and also for the patience to listen and answer all my *infinite* questions during these last two years.

Resumo

Dados uma palavra w e um conjunto de palavras S, denotamos por $e(w)$ o número de pares de letras (a, b) tais que $awb \in S$, por $l(w)$ o número de letras a tais que $aw \in S$ e por $r(w)$ o número de letras a tais que $wa \in S$. Dizemos que uma palavra w é fraca, neutra ou forte, se $e(w)$ é, respetivamente, menor, igual ou maior do que $l(w)+r(w)-1$. Associamos a cada palavra de S um grafo (a que chamamos de grafo de extensão), também baseada nas possíveis extensões da palavra no conjunto. Se o grafo de extensão de qualquer palavra é acíclico ou uma árvore, então dizemos que o conjunto é, respetivamente, acíclico ou uma árvore. O principal resultado deste trabalho é uma prova, que é nova e feita de forma independente, de que um conjunto recorrente que tem apenas um número finito de palavras fortes (que inclui os conjuntos que são acíclicos e recorrentes) é na verdade uniformemente recorrente (este resultado foi também recentemente apresentado em Dolce and Perrin (2018)). Aplicamos este resultado para obter propriedades de descodificações de códigos bifixos maximais de conjuntos acíclicos e resolvemos um problema levantado em Berthé et al. (2015a), que questiona se a descodificação de um código bifixo maximal de um conjunto que é uma árvore e recorrente é um conjunto que também é uma árvore e recorrente.

Abstract

Given a word w of a certain set of words S, we denote by $e(w)$ the number of pairs of letters (a, b) such that $awb \in S$, by $l(w)$ the number of letters a such that $aw \in S$ and by $r(w)$ the number of letters a such that $wa \in S$. We say that a word w is weak, neutral or strong, if $e(w)$ is, respectively, less, equal or bigger than $l(w) + r(w) - 1$. We also associate to each word of S a graph (which is called the extension graph), based also on the possible extensions of the word in the set. If the extension graph of every word is acyclic or a tree, then we say that the set is, respectively, acyclic or a tree. Our main result is a new independent proof that a recurrent set S with only a finite number of strong words (which includes recurrent acyclic sets), is in fact uniformly recurrent (this result has also recently been presented in Dolce and Perrin (2018)). We apply this result to study properties of maximal bifix decoding of acyclic sets and to solve a question raised in Berthé et al. (2015a), where it is asked if the maximal bifix decoding of a recurrent tree set is a recurrent tree set.

Contents

Chapter 1

Introduction

The mathematician Alex Thue, at the beginning of the 20th century, dealt with many combinatorial problems which emerge from the study of sequences of symbols, which we call words. In Berstel (1995) one finds a translation of his work where it is mentioned that he approached problems on words that were directly connected with finitely presented semigroups. Although, his work only became noticed much later, somewhere between the decades of 1970 and 1980. In the 1960's Schützenberger also obtained several results on words related with problems concerning free monoids and free groups (see for example Lyndon and Schützenberger (1962) or Lentin and Schützenberger (1967)). Since then, the interest and research in combinatorics on words has been increasing. Beyond algebra, other fields are also connected to combinatorics on words. In the case of dynamical systems, we can study a smooth dynamical system by analysing a space consisting of infinite words, where each word corresponds to the whole history of the system (an orbit) and each letter of a word corresponds to a state of the system. The study of smooth dynamical systems, by this type of modeling, is made in symbolic dynamics (see §6.5 and §13.6 of Lind and Marcus (1995)).

The theory of codes is an area that emerged from work of Claude Shannon, see Shannon (1948). A code, in a mathematical perspective, is a set of words X where any element of X^* (the monoid generated by X) can be written uniquely as a product of words in X. As it can be seen easily from the definition of a code, problems in this subject are directly related with the field of combinatorics on words. In practice, this notion of code is useful to encode information, i.e., to convert information from a source into symbols, and also to reverse the process of encoding, i.e., to decode encoded information (see for example Shannon (1948)). One type of codes, that have been profoundly studied and which are also considered in this work, are the bifix codes. Several properties of bifix codes can be found in Berstel et al. (2010). In this context, there is a huge interest in bifix codes, since it is noticed a decreasing of the impact of possible errors in the transmission of information by the use of this kind of codes (Gilbert and Moore (1959), Schützenberger (1956)).

In what concerns combinatorics on words, when we analyse an infinite word, what we want is to understand and describe the *behaviour* of that object. To do so, we usually look at the set of its factors, that is, the finite words that can be found *inside* the infinite word. On a set of finite words, it matters to observe the possible extensions of the words with respect to the set, and the frequency with which its elements occur in the infinite word. Given a word w of a certain set of words S, we denote by $e(w)$ the number of pairs of letters (a, b) such that $awb \in S$, by $l(w)$ the number of letters a such that $aw \in S$ and by $r(w)$ the number of letters a such that $wa \in S$. We say that a word w is weak, neutral or strong, if $e(w)$ is, respectively, less, equal or bigger than $l(w)+r(w)-1$. If every factor of an infinite word occurs infinitely often, we say that the word is recurrent. If every factor appears in all factors of sufficiently large length, we say that the infinite word is uniformly recurrent. The authors of Berthé et al. (2015a) associate with each word of a given set of words a graph (which is called the extension graph) that is constructed based on the possible two-sided extensions of that word in the set. Depending on the properties of the extension graph of every word of the set, we call the set acyclic, connected or tree set.

In this work we study properties of a certain type of recurrent sets and whose theory may be considered part of combinatorics on words, symbolic dynamics and theory of codes. More precisely, given a set S of words such that S is recurrent and contains only a finite number of strong words (conditions that are satisfied by recurrent acyclic sets), we show that this set S is in fact uniformly recurrent (Theorem 3.1). This result has also recently been presented in Dolce and Perrin (2018). Furthermore, with this last result, we answer a question raised in Berthé et al. (2015a), which asks if the maximal bifix decoding of a recurrent tree set is a recurrent tree set, by reducing this problem to showing that the maximal bifix decoding of a uniformly recurrent tree set is a uniformly recurrent tree set, which is already proved in Berthé et al. (2015b); in fact, our methods allows us to give an alternative proof of that result. We also show that the maximal bifix decoding of a recurrent acyclic set is a finite union of uniformly recurrent acyclic sets (Theorem 4.11).

In Section 2, we introduce the basic definitions, concerning recurrent sets, bifix codes and acyclic sets.

