Mémoire

Auteur : Backes, Lindsay
Promoteur(s) : Charlier, Emilie
Faculté : Faculté des Sciences
Diplôme : Master en sciences mathématiques, à finalité approfondie
Année académique : 2024-2025
URI/URL : http://hdl.handle.net/2268.2/22976

Faculté des Sciences
Département de Mathématique

Numeration systems and substitutions

Mémoire de fin d'études présenté en vue de l'obtention du titre de
Master en Sciences Mathématiques, à finalité approfondie

Année académique 2024-2025

Auteur :
Lindsay Backes

Promoteur :
Émilie Charlier

Acknowledgements

I would first like to extend a special thanks to my supervisor, Émilie Charlier for her valuable advice and proofreading of this thesis.

I would also like to thank the members of the jury Arnout Van Messen, Julien Leroy and Jean-Pierre Schneiders, as well as the external reviewer, Manon Stipulanti and my supervisor, Émilie Charlier for partitioning in the jury.

I would also like to express my gratitude towards my friends and family for their unwavering support, constant encouragement, and assistance throughout my studies at the University of Liège. In particular, I would also like to thank my mother for all the proofreading she did during my studies and my best friend and my brother for listening to me talking about mathematics even thought they didn't understand the majority of it.

Furthermore, I would like to thank all my friends from the department of mathematics. And a special thank you to Théo for all the projects we completed together, as well as for the help he provided throughout our studies. In addition, I would also like to sincerely thank Pauline and Tiago for their unwavering support throughout our studies and unforgettable memories.

Contents

Introduction

In mathematics, more precisely in discrete mathematics, we are interest in ways to represent integers uniquely. The most famous method is the decimal system, which is used around the world. In the decimal system, the letters are the digits $0, \ldots, 9$. We form words with those letters by placing these letters according to powers of 10. For example, the number seven thousand three hundred nineteen is represented as the word 7319. However, there are other methods. For instance, the Maya Indians used the base 20 and not the base 10. Another example would be the binary system used by computers. This system became very famous with the rise of the technology. In this work, we will study a series of different numeration systems.

In chapter one, we will start by introducing positional numeration systems. Numeration systems enable us to represent numbers by finite words over a suitable alphabet. Afterwards, we introduced another numeration system, namely the Dumont–Thomas numeration system introduced in [9]. The Dumont-Thomas numeration system is associated with a fixed point of a substitution. The idea of this method is for a given $n \in \mathbb{N}$ to decompose the prefix of the fixed point, in a word-greedy way, into the images of words over the alphabet of the substitution. During our studies of the Dumont-Thomas numeration system associated with the fixed point of a substitution, we will encounter the prefix-suffix automaton associated with a substitution. This automaton will be useful tool to prove theorems. A very well known generalisation of a positional numeration system is an abstract numeration system. An abstract numeration system is a numeration system based on an infinite regular language over a totally ordered alphabet. To conclude this chapter, we will introduce abstract numeration systems and we will prove that every Dumont–Thomas numeration system associated with a substitution can be interpreted as an abstract numeration system.

In the middle of the twentieth century, Cobham introduced the notion of k-automatic sequences, which are based on the representation of integers in base k. By feeding the representation of an integer i in base k to a deterministic finite automaton with output, we obtain the i^{th} term of a k-automatic sequence. Cobham proved that there exists an equivalence between k-automatic sequences and the sequences obtained by iterating a uniform morphism of length k. In the second chapter, we will study a generalisation of those sequences using abstract numeration system. We will introduce S-automatic sequences where S is an abstract numeration systems. However, instead of feeding a deterministic finite automaton with output with the representation of i in base k, we will feed it with the representation of i in the abstract numeration system S. We will prove some characterisations of S-automatic sequences, among others that there exists an equivalence between S-automatic sequences and morphic words.

Another very well know positional numeration systems is the numeration system associated with the θ-expansion of 1 of a Parry number. Those Parry numbers are named after Parry, who introduced them in [30]. In chapter 3, we will restrict our studies to the θ-expansion of 1 of a Parry number. Firstly, we will look at the definition of the positional numeration systems as well as the Fabre substitution associated with the θ-expansion of 1. A few links between those notions will be shown. In addition, we can also define an θ-automaton, which is a special automaton verifying certain conditions, among others that this automaton accepts only words which do not contain any factor bigger than or equal to the θ-expansion of 1 in the lexicographic order. After proving some properties of those automata, we will be able to prove that there is a link between a θ-automaton and the fixed point of a conjugate of the Fabre substitution associated with the θ-expansion of 1. This link was first proven by Fabre in [12].

Combinatorially speaking, the fixed points of these particular substitutions associated with a Parry number have been studied in multiple articles. In the fourth chapter we will look at the fixed points of the Fabre substitutions from a θ-expansion of 1 for a simple Parry number. More precisely, we will study the form of those fixed points, in particular its factors. The goal of those studies is to determine all the left special factors and bispecial factors which occur in the studied fixed point. As a result of those studies, we will be able to determine the complexity of the fixed point of a substitution associated with a simple Parry number under specific conditions.

Lastly, we will will show an application of a few results encountered during this thesis in the field of string attractors. The concept of String Attractor was first introduced in the data compression field by Kempa and Prezza [18]. A string attractor is a set of positions within a finite word such that each factor is referenced by at least one position of the string attractor. Even though string attractors were introduced in the data compression field, they also have applications in combinatorial pattern matching. Finding the smallest String Attractor is however NP-hard. In chapter five, we will present the approach of Gheeraert, Romana and Stipulanti in [16] proving the existence of a link between the string attractors of the prefixes of the fixed point of the Fabre substitution under a defined working hypothesis and the base of the positional numeration system introduced in chapter 3.

Chapter 1

Overview of some numeration systems

A numeration system is a unique representation of an integer. A very well known numeration system would be the binary system. At the beginning of this chapter, the notion of two different numeration systems will be introduced; positional numeration system and the Dumont-Thomas numeration system. In the late twentieth century, Dumont and Thomas introduced in [9] a numeration system associated with a fixed point of a substitution prolongable on a given letter. This numeration system is sometimes referred to as the Dumont-Thomas system. In this chapter we will firstly introduce positional numeration systems and then studying the numeration system introduced by Dumont and Thomas.

Furthermore, a link between this numeration system and a particular automaton will be introduced and studied afterwards. Lastly, we will deal with abstract numeration systems. More precisely, we will study the abstract numeration system associated the Dumont-Thomas numeration systems.

1.1 Notions

Before continuing with this chapter, a review of certain concepts and definitions seems appropriate.The purposes of this review is to establish the notions that will be useful in the remainder of this thesis.

An alphabet Σ is a finite set. The elements of an alphabet are called letters. A finite word w in the alphabet Σ is a finite sequence of letters of Σ. For example, a finite word over the alphabet $\Sigma = \{0, 1\}$ is $w = 0101$. An infinite word u in our alphabet Σ is an infinite sequence of letters of Σ. For example, a infinite word over the alphabet $\Sigma = \{0, 1\}$ is $w = 010101\ldots$. If an infinite word w is created by concatenating a finite word u an infinite number of times, we denote this $w = u^\omega$. An infinite word w in the alphabet Σ is ultimately periodic if there are two finite words $u, v \in \Sigma$ such that $w = uv^\omega$. In the special case, where $u = \epsilon$, w is said purely periodic or simply periodic. When an infinite word is not ultimately periodic, it is said to be aperiodic. The set of finite words over the alphabet Σ is denoted Σ^* and the set of infinite words is denoted $\Sigma^{\mathbb{N}}$. If w is a finite word, then u is a factor of w if there exist two words $v, v' \in \Sigma^*$ such that $w = vuv'$. If w is an infinite word, then u is a factor of w if there exist two words $v \in \Sigma^*$ and $v' \in \Sigma^{\mathbb{N}}$ such that $w = vuv'$. The set of all factors of w is denoted $\mathrm{Fact}(w)$. The set of all factor of w of size $n \in \mathbb{N}$ is denoted $\mathrm{Fact}_n(w)$. A factor of w starting at position i and ending at position j in w is noted $w[i, j]$. If $a \in \Sigma$, the number of occurrences of the letter a in our word w is denoted by $|w|_a$. We denote the empty word by ϵ, *i.e.* for all letters a in our alphabet Σ we have $|\epsilon|_a = 0$. Let $(\Sigma, <)$ be a totally ordered alphabet. The order $<$ on Σ extends to an order on $\Sigma^{\mathbb{N}}$ as follows: If u and v are two infinite words over Σ, then u is said to be lexicographically less than v, if the following condition is verified

- there exists $w \in \Sigma^*$, $a, b \in \Sigma$ and $u', v' \in \Sigma^{\mathbb{N}}$ such that $u = wau'$ and $v = wbv'$ and $a < b$.

We write $u <_{lex} v$. This extensions of the order $<$ on Σ extends to an order on $\Sigma^{\mathbb{N}}$ is called lexicographical order. This order extends to $\Sigma^{\mathbb{N}} \cup \Sigma^*$ as follows: Let $\$ \notin \Sigma$ and such that $\$ < a$ for all $a \in \Sigma$ and we replace each finite word w over Σ with the infinite word $w(\$)^\omega$. Similar to lexicographical order, the genealogical order can be defined. Let $(\Sigma, <)$ be a totally ordered alphabet. The order $<$ on Σ extends to Σ^* as follows: If u and v are two finite words over Σ then u is said to be genealogically less than v if one of the following conditions is verified

- $|u| < |v|$;
- $|u| = |v|$, then $u <_{lex} v$.

We write $u <_{gen} v$. Let Σ and Γ be two finite alphabets. A morphism is a map $f\colon \Sigma^* \to \Gamma^*$ such that

$$\forall u, v \in \Sigma^*\colon f(uv) = f(u)f(v).$$

A morphism is prolongable on $a \in \Sigma$ if $\Sigma = \Gamma$, $f(a) = a\Sigma^+$ and $|f^i(a)| \underset{i\to\infty}{\to} \infty$. Let $u, v \in \Sigma^{\mathbb{N}}$ where $u = (u_i)_{i\in\mathbb{N}}$ and $v = (v_i)_{i\in\mathbb{N}}$. The distance between those two words is defined as follows:

$$d(u, v) = \begin{cases} 2^{-\min\{i\in\mathbb{N} \mid \min u_i \neq v_i\}} & \text{if } u \neq v \\ 0 & \text{if } u = v. \end{cases}$$

The sequence $(u_i)_{i\in\mathbb{N}}$ of infinite words over an alphabet Σ converges to $v \in \Sigma^{\mathbb{N}}$, if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i \geq N$, we have $d(u_i, v) < \epsilon$. Let $\$ \notin \Sigma$. We say that the sequence $(u_i)_{i\in\mathbb{N}}$ of finite words over Σ converges to $v \in \Sigma^{\mathbb{N}}$ if the sequence $(u_i\$^\omega)_{i\in\mathbb{N}}$ of infinite words converges to v. Intuitively, the sequence $(u_i)_{i\in\mathbb{N}}$ of finite words over Σ converges to an infinite word $v \in \Sigma^{\mathbb{N}}$ if every prefix of v is a prefix of all but a finite number of the words u_i. A morphism is erasing if there exists a letter a in the alphabet Σ such that $f(a) = \epsilon$, otherwise it is called non-erasing. If f is a non-erasing morphism, it can be extended to a map from $\Sigma^{\mathbb{N}}$ to $\Gamma^{\mathbb{N}}$ as follows: if $u = u_0u_2\dots$ is an infinite word over Σ, then the sequence of words $f(u_0\cdots u_i)_{i\in\mathbb{N}}$ is convergent towards an infinite word over Γ. Its limit is denoted $f(u) = f(u_0)f(u_1)\dots$. Let $f\colon \Sigma^* \to \Sigma^*$ be a morphism. A finite or infinite word w is a fixed point of f if $f(w) = w$. If a morphism is $f: \Sigma^* \to \Sigma^*$ prolongable on a letter $a \in \Sigma$, then for all $i \in \mathbb{N}$ the word $f^i(a)$ is a prefix of $f^{i+1}(a)$ and since $|f^i(a)| \underset{i\to\infty}{\to} \infty$, the sequence $(f^i(a))_{i\in\mathbb{N}}$ converges to an infinite word, denoted

$$f^\omega(a) = \lim_{n\to\infty} f^n(a).$$

If a morphism f is only prolongable on a letter a and on no other letter, there exists one unique fixed point of f. A purely morphic word is an infinite word obtained by iterating a prolongeable substitution. If $w \in \Sigma^{\mathbb{N}}$ is purely morphic and if $g\colon \Sigma \to \Gamma$ is non-erasing, then the word $v = g(w)$ is said to be morphic.

1.2 Positional numeration systems

A finite word can be used to represent an integer in a given numeration system. Analogously, an infinite word can be used to describe a characteristic sequence of integers or a real numbers. A very known numeration system is the positional numeration system.

Definition 1.2.1. A positional numeration system is an increasing sequence of integers $U = (U_i)_{i\in\mathbb{N}}$ (sometimes called the base) with $U_0 = 1$ and such that the quotient of two consecutive terms is bounded *i.e.* $\sup_{i\in\mathbb{N}}\lceil (U_{i+1}/U_i) - 1\rceil < \infty$. The numeration alphabet is the a finite set $\Sigma_U = \{0, \dots, \sup_{i\in\mathbb{N}}\lceil (U_{i+1}/U_i) - 1\rceil\}$. In a numeration system $(U_i)_{i\in\mathbb{N}}$, any integer m admits a unique word $w_{n-1}\dots w_0$ in Σ_U^* such that

$$m = w_0U_0 + \cdots + w_{n-1}U_{n-1} \tag{1.1}$$

with $w_{n-1} \neq 0$ and $\sum_{k=0}^{i} w_kU_k < U_{i+1}$ $\forall i \in \{0, \dots, n-1\}$. The word $w_{n-1}\dots w_0$ is called the (greedy) U-representation of m and is denoted $\mathrm{rep}_U(m)$. The language $\mathrm{rep}_U(\mathbb{N}) = \{\mathrm{rep}_U(n)\colon n \in \mathbb{N}\}$ is called the numeration language. The function $\mathrm{val}_U\colon \Sigma_U^* \to \mathbb{N}$ is defined by $\mathrm{val}_U(w_{n-1}\dots w_0) = \sum_{k=0}^{n-1} w_kU_k$.

A classical example of a positional numeration system, is the Fibonacci numeration.

Example 1.2.2. Let $F = (F_i)_{i\in\mathbb{N}}$ be the Fibonacci sequence defined as follows:

$$\begin{aligned} F_0 &= 1 \\ F_1 &= 2 \\ F_{i+2} &= F_{i+1} + F_i && \forall i \in \mathbb{N}. \end{aligned}$$

We notice that $\sup_{i\in\mathbb{N}}\lceil (F_{i+1}/F_i) - 1\rceil = 1$ because $\dfrac{F_1}{F_0} = 2$ and $\forall i \in \mathbb{N}$, we have

$$\frac{F_{i+2}}{F_{i+1}} = 1 + \underbrace{\frac{F_i}{F_{i+1}}}_{<1}.$$

The numeration alphabet is $\Sigma_F = \{0, 1\}$. If $m = 10$, we have $F_5 = 13 > m \geq F_4 = 8$ and the euclidean division of m by $F_4 = 8$ is

$$m = 1F_4 + r_4$$

where $r_4 = 2$ and $w_4 = 1$. We will iterate this procedure. The euclidean division of r_4 by $F_3 = 5$ is

$$r_4 = 0F_3 + r_3$$

where $r_3 = 2$ and $w_3 = 0$. We will iterate this procedure again. The euclidean division of r_3 by $F_2 = 3$ is

$$r_3 = 0F_2 + r_2$$

where $r_2 = 2$ and $w_2 = 0$. We will iterate this procedure one more time. The euclidean division of r_2 by $F_1 = 2$ is

$$r_2 = 1F_1 + r_1$$

where $r_1 = 0$ and $w_1 = 1$. We have $r_1 = 0$, so $w_0 = 0$. The word 10010 is the greedy F-representation of m.

The following proposition is a direct result of Definition 1.2.1.

Proposition 1.2.3. *Let $U = (U_i)_{i \in \mathbb{N}}$ be a positional numeration system. For all $m, n \in \mathbb{N}$,*

$$m < n \Leftrightarrow \text{rep}_U(m) <_{gen} \text{rep}_U(n)$$

where the natural ordering of the alphabet $\Sigma_U \subset \mathbb{N}$ induces the genealogical ordering $<_{gen}$ on Σ_U^*.

Proof. This is a direct consequence of the greedy condition in the definition of a numeration system. □

Remark 1.2.4. However, it is **not** true that for a positional numeration system $U = (U_i)_{i \in \mathbb{N}}$ all words $u, v \in \Sigma_U$ verify the following

$$u <_{gen} v \Leftrightarrow \text{val}_U(u) < \text{val}_U(v).$$

Let's consider the Fibonacci numeration system, which we already saw in Example 1.3.8. If $u = 111$ and $v = 1000$ then $u <_{gen} v$ as $|u| < |v|$ and

$$\text{val}_F(u) = 1 + 2 + 3 = 6 > 5 = \text{val}_F(v).$$

1.3 Numeration systems associated with a fixed point of a substitution

The numeration system associated with a fixed point of a prolongable non-erasing morphism is often also called the Dumont-Thomas numeration system. This numeration system was first introduced in [9] by Dumont and Thomas. In this section, we will focus ourselves on the construction of this numeration system.

Notation 1.3.1. To simplify the notation, a substitution f is a non-erasing morphism from Σ^* into Σ^*, where Σ is a finite set and f is prolongable on a letter $a \in \Sigma$, we denote this substitution f as a triplet $f = (f, \Sigma, a)$.

To simplify the notation, we will suppose that for the rest of this chapter, f is the substitution $f = (f, \Sigma, a)$.

Definition 1.3.2. A finite sequence of pairs $(p_i, a_i)_{i=0,\ldots,l} \in \Sigma^* \times \Sigma$ is admissible if for all $1 \leq i \leq l$, $p_{i-1}a_{i-1}$ is a prefix of $f(a_i)$. Let $b \in \Sigma$, the sequence is b-admissible if the sequence is admissible and $p_l a_l$ is a prefix of $f(b)$.

The idea of the numeration system associated with a fixed point of a substitution is to be able, for a given substitution $f = (f, \Sigma, a)$, to decompose a prefix of any length of $f^\omega(a)$ into factors such that those factors can be expressed as the image of words over the alphabet of the substitution f.

Using the uniqueness of a limit, we directly remark the following:

Remark 1.3.3. There exists one unique word $w \in \Sigma^{\mathbb{N}}$ such that w is a fixed point of f and starts with the letter a.

We will now present a number of lemmas, which will be useful to prove that a prefix of a fixed point of a substitution can be expressed as concatenation of factors which are iterated images of words by the substitution f.

Lemma 1.3.4. *Let $l \in \mathbb{N}$ and $(p_i, a_i)_{i=0,\ldots,l}$ an admissible sequence, then*

$$\sum_{j=0}^{l} |f^j(p_j)| < |f^l(p_l a_l)|.$$

Proof. We will proceed by induction on $l \in \mathbb{N}$. Base case: if $l = 0$, we have $|p_0| < |p_0 a_0|$ which is always true as $a_0 \in \Sigma$. Induction: We suppose that it is true for all $n < l$. We have

$$\sum_{j=0}^{l} |f^j(p_j)| = \sum_{j=0}^{l-1} |f^j(p_j)| + |f^l(p_l)|.$$

We know by the hypothesis of our recursion and by the definition of an admissible sequence that:

$$\sum_{j=0}^{l-1} |f^j(p_j)| < |f^{l-1}(p_{l-1}a_{l-1})| \leq |f^{l-1}(f(a_l))| = |f^l(a_l)|.$$

So we obtain that:

$$\sum_{j=0}^{l} |f^j(p_j)| < |f^l(a_l)| + |f^l(p_l)| = |f^l(p_l a_l)|.$$

□

Lemma 1.3.5. *Let $b \in \Sigma$ such that $f(b)$ starts with b and let l, $l' \in \mathbb{N}$ so that $(p_i, a_i)_{i=0,\dots,l}$ and $(p'_i, a'_i)_{i=0,\dots,l'}$ are two b-admissible sequences such that*

- $p_l \neq \epsilon$ *and* $p'_l \neq \epsilon$
- $\sum_{j=0}^{l} |f^j(p_j)| = \sum_{j=0}^{l'} |f^j(p'_j)|$.

Then $l = l'$.

Proof. We will proceed by contradiction. By symmetry, we can suppose that $l' > l$. To simplify the notion, we suppose that

$$N = \sum_{j=0}^{l} |f^j(p_j)| = \sum_{j=0}^{l'} |f^j(p'_j)|.$$

We know that by the second point,

$$|f^{l'}(p'_{l'})| \leq N$$

and we know by Lemma 1.3.4 that

$$N < |f^l(p_l a_l)|.$$

Due to $l' > l$ and point 1 and since f is non erasing, we have

$$N \geq |f^{l'}(p'_{l'})| \geq |f^{l+1}(p'_{l'})|.$$

We furthermore notice that $p'_{l'}$ is a prefix of $f(b)$ and that $f(b)$ starts with the letter b so $p'_{l'} = bw$ with $w \in \Sigma^*$. Therefore, we have

$$|f^{l+1}(p'_{l'})| \geq |f^{l+1}(b)|.$$

But $p_l a_l$ is also a prefix of $f(b)$ so

$$N \geq |f^l(f(b))| \geq |f^l(p_l a_l)| > N,$$

which is a contradiction. So we have $l = l'$. □

Lemma 1.3.6. *Let $l \in \mathbb{N}$, $b \in \Sigma$ and $(p_i, a_i)_{i=0,\dots,l}$ and $(p'_i, a'_i)_{i=0,\dots,l}$ be two b-admissible sequences such that*

$$\sum_{j=0}^{l} |f^j(p_j)| = \sum_{j=0}^{l} |f^j(p'_j)|.$$

Then we have $(p_i, a_i) = (p'_i, a'_i)$ for all $i \in \{0, \dots, l\}$.

Proof. We will proceed by induction on l. Base case: if $l = 0$, we know that $|p_0| = |p'_0|$ and as both sequences are b-admissible, $p_0 a_0$ and $p'_0 a'_0$ are both prefixes of $f(b)$. So we can directly conclude that $(p_0, a_0) = (p'_0, a'_0)$. Induction: We suppose that it is true for all $n < l$. We want to prove that $p_l = p'_l$. If $p_l \neq p'_l$, then $|p_l| \neq |p'_l|$ as they are both prefixes of $f(b)$. Without lost of generality, we can suppose that $|p_l| > |p'_l|$ and we notice that $p'_l a'_l$ is therefore a prefix of p_l. Using this and Lemma 1.3.4, we obtain the following:

$$\sum_{j=0}^{l} |f^j(p_j)| \geq |f^l(p_l)| \geq |f^l(p'_l a'_l)| > \sum_{j=0}^{l} |f^j(p'_j)|,$$

which contradicts our hypothesis. So $p_l = p'_l$ and as $p_l a_l$ and $p'_l a'_l$ are both prefixes of $f(b)$ and we obtain $a_l = a'_l$. Finally we are able to conclude by applying the induction hypothesis on the sequences $(p_i, a_i)_{i=0,\dots,l-1}$ and $(p'_i, a'_i)_{i=0,\dots,l-1}$ which are a_l-admissible and such that $\sum_{j=0}^{l-1} |f^j(p_j)| = \sum_{j=0}^{l-1} |f^j(p'_j)|$. □

We are now able to prove one of the main results of [9], on which the definition of the Dumont-Thomas numeration system relies.

Theorem 1.3.7. *Let $f = (f, \Sigma, a)$ be a substitution and let $n \geq 1$. There exists a unique $l \in \mathbb{N}$ and a unique sequence $(p_i, a_i)_{i=0,\dots,l}$ such that*

- *this sequence is a-admissible and $p_l \neq \epsilon$;*
- *$w_0 w_1 \cdots w_{n-1} = f^l(p_l) f^{l-1}(p_{l-1}) \cdots f^0(p_0)$;*

with $w_0 w_1 \cdots w_{n-1}$ being the prefix of size n of $f^\omega(a)$.