In Section 3, we work with sets of words that contain only a finite number of strong words. We start by showing that this kind of sets only have a finite number of non-neutral words. With this property, we observe that right special words of large enough length can be prolonged on the left to right special words. Also, we prove that the number of right special words of a certain length n, as a function of n, is constant for values of n sufficiently large. These observations are essential to prove Theorem 3.1, since the idea of the proof is to analyse the gaps between a certain type of right special words. Theorem 3.1 allows us to reformulate Theorem 3.2 to Theorem 3.3, which is already proved in Berthé et al. (2015b).

In Section 4 we show that the maximal bifix decoding of recurrent acyclic sets is a finite union of factors of certain infinite words. Then we show that these infinite words are in fact uniformly recurrent. Actually, we see that the number of these uniformly recurrent infinite words corresponds to the number of connected components of the Rauzy graph of large enough order (Theorem 4.11). Finally, we give an alternative prove of Theorem 3.3 (is the main result in Berthé et al. (2015b)).

Chapter 2
Preliminaries

2.1 Recurrent sets

We consider $\mathbb{N}$ as the set of positive numbers. Let A be a finite alphabet, A^* the monoid of all finite words on A, A^+ the semigroup of finite non empty words and $A^{\mathbb{N}}$ the set of infinite words on A. We denote by ϵ the empty word. Let $X \subseteq A^*$. We define X^* as the submonoid generated by X in A^*. We define X^ω as the set of infinite words $Y \in A^{\mathbb{N}}$, such that Y can be written as an infinite sequence of words on X. Given a finite word u on A, we denote by u^ω the only element of the set $\{u\}^\omega$ and every infinite word of this kind is said to be *periodic*. An infinite word is *ultimately periodic* if it is a concatenation of a finite word with a periodic word. Given a (finite or infinite) word w on A, we denote by w_n the letter that is in the position n in the word w, and if w is a finite word we denote by $|w|$ the length of w. Given words w and v on A, we say that v is a *factor* of w if v is a finite word and if exist some $N \in \mathbb{N}$ such that $w_{i+N} = v_i$ for all $i \in \{1, \ldots, |v|\}$. We denote by $\text{fact}(w)$ the set of factors of w. Notice that ϵ is a factor of every word on A. A set $S \subseteq A^*$ is said to be *periodic*, if it is the set of factors of a periodic word. In the same way we define an *ultimately periodic set*.

A set of words is said to be *factorial* if it contains the factors of its elements. Let $S \subseteq A^*$ and $w \in S$. We denote

$$L(w) = \{a \in A \mid aw \in S\}$$

$$R(w) = \{a \in A \mid wa \in S\}$$

$$E(w) = \{(a,b) \in A \mid awb \in S\}$$

and $l(w) = \text{Card}(L(w))$, $r(w) = \text{Card}(R(w))$, $e(w) = \text{Card}(E(w))$. We say that $l(w)$ and $r(w)$ are, respectively, the *left* and *right order* of w. We say that w is *weak* if $e(w) < l(w)+r(w)-1$, *neutral* if $e(w) = l(w)+r(w)-1$ and *strong* if $e(w) > l(w)+r(w)-1$.

In this paper, if w is not a strong word, we will say that w is a *non-strong* word. A set $S \subseteq A^*$ is *biextendable* if S is factorial and if every word of S can be extended on the right and on the left to a word in S. The set S is *recurrent* if it is factorial and if for every $u, w \in S$ there is a $v \in S$ such that $uvw \in S$. The set S is *uniformly recurrent* if it is factorial, if any word can be extended on the right to a word in S, and if for every word $u \in S$ there exists an integer $n \geq 1$ such that u is a factor of every word of S of length n. If Y is an infinite word such that $\text{fact}(Y)$ is a recurrent set, then we say that Y is recurrent (analogously, we define a uniformly recurrent word). A uniformly recurrent set is a recurrent set. If S is a recurrent set and $w \in S$, then there exists an infinite word Y such that $\text{fact}(Y) = S$ and $Y_1 \ldots Y_{|w|} = w$ (Proposition 2.2.1 in Berstel et al. (2012)). Notice that a recurrent ultimately periodic set is a periodic set. A word $z \in S$ is *right special* in S if there exist at least two letters $a, b \in A$ such that $za, zb \in S$. We denote by $RS(S)$ the set of right special elements of S.

Let $S \subseteq A^*$ be a factorial set. The Rauzy graph of S of order n is the following labeled graph $G_n(S)$. Its vertices are the words in the set $S \cap A^n$. Its edges are the triples (x, a, y) for all $x, y \in S \cap A^n$ and $a \in A$ such that $xa \in S \cap Ay$. If S is recurrent, then $G_n(S)$ is strongly connected. If S is a union of recurrent sets, then the connected components of $G_n(S)$ are strongly connected.

2.2 Bifix codes

A *prefix code* is a set of nonempty words which does not contain any proper prefix of its elements. A *suffix code* is defined symmetrically. A *bifix code* is a set which is both a prefix and a suffix code. Given $S \subseteq A^*$, a prefix code $X \subseteq S$ is *S-maximal* if it is not properly contained in any prefix code $P \subseteq S$. In the same way we define a *S-maximal suffix code* and an *S-maximal bifix code*. If S is a recurrent set, we have the following result (Theorem 4.2.2 in Berstel et al. (2012)):

Proposition 2.1. *Let S be a recurrent set and let $X \subseteq S$ a finite set. Then the following conditions are equivalent:*

1. *X is an S-maximal bifix code.*
2. *X is an S-maximal prefix code and an S-maximal suffix code.*

A *coding morphism* for a finite prefix code $X \subseteq A^*$ is a monoid morphism $f : B^* \to A^*$ which maps bijectively B onto X (B is a finite alphabet). Let $S \subseteq A^*$ be a factorial set,

$X \subseteq S$ a finite bifix code and f a coding morphism for X. The set $f^{-1}(S)$ is called a *bifix decoding* of S. When X is an S-maximal bifix code, $f^{-1}(S)$ is called a *maximal bifix decoding* of S.

2.3 Acyclic and tree sets

Let $S \subseteq A^*$ and let $U, V \subseteq S$. For $w \in S$, let $U_l(w) = \{l \in U \mid lw \in S\}$ and $V_r(w) = \{r \in V \mid wr \in S\}$. The *generalized extension graph* of w is the following undirected graph $E_{U,V}(w)$. The set of vertices is made of two disjoint copies of $U_l(w)$ and $V_r(w)$. The edges are the pairs (l, r) for $l \in U_l(w)$ and $r \in V_r(w)$ such that $lwr \in S$. The graph $E_{A,A}$ is called the *extension graph* of w. We say that S is an *acyclic set* if it is biextendable (notice that the definition of biextandable requires the set to be factorial) and if for every word $w \in S$, the graph $E_{A,A}(w)$ is acyclic (in the same way we define a *tree set*). The following results are respectively the Proposition 3.7, Proposition 3.9, Theorem 3.11 and Theorem 3.13 of Berthé et al. (2015a):

Proposition 2.2. *Let S be an acyclic set. For any $w \in S$, any finite suffix code U and any finite prefix code V, the generalized extension graph $E_{U,V}(w)$ is acyclic.*

Proposition 2.3. *Let S be a tree set. For any $w \in S$, any finite S-maximal suffix code U and any finite S-maximal prefix code V, the generalized extension graph $E_{U,V}(w)$ is a tree.*

Theorem 2.4. *Any biextendable set which is the bifix decoding of an acyclic set is acyclic.*

Theorem 2.5. *Any maximal bifix decoding of a recurrent tree set is a tree set.*

Chapter 3

Recurrent almost-neutral sets

Let $S \subseteq A^*$ be a biextendable set such that there exists $K \in \mathbb{N}_0$ such that for all words $w \in S$ with length at least K, w is a non-strong word. Then we say S is an *almost-neutral set*. It is easy to see that this class of sets includes all acyclic sets (see Lemma 3.6). The point of this section is to prove the following two theorems:

Theorem 3.1. *If S is a recurrent almost-neutral set then it is uniformly recurrent.*

Theorem 3.2. *The maximal bifix decoding of a recurrent tree set is a recurrent tree set.*

Theorem 3.1 has also recently been presented in Dolce and Perrin (2018). Theorem 3.2 is a direct answer to a question raised in Berthé et al. (2015a). It is easy to see that a tree set is an almost-neutral set (see Lemma 3.6), and so, by Theorem 3.1, a recurrent tree set is in fact a uniformly recurrent tree set. So we can reformulate Theorem 3.2 as follows:

Theorem 3.3. *The maximal bifix decoding of a uniformly recurrent tree set is a uniformly recurrent tree set.*

Furthermore, Theorem 3.3 was already proved in Berthé et al. (2015b), which means that proving Theorem 3.1 allows us to get Theorem 3.2. To prove Theorem 3.1, we need to introduce more terminology and prove other results. First we will prove some properties of almost-neutral sets.

Proposition 3.4. *Let $S \subseteq A^*$ be an almost-neutral set. Then there exists $N \in \mathbb{N}_0$ such that, for all words $w \in S$ with length at least N, w is neutral.*

Proof. Consider the function $g : S \to \mathbb{N}_0$, defined by $g(w) = e(w) - l(w)$ for $w \in S$. Notice that, since S is biextendable, $g(w) \geq 0$. Consider $K \in \mathbb{N}_0$ such that for all words $w \in S$ with length at least K, w is a non-strong word. Suppose w is a word of S with length at least K; Then $l(w) + r(w) - 1 \geq e(w)$. So we have:

$$\begin{aligned} g(w) = e(w) - l(w) = \sum_{a \in L(w)} \big(r(aw) - 1\big) &= \sum_{a \in L(w)} \big(l(aw) + r(aw) - 1 - l(aw)\big) \\ &\geq \sum_{a \in L(w)} \big(e(aw) - l(aw)\big) \\ &= \sum_{a \in L(w)} g(aw) \end{aligned}$$

Notice that the last inequality is an equality if only if $e(aw) = l(aw) + r(aw) - 1$ for all $a \in L(w)$. Let $L_n(w)$ be the set of words z of length n such that $zw \in S$. Applying several times the last inequality we can get $\sum_{z \in L_n(w)} g(zw) \geq \sum_{y \in L_m(w)} g(yw)$, for $m > n$, and it is an equality if and only if $e(tw) = l(tw) + r(tw) - 1$, for all $t \in S$ such that $n < |t| \leq m$ and such that $tw \in S$. Since g is never negative, given $w \in S$, there must exist $N_w \in \mathbb{N}_0$ such that for all $M \in \mathbb{N}$ with $M > N_w$, $\sum_{z \in L_{N_w}(w)} g(zw) = \sum_{y \in L_M(w)} g(yw)$. As we have seen, this equality implies $e(yw) = l(yw) + r(yw) - 1$ for $y \in L_M(w)$. Making $N' = \max_{u \in A^K \cap S}\{N_u\}$ and $N = N' + K + 1$, we conclude that, for every word $z \in S$ with length at least N, z is neutral. □

Corollary 3.5. *Let $S \subseteq A^*$ be an almost-neutral set and the function $g : S \to \mathbb{N}_0$ be defined by $g(w) = e(w) - l(w)$ for $w \in S$. Consider a number N according to Proposition 3.4. Then for all $M \in \mathbb{N}_0$ such that $M \geq N$, we have $\sum_{w \in A^N \cap S} g(w) = \sum_{w \in A^M \cap S} g(w)$.*

Proof. Let $L_n(w)$ be the set of words z of length n such that $zw \in S$. For all $v \in S$ with $|v| \geq N$, v is neutral (by the hypothesis of the corollary). So, from the proof of Proposition 3.4, we have that for all $w \in A^N \cap S$ and for all $n \in \mathbb{N}_0$, $g(w) = \sum_{z \in L_n(w)} g(zw)$. Then, for all $n \in \mathbb{N}_0$, we have:

$$\sum_{w \in A^N \cap S} g(w) = \sum_{w \in A^N \cap S} \left(\sum_{z \in L_n(w)} g(zw) \right) = \sum_{w \in A^{n+N} \cap S} g(w).$$

□

Notice that, given any biextendable set S and any word w of S, the number of edges of $E_{A,A}(w)$ is $e(w)$ and the number of vertices of that graph is $l(w) + r(w)$. So, consider a biextendable set $S \subseteq A^*$ such that there exists $K \in \mathbb{N}_0$ such that for all words $w \in S$ with length at least K, $E_{A,A}(w)$ is an acyclic graph. Then, for any $w \in S$ such that $|w| \geq K$, we must have $e(w) \leq l(w) + r(w) - 1$ (because the graph $E_{A,A}(w)$ is acyclic) and so w is a non-strong word. This means that this set S is almost-neutral. So we have proven the following Lemma:

Lemma 3.6. *Let $S \subseteq A^*$ biextendable set such that there exists $K \in \mathbb{N}_0$ such that for all words $w \in S$ with length at least K, $E_{A,A}(w)$ is an acyclic graph. Then S is an almost-neutral set.*

The last lemma allows us to show the following Corollaries of Proposition 3.4:

Corollary 3.7. *Let $S \subseteq A^*$ be a biextendable set such that there exists $K \in \mathbb{N}_0$ such that for all words $w \in S$ with length at least K, $E_{A,A}(w)$ is an acyclic graph. Then there exists $N \in \mathbb{N}_0$ such that, for all words $w \in S$ with length at least N, $E_{A,A}(w)$ is a tree.*

Proof. Consider $K \in \mathbb{N}_0$ such that for all words $w \in S$ with length at least K, $E_{A,A}(w)$ is an acyclic graph. Since S is almost-neutral by Lemma 3.6, we can choose a number N according to Proposition 3.4, and make $N' = \max\{K, N\}$. Suppose w is a word of S with length at least N'; then the graph $E_{A,A}(w)$ is acyclic and it has $l(w) + r(w) - 1$ edges. Then this graph must be a tree. □

Corollary 3.8. *Let $S \subseteq A^*$ be an acyclic set. Then there exists $N \in \mathbb{N}_0$ such that, for all words $w \in S$ with length at least N, $E_{A,A}(w)$ is a tree.*

Proof. This corollary is clearly a particular case of Corollary 3.7. □

In what follows, we focus on and analyse the right special words of almost-neutral sets. The results that we prove about right special words are essential to prove the Theorem 3.1. Next lemma gives us a sufficient condition for an occurrence of a right special word to be prolongable on the left to another right special word.

Lemma 3.9. *Let S be a biextendable set. If $w \in S$ is a right special word of S such that w is neutral, then there exists $a \in A$ such that aw is a right special word.*

Proof. Consider a word $w \in S$ in the conditions of the statement of the lemma. Since w is neutral, $E_{A,A}(w)$ must have $l(w) + r(w) - 1$ edges. Since w is right special, the graph has at least two vertices in $R(w)$. Then the number of edges of the graph must be at least $l(w) + 1$. This means that must exist some vertex in $L(w)$ that has degree at least 2. If $a \in A$ is the letter that is related with that vertex of degree bigger than 1, then aw is right special. □

Let S be a recurrent almost-neutral set. If S is the set of factors of a periodic word then Theorem 3.1 is obviously true. So we assume that S is not the set of factors of a periodic word. Given $w \in S$, define $RS_w^k(S)$ to be the set of words of $RS(S)$ that have length k and that have w as a factor. Given $w \in S$, consider the function $g_w : \mathbb{N}_0 \to \mathbb{N}_0$ defined by $g_w(k) = |RS_w^k(S)|$. For this set S, fix a number N according to Proposition 3.4. The lemma below describes the behaviour of this function. More precisely, the lemma tells us that in fact the number of right special words of a certain size, that contains a word w, is constant for a large enough length.

Lemma 3.10. *If $m > n \geq N$, then $g_w(m) \geq g_w(n)$. Also, the function g_w is bounded and there exists $p \in \mathbb{N}$ such that $g_w(p) \neq 0$.*

Proof. By Lemma 3.9, every right special word with length at least N can be extended on the left to a right special word. So, if $m > n \geq N$, then $g_w(m) \geq g_w(n)$.

Since $RS_w^k(S) \subseteq RS_\epsilon^k(S)$, we just need to prove that the function g_ϵ is bounded. Consider the function $g : S \to \mathbb{N}_0$, defined by $g(w) = e(w) - l(w)$ for $w \in S$.

Then, given $M \in \mathbb{N}$ such that $M \geq N$, and using Corollary 3.5, we have that:

$$\begin{aligned}
|A^{M+1} \cap S| - |A^M \cap S| &= \sum_{w \in A^M \cap S} (r(w) - 1) \\
&= \sum_{w \in A^M \cap S} (e(w) - l(w)) \\
&= \sum_{w \in A^M \cap S} g(w) \\
&= \sum_{w \in A^N \cap S} g(w) \\
&= |A^{N+1} \cap S| - |A^N \cap S|
\end{aligned}$$

Since $|RS_\epsilon^k(S)| \leq |A^{k+1} \cap S| - |A^k \cap S|$ for all $k \in \mathbb{N}_0$, then

$$\begin{aligned}
g_\epsilon(M) = |RS_\epsilon^M(S)| &\leq |A^{M+1} \cap S| - |A^M \cap S| \\
&= |A^{N+1} \cap S| - |A^N \cap S|
\end{aligned}$$

for all $M \in \mathbb{N}_0$ with $M \geq N$. Hence, g_ϵ is bounded.

Obviously, g_ϵ cannot be the zero function. Suppose now that $w \neq \epsilon$ and assume that g_w is the constant function zero. That means that every word of S that has w as a factor is not right special. Since S is recurrent, we can consider an infinite word Y such that fact$(Y) = S$ and such that w is a prefix of Y. Because S is recurrent, there must exist a prefix of Y of the form wtw. Since any word of S that has w as a factor cannot be right special, Y must be the periodic word $(wt)^\omega$, which is absurd. Then g_w cannot be the zero function for $w \neq \epsilon$. □

As we pointed out before, the last lemma implies that g_w is constant for sufficiently large numbers. So we can choose for every word $w \in S \setminus \{\epsilon\}$ a positive number $k_w \in \mathbb{N}$ such that $k_w \geq N$ and $g_w(k_w) = \max\Big(g_w\big(\mathbb{N}_0 \setminus \{0, \dots, N-1\}\big)\Big)$. Notice that, given $w \in S \setminus \{\epsilon\}$, if $u \in S$ is a word of length at least N that contains w as a factor, then g_u is not the zero constant function by Lemma 3.10. This implies that $g_w(k_w) > 0$. Define a set $J_w \subseteq S$ such that an element $z \in S$ belongs to J_w if the following conditions are satisfied:

1. $|z| > k_w$;
2. $z_1 \cdots z_{k_w}$ is an element of $RS_w^{k_w}(S)$ and is the only proper prefix of z that is right special in S;
3. z is right special in S.