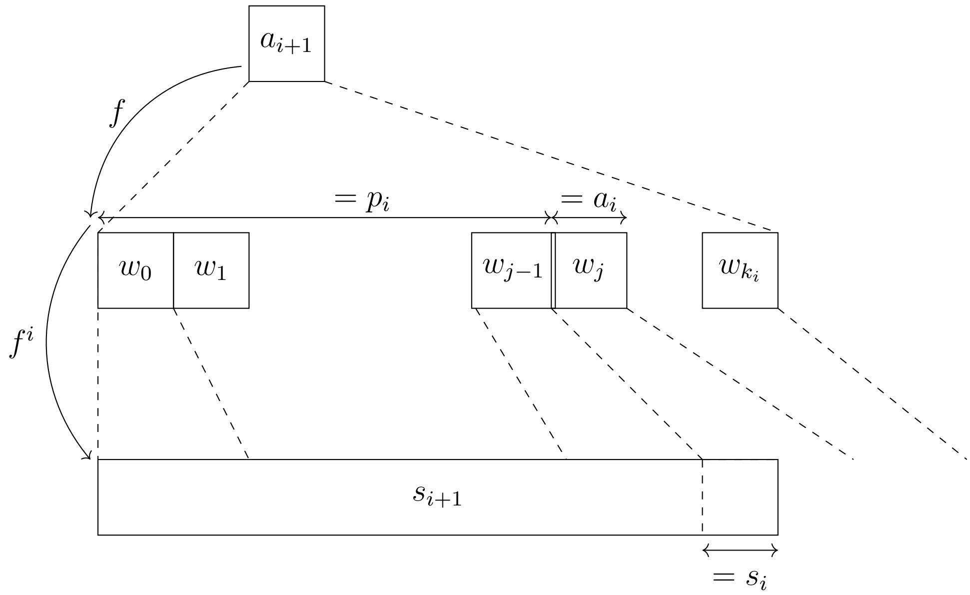
Proof. Since $|f(a)| \geq 2$ and f is non erasing, there exists a unique $l \in \mathbb{N}$ such that

$$|f^l(a)| \leq n < |f^{l+1}(a)|. \tag{1.2}$$

We fix $p_{l+1} = \epsilon, a_{l+1} = a$ and $s_{l+1} = w_0 w_1 \cdots w_{n-1}$. Since the word $w_0 w_2 \cdots w_{n-1}$ and $f^{l+1}(a)$ are prefixes of $f^\omega(a)$ and due to (1.2), we can remark that s_{l+1} is a proper prefix of $f^{l+1}(a_{l+1})$. For all $i = l, \dots, 0$, we define p_i, s_i and a_i recursively as follows. We suppose that s_{i+1} is a proper prefix of $f^{i+1}(a_i)$. We denote:

$$f(a_{i+1}) = w_0 \cdots w_{k_i}.$$

Let j be the smallest integer such that s_{i+1} is a proper prefix of $f^i(w_0 \cdots w_j)$. Such a j exists since s_{i+1} is a proper prefix of $f^i(w_0 \cdots w_{k_i})$. We define $a_i = w_j$ and $p_i = w_0 \cdots w_{j-1}$. Furthermore, we define s_i as the suffix of s_{i+1} such that $s_{i+1} = f^i(p_i)s_i$. This definition of s_i has as consequence that s_i is a proper prefix of $f^i(a_i)$, which allows us to iterate this construction.



We constructed a sequence $(p_i, a_i)_{i\in\{0,\ldots,l\}}$ such that $p_i a_i$ is a prefix of $f(a_{i+1})$ for all $i \in \{0, \ldots, l\}$ and $w_0 \cdots w_{n-1} = f^l(p_l) \cdots f^1(p_1) f^0(p_0)$. Using the Lemma 1.3.5 we know that there exists one unique $l \in \mathbb{N}$ verifying the conditions and using Lemma 1.3.6 we know that there exists a unique sequence $(p_i, a_i)_{i=0,\ldots,l}$ verifying the conditions. □

Example 1.3.8. Consider the following substitution $(f_{Fib}, \{0, 1\}, 0)$, where

$$\begin{aligned} f_{Fib} \colon \{0,1\}^* &\to \{0,1\}^* \\ 0 &\longmapsto 01 \\ 1 &\longmapsto 0 \end{aligned}$$

This substitution is called the Fibonacci substitution. We want to factorise the prefix of $f^\omega_{Fib}(0)$ of length 12 which is $w = 010010100100$. First we need to find $l \in \mathbb{N}$ such that $|f^l_{Fib}(0)| \leq 12 < |f^{l+1}_{Fib}(0)|$. We have $|f^4_{Fib}(0)| = 8 \leq 12 < |f^{4+1}_{Fib}(0)| = 13$. We fix $p_5 = \epsilon$, $a_5 = 0$ and $s_5 = w$. We know that $f_{Fib}(a_5) = f_{Fib}(0) = 01$ and $f^4_{Fib}(0) = 01001010$ is a proper prefix of s_5. So we have $p_4 = 0$, $a_4 = 1$,

$$s_5 = \underbrace{01001010}_{=f^4_{Fib}(0)}\underbrace{0100}_{=s_4}$$

and s_4 is a proper prefix of $f^4_{Fib}(1) = 01001$. We will iterate the factorisation process, we have $f_{Fib}(a_4) = f_{Fib}(1) = 0$ and $s_4 = 0100$ is a proper prefix of $f^3_{Fib}(0) = 01001$. So we have $p_3 = \epsilon$, $a_3 = 0$ and

$$s_4 = 0100 = f_{Fib}^3(\epsilon)s_3 = s_3.$$

We will iterate the process again. We have $f_{Fib}(a_3) = f_{Fib}(0) = 01$ and $f_{Fib}^2(0) = 010$, so s_3 is not a proper prefix of $f_{Fib}^2(0)$. We have $f_{Fib}^2(01) = 01001$ and s_3 is a proper prefix of $f_{Fib}^2(01)$. So we have $p_2 = 0$, $a_2 = 1$ and

$$s_3 = \underbrace{010}_{=f_{Fib}^2(0)} 0.$$

So $s_2 = 0$. We will iterate again. We know that $f_{Fib}(a_2) = f_{Fib}(1) = 0$ and $f_{Fib}(0) = 01$. And we have s_2 is a proper prefix of $f_{Fib}(0)$. So we have $p_1 = \epsilon$ and $a_1 = 0$. Furthermore,

$$s_2 = 0 = f(\epsilon)s_1 = s_1.$$

We will iterate one last time. We have $f_{Fib}(a_1) = f_{fib}(0) = 01$ and s_1 is a proper prefix of 01. So we have $p_0 = 0$ and $a_0 = 1$. Moreover, we have

$$s_2 = 0 = p_0.$$

So we have $s_0 = \epsilon$. We obtain

$$w = 010010100100 = f_{Fib}^4(0)f_{Fib}^3(\epsilon)f_{Fib}^2(0)f_{Fib}^1(\epsilon)0.$$

And we obtain the sequence $(p_i, a_i)_{i=0,\dots,4} = (0,1)(\epsilon,0)(0,1)(\epsilon,0)(0,1)$.

Let's look at another example.

Example 1.3.9. Consider the following substitution $(f, \{0,1,2\}, 0)$, where

$$\begin{aligned} f\colon \{0,1,2\}^* &\to \{0,1,2\}^* \\ 0 &\longmapsto 012 \\ 1 &\longmapsto 12 \\ 2 &\longmapsto 01 \end{aligned}$$

The prefix of $f^\omega(0)$ of length 13 is 0121201120101. We have $f^2(0) = 0121201$ and $f^3(0) = 0121201120101212$, and therefore $|f^2(0)| \leq 13 < |f^3(0)|$. We fix $p_3 = \epsilon$, $a_3 = 0$ and $s_3 = 0121201120101$. We have $f(a_3) = f(0) = 012$. However s_3 is not a proper prefix of $f^2(0) = 0121201$ or $f^2(01) = 01212011201$ but it is a proper prefix of $f^2(012) = 0121201120101212$. So we have $a_2 = 2$ and $p_2 = 01$. Furthermore, we have

$$s_3 = \underbrace{01212011201}_{=f^2(01)}01$$

so $s_2 = 01$. We will iterate this process, we have $f(a_2) = f(2) = 01$ and s_2 is a proper prefix of $f(0)$. So we have $a_1 = 0$, $p_1 = \epsilon$ and $s_2 = s_1$. We will iterate this process one last time, we have $f(a_1) = f(0) = 012$ and s_1 is not a proper prefix of 0 or 01 but it is a proper prefix of 012. So we have $a_0 = 2$, $p_0 = 01$ and

$$s_0 = \underbrace{01}_{=f^0(01)}$$

so $s_0 = \epsilon$. We obtain

$$0121201120101 = f^2(01)01.$$

And we obtain the sequence $(p_i, a_i)_{i=0,\ldots,2} = (01, 2)(\epsilon, 0)(01, 2)$.

Using Theorem 1.3.7, we are able to define the Dumont-Thomas numeration system.

Definition 1.3.10. Let (f, Σ, a) be a substitution and $\Gamma = \{0, \ldots, \max_{\sigma\in\Sigma} |f(\sigma)| - 1\}$. We define the Dumont-Thomas numeration system associated with f as follows

$$\begin{aligned} \operatorname{rep}_{f,a} &: \mathbb{N} \to \Gamma^* \\ n &\mapsto \begin{cases} |p_l| \cdots |p_0| & \text{if } n > 0 \\ \epsilon & \text{otherwise} \end{cases} \end{aligned}$$

where $(p_i, a_i)_{i=0,\ldots,l}$ is the unique sequence of $n \in \mathbb{N}_0$ obtained from Theorem 1.3.7.

Example 1.3.11. Consider the substitution of Example 1.3.8. We have

$$\operatorname{rep}_{f_{Fib},0}(12) = |0||\epsilon||0||\epsilon||0| = 10101.$$

Example 1.3.12. Consider the substitution of Example 1.3.9. We have

$$\operatorname{rep}_{f_{Fib},0}(13) = |01||\epsilon||01| = 202.$$

Lemma 1.3.13. *Let $f = (f, \Sigma, a)$ be a substitution. Let $n, m \in \mathbb{N}$. Then*

1. *$n = m$ if and only if $\operatorname{rep}_{f,a}(n) = \operatorname{rep}_{f,a}(m)$;*
2. *$n < m$ if and only if $\operatorname{rep}_{f,a}(n) <_{gen} \operatorname{rep}_{f,a}(m)$.*

Proof. If either n and m are equal to 0 (resp. n is equal to 0) or $\operatorname{rep}_{f,a}(n)$ and $\operatorname{rep}_{f,a}(m)$ are equal to ϵ (resp. $\operatorname{rep}_{f,a}(n)$ is equal to ϵ), then we can directly conclude. Let $(p_i, a_i)_{i=0,\ldots,l}$ and $(p'_i, a'_i)_{i=0,\ldots,l'}$ be two a-admissible sequences such that
$n = \sum_{i=0}^{l} |f^i(p_i)|$ and $m = \sum_{i=0}^{l'} |f^i(p'_i)|$. Thus we have $\operatorname{rep}_{f,a}(n) = |p_l| \cdots |p_0|$ and $\operatorname{rep}_{f,a}(m) = |p'_{l'}| \cdots |p'_0|$.

1. If $n = m$, we have $\operatorname{rep}_{f,a}(n) = \operatorname{rep}_{f,a}(m)$. If $\operatorname{rep}_{f,a}(n) = \operatorname{rep}_{f,a}(m)$, due to Theorem 1.3.7, we have

$$|p_l| \cdots |p_0| = |p'_{l'}| \cdots |p'_0|.$$

So we have $l = l'$. Furthermore, we can prove recursively that $p_i a_i = p'_i a'_i$ for all $i = l, \dots, 0$. Base case: $p_l a_l = p'_l a'_l$ because they are both prefixes of $f(a)$ of the same length. Let's suppose that $p_{j+1}a_{j+1} = p'_{j+1}a'_{j+1}$ for all $j \in \{0, \dots l-1\}$. We know that $p_j a_j$ and $p'_j a'_j$ are both prefixes of $f(a_{j+1})$ of the same length. Thus we have $p_j a_j = p'_j a'_j$. So we have $(p_i, a_i)_{i=0,\dots,l} = (p'_i, a'_i)_{i=0,\dots,l}$ and therefore we have

$$n = \sum_{i=0}^{l} |f^i(p_i)| = \sum_{i=0}^{l} |f^i(p'_i)| = m.$$

2. We suppose that $\operatorname{rep}_{f,a}(n) <_{gen} \operatorname{rep}_{f,a}(m)$. If $|\operatorname{rep}_{f,a}(n)| < |\operatorname{rep}_{f,a}(m)|$, then $l < l'$ and we know that $f^l(a) \le n < f^{l+1}(a)$ and $f^{l'}(a) \le m < f^{l'+1}(a)$, so $n < m$. If $|\operatorname{rep}_{f,a}(n)| = |\operatorname{rep}_{f,a}(m)|$, then $l = l'$ and we have

$$|p_l| \cdots |p_0| <_{lex} |p'_l| \cdots |p'_0|.$$

Then, there exists an integer k such that $0 \le k \le l$, $|p_j| = |p'_j|$ for all integer j such that $k < j \le l$ and $|p_k| < |p'_k|$. We can prove recursively, that $p_j a_j = p'_j a'_j$ for all $j = l, \dots, k+1$. We know that $p_l a_l = p'_l a'_l$ because they are both prefixes of $f(a)$ of the same length. Let's suppose that $p_{j+1}a_{j+1} = p'_{j+1}a'_{j+1}$ for $j \in \{k+1, \dots, l-1\}$. We know that $p_j a_j$ and $p'_j a'_j$ are both prefixes of $f(a_{j+1})$ of the same length. Thus we have $p_j a_j = p'_j a'_j$. Furthermore, we know that $p_k a_k$ and $p'_k a'_k$ are both prefixes of the image under f of the same letter. This letter is a if $k = l$ and $a_{k+1} = a'_{k+1}$ if $k \ne l$. Since $|p_k| < |p'_k|$, we have that $p_k a_k$ is a prefix of p'_k. In addition, using Lemma 1.3.4 we have

$$\begin{aligned} n - m &= \sum_{i=0}^{l} |f^i(p_i)| - \sum_{i=0}^{l} |f^i(p'_i)| \\ &= \sum_{i=0}^{k} |f^i(p_i)| - \sum_{i=0}^{k} |f^i(p'_i)| \\ &< f^k(p_k a_k) - f^k(p'_k) \le 0. \end{aligned}$$

If $n < m$, we will proceed by contradiction. We suppose that $\operatorname{rep}_{f,a}(n) \not<_{gen} \operatorname{rep}_{f,a}(m)$. However, if $\operatorname{rep}_{f,a}(n) = \operatorname{rep}_{f,a}(m)$, then we know using the first point that $n = m$, which is a contradiction. If $\operatorname{rep}_{f,a}(n) >_{gen} \operatorname{rep}_{f,a}(m)$, then we obtain that $n > m$, which is a contradiction. Therefore, we can conclude that $\operatorname{rep}_{f,a}(n) <_{gen} \operatorname{rep}_{f,a}(m)$.

□

1.4 The prefix-suffix automaton

The decomposition of a prefix of the fixed point of a substitution f can be found in the prefix-suffix automaton associated with this substitution f.

Definition 1.4.1. The prefix-suffix automaton $\mathcal{A}_f$ associated with the substitution $f = (f, \Sigma, a)$ is defined as follows:

- the set of states is Σ
- there exists a directed edge from the state x to y labelled by $e = (p, x, t)$ if and only if $f(y) = pxt$.

A finite sequence $(p_i, a_i, t_i)_{i=0\dots,l}$ labels a path in the prefix-suffix automaton if and only if $f(a_{i+1}) = p_i a_i t_i$ for all $i \in \{0, \dots, l-1\}$.

Example 1.4.2. Consider the following substitution $(g, \{a, b, c\}, a)$:

$$\begin{aligned} g\colon \{a,b,c\}^* &\to \{a,b,c\}^* \\ a &\mapsto abc \\ b &\mapsto c \\ c &\mapsto a \end{aligned}$$

The corresponding prefix-suffix automaton is illustrated in Figure 1.1.

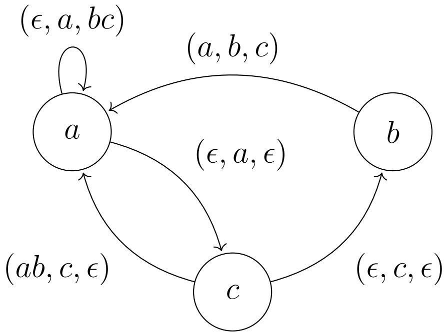


Figure 1.1: Prefix-suffix automaton for $g\colon a \mapsto abc, b \mapsto c, c \mapsto a$

Theorem 1.4.3. *Let $n \geq 1$. In the prefix-suffix automaton associated with $f = (f, \Sigma, a)$. There exists a unique path $(p_i, a_i, t_i)_{i=0,\dots,l}$ such that*

- *the path ends in a;*
- *$p_l \neq \epsilon$;*
- *$n = \sum_{j=0}^{l} |f^j(p_j)|$.*

Proof. We know, due to Theorem 1.3.7, that there exists a unique $l \in \mathbb{N}$ and a unique sequence $(p_i, a_i)_{i=0,\dots,l}$ such that this sequence is a-admissible and $p_l \neq \epsilon$ and $w_0 w_1 \cdots w_{n-1} = f^l(p_l) f^{l-1}(p_{l-1}) \cdots f^0(p_0)$, where $w_0 w_1 \cdots w_{n-1}$ is a prefix of $f^\omega(a)$ of size n.

The sequence $(p_i, a_i)_{i=0,\dots,l}$ is a admissible so for all $0 < i \leq l$, $p_{i-1}a_{i-1}$ is a prefix of $f(a_i)$. So for all $0 < i \leq l$, there exists a unique t_{i-1} such that $f(a_i) = p_{i-1}a_{i-1}t_{i-1}$. The finite sequence $(p_i, a_i, t_i)_{i=0,\dots,l-1}$ labels a path in the prefix-suffix automaton because $f(a_i) = p_{i-1}a_{i-1}t_{i-1}$ for all $0 < i \leq l$. Furthermore $(p_i, a_i)_{i=0,\dots,l}$ is a-admissible so $p_l a_l$ is a prefix of $f(a)$. So $f(a) = p_l a_l t_l$ and the path $(p_i, a_i, t_i)_{i=0,\dots,l}$ ends in a. Lastly, we have $\underbrace{|u_0 u_2 \cdots u_{n-1}|}_{=n} = \sum_{j=0}^{l} |f^j(p_j)|$. To prove the uniqueness of this path, we will proceed by contradiction. Let's suppose that we have two different paths $(p_i, a_i, t_i)_{i=0,\dots,l}$ and $(p'_i, a'_i, t'_i)_{i=0,\dots,l'}$, which verify the conditions. We know that the sequences $(p_i, a_i)_{i=0,\cdots,l}$ and $(p'_i, a'_i)_{i=0,\dots,l'}$ are a-admissible. Using Lemma 1.3.5 we know that $l = l'$ and using Lemma 1.3.6 we know that $(p_i, a_i)_{i=0,\dots,l} = (p'_i, a'_i)_{i=0,\dots,l'}$ and as before, there exists a unique t_{i-1} such that $f(a_i) = p_{i-1}a_{i-1}t_{i-1}$ for all $0 < i \leq l$ and $p_l a_l$ is a prefix of $f(a)$, so $f(a) = p_l a_l t_l$. And we have $(p_i, a_i, t_i)_{i=0,\dots,l} = (p'_i, a'_i, t'_i)_{i=0,\dots,l'}$, which is a contradiction. □

Example 1.4.4. Let us resume Example 1.4.2. We consider the substitution $(g, \{a, b, c\}, a)$, where

$$\begin{aligned} g\colon \{a,b,c\}^* &\to \{a,b,c\}^* \\ a &\longmapsto abc \\ b &\longmapsto c \\ c &\longmapsto a \end{aligned}$$

We can compute $g^\omega(a) = abccaaabcabcabcca\ldots$. The factorisation of the prefix of $g^\omega(a)$ of size 12 is $g^3(ab)g^2(\epsilon)g(a)\epsilon = g^3(ab)g(a) = abccaaabcabc$

In the prefix-suffix automaton, we are able to find the factorisation of the prefix of $g^\omega(a)$ of size 12. It is the following path in our prefix/suffix automaton represented in Figure 1.2.

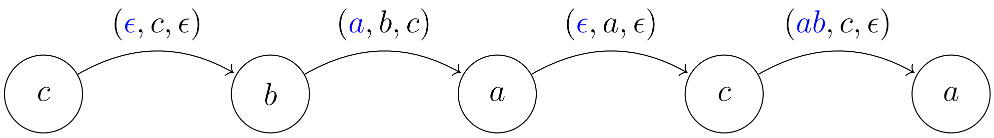


Figure 1.2: Path the prefix of $g^\omega(a)$ of size 12

1.5 Abstract numeration systems

Definition 1.5.1. An abstract numeration system (ANS) is a triple $S = (L, \Sigma, <)$ where L is an infinite regular language, called the numeration language over a totally ordered alphabet $(\Sigma, <)$. We can enumerate the words in L using the genealogical order $<$ induced by the order $<$ on A. Due to this, we have a one-to-one correspondence:

$$\text{rep}_S\colon \mathbb{N} \to L$$

is a mapping of an integer n to the $(n+1)^{th}$ word in the genealogically ordered language L. If $n \in \mathbb{N}$, then $\text{rep}_S(n)$ is the S-representation of n. The map

$$\text{val}_S \colon L \to \mathbb{N}$$

is the inverse map of rep_S. If $w \in L$, then $\text{val}_S(w)$ is the S-numerical value.

Remark 1.5.2. If $S = (L, \Sigma, <)$ is an ANS then the S-representation of 0 is the first word in the genealogically ordered L.

Example 1.5.3. Let us consider the following abstract numeration system $S = (a^*b^*, a < b)$. The genealogical order on a^*b^* gives us the following order:

$$\epsilon < a < b < aa < ab < bb < aaa < aab < abb < bbb < aaaa < \dots.$$

For example, the S-representation of 3 is $\text{rep}_S(3) = aa$ and $\text{val}_S(abb) = 8$. Remark that, the S-value is not defined on ba.

Remark 1.5.4. The Dumont–Thomas numeration system for $\mathbb{N}$ is an abstract numeration system.

Definition 1.5.5. Let $f = (f, \Sigma, a)$ be a substitution. The deterministic finite automaton associated with f is the automaton $\mathcal{A}_{f,a}$ defined as follows:

- Σ is the set of all states;
- a is the initial state;
- all states are final states;
- the alphabet $\Gamma = \{0, \dots, \max_{\sigma \in \Sigma} |f(\sigma)| - 1\}$;
- the transition function $\delta \colon \Sigma \times \Gamma \to \Sigma$ where $\delta(q, i) = f(q)_i$ if $i < |f(q)|$ otherwise, the function is undefined.

Theorem 1.5.6. *Let $f = (f, \Sigma, a)$ be a substitution and $L = \mathcal{L}(A_{f,a}) \setminus 0\,\mathbb{N}^*$. Then $S = (L, \Gamma, <)$ is an abstract numeration system such that $\text{rep}_{f,a} = \text{rep}_S$ where $\Gamma = \{0, \dots, \max_{\sigma \in \Sigma} |f(\sigma)| - 1\}$.*

Proof. The language $L = \mathcal{L}(\mathcal{A}_{f,a}) \setminus 0N^*$ is a regular language. Therefore $S = (L, \Gamma, <)$ is an abstract numeration system. We need to prove that $\text{rep}_{f,a} : \mathbb{N} \to L$ is an increasing bijection with respect to the genealogical order. Let $n, m \in \mathbb{N}$ such that $n < m$. Due to Lemma 1.3.13, we know that $\text{rep}_{f,a}(n) <_{gen} \text{rep}_{f,a}(m)$. So $\text{rep}_{f,a}$ is increasing with respect to the genealogical order and it is therefore injective. We will now prove that it is subjective.

Firstly, we can notice a link between the automaton $\mathcal{A}_{f,a}$and the prefix-suffix automaton associated with the substitution f. If there exists a directed edge from state x to y labelled by $e = (p, x, t)$ in the prefix-suffix automaton, then $f(y) = pxt$. Therefore, we have $\delta(y, |p|) = x$. We have an edge from y to x labelled $|p|$ in $\mathcal{A}_{f,a}$.