Lemma 3.11. *The set J_w is nonempty and finite. More precisely, $|J_w|$ is the number of words of length $k_w + 1$ that have a prefix in $RS_w^{k_w}(S)$.*

Proof. Let $t \in S$ be a word of length $k_w + 1$ that has a prefix in $RS_w^{k_w}(S)$. We are going to prove that must exist a word $z \in J_w$ such that t is a prefix of z. Let us assume otherwise, that is, that there does not exist such a word z. This implies that any word of S that has t as a prefix cannot be a right special word. Since S is recurrent, we can consider an infinite word Y such that fact$(Y) = S$ and such that t is a prefix of Y. Because S is recurrent, there must exist a prefix of Y of the form tut. Since any word of S that has t as a prefix cannot be right special, in particular every prefix with length at least $|t|$ of the word tut cannot be a right special word and so, Y must be the periodic word $(tu)^\omega$, which is absurd. So there must exist a word $z \in J_w$ such that t is a prefix of z. By the definition of J_w, it is easy to see that z is unique.

Conversely, it is obvious that every prefix of length $k_w + 1$ of the words of J_w has a prefix in $RS_w^{k_w}(S)$. Then the lemma is proved. □

Lemma 3.12. *Given $z \in J_w$, if v is the suffix of z of length k_w, then $v \in RS_w^{k_w}(S)$.*

Proof. Choose $z \in J_w$, and let v be the suffix of z of length k_w. Let us assume that v does not belong to $RS_w^{k_w}(S)$. By definition of J_w, z must be right special, and so v must be also right special (because v is a suffix of z). Since we are assuming that v is not in $RS_w^{k_w}(S)$, then v cannot contain w as a factor. By Proposition 3.4 and Lemma 3.9, the words that have length at least N and that are right special can be prolonged on the left to right special words. Then, since $k_w \geq N$, and since v is a suffix of z with length k_w that is right special and that does not contain w as a factor, we must have $g_w(|z|) > g_w(k_w)$, which is an absurd because $g_w(k_w) = \max\Big(g_w\big(\mathbb{N}_0 \setminus \{0,\ldots,N-1\}\big)\Big)$. So v must belong to $RS_w^{k_w}(S)$. □

Finally, we are now in ready to prove the desired result.

Proof of Theorem 3.1: Let $w \in S$. Consider the function $\pi_w : J_w \to S$ that, to each word $z = z_1 \cdots z_{|z|}$ in J_w associates $\pi_w(z) = z_1 \cdots z_{|z|-k_w}$. Since S is recurrent, we can consider an infinite word Y_0 such that $\text{fact}(Y_0) = S$ and such that some prefix of Y_0 is an element of $RS_w^{k_w}(S)$. We are going to prove first that Y_0 is an element of $(\text{Im}(\pi_w))^\omega$. We can write $Y_0 = x_0 Z_0$ such that $x_0 \in RS_w^{k_w}(S)$ and Z_0 is an infinite word. We claim that must exist some prefix v_0 of Z_0 such that $x_0 v_0 \in J_w$. Assume otherwise. Then any prefix of Y_0 with length bigger than $|x_0|$ cannot be a right special word. Because $\text{fact}(Y_0)$ is recurrent, there must exist a prefix of Y_0 of the form $x_0 a u x_0 a$, with $a \in A$ and $u \in S$. Since any prefix of Y_0 that has length bigger than $|x_0|$ cannot be right special, in particular every prefix with length bigger than $|x_0|$ of the word $x_0 a u x_0 a$ cannot be a right special word and so, Y_0 must be the periodic word $(x_0 a u)^\omega$, which is absurd. So there must exist some prefix v_0 of Z_0 such that $x_0 v_0 \in J_w$. Making $z_1 = \pi_w(x_0 v_0)$, we can write $Y_0 = z_1 Y_1$ (where Y_1 is an infinite word). By Lemma 3.12 the suffix of length k_w of any word in J_w is right special and must contain w as a factor, so applying this to $x_0 v_0$, we have that Y_1 has a prefix in $RS_w^{k_w}(S)$. Furthermore, Y_1 is also an infinite word such that $\text{fact}(Y_1) = S$, which means that Y_1 is under the same conditions as Y_0 and so we can repeat this process again (and infinitely). Therefore, Y_0 is an element of $(\text{Im}(\pi_w))^\omega$. Moreover, in this process, we observed that all these words z_i of $(\text{Im}(\pi_w))^\omega$ were preceded by a word of $RS_w^{k_w}(S)$ in that specific occurrence of the z_i in Y_0. Furthermore, $\text{Im}(\pi_w)$ is finite (because J_w is finite), and so, every element of $\text{fact}(Y_0) = S$ with length bigger or equal than $M + k_w - 1$ (where M is the length of the word in $\text{Im}(\pi_w)$ of maximum length) contains some word of $RS_w^{k_w}(S)$ as a factor. Since every element of $RS_w^{k_w}(S)$ contains w as a factor, the theorem is proved. □

Corollary 3.13. *Let $S \subseteq A^*$ be a recurrent set such that there exists $K \in \mathbb{N}_0$ such that for all words $w \in S$ with length at least K, $E_{A,A}(w)$ is an acyclic graph. Then S is uniformly* *recurrent.*

Proof. This corollary is a particular case of Theorem 3.1. □

Corollary 3.14. *If S is a recurrent acyclic set then it is uniformly recurrent.*

Proof. This corollary is a particular case of Corollary 3.13. □

Corollary 3.15. *If U is a recurrent set that is not uniformly recurrent, then for all $N \in \mathbb{N}$, there exists $v, w \in U$ with $|v|, |w| \geq N$, such that v is a strong word and $E_{A,A}(w)$ has a cycle.*

Proof. This is an obvious consequence of Theorem 3.1 and Corollary 3.13. □

Chapter 4

Maximal bifix decoding of acyclic sets

In spite of Theorem 3.3, an analogous theorem for acyclic sets does not hold (see the example below). However, in this section we establish another description for maximal coding morphisms of uniformly recurrent acyclic sets.

Example 1. *Let* $A = \{a, b, c, d\}$*,* $S = \text{fact}\big((abcd)^\omega\big)$*,* $X = \{ab, bc, cd, da\}$*. Is easy to see that* X *is an* S*-maximal bifix code and that* S *is an acyclic set. Let* f *be a coding morphism for* X *defined by* $f(u) = ab$*,* $f(v) = bc$*,* $f(w) = cd$*,* $f(t) = da$*. The set* $f^{-1}(S)$ *is a maximal bifix decoding of* S *and* $f^{-1}(S) = \text{fact}\big((uw)^\omega\big) \cup \text{fact}\big((vt)^\omega\big)$*. Then, there is no word* $q \in f^{-1}(S)$ *such that* $uqv \in f^{-1}(S)$*, and so* $f^{-1}(S)$ *is not a recurrent set.*

From now on, we assume that:

- S is a uniformly recurrent acyclic set;
- Y is an infinite word such that $\text{fact}(Y) = S$;
- $X \subseteq S$ is a finite S-maximal bifix code, where $X = \{X_1, \ldots, X_n\}$;
- $f : B^* \to A^*$ is a maximal coding morphism for X.