Let $i \in \mathbb{N}$ and $c_l \cdots c_0$ be the $S_{f,a}$-representation of i. However, using Theorem 1.4.3, we know that there exists only one unique path $(p_i, a_i, t_i)_{i=0,\ldots,l}$ in the prefix-suffix automaton associated with f such that

- the path ends on a;
- $p_l \neq \epsilon$;
- $i = \sum_{j=0}^{l} |f^j(p_j)|$.

This results in $\mathcal{A}_{f,a}$ in

- a path which starts on a;
- $c_l \neq 0$.

Furthermore, we have

$$f^l(p_l)f^{l-1}(p_{l-1}) \cdots f^1(p_1)p_0$$

is a prefix of length i of $f^\omega(a)$. And moreover, we notice that due to the link between the prefix-suffix automaton and $\mathcal{A}_{f,a}$, that

$$|p_j| = c_j \qquad \forall 0 \leq j \leq i.$$

Lastly, for all $j = l, \ldots, 0$ we define recursively p_j as the prefix of length c_j of $f(\delta(a, c_l \ldots c_{j+1}))$ and a_j the letter after p_j in $f(\delta(a, c_l \ldots c_{j+1}))$. To initialize the process, if $i = l$, then $c_l \ldots c_{j+1} = \epsilon$ and $\delta(a, \epsilon) = a$. The sequence $(p_i, a_i)_{i=0,\ldots,l}$ is a-admissible, so $p_l a_l$ is a prefix of $f(a)$. We know that $|p_l| = c_l$, so p_l is the prefix of $f(\delta(a, \epsilon))$ of size c_l and a_l is the $(c_l + 1)^{th}$ letter in $f(a)$. So $a_l = \delta(a, c_l)$. We can iterate this reasoning, p_{l-1} is a prefix of $f(a_l) = f(\delta(a, c_l))$ of size c_{l-1} and a_{l-1} is the $(c_{l-1} + 1)^{th}$ letter of $f(\delta(a, c_l))$. We can iterate this reasoning until we reach p_0 and a_0. We are able to determine the sequence $(p_i, a_i)_{i=0,\ldots,l}$. And we can conclude that $c_l \cdots c_0 = rep_{f,a}(i)$. □

Example 1.5.7. Let f be a substitution prolongable on a and defined as follows:

$$\begin{aligned} f\colon \{a,b,c\}^* &\to \{a,b,c\}^* \\ a &\mapsto abc \\ b &\mapsto cab \\ c &\mapsto c. \end{aligned}$$

This results in the prefix-suffix automaton illustrated in Figure 1.3:

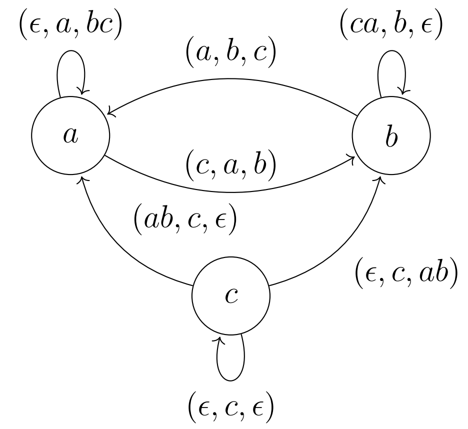


Figure 1.3: Prefix-suffix automaton for $f\colon a \mapsto abc, b \mapsto cab, c \mapsto c$

The automaton created as in the proof of Theorem 1.5.6 is illustrated in Figure 1.4.

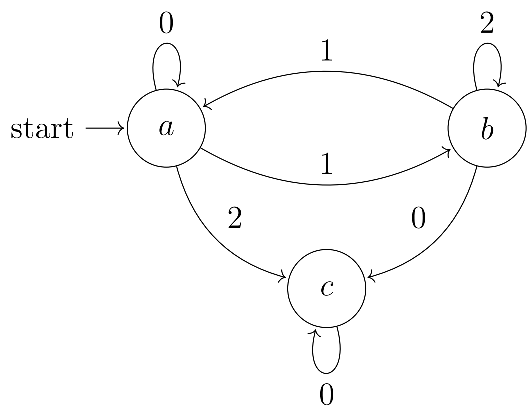


Figure 1.4: Automaton as in proof of Theorem 1.5.6 for $f\colon a \mapsto abc, b \mapsto cab, c \mapsto c$

Chapter 2

Characterisations of S-automatic sequences

In the middle of the twentieth century, Cobham introduced the notion of k-automatic sequences in relation to the classical numeration systems with an integer base $k \geq 2$. Furthermore, Cobham proved that k-automatic sequences can be characterised by sequences obtained by iterating a uniform morphism of length k.

In this chapter, we introduce a generalisation of k-automatic sequences using abstract numeration systems. After introducing the notion of S-automatic sequences for an ANS $S = (L, \Sigma, <)$. We will prove some characterisation of those sequences. We will finish this chapter by proving the equivalence between S-automatic sequences and morphic words.

2.1 First characterisations of an S-automatic sequence

Firstly, we will introduce the concept of S-automatic sequences which naturally generalise the notion of k-automatic sequences introduced by Cobham. Recall that k-automatic sequences are based on the representation of integers in base k. The representation of $i \in \mathbb{N}$ in base k using the greedy algorithm is a word w over the alphabet $\{0, \dots, k-1\}$. This word w is feed to a deterministic finite automaton with output and we obtain the i^{th}term of a k-automatic sequence.

Definition 2.1.1. Let $S = (L, \Sigma, <)$ be an ANS. The sequence $(x_i)_{i \in \mathbb{N}}$ in $\Gamma^{\mathbb{N}}$ is said to be S-automatic if there exists a DFAO $\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ such that for every $i \in \mathbb{N}$,

$$x_i = \tau(\delta(q_0, \text{rep}_S(i))).$$

We say that such a DFAO generates the sequence x .

For the rest of this chapter, we will consider $S = (L, \Sigma, <)$ an abstract numeration system.

Before studying the characterisation of S-automatic sequence using morhic words, we will look at some other characterisations of S-automatic sequence.

Definition 2.1.2. A sequence $(x_i)_{i\in\mathbb{N}}$ is called reversal-S-automatic if there exists a DFAO $\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ such that for every $i \in \mathbb{N}$ we have

$$x_i = \tau(\delta(q_0, \text{rep}_S(i)^R)).$$

Theorem 2.1.3. *If a sequence $x = (x_i)_{i\in\mathbb{N}}$ is S-automatic for some abstract numeration system S then it is also reversal-S-automatic for the same abstract numeration system. In particular, a sequence $(x_i)_{i\in\mathbb{N}}$ is S-automatic if and only if it is reversal-S-automatic.*

Proof. Let $\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ the DFAO that generates the sequence $(x_i)_{i\in\mathbb{N}}$. Due to the definition of a DFAO, we know that Γ is finite, so $\Gamma = \{\gamma_1, \dots, \gamma_r\}$ for $r \in \mathbb{N}_0$. If $r = 1$ then the result is direct. If $r \geq 2$ we define for each $i \in \{1, \dots, r\}$ a regular language L_i accepted by the DFA $\mathcal{A}_i = (Q, q_0, \Sigma, \delta, F_i)$ where $F_i = \{q \in Q | \tau(q) = \gamma_i\}$. We notice that because we are working with complete deterministic automaton that $\forall i, j \in \{1, \dots, r\}$, we have

$$L_i \cap L_j = \emptyset \text{ and } \cup_{i=1}^r L_i = \Sigma^*. \tag{2.1}$$

As L_i is a regular language, $L_i^R = \{w^R | w \in L_i\}$ for all $i \in \{1, \dots, r\}$ is also a regular language as regular languages are stable under reversal. Therefore, for each L_i^R there exists a DFA $\mathcal{A}'_i = (Q_i, q_{0,i}, \Sigma, \delta_i, F'_i)$ accepting this language. Using those languages we create a DFAO $\mathcal{A}' = (Q', q'_0, \Sigma, \delta', \Gamma, \tau')$ as follows:

- $Q' = Q_1 \times \cdots \times \mathbb{Q}_r$;
- $q'_0 = (q_{0,1}, \dots, q_{0,r})$;
- $\delta'((q_1, \dots, q_r), a) = (\delta_1(q_1, a), \dots, \delta_r(q_r, a))$ if $(q_1, \dots, q_r) \in Q'$ and $a \in \Sigma$;
- $\tau'(q_1, \dots, q_r) = \gamma_i$ if $q_i \in F'_i$ with $q_1 \in Q_1, \dots, q_r \in Q_r$.

We need to verify that τ' is well defined. Let $w \in \Sigma^*$ and $(q_1, \dots, q_r) = \delta'(q'_0, w)$, we notice that due to the construction of $\mathcal{A}'$ and (2.1) there exists $i \in \{1, \dots, r\}$ such that $q_i \in F'_i$ and for all $j \in \{1, \dots, r\} \setminus \{i\}$, $q_j \notin F'_j$. Furthermore, the definition of τ' on the other states, $(q_1, \dots, q_r)$ of $\mathcal{A}'$ which can not be expressed as $(q_1, \dots, q_r) = \delta'(q'_0, w)$ with $w \in \Sigma^*$ is irrelevant because these states are not accessible. So τ' is well defined. And it is clear by construction that for each $w \in \Sigma^*$ we have

$$\tau(\delta(q_0, w)) = \tau'(\delta'(q_0, w^R)).$$

Lastly, as the reverse of a reversed word is the word itself, we can obtain the logical connection between the two statements. □

Another characterisation can be obtained using a S-fiber, which is defined as follows.

Definition 2.1.4. Let $a \in \Gamma$ and $S = (L, \Sigma, <)$, the S-fiber $\mathcal{F}_S(x, a)$ of a sequence $x = (x_i)_{i\in\mathbb{N}} \in \Gamma^N$ is defined as

$$\mathcal{F}_S(x, a) = \{\text{rep}_S(i) : x_i = a\}.$$

Theorem 2.1.5. *Let $x = (x_i)_{i\in\mathbb{N}}$ be a sequence over the alphabet Γ. The sequence x is S-automatic if and only if for all $a \in \Gamma$, $\mathcal{F}_S(x, a)$ is a regular subset of L.*

Proof. If x is S-automatic then we have a DFAO $\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ which generates the infinite sequence x. We create a DFA $\mathcal{A}' = (Q, q_0, \Sigma, \delta, F)$, where $F = \{q \in Q | \tau(q) = a\}$ is the set of final states. We denote $\mathcal{L}(\mathcal{A}')$ the language accepted by this DFA. Therefore, $\mathcal{F}_S(x, a)$ is a regular subset of L as it is the intersection between the two regular languages $\mathcal{L}(\mathcal{A}')$ and $\mathcal{A}$. Let $\Gamma = \{a_1, \dots, a_j\}$ for $j \in \mathbb{N}_0$. We notice that if $k, l \in \{1, \dots, j\}$ such that $k \neq l$, then by definition of the S-fiber, we have $\mathcal{F}_S(x, a_k) \cap \mathcal{F}_S(x, a_l) = \emptyset$ and $L = \bigcup_{i=1}^{j} \mathcal{F}_S(x, a_i)$. The S-fiber is a regular language for all a_k where $k \in \{1, \dots, j\}$. So for each of the S-fiber, there exists a DFA $\mathcal{A}_k = (Q_k, q_{0,k}, \Sigma, \delta_k, F_k)$ which accepts this S-fiber. From these automaton we create the DFAO $\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ as follows:

- $Q = Q_1 \times \cdots \times Q_1$;
- $q_0 = (q_{0,1}, \dots, q_{0,j})$;
- $\delta((q_1, \dots, q_j), \sigma) = (\delta_1(q_1, \sigma), \dots, \delta_j(q_j, \sigma))$ if $q_1 \in Q_1, \dots q_j \in Q_j$ and $\sigma \in \Sigma$;
- $\tau(q_1, \dots, q_j) = a_k$ if there is a unique k such that $q_k \in F_k$.

As before, τ is well defined because $\mathcal{F}_S(x, a_k) \cap \mathcal{F}_S(x, a_l) = \emptyset$ and $L = \bigcup_{i=1}^{j} \mathcal{F}_S(x, a_i)$ and if a state can not be reached by a word $w \in L$ then the associated output is meaningless for us. The sequence x is generated from $\mathcal{A}$ and $S = (L, \Sigma, <)$. □

2.2 The equivalence between morphic sequences and S-automatic sequences

We will now characterise S-automatic sequences using morphic sequences. A useful tool for this is the following Lemma.

Lemma 2.2.1. *Let $\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$ be a DFA so that Σ has a fixed order* i.e. *we have $\Sigma = \{a_1 < \ldots < a_j\}$. Let $\$ \notin Q$. We define the following morphism*

$$\varphi \colon Q \cup \{\$\} \to (Q \cup \{\$\})^*$$
$$q \mapsto \begin{cases} \$q_0 & \text{if } q = \$ \\ \delta(q, a_1) \cdots \delta(q, a_j) & \text{otherwise.} \end{cases}$$

This substitution produces the sequence x_φ of the states reached by the words of Σ^* i.e. *for all $k \in \mathbb{N}_0$, $(x_\varphi)_{k+1} = \delta(q_0, w_k)$, where w_k is the k^{th} element of $(\Sigma^*, <)$.*

Proof. We will study the prefixes of $\varphi^\omega(\$)$ of the form $\varphi^i(\$)$ for $i \in \mathbb{N}_0$. We have $\varphi(\$) = \alpha q_0$, and therefore we have

$$\varphi^2(\$) = \varphi(\$)\varphi(q_0)$$

$$=\varphi(\$)\delta(q_0, a_1) \cdots \delta(q_0, a_j).$$

So the treatment is true for $k \in \{1, \ldots, j\}$. Moreover we notice due to the form of φ,

$$\varphi^{i+1}(\$) = \varphi^i(\$)\delta(q_0, w'_{i,1}) \ldots \delta(q_0, w'_{i,j'})$$

where $w'_{i,1}, \ldots, w'_{i,j'}$ are all the words in Σ^* of size i in genealogical order. □

In [35], Rigo and Maes have announced the following lemma

Lemma 2.2.2. *If f and g are two morphisms such that $f(g^\omega(a))$ is an infinite word, then there exist is a non-erasing morphism k prolongable on a and a coding h is a coding such that*

$$f(g^\omega(a)) = h(k^\omega(a)).$$

Theorem 2.2.3. *Every S-automatic sequence is morphic.*

Proof. Let $S = (L, \Sigma, <)$ be an abstract numeration system. Let $\mathcal{A}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ be a DFA accepting the language L and x an S-automatic sequence obtained from the DFAO $\mathcal{A}_x = (Q_x, q'_{0,x}, \Sigma, \delta_x, \Gamma_x, \tau_x)$. Using those two automata, we will create the product automaton $\mathcal{A} = (Q, q_0, \Sigma, \nu, F)$ as follows

- $Q = Q_L \times Q_x$;
- $q_0 = (q_{0,L}, q_{0,x})$;
- the transition function:

$$\begin{aligned} \nu \colon Q \times \Sigma &\to \mathbb{Q} \\ ((q, q'), a) &\mapsto (\delta_L(q, a), \delta_x(q', a)); \end{aligned}$$

- we don't need the final states of $\mathcal{A}$, so we will not explicitly write them.

Let $\$ \notin Q$, we associate to the automaton $\mathcal{A}$ the substitution φ defined as in Lemma 2.2.1:

$$\begin{aligned} \varphi \colon Q \cup \{\$\} &\to (Q \cup \{\$\})^* \\ q &\mapsto \begin{cases} \$q_0 & \text{if } q = \$ \\ \delta(q, a_1) \cdots \delta(q, a_j) & \text{otherwise.} \end{cases} \end{aligned}$$

Using Lemma 2.2.1, we know that this substitution produces the sequence of states reached by the words of Σ^*. However, we are only interested in the words of Σ^* in L. Therefore, we define the erasing substitution h as follows:

$$\begin{aligned} h \colon Q \cup \{\$\} &\to \Gamma \\ q &\mapsto \begin{cases} \epsilon & \text{if } q = \$ \\ \epsilon & \text{if } q = (q', q'') \in Q \text{ and } q' \notin F_L \\ \tau(q'') & \text{if } q = (q', q'') \in Q \text{ and } q' \in F_L. \end{cases} \end{aligned}$$

The sequence x is equal to $h(\varphi^\omega(\$))$ and using Lemma 2.2.2, we can conclude. □

Theorem 2.2.4. *Every morphic sequence is S-automatic for some abstract numeration system S.*

Proof. Let x be a morphic word. Due to Lemma 2.2.2 we can suppose that a non-erasing substitution (k, Σ, a) prolongable on a letter a and a coding h such that

$$x = h(k^\omega(a)).$$

We create the following DFA $\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$ defined as follows:

- the set of all states Q is Σ;
- $q_0 = a$;
- $\Delta = \{\sigma_1, \ldots, \sigma_r\}$ with $r = \max_{b \in A} |k(b)|$;
- the transition function:

$$\begin{aligned} \delta\colon Q \times \Delta &\to Q \\ (u, \sigma_i) &\mapsto k(u)_{i-1} \qquad\qquad \text{if } 1 \leq i \leq |k(u)| \end{aligned}$$

- all states are final states so $F = Q$.

We denote the language L accepted by this DFA. Since k is prolongable on the letter a, we have $\delta(a, \sigma_1) = a$. So if w is a word in L, the word $\sigma_1 w$ is also a word in L. Applying Lemma 2.2.1 to the created DFA, we know that there exists the following morphism

$$\begin{aligned} \varphi\colon Q \cup \{\$\} &\to ((Q \cup \{\$\})^*) \\ q &\longmapsto \begin{cases} \$q_0 & \text{if } q = \$ \\ \delta(q, \sigma_1) \cdots \delta(q, \sigma_r) & \text{otherwise.} \end{cases} \end{aligned}$$

This substitution produces the sequence of states reached by the words Σ^*. We know that k is prolongable on the letter a so $k(a)$ starts with the letter a and x_φ starts, by the construction of φ, with the factor $\$q_0$ and due to the construction of the transition function of the DFA $\mathcal{A}$ the prefix $\$q_0$ is followed by q_0 as $\delta(q_0, \sigma_1) = q_0$. This second occurrence of a q_0 creates throughout x_φ words starting with σ_1 due to the construction of φ. However, those factors do not appear in $k^\omega(a)$ by construction of a fixed point. (An example of this can be found below in Example 2.2.5.) By removing those factors, we obtain the infinite word $\$k^\omega(a)$ due to the construction of δ. Furthermore, if we also remove \$, we obtain $k^\omega(a)$. The language $L \setminus \sigma_1\Delta^*$ is a regular language. Let S be the corresponding abstract numeration system to $L \setminus \sigma_1\Delta^*$ and $\mathcal{A}'$ a DFAO built on the $\mathcal{A}$ such that for each state $a \in \Sigma$ the output is $h(a)$ and we obtain that x is S-automatic. □

To better understand the reasoning in the proof of Theorem 2.2.4 we have the following example.

Example 2.2.5. Let's suppose that $\Sigma = \{a, b, c\}$ and $\Gamma = \{0, 1, 2\}$. We define the substitutions:

$$k\colon \Sigma \to \Sigma^+\colon \begin{cases} a \mapsto abc \\ b \mapsto bc \\ c \mapsto aac \end{cases}$$

$$h\colon \Sigma \to \Gamma\colon \begin{cases} a \mapsto 0 \\ b \mapsto 1 \\ c \mapsto 2. \end{cases}$$

We notice that k is non-erasing and prolongable on the letter a and h is 1-uniform. We create the automaton as in the proof of Theorem 2.2.4.

We have $\mathcal{A} = (Q, q_0, \Delta, \delta, F)$:

- the set of all states is $Q = \{a, b, c\}$;
- $q_0 = a$;
- $\Delta = \{\sigma_1, \sigma_2, \sigma_3\}$ because $\max_{b\in A} |k(b)| = 3$;
- the transition function:

$$\begin{aligned} \delta\colon Q \times \Delta &\to Q \\ (u, \sigma_i) &\mapsto k(u)_{i-1} \qquad\qquad \text{if } 1 \leq i \leq |k(u)|; \end{aligned}$$

- all states are final states so $F = Q$.

We have that automaton represented in Figure 2.1

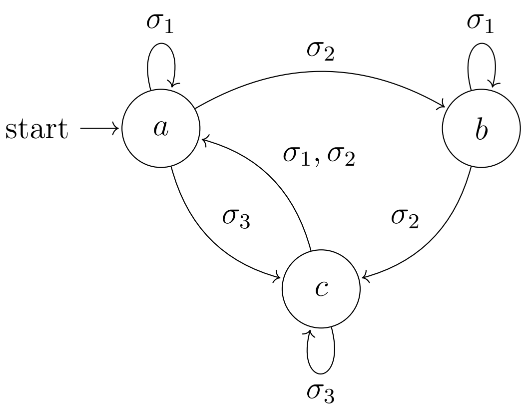


Figure 2.1: The automaton associated to $\mathcal{A}$

The substitution obtained from $\mathcal{A}$ as in Lemma 2.2.1 is

$$\varphi\colon (\{a,b,c\}\cup\{\$\}) \to ((\{a,b,c\}\cup\{\$\})^{*})\colon \begin{cases} \$ \mapsto \$a \\ a \mapsto k(a) = abc \\ b \mapsto k(b) = bc \\ c \mapsto k(c) = aac. \end{cases}$$

If we compare $\varphi^{\omega}(\$)$ and $k^{\omega}(a)$.

$$k^{\omega}(a) = abcbcaacbcaacabcabca\ldots$$
$$\varphi^{\omega}(\$) = \$aabcabcbcaacabcncaacbcaacabcabca\ldots$$

We notice that if we remove the blue factors of $\varphi^{\omega}(\$)$ that it would be equal to $k^{\omega}(a)$. On this example it is also easy to see, that the 3^{rd} blue factor is the image of the 2^{nd} factor in blue of k and the 4^{th} is the image of the 3^{rd}.

As a consequence of Theorem 2.2.3 and Theorem 2.2.4, we have the following Theorem.

Theorem 2.2.6. *A sequence $x = (x_i)_{i\in\mathbb{N}}$ is generated by a substitution f if and only if it is $S-$automatic for some* ANS $S = (L, \Sigma, <)$.

Chapter 3

A study of substitutions, numeration systems and automata associated with a Parry number

In the first chapter 1, we introduced a few numeration systems. In this chapter, we will now focus ourselves on specific positional numeration system, namely the θ-numeration system. To be able to define this positional numeration system, we will define a representation of a real number strictly bigger than 1. In this chapter, θ will be such a real number and we will explore the θ-expansion of 1 denoted $D_\theta(1)$, which is an infinite sequence of integers. If θ is a Parry number then this sequence can have one of the following forms:

- it can either be finite, meaning that after $(n+1)^{th}$ element in our sequence, all the elements following it will be equal to 0, with $n \in \mathbb{N}$;
- or it can be ultimately periodic.

3.1 Introduction to positional numeration systems associated with a Parry number

Definition 3.1.1. Let $\theta \in \mathbb{R}$ such that $\theta > 1$. The θ-expansion of 1 (noted $D_\theta(1)$) is the infinite sequence of integers $(\alpha_i)_{i\in\mathbb{N}}$ such that

$$\begin{aligned}
&\alpha_0 = \lfloor \theta \rfloor \\
&r_0 = \{\theta\} \\
&\text{and } \forall i \in \mathbb{N}, \alpha_{i+1} = \lfloor \theta r_i \rfloor, r_{i+1} = \{\theta r_i\}
\end{aligned}$$

where $\lfloor \delta \rfloor$ is the integer part of δ and $\{\delta\}$ is the fractional part of δ.

Definition 3.1.2. A real number $\theta > 1$ is called a *Parry*-number if the θ-expansion of 1 is ultimately periodic or finite. Furthermore, it is called a simple Parry number if $D_\theta(1)$ is finite.

In [30], Parry studied Parry numbers and their corresponding θ-expansion of 1. In particular, he was able to prove that if a real number $\theta > 1$ with a θ-expansion of 1 $D_\theta(1) = \alpha_0\alpha_1\cdots$ with specific conditions, then a sufficient condition for the existence of θ' with $\theta'-$expansion of 1 being $D_{\theta'}(1) = \beta_0\beta_1\cdots$ is that $\beta_i\beta_{i+1}\cdots <_{lex} \alpha_0\alpha_1\cdots$ for all $i \in \mathbb{N}_0$. As a result of this theorem we have the following theorem.

Theorem 3.1.3. *Let θ be a Parry number. The θ-expansion of 1 (noted $D_\theta(1)$) verifies the conditions of Parry which are the following:*

- *If $D_\theta(1) = \alpha_0\cdots\alpha_n$ for $n \in \mathbb{N}$ then*

$$\begin{aligned}
&\alpha_0\cdots\alpha_n >_{lex} \alpha_1\cdots\alpha_n\\
&\alpha_0\cdots\alpha_n >_{lex} \alpha_2\cdots\alpha_n\\
&\vdots\\
&\alpha_0\cdots\alpha_n >_{lex} \alpha_n;
\end{aligned}$$

- *If $D_\theta(1) = \alpha_0\ldots\alpha_n(\alpha_{n+1}\ldots\alpha_{n+m})^\omega$ for $n \in \mathbb{N}$ then*

$$\begin{aligned}
&\alpha_0\cdots\alpha_{n+m} >_{lex} \alpha_1\cdots\alpha_{n+m}\alpha_{n+1}\\
&\alpha_0\cdots\alpha_{n+m} >_{lex} \alpha_2\cdots\alpha_{n+m}\alpha_{n+1}\alpha_{n+2}\\
&\vdots\\
&\alpha_0\cdots\alpha_{n+m} >_{lex} \alpha_k\cdots\alpha_{n+m}\alpha_{n+1}\cdots\alpha_{n+k} \pmod{m}\\
&\vdots\\
&\alpha_0\cdots\alpha_{n+m} >_{lex} \alpha_{n+m}\alpha_{n+1}\cdots\alpha_{n+m} \pmod{m}
\end{aligned}$$

with $>_{lex}$ being the lexicographic order.

Definition 3.1.4. Let θ be a Parry number.
The polynomial $Q_\theta(x) \in \mathbb{Z}[x]$ is defined from the θ-expansion of 1 as follows:

- If $D_\theta(1) = \alpha_0\cdots\alpha_n$ then $Q_\theta(x) = x^{n+1} - \alpha_0x^n - \cdots - \alpha_n$.
  The linear recurrent sequence of integers $U = (U_i)_{i\in\mathbb{N}}$ associated with $Q_\theta(x)$ is defined by:

$$\begin{aligned}
&U_0 = 1,\\
&U_i = \alpha_0U_{i-1} + \alpha_1U_{i-2} + \cdots + \alpha_{i-1}U_0 + 1 && \text{if } 1 \le i \le n\\
&U_i = \alpha_0U_{i-1} + \alpha_1U_{i-2} + \cdots + \alpha_nU_{i-(n+1)} && \text{if } i > n;
\end{aligned}$$

- If $D_\theta(1) = \alpha_0\cdots\alpha_n(\alpha_{n+1}\ldots\alpha_{n+m})^\omega$ then

$$Q_\theta(x) = x^{n+m+1} - \alpha_0 x^{n+m} - \cdots - \alpha_{n+m} - (x^{n+1} - \alpha_0 x^n - \cdots - \alpha_n)$$

The linear recurrent sequence of integers $U = (U_i)_{i \in \mathbb{N}}$ associated with $Q_\theta(x)$ is defined by:

$$\begin{aligned} &U_0 = 1, \\ &U_i = \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{i-1} U_0 + 1 && \text{if } i \geq 1 \end{aligned}$$

Remark 3.1.5. Notice that $(U_i)_{i \in \mathbb{N}}$ is a position numeration system (as introduced in chapter 1), with $\Sigma = \{0, \cdots, [\theta]\}$ being the numeration alphabet. By definition, each integer $j \in \mathbb{N}$ admits one unique U-representation w. Furthermore, w does not start with the letter 0 and does not contain any factor which are equal or exceed $D_\theta(1)$ in the lexicographic order. The numeration language is denoted $\text{rep}_U(\mathbb{N})$. We call this positional numeration system the θ numeration system.

Example 3.1.6. Let θ be the golden ratio $\dfrac{1+\sqrt{5}}{2}$. Firstly, we want to calculate the θ-expansion of 1. We have

$$\begin{aligned} \alpha_0 &= \lfloor \frac{1+\sqrt{5}}{2} \rfloor = 1 \\ r_0 &= \left\{ \frac{1+\sqrt{5}}{2} \right\} = \frac{-1+\sqrt{5}}{2} \end{aligned}$$

and

$$\begin{aligned} \alpha_1 &= \lfloor \frac{1+\sqrt{5}}{2} r_0 \rfloor = 1 \\ r_1 &= \left\{ \frac{1+\sqrt{5}}{2} r_0 \right\} = 0 \end{aligned}$$

We obtain the simple θ-expansion $D_\theta(1) = 11$. Furthermore, we want to obtain the base of the positional numeration system.

$$\begin{aligned} &U_0 = 1 \\ &U_1 = \alpha_0 U_0 + 1 = 2 \\ &U_{i+2} = \alpha_0 U_{i+1} + \alpha_1 U_i = U_{i+1} + U_i && \forall i \in \mathbb{N} \end{aligned}$$

This is the Fibonacci sequence, which we already encountered in Example 1.3.8.

Example 3.1.7. Let θ be the golden ratio to the power of 2, we have $\theta = \left(\frac{1+\sqrt{5}}{2}\right)^2 > 1$. Firstly, we want to calculate the θ-expansion of 1. We have

$$\alpha_0 = \lfloor \left(\frac{1+\sqrt{5}}{2}\right)^2 \rfloor = \lfloor \frac{3+\sqrt{5}}{2} \rfloor] = 2$$
$$r_0 = \left\{\frac{3+\sqrt{5}}{2}\right\} = \frac{-1+\sqrt{5}}{2}$$

and

$$\alpha_1 = \lfloor \frac{3+\sqrt{5}}{2} r_0 \rfloor = 1$$
$$r_1 = \left\{\frac{3+\sqrt{5}}{2} r_0\right\} = \frac{-1+\sqrt{5}}{2}$$

Since $r_1 = r_0$, we obtain the θ-expansion $D_\theta(1) = 2(1)^\omega$. Furthermore, we want to obtain the base of the positional numeration system.

$$U_0 = 1$$
$$U_1 = \alpha_0 U_0 + 1 = 3$$
$$U_{i+2} = \alpha_0 U_{i+1} + \cdots + \alpha_{i+1} U_0 + 1 = 2U_{i+1} + U_i + \cdots U_0 + 1 \qquad \forall i \in \mathbb{N}\,.$$

3.2 Some properties of a θ-automaton

The θ-expansion has been studied in a lot of articles, among others in [12] and [15]. In [12], the link between a θ-automaton and a θ-substitution has been studied.

In this section, we will introduce the notion of a θ-automaton and some of its properties.

Definition 3.2.1. Let θ be a real number bigger than 1 such that $D_\theta(1) = \alpha_0 \cdots \alpha_n$. The automaton associated with θ is $\Delta_\theta = (Q, q_0, F, \Sigma, \delta)$ where

- Q is a set containing $n+1$ states which we will denote $q_0, q_1, \ldots, q_n$;
- q_0 is the initial state;
- $F = \{q_0, q_1, \ldots q_n\}$;
- the transition function is defined by

$$\delta\colon Q \times \Sigma \to Q$$
$$(q_i, a) \longmapsto \begin{cases} q_0 \text{ if } a < \alpha_{i+1} \\ q_{i+1 \pmod n} \text{ if } a = \alpha_{i+1} \\ p \text{ otherwise.} \end{cases}$$

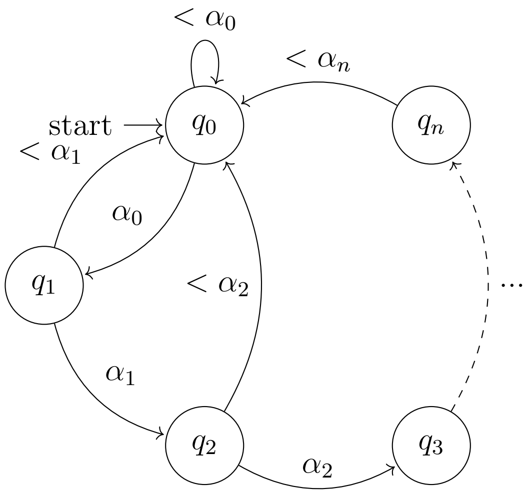


Figure 3.1: The Parry automaton associated with θ where $D_\theta(1) = \alpha_0 \cdots \alpha_n$

The corresponding automaton is illustrated in Figure 3.1. This automaton is called the Parry automaton.

Similarly, we are able to define the automaton associated with θ if

$$D_\theta(1) = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^\omega.$$

Definition 3.2.2. Let θ be a real number bigger than 1 such that

$$D_\theta(1) = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^\omega.$$

The automaton associated with θ is the automaton $\Delta_\theta = (Q, q_0, F, \Sigma, \delta)$ where

- Q is a set containing $n + m + 1$ states which we will denote $q_0, q_1, \ldots q_{n+m}$;
- q_0 is the initial state;
- $F = \{q_0, q_1, \ldots, q_{n+m}\}$;
- the transition function is defined by

$$\delta\colon Q \times \Sigma \to Q$$
$$(q_i, a) \longmapsto \begin{cases} q_0 \text{ if } a < \alpha_{i+1} \\ q_{(i+1)} \text{ if } a = \alpha_{i+1} \text{ and } i < n+m \\ q_{(n+1)} \text{ if } a = \alpha_{n+1} \text{ and } i < n+m \\ q \text{ otherwise .} \end{cases}$$

The corresponding automaton is illustrated in Figure 3.2. This automaton is called the Parry automaton.

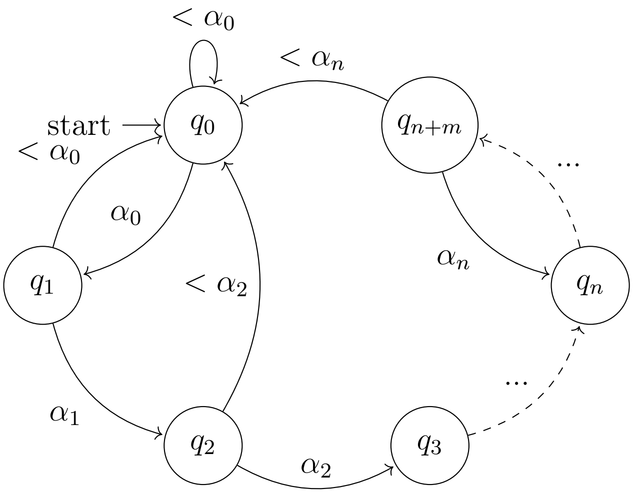


Figure 3.2: The Parry automaton associated with θ where $D_\theta(1) = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^\omega$

Definition 3.2.3. Let T be a set of non-negative integers. The language created with the set of U-representations of T in the θ numeration system is denoted $rep_U(T)$.

The set T is said to be U-recognisable if $rep_U(T)$ is accepted by an automaton with a finite number of states.

By theorem 3.1.3, we have

Proposition 3.2.4. *The language* $0^* \operatorname{rep}_U(\mathbb{N})$ *is accepted by the Parry automaton. In particular, this language is recognisable by a finite automaton and therefore* $\mathbb{N}$ *is* U*-recognisable.*

Let θ be a Parry number and T a set of non-negative integers. We consider the following two automata

- the Parry automaton associated with θ, $\Delta_\theta = (Q', q_0', F', \Sigma, \delta)$
- an automaton $\mathcal{A}_T = (P, p_0, F'', \Sigma, \delta')$ such that $\delta(p_0, 0) = p_0$ and $\mathcal{A}_T$ accepts $\operatorname{rep}_U(T)$.

Using those two automata, we create the product automaton $\mathcal{A} = (Q, q_0, F, \Sigma, \delta_\theta)$ where

- $Q = Q' \times P$
- $q_0 = (q_0', p_0)$
- $F = (q, p)$, where $(q, p) \in Q' \times F''$.
- the transition function $\delta((q, p), a) = (\delta(q, a), \delta'(p, a))$ for $a \in \Sigma$ if $\delta(q, a)$ and $\delta'(p, a))$ are defined.

We obtain an automaton $\mathcal{A}$ recognising $\operatorname{rep}_U(T)$ and verifying

- $\delta_\theta(q_0, w)$ is defined in $\mathcal{A}$ if and only if w is a word in $0^* rep_U(\mathbb{N})$;
- $\delta_\theta(q_0, 0) = q_0$.

This construction motivated the definition of a θ-automaton.

Definition 3.2.5. An automaton $\mathcal{A} = (Q, q_0, F, \Sigma, \delta_\theta)$ is a θ-automaton if

- $\delta_\theta(q_0, w)$ is defined in $\mathcal{A}$ if and only if w is a word in $0^* rep_U(\mathbb{N})$;
- $\delta_\theta(q_0, 0) = q_0$.

We notice that each U-recognisable set of $\mathbb{N}$ is accepted by a θ-automaton because we can always construct a product automaton as above for each U-recognisable set of integers. And the constructed product automaton verifies the conditions of a θ-automaton.

Example 3.2.6. Let's reconsider Example 3.1.6 where θ is the golden ratio $\dfrac{1+\sqrt{5}}{2}$ and the θ-expansion of 1 is $D_\theta(1) = 11$. And we consider the set T of integers such that their U-representation contains an odd number of 1. The Parry automaton associated with θ is represented in Figure 3.3 and

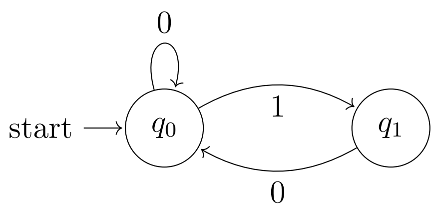


Figure 3.3: The Parry automaton associated with θ being the golden ratio

the automaton, which accepts $\text{rep}_U(T)$ is represented in Figure 3.4.

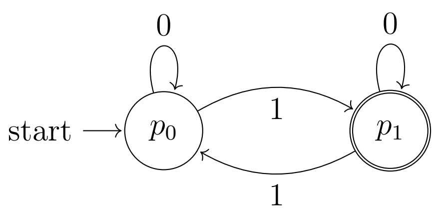


Figure 3.4: The automaton accepting $\text{rep}_U(T)$

We obtain a θ-automaton represented in Figure 3.5.

Example 3.2.7. Let's reconsider Example 3.1.7. Let θ be the golden ratio to the power of 2, we have $\theta = \left(\dfrac{1+\sqrt{5}}{2}\right)^2 > 1$ and the θ-expansion of 1 is $D_\theta(1) = 2(1)^\omega$. And we

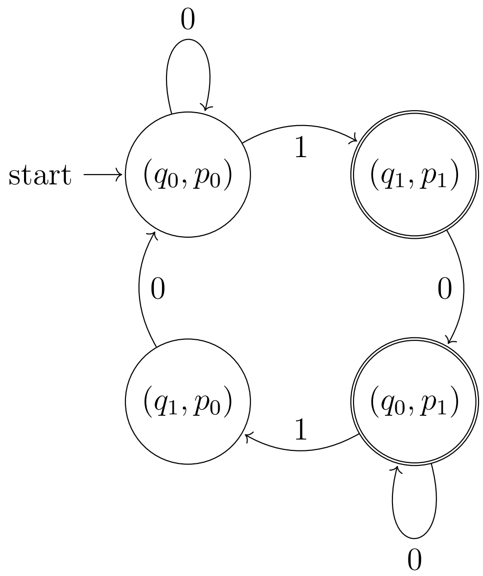


Figure 3.5: A θ-automaton for he golden ratio

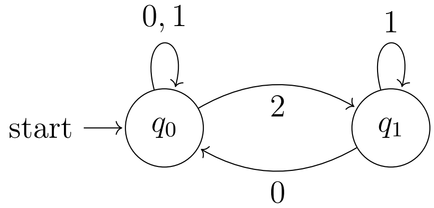


Figure 3.6: The Parry automaton associated with θ being the golden ratio to the power of 2

consider the set T of integers such that their U-representation contains an odd number of 1. The Parry automaton associated with θ is represented in Figure 3.6. And a automaton, which accepts $\mathrm{rep}_U(T)$ is represented in Figure 3.7. We obtain a θ-automaton represented in Figure 3.8.

A direct consequence of the definition of a θ-automaton, more precisely the fact that a θ-automaton does not accept any word which contain a factor which is equal to or bigger than D_θ in lexicographic order, is the following Lemma.

Lemma 3.2.8. *Let θ be a simple Parry number ($D_\theta(1) = \alpha_0 \cdots \alpha_n$) and $\mathcal{A}$ a corresponding θ-automaton. Let p be a state of $\mathcal{A}$:*

1. *if there exists a state p' and a word $w = v\alpha_0 \cdots \alpha_i \in 0^* \mathrm{rep}_U(\mathbb{N})$ such that $i \in \{0, \ldots, n-2\}$ and $\delta_\theta(p', w) = p$, then the transition $\delta_\theta(p, a)$ exists in $\mathcal{A}$ if $a \in \{0, \ldots, \alpha_{i+1}\}$;*
2. *if there exists a state p' and a word $w = v\alpha_0 \cdots \alpha_{n-1} \in 0^* \mathrm{rep}_U(\mathbb{N})$ such that $\delta_\theta(p', w) = p$, then the transition $\delta_\theta(p, a)$ exists in $\mathcal{A}$ if $a \in \{0, \ldots, \alpha_n - 1\}$;*

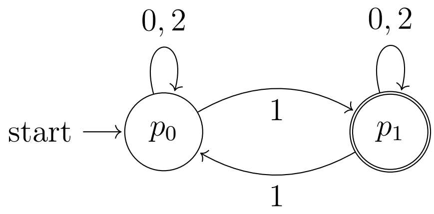


Figure 3.7: Automaton accepting $\mathrm{rep}_U(T)$ where θ is the golden ratio to the power of 2

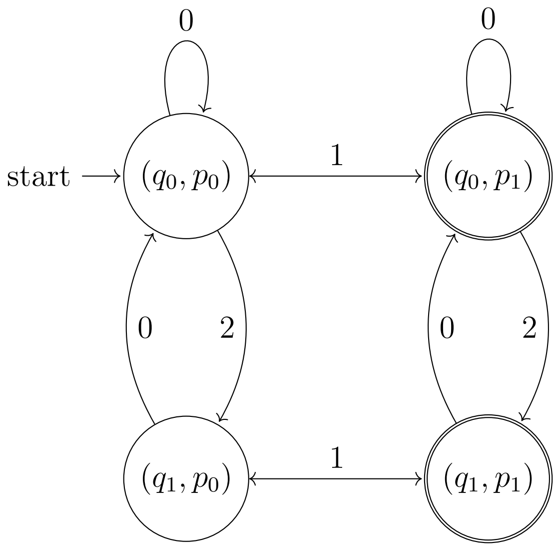


Figure 3.8: A θ-automaton for the golden ratio to the power of 2

3. *if for all states p' and all words $w \in 0^* \mathrm{rep}_U(\mathbb{N})$ such that $\delta_\theta(p', w) = p$ and w does not contain any suffixes of the form $\alpha_0 \cdots \alpha_i$ with $i \in \{0, \dots, n-1\}$, then for all $a \in \Sigma$, the transition $\delta_\theta(p, a)$ exists.*

Similarly, if D_θ is eventually periodic, we have:

Lemma 3.2.9. *Let θ be a non simple Parry number ($D_\theta = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^\omega$) and $\mathcal{A}$ a corresponding θ-automaton. Let p be a state of $\mathcal{A}$:*

1. *if there exists a state p' and a word $w = v\alpha_0 \cdots \alpha_i \in 0^* \mathrm{rep}_U(\mathbb{N})$ such that $i \in \{0, \dots, n+m-1\}$ (resp. $w = v\alpha_0 \dots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^k \alpha_1 \cdots \alpha_j \in 0^* \mathrm{rep}_U(\mathbb{N})$, $k \in \mathbb{N}$ such $j \in \{1, \dots, m-1\}$) and $\delta_\theta(p', w) = p$, then the transition $\delta_\theta(p, a)$ exists in $\mathcal{A}$ if $a \in \{0, \dots, \alpha_{i+1}\}$ (resp. $a \in \{0, \dots, \alpha_{n+j+1}\}$);*

2. *if there exists a state p' and a word $w = v\alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^k \in 0^* \mathrm{rep}_U(\mathbb{N})$, $k \in \mathbb{N}$ such that $\delta_\theta(p', w) = p$, then the transitions $\delta_\theta(p, a)$ exists in $\mathcal{A}$ if $a \in \{0, \dots, \alpha_{n+1}\}$;*

3. *if for all states p' and and all words $w \in 0^* \mathrm{rep}_U(\mathbb{N})$ such that $\delta_\theta(p', w) = p$ and w does not contain any suffixes of the form $\alpha_0 \dots \alpha_i$, then for all $a \in \Sigma$, the transition $\delta_\theta(p, a)$ exists.*

Proposition 3.2.10. *Let θ be a simple Parry number ($D_\theta = \alpha_0 \dots \alpha_n$) and $\mathcal{A} = (Q, q_0, F, \Sigma, \delta_\theta)$ a corresponding θ-automaton. Then*

1. *if p, p' and p'' are states of $\mathcal{A}$ and i, $j \in \{0, \dots, n-2\}$ such that*
$$\delta_\theta(p, \alpha_0 \cdots \alpha_i) = \delta_\theta(p', \alpha_0 \dots \alpha_j) = p''$$
then $i = j$;
2. *if p, p' and p'' are states of $\mathcal{A}$ and $i \in \{0, \dots, n-1\}$ such that*
$$\delta_\theta(p, \alpha_0 \cdots \alpha_i) = \delta_\theta(p', \alpha_0 \cdots \alpha_{n-1}) = p''$$
then $i = n-1$;
3. *if p and p' are states of $\mathcal{A}$ and $i \in \{0, \dots, n-1\}$ such that $\delta_\theta(p, \alpha_0 \dots \alpha_i) = p'$ than there does not exist a state p'' in $\mathcal{A}$ and a word $w \in 0^* \operatorname{rep}_U(\mathbb{N})$ which does not contain any suffices of the form $\alpha_0 \cdots \alpha_j$ where $j \in \{0, \dots, n-1\}$ such that $\delta_\theta(p'', w) = p'$.*

Proof. The proof of those statements are similarly. Firstly, we prove the first point. We will proceed by contradiction. By symmetry we suppose that $i < j$ and $\delta_\theta(p, \alpha_0 \cdots \alpha_i) = \delta_\theta(p', \alpha_0 \cdots \alpha_j) = p''$. Let $w_1, w_2 \in \operatorname{rep}_U(\mathbb{N})$ be two words such that $\delta(q_0, w_1) = p$ and $\delta(q_0, w_2) = p'$. We have $\delta(q_0, w_1\alpha_0 \cdots \alpha_i) = \delta(q_0, w_2\alpha_0 \cdots \alpha_j) = p''$. Due to Lemma 3.2.8, we know that $\delta_\theta(p'', \alpha_{i+1})$ and $\delta_\theta(p'', \alpha_{j+1})$ exist in a θ-automaton $\mathcal{A}$. Reiterating this reasoning, we obtain that

$$\delta_\theta(p'', \alpha_{i+1} \cdots \alpha_{n-1}(\alpha_n - 1)\alpha_0 \cdots (\alpha_n - 1)$$

and

$$\delta_\theta(p'', \alpha_{j+1} \cdots \alpha_{n-1}(\alpha_n - 1)\alpha_0 \cdots (\alpha_n - 1)$$

are defined in $\mathcal{A}$. Therefore,

$$w_1\alpha_0 \dots \alpha_i\alpha_{j+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_n - 1)$$

$$w_2\alpha_0 \dots \alpha_j\alpha_{i+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_n - 1)$$

are accepted by $\mathcal{A}$ and are therefore words of the language $\operatorname{rep}_U(\mathbb{N})$. As a consequence, we have

$$\alpha_0 \cdots \alpha_i\alpha_{j+1} \cdots (\alpha_n - 1)\alpha_0 \cdots (\alpha_{j-i-1}) <_{lex} \alpha_0 \cdots \alpha_i\alpha_{i+1} \cdots \alpha_{n-j+i}\alpha_{n-j+i+1} \cdots \alpha_n \quad (3.1)$$

and

$$\alpha_0 \dots \alpha_j\alpha_{i+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_{i+n-j}) <_{lex} \alpha_0 \dots \alpha_j\alpha_{j+1} \dots \alpha_n. \quad (3.2)$$