We denote by P_Y the set of ordered pairs (u, Z), where $u \in A^*$, such that $uZ = Y$ and no element of X is a suffix of u.

Note that, by Proposition 2.1, X is a finite S-maximal prefix code. So, if Z is an infinite word such that $\text{fact}(Z) \subseteq \text{fact}(Y)$ then some X_j is a prefix of Z. This also means that such Z belongs to X^ω.

Lemma 4.1. *The set* P_Y *is nonempty and finite.*

Proof. Note that $(\epsilon, Y) \in P_Y$, because ϵ is the unique suffix of ϵ and $\epsilon \notin X$. So $P_Y \neq \emptyset$.

By Proposition 2.1, X is a finite S-maximal suffix code. So, if $u \in S$ is such that $|u| \geq |X_i|$, for all $i \in \{1, \ldots, n\}$, then some X_i is a suffix of u. So the number of prefixes of

Y that don't have any X_i as suffix is finite. Consider now $u \in S$ such that u is a prefix of Y and no X_i is a suffix of u. There's a unique infinite word Z such that $uZ = Y$. This word Z obviously satisfies the condition $\text{fact}(Z) \subseteq \text{fact}(Y)$, and so $Z \in X^\omega$. This means that there is a unique infinite word Z such that $(u, Z) \in P_Y$. So P_Y is finite. □

Denote by Q_Y the elements $Z \in X^\omega$ such that there exists $u \in A^*$ with $(u, Z) \in P_Y$. Since P_Y is finite (by the last lemma), Q_Y is also finite.

Notice that we can extend the coding morphism $f : B^* \to A^*$ to a unique function $f^* : B^* \cup B^\omega \to A^* \cup A^\omega$ such that, given $b \in B$ and $Z \in B^\omega$, $f^*(bZ) = f^*(b)f^*(Z)$. Obviously f^* satisfies the equality $f^*(wZ) = f^*(w)f^*(Z)$ for all $w \in B^*$ and all $Z \in B^\omega$. For notation convenience, we denote the function f^* just by f. It is easy to see that this new function f is also injective. Then, since Q_Y is finite, $f^{-1}(Q_Y)$ must be finite. Notice that we have also $f(B^*) = X^*$ and $f(B^\omega) = X^\omega$.

The following sequence of lemmas is used to prove that any infinite word of $f^{-1}(Q_Y)$ is uniformly recurrent.

Lemma 4.2. *Given $w \in X^*$, there exists $Z \in f^{-1}(Q_Y)$ such that $f^{-1}(w)$ occurs infinitely often in Z.*

Proof. Let $w \in X^*$. Consider the set $C_w = \{i \in \mathbb{N} \mid Y_i \dots Y_{i+|w|-1} = w\}$. Since $\text{fact}(Y)$ is a recurrent set, C_w is an infinite set. Given $i \in C_w$, there exists a word $t_i \in \text{fact}(Y)$, a word $v_i \in X^*$, and an infinite word T_i such that $(t_i, v_i w T_i) \in P_Y$, $|t_i v_i| = i - 1$. Let $h : C_w \to P_Y$ be the function defined by $h(i) = (t_i, v_i w T_i)$. Since C_w is infinite and P_Y is finite, there exists $(u, W) \in P_Y$ such that $h^{-1}\big((u, W)\big)$ is an infinite set. Let $i, j \in h^{-1}\big((u, W)\big)$ be such that $i < j$. Since, X is a finite S-maximal prefix code (Proposition 2.1), $v_i w, v_j w \in X^*$ and $v_i w$ is a proper prefix of $v_j w$, there must exist $p \in X^*$, with $|p| > 0$, such that $v_i w p = v_j w$. This means that $f^{-1}(w)$ occurs infinitely often in $f^{-1}(W)$. □

It is easy to see that any factor of any infinite word of $f^{-1}(Q_Y)$ is in $f^{-1}(\text{fact}(Y))$. Conversely, Lemma 4.2 implies that any element of $f^{-1}(\text{fact}(Y))$ is a factor of some infinite word of $f^{-1}(Q_Y)$. Then, if we make $H = \{\text{fact}(Z) \mid Z \in f^{-1}(Q_Y)\}$, $f^{-1}(\text{fact}(Y))$ is the union of the elements of H. So we have proven:

Lemma 4.3. *Let $H = \{\text{fact}(Z) \mid Z \in f^{-1}(Q_Y)\}$. Then $f^{-1}(\text{fact}(Y))$ is the union of the elements of H.*

Lemma 4.4. $f^{-1}(\text{fact}(Y))$ *is a biextendable set.*

Proof. This is an obvious consequence of Lemma 4.2. □

Lemma 4.5. *Let* $Z \in f^{-1}(Q_Y)$*. If* $\text{fact}(Z)$ *is a biextendable set, then* $\text{fact}(Z)$ *is an acyclic set.*

Proof. Let $Z \in f^{-1}(Q_Y)$ such that $\text{fact}(Z)$ is a biextendable set and let $w \in \text{fact}(Z)$. The set $f^{-1}(\text{fact}(Y))$ is an acyclic set by Lemma 4.4 and Theorem 2.4. Then, the extension graph of w in $f^{-1}(\text{fact}(Y))$ does not have any cycle. Since the extension graph of w in $\text{fact}(Z)$ is a subgraph of the extension graph of w in $f^{-1}(\text{fact}(Y))$, it must be an acyclic graph. □

Lemma 4.6. *Given* $Z \in f^{-1}(Q_Y)$*, there exists* $W \in f^{-1}(Q_Y)$*, such that* $\text{fact}(W)$ *is uniformly recurrent and* $\text{fact}(Z) \subseteq \text{fact}(W)$*.*