Moreover, we know that $\alpha_0 \dots \alpha_n$ verifies the Parry condition, so

$$\alpha_{n-j+i+1} \dots \alpha_n <_{lex} \alpha_0 \dots \alpha_{j-i-1}$$

or

$$\alpha_{n-j+i+1} \dots \alpha_n = \alpha_0 \dots \alpha_{j-i-1}.$$

So we have due to that and(3.1) that

$$\alpha_{j+1} \dots (\alpha_n - 1) <_{lex} \alpha_{i+1} \dots \alpha_{n-j+i}$$

and due to (3.2) that

$$\alpha_{i+1} \dots \alpha_{n-j+i} <_{lex} \alpha_{j+1} \dots \alpha_n.$$

As those two inequalities can not be true at the same time if $j > i$, we have proven that $j = i$. □

As before there exists a similar result if the θ-expansion is ultimately periodic. In this case we have

Proposition 3.2.11. *Let θ be a Parry number such that the θ-expansion is eventually periodic ($D_\theta(1) = \alpha_0 \dots \alpha_n(\alpha_{n+1} \dots \alpha_{n+m})$) and $\mathcal{A}$ a corresponding θ-automaton. Then*

1. *if p, p' and p'' are states of $\mathcal{A}$ and $i, j \in \{0, \dots, n\}$ such that $\delta_\theta(p, \alpha_0 \dots \alpha_i) = \delta_\theta(p', \alpha_0 \dots \alpha_j) = p''$ then $i = j$;*

2. *if p, p' and p'' are states of $\mathcal{A}$ and $i \in \{0, \dots, n\}$ and $j \in \{n+1, \dots, n+m\}$ such that $\delta_\theta(p, \alpha_0 \dots \alpha_i) = \delta_\theta(p', \alpha_0 \dots \alpha_j) = p''$ then $i = n$ and $j = n + m$. So in particular, for all $k \in \mathbb{N}$ and $l \in \{1, \dots, m\}$, it is possible, that there exists three states p, p' and p'' of $\mathcal{A}$ such that*

$$\delta_\theta(p, \alpha_0 \dots \alpha_{n+l}) = \delta_\theta(p', \alpha_0 \dots \alpha_n(\alpha_{n+1} \dots \alpha_{n+m})^k \alpha_{n+1} \dots \alpha_{n+l} = p'';$$

3. *if p, p' and p'' are states of $\mathcal{A}$ and $i, j \in \{n+1, \dots, n+m\}$ such that $\delta_\theta(p, \alpha_0 \dots \alpha_i) = \delta_\theta(p', \alpha_0 \dots \alpha_j) = p''$ then $i = j$;*

4. *if p,p' and p'' are states of $\mathcal{A}$ and $i \in \{0, \dots, n+m\}$ such that $\delta_\theta(p, \alpha_0 \dots \alpha_i) = p'$ then there does not exist a state p'' in $\mathcal{A}$ and a word $w \in 0^* \operatorname{rep}_U(\mathbb{N})$ which does not contain any suffixes of the form $\alpha_0 \dots \alpha_i$ where $i \in \{0, \dots, n\}$ and is not of the form $\alpha_0 \dots \alpha_n(\alpha_{n+1} \dots \alpha_{n+m})^k \alpha_{n+1} \dots \alpha_{n+l}$ where $k \in \mathbb{N}$ and $l \in \{1, \dots, m\}$ such that $\delta_\theta(p'', w) = p'$.*

Proof. Firstly, we prove the first point. We will proceed by contradiction. By symmetry we suppose that $i < j$ and $\delta_\theta(p, \alpha_0 \dots \alpha_i) = \delta_\theta(p', \alpha_0 \dots \alpha_j) = p''$. Let $w_1, w_2 \in \text{rep}_U(\mathbb{N})$ be two words such that $\delta(q_0, w_1) = p$ and $\delta(q_0, w_2) = p'$. We have $\delta(q_0, w_1\alpha_0 \dots \alpha_i) = \delta(q_0, w_2\alpha_0 \dots \alpha_j) = p''$. Due to Lemma 3.2.9, we know that $\delta_\theta(p'', \alpha_{i+1})$ and $\delta_\theta(p'', \alpha_{j+1})$ exists in a θ-automaton $\mathcal{A}$. Reiterating this reasoning, we can conclude that

$$\delta_\theta(p'', \alpha_{i+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_n - 1))$$

and

$$\delta_\theta(p'', \alpha_{j+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_n - 1))$$

are defined in $\mathcal{A}$. Therefore,

$$w_1\alpha_0 \dots \alpha_i\alpha_{j+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_n - 1)$$

and

$$w_2\alpha_0 \dots \alpha_j\alpha_{i+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_n - 1)$$

are accepted by a automaton $\mathcal{A}$ and are therefore words of the language $\text{rep}_U(\mathbb{N})$. As a consequence, we have

$$\alpha_0 \dots \alpha_i\alpha_{j+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_{j-i-1}) <_{lex} \alpha_0 \dots \alpha_i\alpha_{i+1} \dots \alpha_n$$

and

$$\alpha_0 \dots \alpha_j\alpha_{i+1} \dots (\alpha_n - 1)\alpha_0 \dots (\alpha_{i+n-j}) <_{lex} \alpha_0 \dots \alpha_j\alpha_{j+1} \dots \alpha_n.$$

Moreover, we know that $\alpha_0 \dots \alpha_n$ verifies the Parry condition, so

$$\alpha_{n-j+i+1} \dots \alpha_n <_{lex} \alpha_0 \dots \alpha_{j-i-1}$$

or

$$\alpha_{n-j+i+1} \dots \alpha_n = \alpha_0 \dots \alpha_{j-i-1}.$$

So we have

$$\alpha_{j+1} \dots (\alpha_n - 1) <_{lex} \alpha_{i+1} \dots \alpha_{n-j+i}$$
$$\alpha_{i+1} \dots \alpha_{n-j+i} <_{lex} \alpha_{j+1} \dots \alpha_n.$$

As those two inequalities can not be true at the same time if $j > i$, we have proven that $j = i$. The other statements can be proven similarly.

□

3.3 Some properties of the Fabre substitution and of a θ-substitution

To be able to study the link between θ-automaton and θ-substitution we will introduce in this section, the notion of a θ-substitution. To be able to introduce this notion, we first need to define a conjugate of a substitution.

Definition 3.3.1. Let $f = (f, \Sigma, a_0)$ and $g = (g, \Gamma, b_0)$ be two substitutions. A morphism from f to g is a non-erasing application h from Σ to Γ such that:

- $h(a_0) = b_0$;
- for all $\sigma \in \Sigma$, $h(f(\sigma)) = g(h(\sigma))$ This will be noted $f \to g$.

Furthermore, $f = (f, \Sigma, a_0)$ and $g = (g, \Gamma, b_0)$ are said to be conjugates if there exists a substitution $l = (l, \Delta, c_0)$ such that $l \to f$ and $l \to g$.

Definition 3.3.2. Let $f = (f, \Sigma, a_0)$ be a substitutions, the set of substitution which are conjugates to f is denoted C_f.

However, it is often complicated to verify the condition proving that two substitutions are conjugate. Therefore, the following theorem is extremely useful.

Theorem 3.3.3. *Let $f = (f, \Sigma, a_0)$ and $g = (g, \Gamma, b_0)$ be two substitutions. Then f and g are conjugates if and only if for all $i \in \mathbb{N}$ $|f(a_i)| = |g(b_i)|$(with a_i and b_i being the $(i+1)^{th}$ letter of $f^\omega(a_0)$ and $g^\omega(b_0)$ respectively). In this case, there exists a substitution noted $f \times g = (f \times g, \Lambda, d_0)$ with $\Lambda \subset \Sigma \times \Gamma$ and $d_0 = (a_0, b_0)$ such that*

- *$f \times g \to f$ and $f \times g \to g$;*
- *for all $l = (l, \Delta, c_0)$ that verifies $l \to f$ and $l \to g$ we have $l \to f \times g$.*

Let's look at the following example, which illustrates the use of this proposition.

Example 3.3.4. Let $\Sigma = \{a, b, c, d\}$ and $\Gamma = \{0, 1, 2\}$ and let's consider the following substitutions:

$$\begin{aligned} f\colon \Sigma &\to \Sigma^* \\ a &\mapsto abc \\ b &\mapsto abd \\ c &\mapsto ac \\ d &\mapsto bc \end{aligned}$$

and

$$g\colon \Gamma \to \Gamma^*$$

$$0 \mapsto 012$$
$$1 \mapsto 002$$
$$2 \mapsto 12$$

Are $f = (f, \Sigma, a)$ and $g = (g, \Gamma, 0)$ conjugates? To be able to determine this, we will look if for all $i \in \mathbb{N}$ $|f(a_i)| = |g(b_i)|$(with a_i and b_i being the $(i+1)^{th}$ letter of $f^\omega(a_0)$ and $g^\omega(b_0)$ respectively). Firstly, we have

$$f(a) = abc$$
$$g(0) = 012$$

so $|f(a)| = |g(0)|$. Furthermore, we observe that a and 0 are on the first position of their respective fixed points, b and 1 on the second, c and 2 on the third. So we need to verify that $|f(b)|$ and $|g(1)|$ are equal and $|f(c)|$ and $|g(2)|$ are equal. We have

$$f(b) = abd$$
$$g(1) = 002$$

so $|f(b)| = |g(0)|$. In addition, we notice that a and 0 can be on the same position in their respective fixed points as can b and 0 and d and 2. We already verified that $|f(a)| = |g(0)|$ but we also need to verify that $|f(b)|$ and $|g(0)|$ are equal and $|f(d)|$ and $|g(2)|$ are equal. Furthermore, we have

$$f(c) = ac$$
$$g(2) = 12$$

so $|f(c)| = |g(2)|$. Again, we can conclude that a and 1 can be found on the same position in their respective fixed points as can c and 2. We already verified that $|f(c)| = |g(2)|$ but we also need to verify that $|f(a)|$ and $|g(1)|$ are equal. We have

$$f(b) = abd$$
$$g(0) = 012$$

so $|f(b)| = |g(0)|$. We can conclude a and 0, b and 1 and d and 2 can be on the same position in their respective fixed point. We already verified that $|f(a)| = |g(0)|$ and $|f(b)| = |g(1)|$ but we still need to verify that $|f(d)|$ and $|g(2)|$ are equal. We have

$$f(d) = bc$$
$$g(2) = 12$$

so $|f(d)| = |g(2)|$. And b and 1, c and 2 can be on the same position in their respective fixed points. We already verified that $|f(c)| = |g(2)|$ and $|f(b)| = |g(1)|$. We have

$$f(a) = abc$$

$$g(1) = 002$$

so $|f(a)| = |g(1)|$. Again, we can conclude that a and 0, b and 0 and c and 2 are on the same position in their respective fixed point. We already verified that $|f(a)| = |g(0)|$, $|f(b)| = |g(0)|$ and $|f(c)| = |g(2)|$. So we can conclude that f and g are conjugate. Furthermore, we can create the substitution

$$\begin{aligned} f \times g \colon (\Sigma, \Gamma) &\to (\Sigma, \Gamma)^* \\ (a, 0) &\mapsto (a, 0)(b, 1)(c, 2) \\ (b, 1) &\mapsto (a, 0)(b, 0)(d, 2) \\ (c, 2) &\mapsto (a, 1)(c, 2) \\ (b, 0) &\mapsto (a, 0)(b, 1)(d, 2) \\ (d, 2) &\mapsto (b, 1)(c, 2) \\ (a, 1) &\mapsto (a, 0)(b, 0)(c, 2) \end{aligned}$$

prolongable on $(a, 0)$. In the proof of theorem 3.3.3, we show that a substitution create like $f \times g$ verifies the conditions

- $f \times g \to f$ and $f \times g \to g$;
- for all $l = (l, \Delta, c_0)$ that verifies $l \to f$ and $l \to g$ we have $l \to f \times g$.

Proof. If are $f = (f, \Sigma, a_0)$ and $g = (g, \Gamma, b_0)$ are two substitutions which are conjugates, then there exists a substitution $l = (l, \Delta, c_0)$ such that $l \to f$ and $l \to g$ and two applications

$$s : \Delta^* \to \Sigma^*$$

and

$$t : \Delta^* \to \Gamma^*$$

such that

- $s(c_0) = a_0$;
- for all $c \in \Delta$, $s(l(c)) = f(s(c))$

and

- $t(c_0) = a_0$;
- for all $c \in \Delta$, $t(l(c)) = g(t(c))$.

Due to this, we can prove that

$$s(l^{\omega}(c_0)) = f^{\omega}(a_0). \tag{3.3}$$

We know that $l^i(c_0)$ is a prefix of $l^{\omega}(c_0)$ for all $i \in \mathbb{N}$ and $|l^i(c_0)| < |l^{i+1}(c_0)$. So we will prove (3.3) by proving that

$$s(l^i(c_0)) = f^i(a_0) \tag{3.4}$$

for all $i \in \mathbb{N}$.We will proceed by induction. Base case: By definition, we know that $s(c_0) = a_0$. Induction, let's suppose that (3.4) is verified for all integer smaller than $i \in \mathbb{N}$. We know that $\forall c \in \Delta$, $s(l(c)) = f(s(c))$ so we have

$$\begin{aligned} s(l^i(c_0)) = s(l(l^{i-1}(c_0)) &= f(s(l^{i-1}(c_0))) \\ &= f(f^{i-1}(a_0)) = f^i(a_0). \end{aligned}$$

By symmetry, we also have $t(l^{\omega}(c_0) = g^{\omega}(b_0)$. In particular, we remark that for all $i \in \mathbb{N}$ we have

- $s(l(c_i)) = f(s(c_i)) = f(a_i)$;
- $t(l(c_i)) = g(t(c_i)) = g(b_i)$

where c_i (resp. a_i and b_i) is $(i+1)^{th}$ letter in $l^{\omega}(c_0)$ (resp. $f^{\omega}(a_0)$ and $g^{\omega}(b_0)$). Furthermore, we remark that

$$|g(b_i)| = |t(l(c_i))| = |s(l(c_i))| = |f(a_i)| \tag{3.5}$$

as s (resp. t) is a non-erasing application from Δ to Σ (resp. Γ).

We define the set $\Lambda = \{(a,b) \in \Sigma \times \Gamma | \exists i \in \mathbb{N}, a = a_i, b = b_i\}$ and the substitution $f \times g = \{\Sigma \times \Gamma, \Lambda, (a_0, b_0)\}$ such that

$$\forall (a,b) \in \Lambda, f \times g(a,b) = (a'_0, b'_0) \cdots (a'_k, b'_k) \text{ where } f(a) = a'_0 \cdots a'_k \text{ and } g(b) = b'_0 \cdots b'_k$$

this definition is valid, as we have proven (3.5). Furthermore, we define the following two applications

$$\begin{aligned} \Pi_1 \colon \Lambda &\to \Sigma \\ (a,b) &\mapsto a \end{aligned}$$

and

$$\begin{aligned} \Pi_2 \colon \Lambda &\to \Gamma \\ (a,b) &\mapsto b. \end{aligned}$$

Those are applications from $f \times g$ to f (resp. g) by their definition. So we have $f \times g \to f$ and $f \times g \to g$. We now want to show that $\forall l = (l, \Delta, c_0)$ that verify $l \to f$ and $l \to g$ we have $l \to f \times g$. To prove this, we show that the application constructed as follows

$$h\colon \Delta \to \Lambda$$
$$c \mapsto h(c) = (s(c), t(c))$$

is a non-erasing application from l to $f \times g$. By definition h is non-erasing. We also notice that

$$h(c_0) = (s(c_0), t(c_0)) = (a_0, b_0).$$

So the first condition is verified. Secondly, if $c \in \Delta$ and $l(c) = c'_0 \dots c'_{k'}$ we have

$$h(l(c)) = h(c'_0 \dots c'_{k'}) = h(c'_0) \dots h(c'_{k'}) = (s(c'_0), t(c'_0)) \cdots (s(c_{k'}), t(c'_{k'})).$$

We have

$$\Pi_1(h(l(c)) = s(l(c)) = f(s(c)) = \Pi_1((f \times g)(h(c)))$$

and similar we have

$$\Pi_2(h(l(c)) = \Pi_2((f \times g)(h(c)))$$

and therefore $h(l(c) = (f \times g)(h(c))$.

Vice versa, if for all $i \in \mathbb{N}$ we have $|f(a_i)| = |g(b_i)|$, we can construct $f \times g$ as previously and therefore f and g are conjugate because we have $f \times g \to f$ and $f \times g \to g$.

□

The definition of a θ-substitution relies on the following substitution.

Definition 3.3.5. Let θ be a Parry number. The substitution associated with θ (more precisely with $D_\theta(1)$)is defined as follows:

- If $D_\theta(1) = \alpha_0 \dots \alpha_n$ then the substitution associated with θ is $f_\theta = (f_\theta, \{0, \cdots, n\}, 0)$ with

$$f_\theta(0) = 0^{\alpha_0}1$$
$$f_\theta(1) = 0^{\alpha_1}2$$
$$\vdots$$
$$f_\theta(n-1) = 0^{\alpha_{n-1}}n$$
$$f_\theta(n) = 0^{\alpha_n}.$$

  We will denote X_θ the fixed point $f_\theta^\omega(0)$;

- If $D_\theta(1) = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^\omega$ then the substitution associated with θ is $f_\theta = (f_\theta, \{0, \cdots, n+m\}, 0)$ with

$$
\begin{aligned}
&f_\theta(0) = 0^{\alpha_0}1\\
&f_\theta(1) = 0^{\alpha_1}2\\
&\vdots\\
&f_\theta(n+m-1) = 0^{\alpha_{n+m-1}}(n+m)\\
&f_\theta(n+m) = 0^{\alpha_{n+m}}(n+1)
\end{aligned}
$$

  We will denote X_θ the fixed point $f_\theta^\omega(0)$.

We will call this substitution the Fabre substitution.

Example 3.3.6. In Example 3.1.6, we had $\theta = \dfrac{1+\sqrt{5}}{2}$ and $D_\theta(1) = 11$. We define the Fabre substitution as $(f_\theta, \{0, 1\}, 0)$ with

$$
\begin{aligned}
f_\theta :&\{0,1\}^* \to \{0,1\}^*\\
&0 \mapsto 01\\
&1 \mapsto 0.
\end{aligned}
$$

This substitution is know as the Fibonacci substitution.

Example 3.3.7. In Example 3.1.7, we had $\theta = \left(\dfrac{1+\sqrt{5}}{2}\right)^2$ and $D_\theta(1) = 2(1)^\omega$. By definition of the Fabre substitution, we have $(f_\theta, \{0, 1\}, 0)$ with

$$
\begin{aligned}
f_\theta :&\{0,1\}^* \to \{0,1\}^*\\
&0 \mapsto 001\\
&1 \mapsto 01.
\end{aligned}
$$

Before, defining a θ-substitution, we will first explore some of the properties of the Fabre substitution. Firstly, after having defined the Fabre substitution, we are able to associate a matrix to Fabre substitution as follows.

Definition 3.3.8. Let θ be a Parry number.

- If $D_\theta(1) = \alpha_0 \ldots \alpha_n$ then the matrix $M_\theta = (m_{i,j})_{0 \le i,j \le n}$ where $m_{ij} = |f_\theta(i)|_j$, is called the matrix of occurrences of f_θ.

- If $D_\theta(0) = \alpha_1 \ldots \alpha_n(\alpha_{n+1} \ldots \alpha_{n+m})^\omega$ then the matrix $M_\theta = (m_{i,j})_{0 \le i,j \le n+m}$ where $m_{ij} = |f_\theta(i)|_j$, is called the matrix of occurrences of f_θ

Example 3.3.9. The matrix obtained using the substitution of Example 3.3.6 is:

$$M_\theta = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Example 3.3.10. The matrix obtained using the substitution of Example 3.3.7 is:

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

We will now study a few properties which result from this definition. To be able to do this, we need the following definitions.

Definition 3.3.11. The characteristic polynomial of a matrix A of size $i \times i$ is the polynomial

$$p_A(x) = \det(xI_i - A)$$

where I_i denotes the $i \times i$ identity matrix.

The first property, which we will be studying is the following:

Property 3.3.12. *Let θ be a Parry number. The characteristic polynomial of M_θ is equal to $Q_\theta(x)$.*

Proof. We will treat the two different types of θ-expansions separately. First, let's suppose that $D_\theta(1) = \alpha_0 \dots \alpha_n$ and the Fabre substitution $f_\theta = (f_\theta, \{0, \dots, n\}, 0)$:

$f_\theta(0) = 0^{\alpha_0}1$

$f_\theta(1) = 0^{\alpha_1}2$

$\vdots$

$f_\theta(n-1) = 0^{\alpha_{n-1}}n$

$f_\theta(n) = 0^{\alpha_n}.$

We will proceed by induction on the size of the occurrence matrix. Base case: If the occurrence matrix is of size $n = 1$, noted M_{θ_1} then we have

$$p_{M_{\theta_1}}(x) = \det(xI_1 - M_{\theta_1}) = \det(x - \alpha_0) = x - \alpha_0 = Q_{\theta_1}(x)$$

Induction: Let's suppose that this property is verified for all occurrence matrices of size smaller than or equal to some $n \in \mathbb{N}_0$. We denote $M_{\theta_{n+1}}$ the occurrence matrix of size $n+1$. We obtain

$$p_{M_{\theta_{n+1}}}(x) = \det(xI_{n+1} - M_{\theta_{n+1}})$$

$$= \det\left(\begin{pmatrix} x & 0 & 0 & \cdots & 0 \\ 0 & x & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & x & 0 \\ 0 & \cdots & \cdots & 0 & x \end{pmatrix} - \begin{pmatrix} \alpha_0 & 1 & 0 & \cdots & 0 \\ \alpha_1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & 1 \\ \alpha_n & 0 & & \cdots & 0 \end{pmatrix}\right)$$

$$= \det\begin{pmatrix} x-\alpha_0 & -1 & 0 & \cdots & 0 \\ -\alpha_1 & x & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & -1 \\ -\alpha_n & 0 & \cdots & 0 & x \end{pmatrix}$$

$$= (-1)^{n+1+1}(-\alpha_n)\det\begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ x & -1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x & -1 \end{pmatrix}$$

$$+ (-1)^{2(n+1)}x\det\begin{pmatrix} x-\alpha_0 & -1 & 0 & \cdots & 0 \\ -\alpha_1 & x & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & x & -1 \\ -\alpha_{n-1} & 0 & \cdots & 0 & x \end{pmatrix}$$

$$= (-1)^{n+2}(-\alpha_n)(-1)^n + xQ_{\theta_n}$$

$$= -\alpha_n + xQ_{\theta_n} = Q_{\theta_{n+1}}$$

We will now suppose that $D_\theta(1) = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})^\omega$ and $(f_\theta, \{0, \ldots, n{+}m\}, 0)$:

$f_\theta(0) = 0^{\alpha_0}1$

$f_\theta(1) = 0^{\alpha_1}2$

$\vdots$

$f_\theta(n+m-1) = 0^{\alpha_{n+m-1}}(n+m)$

$f_\theta(n+m) = 0^{\alpha_{n+m}}(n+1)$

For the second type, we will proceed, as in the first case, by induction on the size of the occurrence matrix. Base case: If $n = 0$ and $m = 1$, the matrix of occurrence $M_{\theta_{1,1}}$ is of size 2, then we have

$$p_{M_{\theta_{1,1}}}(x) = \det(xI_2 - M_{\theta_{1,1}}) = (x-\alpha_0)(x-1) - \alpha_1$$

$$= x^2 - x - x\alpha_0 - \alpha_0 - \alpha_1$$

and

$$\begin{aligned} Q_{\theta_{1,1}}(x) &= (x^2 - \alpha_0 x - \alpha_1) - (x - \alpha_0) \\ &= x^2 - x - x\alpha_0 - \alpha_0 - \alpha_1. \end{aligned}$$

Induction on n: Let's suppose that this is verified for all positive integers smaller than $n \in \mathbb{N}$, $m = 1$ and we prove it for n. We have $D_{\theta_{n,1}}(1) = \alpha_0 \ldots \alpha_n (\alpha_{n+1})^\omega$ and we obtain

$$\begin{aligned} p_{M_{\theta_{n,1}}}(x) &= \det(xI_{n+2} - M_{\theta_{n,1}}) \\ &= \det \begin{pmatrix} x - \alpha_0 & -1 & 0 & \cdots & 0 \\ -\alpha_1 & x & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & x & -1 \\ -\alpha_{n+1} & 0 & \cdots & 0 & x-1 \end{pmatrix} \\ &= -\alpha_n (-1)^{n+2+1} \det \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ x & -1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x & -1 \end{pmatrix} \\ &+ (x-1)(-1)^{2(n+2)} \det \begin{pmatrix} x - \alpha_0 & -1 & 0 & \cdots & 0 \\ -\alpha_1 & x & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & x & -1 \\ -\alpha_n & 0 & \cdots & 0 & x \end{pmatrix} \\ &= -\alpha_n + (x-1) Q_{\theta_{n+1}}(x) \\ &= Q_{\theta_{n,1}}(x) \end{aligned}$$

Induction on m: Let's suppose that this is verified for all positive integers smaller than $m \in \mathbb{N}_0$ and all $n \in \mathbb{N}$ and we prove it for m:

$$\begin{aligned} p_{M_{\theta_{n,m}}}(x) &= \det(xI_{n+m+1} - M_{\theta_{n,m}}) \\ &= \begin{pmatrix} x - \alpha_0 & -1 & 0 & \cdots & 0 \\ -\alpha_1 & x & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & x & -1 \\ -\alpha_{n+m} & 0 & \cdots -1 \cdots & 0 & x \end{pmatrix} \end{aligned}$$

$$= -\alpha_{n+m}(-1)^{n+m+2} \det \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ x & -1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x & -1 \end{pmatrix}$$

$$+ x(-1)^{2(n+m+1)} \begin{pmatrix} x-\alpha_0 & -1 & 0 & \cdots & 0 \\ -\alpha_1 & x & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & x & -1 \\ -\alpha_{n+m-1} & 0 & \cdots & 0 & x \end{pmatrix}$$

$$+ (-1)(-1)^{n+m+1+n+2} \begin{pmatrix} x-\alpha_0 & -1 & 0 & \cdots & & \cdots & & & 0 \\ -\alpha_1 & x & -1 & \ddots & & & & & \vdots \\ \vdots & 0 & \ddots & \ddots & & & & & 0 \\ -\alpha_{n-2} & \vdots & \ddots & \ddots & -1 & 0 & & & 0 \\ -\alpha_{n-1} & 0 & & 0 & x & 0 & & & 0 \\ -\alpha_{n+1} & 0 & & 0 & 0 & -1 & & & 0 \\ -\alpha_{n+2} & 0 & & & & x & \ddots & & 0 \\ \vdots & \vdots & \ddots & & & & \ddots & \ddots & 0 \\ -\alpha_{n+m-1} & 0 & \cdots & & & \cdots & 0 & x & -1 \end{pmatrix}$$

So we obtain

$$\begin{aligned} &-\alpha_{n+m} + x(x^{m+n} - \alpha_0 x^{m+n-1} - \ldots - \alpha_{n+m-2}x - \alpha_{n+m-1}) + \\ &(-1)^{m+2}(-1)^{m+1}(x^n - \alpha_0 x^{n-1} - \ldots - \alpha_{m-1}) \\ &= (x^{m+n+1} - \alpha_0 x^{m+n} - \ldots - \alpha_{n+m-1}x - \alpha_{n+m}) - (x^{n+1} - \alpha_0 x^n - \ldots - \alpha_n) \end{aligned}$$

□

Definition 3.3.13. The Parikh vector, also known as the composition vector of the finite word w over an alphabet $\Sigma = \{0, 1, \ldots, i\}$ is the vector

$$P(w) = \begin{pmatrix} |w|_0 \\ |w|_1 \\ \vdots \\ |w|_i \end{pmatrix}.$$

Example 3.3.14. Let's consider the set $\Sigma = \{a, b, c\}$ and the word $w = abcabcab$. The Parikh vector of w is

$$P(w) = \begin{pmatrix} |w|_a \\ |w|_b \\ |w|_c \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 2 \end{pmatrix}.$$

The following two properties can also be found as a result from Definition 3.3.5.