Proof. Let $Z \in f^{-1}(Q_Y)$. Let F_k be the prefix of length k of the infinite word Z. Notice that $f(F_k) \in X^*$. Then, by Lemma 4.2, there exists $L_k \in f^{-1}(Q_Y)$ such that F_k occurs infinitely often in L_k. Let $h : \mathbb{N} \to f^{-1}(Q_Y)$ be the function defined by $h(k) = L_k$. Since $f^{-1}(Q_Y)$ is finite there must exist some $V_1 \in f^{-1}(Q_Y)$ such that $h^{-1}(V_1)$ is an infinite set. Since F_i is a prefix of F_j for $j > i$, then all prefixes of Z must appear infinitely often in V_1. This also means that all factors of Z must appear infinitely often in V_1. If V_1 is recurrent then we end the process. Otherwise, we must have $\text{fact}(Z) \subsetneq \text{fact}(V_1)$, and we apply the same method to find $V_2 \in f^{-1}(Q_Y)$ such that the factors of V_1 appear infinitely often in V_2. If V_2 is recurrent we end the process. Otherwise, $\text{fact}(V_1) \subsetneq \text{fact}(V_2)$, and so apply the same method to find $V_3 \in f^{-1}(Q_Y)$ such that the factors of V_2 appear infinitely often in V_3... This process must end because $f^{-1}(Q_Y)$ is finite. Hence, there exists $W \in f^{-1}(Q_Y)$ such that W is recurrent and the factors of Z appear infinitely often in W. By Lemma 4.5, $\text{fact}(W)$ is an acyclic set, so, by Corollary 3.14, $\text{fact}(W)$ is uniformly recurrent. □

The following elementary lemmas (4.7 and 4.9) may be considered folklore.

Lemma 4.7. *Let* $U \subseteq A^*$ *be an infinite and factorial set and let* $V \subseteq A^*$ *be a uniformly recurrent set. If* $U \subseteq V$*, then* $U = V$*.*

Proof. Let $U \subseteq A^*$ be an infinite and factorial set and let $V \subseteq A^*$ be a uniformly recurrent set such that $U \subseteq V$. Let $w \in V$. Since V is uniformly recurrent and $w \in V$, there exists $M \in \mathbb{N}$ such that every word of V with length M contains w as a factor. Then, since $U \subseteq V$, every word of U with length M contains w as a factor. So $w \in U$. Since w is an arbitrary word of V, then $U = V$. □

Proposition 4.8. *Let* $Z \in f^{-1}(Q_Y)$*. Then* $\text{fact}(Z)$ *is uniformly recurrent.*

Proof. This is true by Lemma 4.6 and Lemma 4.7. □

As we have seen in the beginning of this section, the maximal bifix decoding of a uniformly recurrent acyclic set is not necessarily a uniformly recurrent acyclic set. However, notice that by Lemma 4.3, Proposition 4.8 and Lemma 4.5 the maximal bifix decoding of a uniformly recurrent acyclic set is a finite union of uniformly recurrent acyclic sets. Furthermore, the lemma below tells us that the decomposition of any set as a finite union of different uniformly recurrent sets is unique.

Lemma 4.9. *If* $U \subseteq A^*$ *can be written as a finite union of different uniformly recurrent sets, then that union is the only decomposition of the set* U *as a (finite or infinite) union of different uniformly recurrent sets.*

Proof. Let $U \subseteq A^*$ and let $V_1, \ldots, V_q \subseteq U$ be different uniformly recurrent sets such that $U = \bigcup_{i \in \{1,\ldots,q\}} V_i$. Notice that any recurrent subset of A^* is a set of factors of some infinite word. Then suppose that Z is an infinite word such that $\text{fact}(Z)$ is uniformly recurrent and $\text{fact}(Z) \subseteq U$. Let F_k be the prefix of length k of the infinite word Z. Then, F_k belongs to some V_{m_k}. Let $h : \mathbb{N} \to \{V_1, \ldots, V_q\}$ be the function defined by $h(k) = V_{m_k}$. Since $\{V_1, \ldots, V_q\}$ is finite there must exist some V_p such that $h^{-1}(V_p)$ is an infinite set. Since F_i is a prefix of F_j for $j > i$, then all prefixes of Z must be in V_p. This also means that all factors of Z must be in V_p, or, in other words, $\text{fact}(Z) \subseteq V_p$. Since $\text{fact}(Z)$ and V_p are uniformly recurrent, then, by Lemma 4.7, $\text{fact}(Z) = V_p$. It follows that the decomposition of U as a union of different uniformly recurrent sets is unique. □

We saw that $f^{-1}(\text{fact}(Y))$ can be decomposed as a unique finite union of different uniformly recurrent acyclic sets. However, we want to describe a little more the maximal bifix decoding of S. More precisely, we relate the number of uniformly recurrent acyclic sets that are part of decomposition of $f^{-1}(\text{fact}(Y))$ with the number of connected components of the Rauzy graphs of sufficiently large order (in fact, we prove that these numbers are the same). But before that, we need the following result:

Lemma 4.10. *Let* $Z, V \in f^{-1}(Q_Y)$*. Then* $\text{fact}(Z) \neq \text{fact}(V)$ *if and only if there exists* $D \in \mathbb{N}$ *such that* $\text{fact}(Z) \cap \text{fact}(V) \cap B^D$ *is an empty set.*

Proof. Suppose that $\text{fact}(Z) \neq \text{fact}(V)$. Assume that such D does not exist. Let $K \in \mathbb{N}$ and let $w \in \text{fact}(Z) \cap B^K$. By Proposition 4.8, $\text{fact}(Z)$ and $\text{fact}(V)$ are uniformly recurrent. So, there must exist $M \in \mathbb{N}$ such that every word of $\text{fact}(Z)$ with length M

contains w as a factor. Let $u \in \text{fact}(Z) \cap \text{fact}(V) \cap B^M$. Since $u \in \text{fact}(Z)$, u must contain w as a factor. Since $u \in \text{fact}(V)$, w must be in $\text{fact}(V)$. This implies that $\text{fact}(Z) \cap B^K \subset \text{fact}(V)$. Since K is an arbitrary number, we have $\text{fact}(Z) \subseteq \text{fact}(V)$. Analogously, it can be shown that $\text{fact}(V) \subseteq \text{fact}(Z)$ so, we conclude that $\text{fact}(V) = \text{fact}(Z)$ which is a contradiction.
The other implication is obvious. □

Since $f^{-1}(Q_Y)$ is a finite set, then, using Lemma 4.10, it is easy to see that there must exist $D \in \mathbb{N}$ such that, given any $D' \in \mathbb{N}$ with $D' \geq D$ and any $Z, V \in f^{-1}(Q_Y)$, $\text{fact}(Z) \neq \text{fact}(V)$ is equivalent to $\text{fact}(Z) \cap \text{fact}(V) \cap B^{D'}$ being an empty set. Fix such a number D.