Property 3.3.15. *Let θ be a Parry number. The sequence $|f_\theta^i(0)|_{i\in\mathbb{N}}$ is equal to the base of the positional numeration system $(U_i)_{i\in\mathbb{N}}$ associated with θ.*

Proof. We will threat the two types of the θ-expansion separately. If $D_\theta(1) = \alpha_0 \cdots \alpha_n$ the base of the numeration system $(U_i)_{i\in\mathbb{N}}$ is given by

$$\begin{aligned} &U_0 = 1, \\ &U_i = \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{i-1} U_0 + 1 && \text{if } 1 \le i \le n \\ &U_i = \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_n U_{i-(n+1)} && \text{if } i > n \end{aligned}$$

Firstly, we define the following sequence of integers by recursion: $T_0 = 1$ and for all $1 \le i \le n$, we define $T_i = \alpha_0 T_{i-1} + \ldots + \alpha_{i-1} T_0$. We observe that $T_0 = U_0 = 1$. Furthermore, we can prove by induction on $1 \le i \le n$ that

$$T_i = U_i - U_{i-1}. \tag{3.6}$$

Base case: we have $U_1 - U_0 = \alpha_0 U_0 + 1 - 1 = \alpha_0 T_0 = T_1$. Induction: Let's suppose that the hypothesis is true for all positive integers smaller than or equal to $i < n$. We have

$$\begin{aligned} U_{i+1} - U_i &= \alpha_0 U_i + \alpha_1 U_{i-1} + \cdots + \alpha_i U_0 + 1 - \alpha_0 U_{i-1} - \cdots - \alpha_{i-1} U_0 - 1 \\ &= \alpha_0 (U_i - U_{i-1}) + \alpha_1 (U_{i-1} - U_{i-2}) + \cdots + \alpha_{i-1} (U_1 - U_0) + \alpha_i U_0 \\ &= \alpha_0 T_i + \alpha_1 T_{i-1} + \cdots + \alpha_{i-1} T_1 + \alpha_i T_0 \\ &= T_{i+1}. \end{aligned}$$

We will prove by induction on $0 \le i \le n$ that $|f_\theta^i(0)| = U_i$. If $i = 0$, $|f_\theta^0(0)| = |0| = 1$ and $U_0 = 1$ by definition.

If $1 \le i \le n$, we can prove by induction that

$$P(f_\theta^i(0)) = \begin{pmatrix} T_i \\ T_{i-1} \\ \vdots \\ T_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Base case: if $i = 0$, we have $f_\theta^0(0) = 0$ and $T_0 = 1$. Induction: Let's suppose the hypothesis is true for all positive integers smaller than or equal to $1 \le i < n$, then we have

$$\begin{aligned} f_\theta(0^{T_i}) &= (0^{\alpha_0}1)^{T_i} \\ f_\theta(1^{T_{i-1}}) &= (0^{\alpha_1}2)^{T_{i-1}} \\ &\vdots \\ f_\theta(i^{T_0}) &= (0^{\alpha_i}(i+1))^{T_0} \end{aligned}$$

adding all the zeros up, we obtain $\alpha_0 T_i + \alpha_1 T_{i-1} + \cdots + \alpha_i T_0 = T_{i+1}$ zeroes. And using (3.6), we conclude that

$$|f_\theta^i(0)| = \sum_{r=0}^{i} T_r = U_i.$$

If $i > n$, we have suppose that the for every integer $j < i$, that $|f_\theta^j(0)| = U_j$. We have

$$\begin{aligned} |f_\theta^i(0)| &= |(f_\theta^{i-1}(0))^{\alpha_0}| + |f_\theta^{i-1}(1)| \\ &= \alpha_0 |f_\theta^{i-1}(0)| + |f_\theta^{i-1}(1)| \\ &= \alpha_0 U_{i-1} + |(f_\theta^{i-2}(0))^{\alpha_1}| + |f_\theta^{i-2}(2)| \\ &= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + |f_\theta^{i-2}(2)| \\ &\vdots \\ &= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{n-1} U_{i-n} + |f_\theta^{i-n-1}(n)| \\ &= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{n-1} U_{i-n} + \alpha_n |f_\theta^{i-n-2}(0)| \\ &= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{n-1} U_{i-n} + \alpha_n U_{i-n-1}. \end{aligned}$$

By definition, we have $U_i = \alpha_0 U_{i-1} + \cdots + \alpha_n U_{i-n-1}$ and therefore $|f_\theta^i(0)| = U_i$.

If $D_\theta(1) = \alpha_0 \cdots \alpha_n (\alpha_{n+1} \cdots \alpha_{n+m})^\omega$ for $n \in \mathbb{N}$, then

$$\begin{aligned} U_0 &= 1, \\ U_i &= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{i-1} U_0 + 1 \qquad \text{if } i \ge 1. \end{aligned}$$

By definition, we have $U_0 = |f_\theta^0(0)|$. As before, we will proceed by induction. Let's suppose that for all integer j smaller than i, we have $f_\theta^j(0) = U_j$. We have

$$\begin{aligned} |f_\theta^i(0)| &= |(f_\theta^{i-1}(0))^{\alpha_0}| + |f_\theta^{i-1}(1)| \\ &= \alpha_0 |f_\theta^{i-1}(0)| + |f_\theta^{i-1}(1)| \\ &= \alpha_0 U_{i-1} + |(f_\theta^{i-2}(0))^{\alpha_1}| + |f_\theta^{i-2}(2)| \\ &= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} |f_\theta^{i-2}(2)| \end{aligned}$$

$$\vdots$$

$$= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{i-2} U_{i-n} + |f_\theta(i-1)|$$
$$= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{i-2} U_{i-n} + \alpha_{i-1}|0| + |i|$$
$$= \alpha_0 U_{i-1} + \alpha_1 U_{i-2} + \cdots + \alpha_{i-2} U_{i-n} + \alpha_{i-1} U_0 + 1.$$

So we have $|f_\theta^i(0)| = U_i$. □

After having studied the Fabre substitution, we will now define a θ-substitution.

Definition 3.3.16. Let θ be a Parry number and f_θ the Fabre substitution associated with θ. In this case, the substitutions in the set C_{f_θ} defined as in Definition 3.3.2 are called θ-substitutions.

Remark 3.3.17. As C_{f_θ} contains the substitution f_θ, f_θ is a θ-substitution.

In chapter 1, we proved in Theorem 1.3.7 that if (f, Σ, a) is a substitution and $N \geq 1$, then there exists of a unique $l \in \mathbb{N}$ and a unique sequence $(p_i, a_i)_{i=0\ldots,l}$ such that

- this sequence is a-admissible and $p_l \neq \epsilon$;
- $w_0 w_1 \cdots w_{N-1} = f^l(p_l) f^{l-1}(p_{l-1}) \cdots f^0(p_0)$

with $w_0 w_2 \cdots w_{N-1}$ being the prefix of size N of $f^\omega(a)$. We can prove that there is a link between the word $p_l \ldots p_0$ and the U-representation of N in the θ numeration system.

Theorem 3.3.18. *Let θ be a Parry number, $f = (f, \Sigma, a_0)$ be a θ-substitution and $N \geq 1$. Let $p_l \ldots p_0$ be the word such that $p_l \neq 0$, a_0-admissible and*

$$w_0 w_1 \cdots w_{N-1} = f_\theta^l(p_l) f_\theta^{l-1}(p_{l-1}) \cdots f_\theta^0(p_0)$$

with $w_0 w_2 \cdots w_{N-1}$ being the prefix of size N of $f_\theta^\omega(a_0)$. Then the word $|p_l||p_{l-1}|...|p_0|$ is the U-representation of N in the θ-numeration system.

Proof. If f_θ is the Fabre substitution corresponding to $D_\theta(1)$, we know that for each θ-substitution $g = (g, \Gamma, b_0)$, g is a conjugate of f_θ. Due to Theorem 3.3.3, we know for each $i \in \mathbb{N}_0$, $|f(a_i)| = |g(b_i)|$ (with a_i (resp. b_i) being the $(i+1)^{th}$ letter of X_{f_θ} (resp. $g^\omega(b_0)$)). So we only have to prove the statement for f_θ.

If $D_\theta(1) = \alpha_0 \cdots \alpha_n$, then by definition of f_θ, we have

$$\forall i \in \{0, \cdots n\}, |f_\theta(i)| \leq |f_\theta(0)| = \alpha_0 + 1.$$

Therefore, for each proper prefix u of $f_\theta(i)$ for $i \in \{0, \cdots n\}$, we know that $|u| \leq \alpha_0$ and that u is either the empty word ϵ or a word containing only the letter 0.

We denote w the prefix of X_{f_θ} of length N. We have

$$w = f_\theta^l(p_l) f_\theta^{l-1}(p_{l-1}) \ldots p_0$$

where

$$N = |f_\theta^l(p_l)| + |f_\theta^{l-1}(p_{l-1})| + \cdots + |p_0|.$$

We notice that $p_j a_j$ is a prefix of

- $f_\theta(a_{j+1})$ if $0 \le j < l$;
- $f_\theta(0)$ if $j = l$.

So for each $j \in \{0, \dots, l\}$, p_j is a proper prefix of $f_\theta(a)$ for some $a \in \{0, \dots, n\}$. By construction of f_θ, p_j only contains the letter 0. So we have for all $j \in \{0, \cdots l\}$, that $|f_\theta^j(p_j)| = |p_j||f_\theta^j(0)|$. Furthermore, using Property 3.3.15, we have $|f_\theta^j(p_j)| = |p_j|U_j$. So we have

$$N = |p_l|U_l + |p_{l-1}|U_{l-1} + \cdots + |p_0|U_0.$$

Furthermore, due to the definition of f_θ and the definition of the positional numeration system associated with a fixed point, we have

- If $j \in \{1, \cdots, l\}$ and the word p_j is a proper prefix of $f_\theta(i)$ for $i \in \{0, \cdots n-1\}$, then
  - either $|p_j| < \alpha_i$ and p_{j-1} is a proper prefix of $f_\theta(0)$;
  - or $|p_j| = \alpha_i$ and p_{j-1} is a proper prefix of $f_\theta(i+1)$.
- If $j \in \{1, \cdots, l\}$, the word p_j is a proper prefix of $f_\theta(n)$, then
  - $|p_j| < \alpha_n$ and p_{j-1} is a proper prefix of $f_\theta(0)$.

We notice that, due this there exists no factor equal or larger than $D_\theta(1)$ in

$$|p_l||p_{l-1}| \cdots |p_1||p_0|.$$

So by uniqueness of the U-representation, we can conclude. If $D_\theta(1) = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \cdots \alpha_{n+m})$, the proof is similar. □

Example 3.3.19. Reconsider the Example 3.3.6. We defined the Fabre substitution as $(f_\theta, \{0,1\}, 0)$ with

$$\begin{aligned} f_\theta : &\{0,1\}^* \to \{0,1\}^* \\ &0 \mapsto 01 \\ &1 \mapsto 0. \end{aligned}$$

This substitution is know as the Fibonacci substitution. In Example 1.3.8, we proved that the factorisation of the prefix of its fixed point of size 12 is $f_\theta^4(0)f_\theta^2(0)0$. The U-representation of 12 in the θ numeration system (Example 3.1.6) is 10101. We notice that $10101 = |0||\epsilon||0||\epsilon||0|$.

Remark 3.3.20. Due to Definition 1.3.10 and Theorem 3.3.18 for each $N \in \mathbb{N}$, the U-representation of N in the θ numeration system is equal to the representation of N in the Dumont-Thomas numeration system associated with f_θ.

3.4 Link between θ-substitutions and θ-automata

Before we are able to prove the main result of [12], we need the following Lemma. This Lemma allows us to prove more easily that some property P is verified for each letter in a fixed point.

Lemma 3.4.1. *Let $f = (f, \Sigma, a)$ be an increasing substitution. We define σ_f by*

$$\sigma_f(0) = 0$$
$$\sigma_f(i) = |f(f^\omega(a)[0, i-1])| \qquad \forall i \in \mathbb{N}.$$

Then a property P is verified for a_i, where a_i is the $(i+1)^{th}$ letter of $f^\omega(a)$, if the following conditions are verified:

- *the property P is verified for a_0;*
- *if the property P is verified for a_i, $i \in \mathbb{N}$, then the property is verified for all a_k, where $k \in \{\sigma_f(i), \cdots, \sigma_f(i+1)-1\}$.*

Proof. Let P be a property that verifies the two conditions of the statement and we prove that the property holds for all a_i, where $i \in \mathbb{N}$. We know that the property is true for a_0, so the property is also verified for all a_k, where $k \in \{\sigma_f(0), \cdots, \sigma_f(1)-1\}$. If we fix $N \in \mathbb{N}$ such that the property is true for all a_k with $k < \sigma_f(N)$, we notice that the property is also true for a_N because f is increasing and non-erasing and therefore $N < \sigma_f(N)$. We can conclude using the second point because the property P is verified for all a_i, where $i < \sigma_f(N+1)$. □

Definition 3.4.2. Let $\mathcal{A} = (Q, q_{0,}, F, \Sigma, \delta_\theta)$ be a θ-automaton, the infinite sequence of states of a θ-automaton is $(\delta_\theta(q_0, \upsilon_\theta(i)))_{i\in\mathbb{N}}$, where $\upsilon_\theta(i)$ is the U-representation of $i \in \mathbb{N}$ in the θ numeration system.

We are now able to prove the main theorem of [12], which explains a link between a θ-automaton and the fixed point of $f_\theta^\omega(0)$.

Theorem 3.4.3. *Let θ be a Parry number.*

1. *The infinite sequence of states of a θ-automaton is the fixed point of a θ-substitution.*
2. *The fixed point of a θ-substitution is the infinite sequence of states of a θ-automaton.*

Proof. If $D_\theta(1) = \alpha_0 \ldots \alpha_n$. Let $\mathcal{A} = (Q, q_{0,}, F, \Sigma, \delta_\theta)$ be a θ-automaton. Using this automaton, we define the substitution $g = (g, \Gamma, (q_0, 0))$:

- $\forall i \in \{0, \ldots, n-1\}$, we define

$$g(p, i) = (\delta_\theta(p, 0), 0)(\delta_\theta(p, 1), 0)\ldots(\delta_\theta(p, (\alpha_i - 1)), 0)(\delta_\theta(p, \alpha_i), i+1)$$

with $p \in Q$ if $\delta_\theta(p, 0), \ldots, \delta_\theta(p, \alpha_i)$, are defined;

- if $i = n$

$$g(p, n) = (\delta_\theta(p, 0), 0)(\delta_\theta(p, 1), 0) \dots (\delta_\theta(p, (\alpha_n - 1)), 0)$$

  with $p \in Q$ if $\delta_\theta(p, 0), \dots, \delta_\theta(p, \alpha_n - 1)$, are defined.

We define $\Gamma \subset \Sigma \times \{0, \dots, n\}$ as the set of letter (p, j) of $g^\omega(q_0, 0)$. Firstly, we want to prove that if (p, j) is a state of Γ, then there exists a word $w \in rep_U(\mathbb{N})$ such that $\delta_\theta(q_0, w) = p$ and if $j \geq 1$, then the longest suffix of w which is a prefix of $\alpha_0 \cdots \alpha_n$ is $\alpha_0 \cdots \alpha_{j-1}$ and if $j = 0$, then w does not contain a suffix of this form. We will prove that this is the case for each letter in $g^i(q_0, 0)$ with $i \in \mathbb{N}$. We will proceed by induction. Base case: if $i = 0$, we have $g^0(q_0, 0) = (q_0, 0)$. By definition of a θ-automaton there exists $\delta_\theta(q_0, 0) = q_0$ in the automaton. So there exists a word such that $\delta_\theta(q_0, w) = q_0$ and w does not contain any suffixes which is a prefix of $\alpha_0 \cdots \alpha_n$. Induction, we suppose that the statement holds for all integer smaller than or equal to $i \in \mathbb{N}$ and we will prove that the statement holds for $i + 1$. By the induction hypothesis, we know that for each letter (p, j) in the word $g^i(q_0, 0)$ there exist a word w such that $\delta_\theta(q_0, w) = p$ and if $j \geq 1$ the longest suffix of w which is a prefix of $\alpha_0 \cdots \alpha_n$ is $\alpha_0 \cdots \alpha_{j-1}$ and if $j = 0$, then w does not contain a suffix of this form. By Lemma 3.2.8, we know that

- if $j \in \{1, \dots, n-1\}$ then $\delta_\theta(q_0, w0), \dots, \delta_\theta(q_0, w\alpha_j)$ are defined in the θ-automaton;
- if $j = n$ then $\delta_\theta(q_0, w0), \dots, \delta_\theta(q_0, w\alpha_j - 1)$ are defined in the θ-automaton;
- if $j = 0$ then $\delta_\theta(q_0, w0), \dots, \delta_\theta(q_0, w\alpha_0)$ are defined in the θ-automaton.

By the construction of the substitution, we can conclude. This also proves that $\forall (p, j) \in \Gamma$, $g(p, j)$ is well defined. We also notice that due to definition of a θ-automaton, we have $g(q_0, 0) = (q_0, 0)w$ where $w \in \Gamma^* \setminus \{\epsilon\}$, so g is prolongable on $(q_0, 0)$. (An Example of this can be found in Example 3.4.4)

We now want to prove that the substitution $g = (g, \Gamma, (q_0, 0))$ is a θ-substitution. To do this, we will prove that

$$\begin{aligned} h \colon \Gamma &\to \{0, \dots, n\} \\ (p, i) &\longmapsto h(p, i) = i \end{aligned}$$

is a morphism from g to f_θ. Firstly we notice, that $h(q_0, 0) = 0$ and 0 is the letter on which f_θ is prolongable. Furthermore, we have $\forall i \in \{0, \dots, n-1\}$ and p such that $(p, i) \in \Gamma$:

$$\begin{aligned} & h(g(p, i)) = h((\delta_\theta(p, 0), 0)(\delta_\theta(p, 1), 0) \cdots (\delta_\theta(p, (\alpha_i - 1)), 0)(\delta_\theta(p, \alpha_i), i + 1)) \\ \Leftrightarrow\ & h(g(p, i)) = 0^{\alpha_i}(i + 1) \\ \Leftrightarrow\ & h(g(p, i)) = f_\theta(i) \\ \Leftrightarrow\ & h(g(p, i)) = f_\theta(h(p, i)). \end{aligned}$$

If $i = n$ and p such that $(p, i) \in \Gamma$, we have:

$$\begin{aligned}
& h(g(p, n-1)) = h((\delta_\theta(p,0),0)(\delta_\theta(p,1),0)\dots(\delta_\theta(p,(\alpha_n - 1)),0)) \\
\Leftrightarrow\ & h(g(p,n)) = 0^{\alpha_n} \\
\Leftrightarrow\ & h(g(p,n)) = f_\theta(n) \\
\Leftrightarrow\ & h(g(p,n)) = f_\theta(h(p,n)).
\end{aligned}$$

So h is a morphism from g to f_θ and therefore, g is a θ-substitution. Moreover, we can prove that if (p_m, i) is the $(m+1)^{th}$ letter of $g^\omega(q_0, 0)$, then $p_m = \delta_\theta(q_0, \upsilon_\theta(m))$. If $m = 0$, then $p_0 = q_0$ and $q_0 = \delta_\theta(q_0, 0)$ because of the definition of a θ-automaton. We suppose that for all integers, smaller than or equal to $m \in N$ the statement holds and we will prove it $\forall k \in \{\sigma_g(m), \cdots, \sigma_g(m+1) - 1\}$ where σ_g is defined as in Lemma 3.4.1. If $k = \sigma_g(m)$, then we have by definition, that $k = \sigma_g(m) = |g(g^\omega(q_0,0)[0, m-1])|$. If $g^\omega(q_0, 0)[0, m-1] = g^l(w_1)g^{l-1}(w_2)\dots g^1(w_l)w_0$, then we have $|w_l||w_{l-1}|\cdots|w_0| = \upsilon_\theta(m)$. So we have $g(g^\omega(q_0,0)[0, m-1]) = g^{l+1}(w_1)g^l(w_2)\cdots g^2(w_l)g^1(w_0)$ and therefore, we have $|w_l||w_{l-1}|\cdots|w_0||\epsilon| = |w_l||w_{l-1}|\cdots|w_0|0 = \upsilon_\theta(k)$. In $\mathcal{A}$, we have

$$\delta_\theta(p_0, \upsilon_\theta(k)) = \delta_\theta(p_0, \upsilon_\theta(m)0) = \delta_\theta(p_m, 0).$$

As $k = |g(g^\omega(q_0,0)[0, m-1])|$, the $(k+1)^{th}$ letter in $g^\omega(q_0, 0)$ is the first letter of $g(p_m, i)$. By definition of g, this letter is $(\delta(p_m, 0), 0)$ or $(\delta(p_m, 0), i+1)$ if $i \in \{0, \dots, n-1\}$ and $\alpha_i = 0$. In both cases, we have $\delta(p_m, 0) = p_k$ and we have for $k = \sigma_g(m)$, $\delta_\theta(q_0, \upsilon_\theta(k)) = p_k$. If $k \in \{\sigma_g(m)+1, \dots, \sigma_g(m+1)-1\}$, we have $k = |g(g^\omega(q_0,0)[0, m-1])u|$, where u is a proper prefix of $g(p_m, i)$ of length s. We have $\upsilon_\theta(k) = |w_l||w_{l-1}|\cdots|w_0|s$. The $(k+1)^{th}$ letter in $g^\omega(q_0, 0)$ is the $(s+1)^{th}$ letter of $g(p_m, i)$ and we can prove as before that $\delta_\theta(p_0, \upsilon_\theta(k)) = q_k$. A θ-substitution g is increasing as f_θ is increasing by definition, so using Lemma 3.4.1, we have $\delta_\theta(q_0, \upsilon_\theta(m)) = p_m$ if (p_m, i) is $(m+1)^{th}$ letter of $g^\omega(q_0, 0)$. Furthermore, we are able to prove that if $(p, i) \in \Gamma$ then there does not exist $j \neq i$ such that $(p, j) \in \Gamma$. We will proceed by contradiction. First, we notice that if $(p, i), (p, j) \in \Gamma$ such that $j \neq i$ and $i = 0$. Then there exists a word $w \in rep_U(\mathbb{N})$ such that $\delta_\theta(q_0, w) = p$ such that w does not have a suffix of the form $\alpha_0 \cdots \alpha_k$ for $k \in \{0, \dots, n\}$ and a word $w' \in rep_U(\mathbb{N})$ such that $\delta_\theta(q_0, w') = p$ and the longest suffix of w' which is a prefix of $\alpha_0 \cdots \alpha_n$ is $\alpha_0 \cdots \alpha_{j-1}$. This contradicts Proposition 3.2.10. By symmetry, if $j \neq i$ and $j = 0$, we obtain the same contradiction. Let's suppose that $(p, i), (p, j) \in \Gamma$ such that $j \neq i$, $i \neq 0$ and $j \neq 0$. So then there exists a word $w \in rep_U(\mathbb{N})$ such that $\delta_\theta(q_0, w) = p$ and the longest suffix of w which is a prefix of $\alpha_0 \cdots \alpha_n$ is $\alpha_0 \cdots \alpha_{i-1}$. There also exists a word $w' \in rep_U(\mathbb{N})$ such that $\delta_\theta(q_0, w') = p$ and the longest suffix of w' which is a prefix of $\alpha_0 \cdots \alpha_n$ is $\alpha_0 \cdots \alpha_{j-1}$. Let's suppose that $w = u\alpha_0 \cdots \alpha_{i-1}$ and $w' = u'\alpha_0 \cdots \alpha_{j-1}$. So we have $\delta_\theta(q_0, u) = p'$ and $\delta_\theta(q_0, u') = p''$ where $\delta_\theta(p', \alpha_0 \cdots \alpha_{i-1}) = p$ and $\delta_\theta(p'', \alpha_0 \cdots \alpha_{j-1}) = p$. Using Proposition 3.2.10, we obtain a contradiction. So each letter (p, i) in Γ is completely determined by p. This means that there exists a isomorphism between Q and Γ and we can define our substitution g on Q instead of Γ. We have a substitution $g = (g, Q, q_0)$ which is a θ-substitution such that

$g^\omega(q_0) = [\delta_\theta(q_0, v_\theta(m))]_{m\in\mathbb{N}}$. We now prove the second statement of the theorem. We now suppose that $g = (g, \Gamma, b_0)$ is a θ-substitution and $\mathcal{A} = (\Gamma, b_0, F, \Sigma, \delta)$ where

- $\delta(b, i) = b'$ if b' is the $(i+1)^{th}$ letter in $g(b)$;
- $F = \Gamma$;
- $\Sigma = \{0, \ldots, \alpha_0\}$.

We want to prove that the automaton $\mathcal{A}$ is a θ-automaton. We first notice that Σ labels the automaton $\mathcal{A}$ if it contains the set $\{0, \ldots, \alpha_0\}$ because g is a θ-substitution, the image of a letter in Γ is smaller than or equal to $(\alpha_0 + 1)$. Furthermore g is prolongable on the letter b_0, so we have $\delta(b_0, 0) = b_0$. To prove that $\mathcal{A}$ is a θ-automaton we need to prove that $\mathcal{A}$ does not except any words which contain a factor which is bigger or equal to $D_\theta(1) = \alpha_0 \cdots \alpha_n$ in the lexicographical order. Let's suppose that h is an application form g to f_θ. So we have

- $h(b_0) = a_0$;
- for all $\sigma \in \Sigma$, $h(g(\sigma)) = f_\theta(h(\sigma))$.

We have

- if $h(\sigma) = i$ with $i \in \{0, \ldots, n-1\}$ then $|g(\sigma)| = \alpha_i + 1$
    - if b is a letter of $g(\sigma)$ then $h(b) = i + 1$ if b is the last letter or $h(b) = 0$ if b is not the last letter;
- if $h(\sigma) = n$ then $|g(\sigma)| = \alpha_n$
    - if b is a letter of $g(\sigma)$ then $h(b) = 0$.

So the automaton $\mathcal{A}$ is a θ-automaton. To prove the second statement, we need to prove that the sequences of states of $\mathcal{A}$ is the fixed point of g. As in point 1, we can associate a θ-substitution to $\mathcal{A}$. However, due to the construction, this substitution is g, which will be isomorphic to g, we can conclude as $g^\omega(b_0) = [\delta_\theta(b_0, v_\theta(m))]_{m\in\mathbb{N}}$ using point 1. The demonstration for $D_\theta(1) = \alpha_0 \cdots \alpha_n(\alpha_{n+1} \ldots \alpha_{n+m})^\omega$ is similar, a θ-substitution on which its proof relies is $g = (g, \Gamma, (q_0, 0))$:

- $\forall i \in \{1, \ldots, n+m-1\}$, we define

  $$g(p, i) = (\delta_\theta(p, 0), 0)(\delta_\theta(p, 1), 0) \ldots (\delta_\theta(p, (\alpha_i - 1)), 0)(\delta_\theta(p, \alpha_i), i + 1)$$

  with $p \in Q$;
- $g(p, n+m) = (\delta_\theta(p, 0), 0)(\delta_\theta(p, 1), 0) \ldots (\delta_\theta(p, (\alpha_{n+m} - 1)), 0)(\delta_\theta(p, (\alpha_{n+m})), n + 1)$ with $p \in Q$.

As the proof is nearly identical to the previous one, we will not detail it. □

Let's look at an example of the construction of a substitution done in the beginning of the proof Theorem 3.4.3

Example 3.4.4. Let's consider $D_\theta = 21201$. A θ-automaton of $D_\theta = 21201$ is illustrated in Figure 3.9.

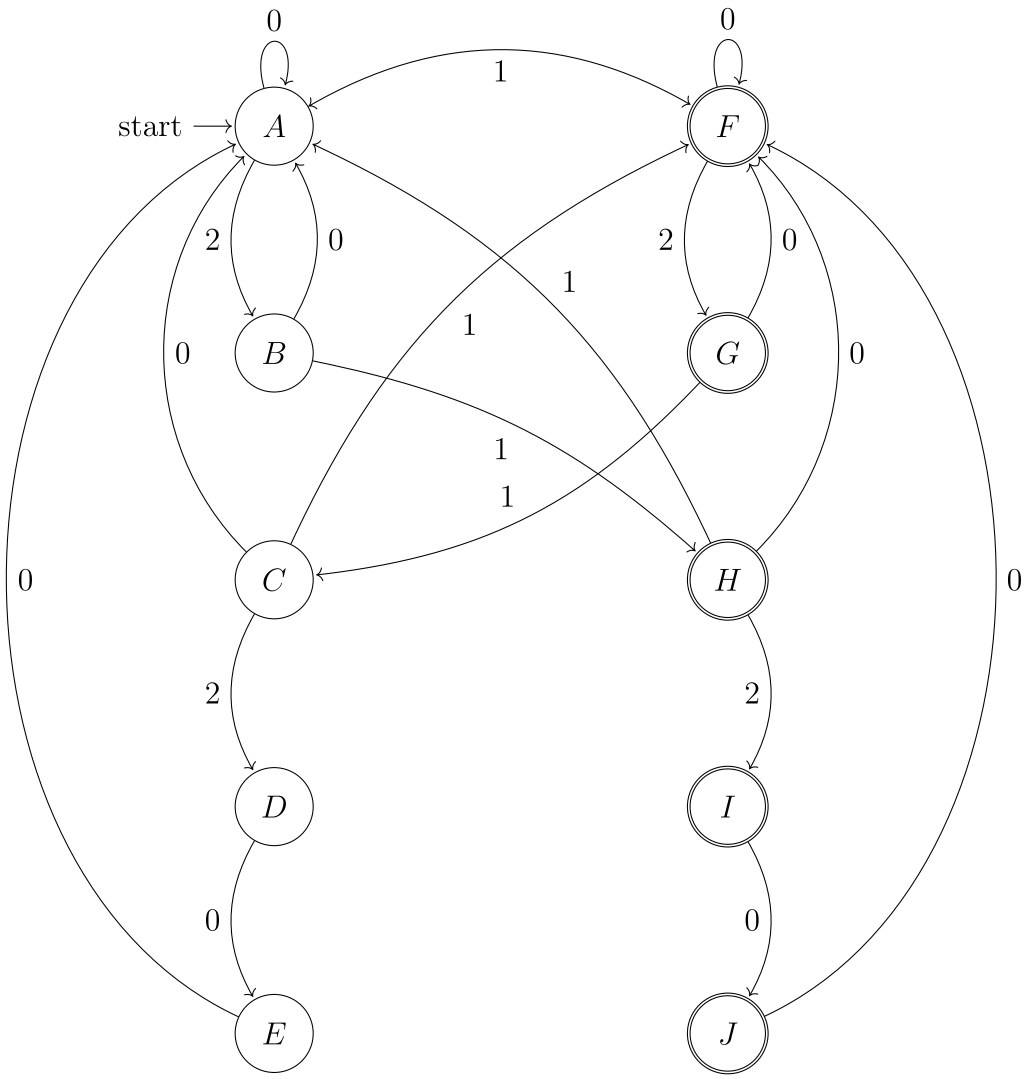


Figure 3.9: a θ-automaton for $D_\theta = 21201$

Using this θ-automaton, we create the substitution $g = (g, \Gamma, (A, 0))$ as in the proof of Theorem 3.4.3:

$$g((A,0)) = (A,0)(F,0)(B,1)$$
$$g((F,0)) = (F,0)(A,0)(G,1)$$

$$
\begin{aligned}
g((B,1)) &= (A,0)(H,2)\\
g((G,1)) &= (F,0)(C,2)\\
g((H,2)) &= (F,0)(A,0)(I,3)\\
g((C,2)) &= (A,0)(F,0)(D,3)\\
g((I,3)) &= (J,4)\\
g((D,3)) &= (E,4)\\
g((J,4)) &= (A,0)\\
g((E,4)) &= (F,0)
\end{aligned}
$$

And we have $\Gamma = \{(A,0)(F,0)(B,1),(F,0),(G,1),(H,2),(C,2),(I,3),(D,3),(J,4),(E,4)\}$.

Chapter 4

Complexity of a fixed point the Fabre substitution associated with a simple Parry number

Frougny, Masáková and Pelantová concerned themselves in [15] with the study of the complexity of the infinite word X_θ, where θ is a simple Parry number. The complexity of an infinite word is determined by the complexity function $\mathcal{C}$ defined for every $n \in \mathbb{N}$ such that $C(n) = |\operatorname{Fact}_n(w)|$. To study the complexity function, it is useful to know how to find all the factor of length $n+1$ starting from the all factors of length n. Therefore, we will first focus on the study of factors with more than one extension.

In this section, we fix θ being a simple Parry number, we study the fixed point of the Fabre substitution f_θ defined using the θ-expansion $D_\theta = \alpha_0 \cdots \alpha_n$.

4.1 Form of the fixed point of the Fabre substitution associated with a simple Parry number

We will first be interested in the form of the fixed point of the Fabre substitution f_θ. Using those results, we will then be able to study the structure of some particular factors of X_{f_θ}

Lemma 4.1.1. *The word $f_\theta^i(0)$ ends with the letter $i \bmod (n+1)$ for every $i \in \mathbb{N}$.*

Proof. We will proceed by induction on the exponent i. Base case: if $i = 0$, by definition $f_\theta^0(0) = 0$ and if $i = 1$ we have by definition $f_\theta^1(0) = 0^{\alpha_0}1$. Let's suppose that for every integer smaller than or equal to $i \in \mathbb{N}$ the property holds. We have

$$f_\theta^i(0) = w(i \bmod (n+1))$$

for some word w in our alphabet $\{0, 1, \ldots, n\}^+$. We obtain

$$f_\theta^{i+1}(0) = f_\theta(w(i \bmod (n+1))) = f_\theta(w) f_\theta(i \bmod (n+1)).$$

If $i \bmod (n+1) \leq n-1$ then we have $f_\theta^{i+1}(0) = f_\theta(w)0^{\alpha_{i \bmod (n+1)}} i+1 \bmod (n+1)$ and if $i \bmod (n+1) = n$ then we have $f_\theta^{i+1}(0) = f_\theta(w)0^{\alpha_{i \bmod (n+1)}}$. □

Lemma 4.1.2. *For every positive integer i less than or equal to n, we have*

$$f_\theta^i(0) = (f_\theta^{i-1}(0))^{\alpha_0}(f_\theta^{i-2}(0))^{\alpha_1} \cdots (f_\theta^1(0))^{\alpha_{i-2}}(0^{\alpha_{i-1}}i)$$

and for every i bigger than n, we have

$$f_\theta^i(0) = (f_\theta^{i-1}(0))^{\alpha_0}(f_\theta^{i-2}(0))^{\alpha_1} \cdots (f_\theta^{i-(n+1)}(0))^{\alpha_n}.$$

Proof. First if $i \leq n$, we proceed by recursion on i. Base case: If $i = 0$ then $f_\theta^0(0) = 0$ and if $i = 1$, we have $f_\theta^1(0) = 0^{\alpha_0}1$. Induction: Let's suppose that the equality hold for all integers smaller than $i \leq n$. We have

$$\begin{aligned}
f_\theta^i(0) &= f_\theta(f_\theta^{i-1}(0)) \\
&= f_\theta((f_\theta^{i-2}(0))^{\alpha_0}(f_\theta^{i-3}(0))^{\alpha_1} \cdots (f_\theta^1(0))^{\alpha_{i-3}}(0^{\alpha_{i-2}}(i-1)) \\
&= f_\theta((f_\theta^{i-2}(0))^{\alpha_0})f_\theta(f_\theta^{i-3}(0))^{\alpha_1}) \cdots f_\theta(f_\theta^1(0))^{\alpha_{i-3}})f_\theta(0^{\alpha_{i-2}}(i-1)) \\
&= (f_\theta^{i-1}(0))^{\alpha_0}(f_\theta^{i-2}(0))^{\alpha_1} \cdots (f_\theta^2(0))^{\alpha_{i-3}}(f_\theta(0))^{\alpha_{i-2}}(f_\theta(i-1)) \\
&= (f_\theta^{i-1}(0))^{\alpha_0}(f_\theta^{i-2}(0))^{\alpha_1} \cdots (f_\theta^1(0))^{\alpha_{i-2}}0^{\alpha_{i-1}}i.
\end{aligned}$$

Secondly, if $i > n$ we proceed by recursion on i. Base case: If $i = n+1$ then using the previously proven equality, we have:

$$\begin{aligned}
f_\theta^{n+1}(0) &= f_\theta(f_\theta^n(0)) \\
&= f_\theta(f_\theta^{n-1}(0))^{\alpha_0}(f_\theta^{n-2}(0))^{\alpha_1} \cdots (f_\theta^1(0))^{\alpha_{n-2}}(0^{\alpha_{n-1}}(n)) \\
&= f_\theta^n(0))^{\alpha_0}(f_\theta^{n-1}(0))^{\alpha_1} \cdots (f_\theta^2(0))^{\alpha_{n-2}}f_\theta(0^{\alpha_{n-1}}(n)) \\
&= f_\theta^n(0))^{\alpha_0}(f_\theta^{n-1}(0))^{\alpha_1} \cdots (f_\theta^2(0))^{\alpha_{n-2}}(f_\theta(0))^{\alpha_{n-1}}f_\theta(n) \\
&= f_\theta^n(0))^{\alpha_0}(f_\theta^{n-1}(0))^{\alpha_1} \cdots (f_\theta^2(0))^{\alpha_{n-2}}(f_\theta(0))^{\alpha_{n-1}}(0)^{\alpha_n}.
\end{aligned}$$

Induction: Let's suppose that the equality hold for all integers smaller than $i \geq n+1$. We have

$$\begin{aligned}
f_\theta^{i+1}(0) &= f_\theta(f_\theta^i(0)) \\
&= f_\theta(f_\theta^{i-1}(0))^{\alpha_0}(f_\theta^{i-2}(0))^{\alpha_1} \cdots (f_\theta^{i-(n+1)}(0))^{\alpha_n}) \\
&= f_\theta^i(0))^{\alpha_0}(f_\theta^{i-1}(0))^{\alpha_1} \cdots (f_\theta^{i+1-(n+1)}(0))^{\alpha_n}.
\end{aligned}$$

□

Remark 4.1.3. Using Lemma 4.1.2, we can observe another way to prove the previously proven Property 3.3.15 for the case if $D_\theta(1) = \alpha_0 \cdots \alpha_n$. Proceeding by induction on i, we can easily see that sequence $(|f_\theta^i(0)|)_{i \in \mathbb{N}}$ is equal to the base of the positional numeration system associated with θ *i.e.* $|f_\theta^i(0)| = U_i$ for all $i \in \mathbb{N}$.

Definition 4.1.4. Let w be a word over the alphabet Σ. A factor v of w is called a left special factor of w if there exist at least two distinct letters a and b of Σ such that av and bv are factors of w. In this case, we say that a and b are possible left extensions of v. In the same way, we define right special factors, v is a right special factor of w, if va and vb are factors of w and a and b are two distinct letters in Σ. A word v is a bispecial factor of w if it is right special and left special factor.

Example 4.1.5. Let's consider the following word $w = a\underbrace{ab}_{=u}ab\underbrace{ab}_{=v}b$. Notice that the factor u has a left extension a and v has a left extension b. So the factor ab is a left special factor of w. Furthermore, the factor ab is a bispecial factor of w because u has a right extension a and v has a right extension b.

We will interest ourselves on the structure of special factors of the word X_θ and the structure of special factors of the θ-expansion $D_\theta = \alpha_0 \cdots \alpha_n$.

A direct Corollary of Lemmas 4.1.1 and Lemma 4.1.2 is:

Corollary 4.1.6. *For every $i \in \mathbb{N}$, the word $f_\theta^i(0)$ is a left special factor of X_θ with $n+1$ distinct left extensions.*

Proof. For every $j \in \mathbb{N}_0$, we notice that each word $f_\theta^j(0)$ is a word of the form $f_\theta^{j-1}(0)w_j$, where w_j is a finite word over our alphabet $\{0, \cdots, n\}$ and this reasoning can be iterated for the prefix $f_\theta^{i-1}(0)$. Furthermore, for every integer $i \in N$, using Lemma 4.1.2 we know also that

$$f_\theta^{i+n+1}(0) = (f_\theta^{i+n}(0))^{\alpha_0}\Big(\underbrace{f_\theta^{i+n-1}(0)}_{(\underbrace{f_\theta^{i+n-3}(0)}_{\underbrace{\vdots}_{=(f_\theta^i(0))^{\alpha_0}w'_{i+n-2}}})^{\alpha_0}w_{i+n-2}}\Big)^{\alpha_1}\cdots\Big(\underbrace{f_\theta^{i+2}(0)}_{=(\underbrace{f_\theta^{i+1}(0)}_{=(f_\theta^i(0))^{\alpha_0}w'_{i+2}})^{\alpha_0}w_{i+2}}\Big)^{\alpha_{n-2}}\Big(\underbrace{f_\theta^{i+1}(0)}_{=(f_\theta^i(0))^{\alpha_0}w_{i+1}}\Big)^{\alpha_{n-1}}(f_\theta^i(0))^{\alpha_n}$$

Using Lemma 4.1.1, we can conclude because

- $f_\theta^{i+n}(0)$ ends with $i+n \mod (n+1)$;
- $f_\theta^{i+n-1}(0)$ ends with $i+n-1 \mod (n+1)$;

  $\vdots$
- $f_\theta^{i+1}(0)$ ends with $i+1 \mod n+1$.

Those factors are followed by the factor $f_\theta^i(0)$.

□

4.2 Left special factors in the fixed point of the Fabre substitution associated with a simple Parry number

We now have all the tools to be able to study the structure of left special factors of X_θ.

Definition 4.2.1. An infinite word w is called an infinite left special factor of v, if for all $i \in \mathbb{N}$, the prefix of w of length i is a left special factor of v.

Corollary 4.2.2. *The fixed point X_θ is an infinite left special factor of itself.*

Proof. This is a direct consequence of Corollary 4.1.6 as $\forall i \in \mathbb{N}$, $f_\theta^i(0)$ is a prefix of X_θ and it is a left special factor of X_θ. So each prefix of $f_\theta^i(0)$ is a left special factor of X_θ. □

We will now study the left special factors of X_θ in more detail.

Lemma 4.2.3. *The image of a left special factor of X_θ with p left extensions under f_θ is a left special factor with at least p left extensions.*

Proof. Let w be a left special factor of X_θ with p left extensions. There exists $a_1, \ldots, a_p$ distinct letters of $\{0, \ldots, n\}$ such that $a_1 w, \ldots, a_p w$ are factors of X_θ. Then

$$f_\theta(a_1) f_\theta(w), \ldots, f_\theta(a_p) f_\theta(w)$$

are also factors of X_θ. By the definition of the Fabre substitution f_θ, we know that the image of two distinct letters a and b of the alphabet $\{0, \ldots, n\}$ by f_θ finish with two distinct letters. In particular, the last letters of the factors $f_\theta(a_1), \ldots, f_\theta(a_p)$ are pairwise distinct. So we can conclude. □

Lemma 4.2.4. *A left special factor with p left extensions ending in a letter different than 0 is the image of a uniquely determined left special factor with p left extensions.*

Proof. Let w be a left special factor of X_θ with p left extensions which does not end in 0. There exists $a_1, \ldots, a_p$ distinct letters of $\{0, \ldots, n\}$ such that $a_1 w, \ldots, a_p w$ are factors of X_θ. There exists $m \in \mathbb{N}$ such that $a_1 w, \ldots, a_p w$ are factors of $f_\theta^m(0)$. Let $v_1, \ldots, v_p$ be factors of $f_\theta^{m-1}(0)$ such that for every $i \in \{1, \ldots, p\}$, $a_i w$ is a factor of $f_\theta(v_i)$. We suppose that v_i is of minimal length, for every $i \in \{1, \ldots, p\}$. There are at least two distinct left extensions of w, meaning at least one of them is different than 0. Without lost of generality, we can suppose that $a_1 \neq 0$. By definition of v_1, $f_\theta(v_1)$ contains the factor $a_1 w$, which starts and ends in a letter different from 0. By definition of f_θ and the minimality of v_1, we deduce that v_1 is of the form $v_1 = a_1' w'$ where $f_\theta(w') = w$. Hence $v_2, \cdots, v_p$ have the form $a_2' w', \cdots, a_p' w'$ where $a_2', \cdots, a_p'$ are pairwise distinct. This means that the preimage of w' of w is a left special factor with at least p distinct extensions. Using Lemma 4.2.3, we have the equality. □

Remark 4.2.5. Using Lemma 4.2.4 we can improve the statement of Lemma 4.2.3. Even though, we are not sure that the image of a left special factor w does not end with 0, we know that there exists a letter $a_1 \neq 0$ such that $a_1 f_\theta(w)$ is a factor of X_θ. If we construct v_1 as before, due to the fact that v_1 is of minimal length such that $a_1 f_\theta(w)$ is a factor of $f_\theta(v_1)$, we know that $v_1 = a_1' w$ and we can conclude as in Lemma 4.2.4. The improved statement of Lemma 4.2.3 we obtain is: The image of a left special factor of X_θ with p left extensions under f_θ is a left special factor with p left extensions.