Theorem 4.11. *Consider any $D' \in \mathbb{N}$ with $D' \geq D$. Let $c \in \mathbb{N}$ be the number of connected components of the graph $G_{D'}(f^{-1}(S))$. The maximal bifix decoding of S is a union of c different uniformly recurrent acyclic sets. The decomposition of this set as a union of uniformly recurrent sets is unique.*

Proof. Let $H = \{\text{fact}(Z) \mid Z \in f^{-1}(Q_Y)\}$. Then, since $f^{-1}(S)$ is the union of the elements of the set H, $f^{-1}(S)$ is the union of $|H|$ different uniformly recurrent acyclic sets. We are going to show that $|H|$ is equal to c. If u_1, u_2 are elements of $B^{D'} \cap f^{-1}(S)$ such that (u_1, u_2) is an edge of the graph $G_{D'}(f^{-1}(S))$, then there exists an infinite word $W \in f^{-1}(Q_Y)$ such that $u_1, u_2 \in \text{fact}(W)$, in other words, there exists one element of H that contains u_1 and u_2. Furthermore, by definition of D, for each word of $B^{D'} \cap f^{-1}(S)$ there is only one element of H that contains it. Then, since any two words that are connected by an edge in $G_{D'}(f^{-1}(S))$ belong to the same element of H, any two words that are connected by a path in $G_{D'}(f^{-1}(S))$ belong to the same element of H. Conversely, if two words $u_1, u_2 \in B^{D'} \cap f^{-1}(S)$ belong to the same element of H, it is obvious that these words are connected by a path in $G_{D'}(f^{-1}(S))$. Hence, we have $|H| = c$.
By Lemma 4.9 the decomposition of $f^{-1}(S)$ as a union of different uniformly recurrent sets is unique. □

The above results and methods also allow us to give a short proof of the main theorem of Berthé et al. (2015b) (Theorem 3.3).

Proof of Theorem 3.3: Assume that S is a uniformly recurrent tree set. By Theorem 2.5, $f^{-1}(S)$ is a tree set. Notice that $B^D \cap f^{-1}(S)$ is an $f^{-1}(S)$-maximal bifix code.

So, by Propositions 2.1 and 2.3, the graph $E_{B^D \cap f^{-1}(S), B^D \cap f^{-1}(S)}(\epsilon)$ is a tree. Let $H = \{\text{fact}(Z) \mid Z \in f^{-1}(Q_Y)\}$. If u_1, u_2 are elements of $B^D \cap f^{-1}(S)$ such that (u_1, u_2) is an edge of the graph $E_{B^D \cap f^{-1}(S), B^D \cap f^{-1}(S)}(\epsilon)$ then there exists one element of H that contains $u_1 u_2$. Furthermore, by definition of D, for each word of $B^D \cap f^{-1}(S)$ there is only one element of H that contains it. Then, since any two words that are connected by an edge in $E_{B^D \cap f^{-1}(S), B^D \cap f^{-1}(S)}(\epsilon)$ belong to the same element of H, any two words that are connected by a path in $E_{B^D \cap f^{-1}(S), B^D \cap f^{-1}(S)}(\epsilon)$ belong to the same element of H. Since this graph is connected (because it is a tree), the last observation implies that all the words in $B^D \cap f^{-1}(S)$ must belong to the same element of H. But, by definition of D, this is only possible if $|H| = 1$. This means that $f^{-1}(S) = \text{fact}(W)$ for any $W \in f^{-1}(Q_Y)$. So $f^{-1}(S)$ is uniformly recurrent. □

Chapter 5

Conclusion

We defined the class of the almost-neutral sets, which contains the acyclic sets, and we showed that each set of this class only contains a finite number of non-neutral words. This fact allowed us to show that, in this kind of sets, the right special words of large enough length can be prolonged on the left to right special words. Also, we saw that the number of right special words of a certain length n, as a function of n, is constant for sufficiently large values of n. Then we restricted our analysis to the class of recurrent almost-neutral sets, whose sets are in fact sets of factors of infinite words (by the property of being recurrent). We considered these infinite words and, with our observations about right special words, we proved that the gaps between two consecutive occurrences of any word in any infinite word have limited length, which implies that recurrent almost-neutral sets are in fact uniformly recurrent. Consequently, this reduces the open problem of Berthé et al. (2015a) to Theorem 3.3, which is also established in Berthé et al. (2015b).
Furthermore, we studied the maximal bifix decoding of uniformly recurrent acyclic sets and, with Theorem 3.1, we concluded that they are a unique finite union of different uniformly recurrent acyclic sets. In fact, it was shown that the number of sets of this finite union is the number of connected components of any Rauzy graph of sufficiently large order. Since we obtained this result for acyclic sets using Theorem 3.1, it is natural to ask if this result is also true for uniformly recurrent almost-neutral sets, i.e., is the maximal bifix decoding of uniformly recurrent almost-neutral sets a finite union of uniformly recurrent almost-neutral sets? Looking carefully for the arguments that were used to prove Theorem 4.11, the raised question can be reduced to the following problem: is the maximal bifix decoding of any uniformly recurrent almost-neutral set an almost-neutral set?

Bibliography

J. Berstel. Axel Thue's papers on repetitions in words: a translation. Number 20 in Publications du LaCIM. Université du Québec à Montréal, 1995.

J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata*. Cambridge Univ Press, 2010.

J. Berstel, C. De Felice, D. Perrin, C. Reutenauer, and G. 3 Rindone. Bifix codes and Sturmian words. *J. Algebra*, pages 146–202, 2012.

V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Acyclic, connected and tree sets. *Monatsh Math*, pages 521–550, 2015a.

V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Maximal bifix decoding. *Discrete Mathematics*, pages 725–742, 2015b.

F. Dolce and D. Perrin. Eventually dendric subshifts. Technical report, 2018. arXiv:1807.05124.

Edgar N. Gilbert and Edward F. Moore. Variable length binary encodings. *The Bell System Technical Journal*, pages 933–967, 1959.

A. Lentin and M. P. Schützenberger. A combinatorical problem in the theory of free monoids. *Combinatorial Mathematics*, pages 112–144, 1967.

D. Lind and B. Marcus. *An introduction to symbolic dynamics and coding*. Cambridge Univ Press, 1995.

R. C. Lyndon and M. P. Schützenberger. The equation $a^m = b^n c^p$ in a free group. *Michigan Math. J.*, pages 289–298, 1962.

M. P. Schützenberger. On an application of semigroup methods to some problems in coding. *IRE Trans. Inform. Theory*, pages 933–967, 1956.

C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, pages 379–423, 1948.