A nearly direct consequence of Lemma 4.2.4 is the following theorem.

Theorem 4.2.6. *Let w be an infinite left special factor of X_θ. Then there exists an infinite left special factor v of X_θ such that $f_\theta(v) = w$.*

Proof. Let w be an infinite left special factor of X_θ. By definition, $\forall i \in \mathbb{N}$, $w[0, i]$ is a left special factor of X_θ. By definition of f_θ, w contains an infinite number of letters different from 0. Using Lemma 4.2.4, we know that every finite prefix of w ending with a letter different than 0 is the image of a uniquely determined left special factor of X_θ. □

We are now able to study the left special factors that are prefixes of an infinite left special factor of X_θ.

Theorem 4.2.7. *The infinite word X_θ has a unique infinite left special factor, namely X_θ itself. Furthermore, each prefix of X_θ is a left special factor with $n+1$ left extensions.*

Proof. We proceed by contradiction, let's suppose that X_θ has two or more infinite left special factors. Let w and v be two distinct infinite left special factor of X_θ such that $d(w, v)$ is maximal. We know, due to Theorem 4.2.6, that there exists two infinite left special factor w' and v' of X_θ such that $f_\theta(w') = w$. and $f_\theta(v') = v$. Due to the construction of f_θ, $d(w', v') > d(w, v)$, which contradicts the maximality between w and v. Furthermore, for each prefix w of X_θ, there exists a $i \in \mathbb{N}$ such that w is a prefix of $f_\theta^i(0)$. Corollary 4.1.6, implies that $f_\theta^i(0)$ is a left special factor of X_θ with $n+1$ distinct left extensions and therefore w is a left special factor of X_θ with at least $n+1$ distinct left extensions. As there are at most $n+1$ distinct left extensions for a factor in X_θ, we can conclude that each prefix of X_θ is a left special factor with $n+1$ left extensions. □

We studied left special factors that are prefixes of an infinite special factor. We will now look at those left special factors that are not prefixes of any left special factors.

Definition 4.2.8. A left special factor w of the infinite word X_θ is called a maximal left special factor of X_θ if for all $a \in \{0, ..., n\}$, wa is not a left special factor of X_θ.

Example 4.2.9. We consider $\theta = 1 + \sqrt{3}$. The θ-expansion of 1 is $D_\theta(1) = 22$. The corresponding Fabre substitution is $(f_\theta, \{0, 1\}, 0)$ where:

$$f_\theta : \{0,1\}^* \to \{0,1\}^*$$
$$0 \mapsto 001$$

$$1 \mapsto 00.$$

The fixed point of this substitution is

$$X_\theta = 00100100001001000010010010010 01\ldots$$

We notice that, due to the definition of f_θ, the letter 1 is always between the letters 0. So the maximal number of 0 behind each other is 4. So the factor 000 is a maximal left special factor of X_θ because the only left extension of 0001 is 0 and the only left extension of 0000 is 1.

Lemma 4.2.10. *A left special factor of X_θ which has exactly one right extension is not a maximal left special factor.*

Proof. This is a direct result of the definition of a maximal left special factor of X_θ. $\square$

Definition 4.2.11. We denote

$$j_k := \min\{i \in \mathbb{N}_0 \,|\, i \leq k, \alpha_{k-i} \neq 0\} \qquad \text{for } 1 \leq k \leq n.$$

This definition is valid because $\alpha_0 > 0$ and therefore such a j_k always exists.

Lemma 4.2.12. *Let $i \in \mathbb{N}$. If $f_\theta^i(0)$ ends with the letter k*

1. *if $k \geq 2$, then $f_\theta^i(0)$ has the suffix $j_{k-1}0^{\alpha_{k-1}}k$;*
2. *if $k = 0$, then $f_\theta^i(0)$ has the suffix $j_n 0^{\alpha_n}$.*

Proof. Let $i \in \mathbb{N}$.

1. If $f_\theta^i(0)$ ends with the letter $k \geq 2$, then we know using Lemma 4.1.1 that $f_\theta^{i-j_{k-1}-1}(0)$ ends with the letter $k - j_{k-1} - 1$. Furthermore, we have

$$f_\theta^{i-j_{k-1}}(0) = f_\theta(f_\theta^{i-j_{k-1}-1}(0))$$

and therefore $f_\theta^{i-j_{k-1}}(0)$ ends with $0^{\alpha_{k-j_{k-1}-1}}k - j_{k-1}$. Furthermore, by definition of j_{k-1}, we know that $\alpha_{k-1-j_{k-1}} \neq 0$ and $\alpha_{k-j_{k-1}} = 0, \ldots, \alpha_{k-2} = 0$. Therefore we obtain

$f_\theta^{i-j_{k-1}+1}(0)$ ends with the suffix $f_\theta(0(k - j_{k-1}))$ which in turn ends with

$$10^{\alpha_{k-j_{k-1}}}(k - j_k + 1) = 1(k - j_{k-1} + 1)$$

$f_\theta^{i-j_{k-1}+2}(0)$ ends with the suffix $f_\theta(1(k - j_{k-1} + 1))$ which in return ends with the suffix

$$20^{\alpha_{k-j_{k-1}+1}}(k - j_{k-1} + 2) = 2(k - j_{k-1} + 2)$$

$\vdots$

$f_\theta^{i-1}(0)$ which ends with the suffix $f_\theta((j_{k-1}-2)(k-2))$ which in return ends with the suffix

$$(j_{k-1}-1)0^{\alpha_{k-2}}(k-1) = (j_{k-1}-1)(k-1))$$

$f_\theta^{i}(0)$ which ends with the suffix $f_\theta((j_{k-1}-1)(k-1))$ which in return ends with the suffix

$$j_{k-1}0^{\alpha_{k-1}}k.$$

2. If $f_\theta^i(0)$ ends with the letter $k=0$ then we know using Lemma 4.1.1 that $f_\theta^{i-1}(0)$ ends with n and $f_\theta^{i-1-j_n}(0)$ ends with the letter $n-j_n$. Furthermore, we have

$$f_\theta^{i-j_n}(0) = f_\theta(f_\theta^{i-j_n-1}(0))$$

and therefore end in $0^{\alpha_{n-j^n}}(n-j_n+1)$ By definition of j_n, we know that $\alpha_{n-j_n} \neq 0$ and $\alpha_{n-j_n+1}=0,\cdots,\alpha_{n-1}=0$. Therefore we obtain

$f_\theta^{i-j_n+1}(0)$ ends with $1(n-j_n+2)$
$f_\theta^{i-j_n+2}(0)$ ends with $2(n-j_n+3)$
$\vdots$
$f_\theta^{i-2}(0)$ ends with $(j_n-2)(n-1)$
$f_\theta^{i-1}(0)$ ends with $(j_n-1)(n)$
$f_\theta^{i}(0)$ ends with $(j_n)0^{\alpha_n}$.

□

Using this lemma, we can prove the following useful lemma.

Lemma 4.2.13. *All factors of X_θ of the form $a0^rb$, where a and b are two letters different than 0 and $r \in \mathbb{N}$ are the following:*

1. $j_{k-1}0^{\alpha_{k-1}}k$ *for* $2 \le k \le n$*;*
2. $k0^{\alpha_0}1$ *for* $1 \le k \le n$*;*
3. $j_n0^{\alpha_0+\alpha_n}1$.

Proof. We want to prove that for every $i \in N$, the only factors of the form

$$a0^rb \tag{4.1}$$

with $a \neq 0$, $b \neq 0$ and $r \in \mathbb{N}$ in $f_\theta^i(0)$ are the factors given above. We will proceed by induction on i. Base case: if $i=0$, $f_\theta^0(0)=0$ does not contain any factor of the form (4.1) and if $i = 1$, $f_\theta^1(0) = 0^{\alpha_0}1$ does not contain any either. However, $f_\theta^2(0) = (0^{\alpha_0}1)^{\alpha_0}(0^{\alpha_1}2)$ does contain the factors of the form (4.1), namely $10^{\alpha_1}2$ and possibly $10^{\alpha_0}1$. The factor $10^{\alpha_1}2$ is of the first form given above and $10^{\alpha_0}1$ is the second. Induction: We suppose that the condition is verified for all integers smaller than $i \leq n$ and we will prove it for i. Using Lemma 4.1.2, we know that

$$f_\theta^i(0) = (f_\theta^{i-1}(0))^{\alpha_0}(f_\theta^{i-2}(0))^{\alpha_1} \cdots (f_\theta^1(0))^{\alpha_{i-2}}0^{\alpha_{i-1}}i.$$

New factor of the form (4.1) can only be created starting in an occurrence of a factor $f_\theta^{i'}(0)$ and ending in another one $f_\theta^{i''}(0)$ where $i > i' \geq i''$. Those factors will be of the form (4.1) and are the first form above except the factor created with the suffix of $f_\theta^i(0)$. This factor will be of the second form above. Let's suppose that $i > n$. We will again proceed by induction on i. Base case: If $i = n+1$, using Lemma 4.1.2, we know that

$$f_\theta^{n+1}(0) = (f_\theta^n(0))^{\alpha_0}(f_\theta^{n-1}(0))^{\alpha_1} \cdots (f_\theta^0(0))^{\alpha_n}.$$

New factor of the form (4.1) can only be created starting in an occurrence of a factor $f_\theta^{i'}(0)$ and ending in another one $f_\theta^{i''}(0)$ where $n \geq i' \geq i''$. Those factors will be of the form (4.1) and are of the first form given above. Induction: We suppose that the condition is verified for all integer smaller than $i > n+1$. Using Lemma 4.1.2, we notice again that the only new factors created of the searched form start in an occurrence of a factor $f_\theta^{i'}(0)$ and ending in another one $f_\theta^{i''}(0)$ where $i > i' \geq i'' \geq i - n - 1$. All newly created factors of the form (4.1) are of the first form above except the factor created which were created by starting in $f_\theta^{i'}(0)$ where $i' \mod n+1 \equiv 0$. In this case the factor will be of the third form above.

□

Example 4.2.14. Let's consider $D_\theta = 2101$. The corresponding Fabre substitution is $(f_\theta, \{0, 1, 2\}, 0)$ where

$$\begin{aligned} f_\theta :&\{0,1,2\}^* \to \{0,1,2\}^* \\ &0 \mapsto 001 \\ &1 \mapsto 02 \\ &2 \mapsto 3 \\ &3 \mapsto 0. \end{aligned}$$

We obtain the fixed point

$$X_\theta = 00100102001001020013001001020010010020013001001000 1 \ldots$$

For example, we can directly see the factors of the form

- $j_{k-1}0^{\alpha_{k-1}}k$ for $2 \leq k \leq n$ in green;

- $k0^{\alpha_0}1$ for $1 \leq k \leq n$ in blue;
- $j_n 0^{\alpha_0+\alpha_n}1$ in orange.

A direct consequence of Lemma 4.2.13 is the following Corollary.

Corollary 4.2.15. *Let w be a left special factor of X_θ of the form $w = v0^r$, where v does not end with the letter 0 and $r \in \mathbb{N}$. If $r \notin \{\alpha_0, \dots, \alpha_{n-1}\}$, then w has a unique right extension.*

Proof. We know, due to Lemma 4.2.13 that r is either equal to or smaller than $\alpha_0 + \alpha_n$ because due to the Parry condition in Theorem 3.1.3 we know that $\alpha_0 \geq \alpha_k$ for $1 \leq k \leq n$. If it's smaller, the only letter that can follow w is 0 because otherwise we would have a factor of the form $a0^r b$, where a and b are two letters different than 0 and $r \notin \{\alpha_0, \cdots, \alpha_n, \alpha_0+\alpha_n\}$, which would contradict Lemma 4.2.13. If $r = \alpha_0 + \alpha_n$ the letter following w has to be a 1 because of Lemma 4.2.13. □

Lemma 4.2.16. *The word 0^r, for $1 \leq r \leq \alpha_0$ is a left special factor of X_θ with $n+1$ left extensions.*

Proof. If $1 \leq r \leq \alpha_0$, then 0^r is a prefix of X_θ and we know due Theorem 4.2.7 that each prefix of X_θ is a left special factor with $n+1$ distinct left expansions. □

The following tree statements are a result from Lemma 4.2.13.

Lemma 4.2.17. *The following three statements are true.*

1. *The word 0^r, for $\alpha_0 < r \leq \alpha_0 + \alpha_n - 1$ is a left special factor of X_θ with 2 left extensions, namely 0 and j_n;*
2. *The word $0^{\alpha_0+\alpha_n-1}$ is a maximal left special factor of X_θ if $\alpha_n > 1$;*
3. *If $\alpha_n = 1$, then X_θ does not have a maximal left special factor of the form 0^r with $r \in \mathbb{N}_0$.*

Proof.
1. If $\alpha_0 < r \leq \alpha_0 + \alpha_n - 1$, 0^r is a factor in $j_n 0^{\alpha_0+\alpha_n}1$ because X_θ contains an infinite number of letters which are different from 0. So the only possible extensions to left are either 0 and j_n.
2. We note that a consequence of the condition of Parry in Theorem 3.1.3 is that $\alpha_0 \geq \alpha_k$ for $1 \leq k \leq n$. So $\alpha_0 + \alpha_n > \alpha_k$ for $1 \leq k \leq n$ if $\alpha_n > 1$. Furthermore, using Lemma 4.2.13, the only right extensions of $0^{\alpha_0+\alpha_n-1}$ are 0 and 1. However, both $0^{\alpha_0+\alpha_n-1}0$ and $0^{\alpha_0+\alpha_n-1}1$ have one left extension so $0^{\alpha_0+\alpha_n-1}$ is a maximal left special factor of X_θ.
3. If $\alpha_n = 1$, for all α_k with $1 \leq k \leq n-1$ using Lemma 4.2.13 we can observe that there are at least two distinct left extensions for the factor 0^{α_k}. The only possibility to have a maximal left special factor of the form 0^r with $r \in \mathbb{N}_0$ is with $0^{\alpha_0+\alpha_n-1} = 0^{\alpha_0}$ but 0^{α_0} has at least two distinct left extensions.

□

Lemma 4.2.18. *Let w be a factor of X_θ ending in nc. Then $c = 0$.*

Proof. We proceed by contradiction. If $c \neq 0$, then nc is a factor of X_θ of the form $a0^r b$ where a and b are different from the letter 0 and $r \in \mathbb{N}$. In this case, $a = n$, $b = c$ and $r = 0$. Since $\alpha_0 > 1$ and $r = 0$, we know due to Lemma 4.2.13 that $a = j_{k-1}$ for $2 \leq k \leq n$. But $j_{k-1} \leq k - 1 \leq n - 1$, which leads to our contradiction. □

Lemma 4.2.19. *Let $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$. Let w be a right special factor of X_θ with at least 3 distinct right extensions a, b, c such that w contains at least one letter different from 0, wa is a left special factor and $a \neq 0$. Then there exists a word $\tilde{w}$ such that $\tilde{w}$ is a right special factor of X_θ with at least 3 distinct right extensions $\tilde{a}, \tilde{b}, \tilde{c}$ and $\tilde{w}\tilde{a}$ is a left special factor, $\tilde{a} \neq 0$ and $aw = f_\theta(\tilde{w}\tilde{a})$.*

Proof. Let i be the last letter in w different from 0. We have $w = w'i0^r$ where $r \in \mathbb{N}$ and w' is a prefix of w of size $|w| - r - 1$. As wa, wb and wc are factors of X_θ, so are $i0^r a$, $i0^r b$ and $i0^r c$. Furthermore as $a \neq b \neq c$ and $a, b, c \geq 0$, at least one of these letters is bigger than 2. Without lost of generality, we will suppose that $a \geq 2$. Using Lemma 4.2.13 and the hypothesis that $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$, we know that $r < \alpha_0$. Furthermore $w'i$ is a left special factor because wa is a left special factor. Using Lemma 4.2.4, there exists a left special factor $\tilde{w}$ such that $w'i = f_\theta(\tilde{w})$. So

$$\begin{aligned} wa &= f_\theta(\tilde{w})0^r a \\ wb &= f_\theta(\tilde{w})0^r b \\ wc &= f_\theta(\tilde{w})0^r c \end{aligned}$$

are distinct factors of X_θ. By definition of f_θ and Lemma 4.2.13, there exist 3 distinct letters $\tilde{a}, \tilde{b}, \tilde{c}$ such that $\tilde{w}\tilde{a}$, $\tilde{w}\tilde{b}$ and $\tilde{w}\tilde{c}$ are factors of X_θ. We know that $a \neq 0$ and $r < \alpha_0$. Since $f_\theta(\tilde{a}) = 0^r a$, we know that $\tilde{a} \neq 0$. Furthermore, as $f_\theta(\tilde{w}\tilde{a}) = wa$ and wa is a left special factor due to Lemma 4.2.4, we know that $\tilde{w}\tilde{a}$ is also a left special factor. □

Proposition 4.2.20. *Let $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$ or $\alpha_0 = \ldots = \alpha_{n-1}$. Then for every maximal left special factor w containing a letter different from 0, there exists a maximal left special factor v and $r \in \{\alpha_0, \ldots, \alpha_{n-1}\}$ such that $w = f_\theta(v)0^r$.*

Proof. Let $w = w_0 \cdots w_k$ be a maximal left special factor of X_θ containing a letter different than 0, with $k \in \mathbb{N}$. We define $k' = \max\{0 \leq l \leq k : w_l \neq 0\}$. Due to Lemma 4.2.4, we know that there exists a left special factor $v = v_0 \cdots v_{k''}$ such that $w_0 \cdots w_{k'} = f_\theta(v_0 \ldots v_{k''})$ and therefore, we have

$$w = w_0 \cdots w_{k'}0^r = f_\theta(v_0 \cdots v_{k''})0^r \qquad \text{where } r = k - k'.$$

If $r \notin \{\alpha_0, \ldots, \alpha_{n-1}\}$, then w has a unique right extension, according to Corollary 4.2.15. Using Lemma 4.2.10, we know that w is not a maximal left special factor, which is a contradiction. We conclude that $r \in \{\alpha_0, \ldots, \alpha_{n-1}\}$. If v is not a maximal left special factor then there exists $a \in \{0, \cdots, n\}$ such that va is a left special factor of X_θ. We need to distinguish between two cases.

- if $\alpha_0 = \ldots = \alpha_{n-1}$, using Lemma 4.2.18, we know that if $a = n$, then $vn0$ is a left special factor. Using Lemma 4.2.3 we know that either

$$f_\theta(va) = f_\theta(v)0^{\alpha_a}a + 1 = f_\theta(v)0^{\alpha_0}a + 1$$

if $a \neq n$ or

$$f_\theta(vn0) = f_\theta(v)0^{\alpha_0+\alpha_n}1$$

if $a = n$, is a left special factor. Since $r \in \{\alpha_0, \ldots, \alpha_{n-1}\}$ and $\alpha_0 = \ldots = \alpha_{n-1}$ we have $r = \alpha_0$. In both cases, w is a proper prefix of them, which contradicts the maximality of w;

- if $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$, using Lemma 4.2.3 we know that $f_\theta(va)$ is also a left special factor. Since $w = f_\theta(v)0^r$ is a maximal left special factor, w is not a proper prefix of $f_\theta(va)$. Therefore and because $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$ and $r \in \{\alpha_0, \ldots, \alpha_{n-1}\}$, we have $a \neq 0$. Furthermore, as w is a maximal left special factor, there exist two distinct letter $\tilde{b}$ and $\tilde{c}$ such that $w\tilde{b}$ and $w\tilde{c}$ are not left special factors of X_θ because if w would have just one right extension, w would not be a maximal left special factor due to Lemma 4.2.10. So $f_\theta(v)0^r\tilde{b}$ and $f_\theta(v)0^r\tilde{c}$ are factors of X_θ (but not left special factors of X_θ). So by definition of a fixed point, there exist two distinct letter b and c such that vb and vc are factors of X_θ. So v has at least 3 distinct extensions to the right, namely a, b and c and va is a left special factor and $a \neq 0$. Using Lemma 4.2.19, we know that there exists a word $\widetilde{v^{(1)}}$ such that $\widetilde{v^{(1)}}$ is a right special factor of X_θ with at least 3 right extensions $\widetilde{a^{(1)}}, \widetilde{b^{(1)}}, \widetilde{c^{(1)}}$ and $\widetilde{v^{(1)}}\widetilde{a^{(1)}}$ is a left special factor, $\widetilde{a^{(1)}} \neq 0$ and $av = f_\theta(\widetilde{v^{(1)}}\widetilde{a^{(1)}})$. Reiterating the Lemma and because of the definition of f_θ, we obtain $\widetilde{v^{(q)}} = 0^t$ with $q, t \geq 1$ which is a right special factor of X_θ with at least 3 distinct right extensions $\widetilde{a^{(q)}}$, $\widetilde{b^{(q)}}$ and $\widetilde{c^{(q)}}$ such that $\widetilde{v^{(q)}}\widetilde{a^{(q)}}$ is a left special factor and $\widetilde{a^{(q)}} \neq 0$. Using Lemma 4.2.13, we can conclude that $t = \alpha_0$ and $\widetilde{a^{(q)}} = 1$. Either $\widetilde{b^{(q)}}$ or $\widetilde{c^{(q)}}$ is bigger than 1. Let's suppose that $\widetilde{b^{(q)}} > 1$. However $\widetilde{v^{(q)}}\widetilde{b^{(q)}} = 0^{\alpha_0}\widetilde{b^{(q)}}$ is due to Lemma 4.2.13 not a factor of X_θ, which is a contradiction.

□

Corollary 4.2.21. *Let $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$ or $\alpha_0 = \ldots = \alpha_{n-1}$. If $\alpha_n = 1$ then the fixed point X_θ does not have any maximal left special factors.*

Proof. Let's suppose that X_θ does have a maximal left special factor. If $\alpha_n = 1$, X_θ does not contain any maximal left special factor of the form 0^r, $r \in \mathbb{N}^+$, due to Lemma 4.2.17. So every maximal left special factor contains at least one letter different from 0. We know due to Proposition 4.2.20 that for a given maximal left special factor w there exists a maximal left special factor v and $r \in \{\alpha_0, \ldots, \alpha_{n-1}\}$ such that $w = f_\theta(v)0^r$. We know that $|w| > |v|$. As v is also a maximal left special factor, we can iterate this process. As w is a finite word, we will reach the empty word ϵ. However, this means, that there exists a left special factor v' such that $v' = f_\theta(\epsilon)0^r = 0^r$ with $r \in \mathbb{N}^+$. This is a contradiction because there does not exists a maximal left special factor of this form. □

Theorem 4.2.22. *Let $\alpha_0 > \max\{\alpha_1, \dots, \alpha_{n-1}\}$ or $\alpha_0 = \dots = \alpha_{n-1}$ and $\alpha_n > 1$, then for any $i \in \mathbb{N}$ there exists an integer sequence $(s_i)_{i\in\mathbb{N}}$ such that*

$$V_0 = 0^{\alpha_0+\alpha_n-1}$$
$$V_i = f_\theta\left(V_{i-1}\right) 0^{s_{i-1}} \qquad \text{for } i \in \mathbb{N}_0$$

are maximal left special factors of X_θ. And for every maximal left special factor v of X_θ there exists $i \in \mathbb{N}$ such that $v = V_i$.

Proof. This is a direct consequence of Proposition 4.2.20 and Lemma 4.2.17. □

We will now try to describe the sequence $(V_i)_{i\in\mathbb{N}}$. Due to Theorem 4.2.22, we know that describing the sequence $(V_i)_{i\in\mathbb{N}}$ means to describe the sequence $(s_i)_{i\in\mathbb{N}}$. The description of the sequence $(s_i)_{i\in\mathbb{N}}$ is generally complicated. However, in a few particular cases, we are able to describe it.

The first particular case, we will study is if we have $\alpha_0 = \alpha_1 = \dots = \alpha_{n-1}$.

Lemma 4.2.23. *If $\alpha_0 = \alpha_1 = \dots = \alpha_{n-1}$, then the sequence $(s_i)_{i\in\mathbb{N}}$ is a constant sequence, where each term is equal to α_0.*

Proof. This is a direct result of the definition of $(s_i)_{i\in\mathbb{N}}$ in Theorem 4.2.22 and Proposition 4.2.20. We know that V_i contains at least one letter different from 0, for all $i \in \mathbb{N}_0$ because of its construction in Theorem 4.2.22. Using Proposition 4.2.20, we know that for every maximal left special factor w which contains at least one letter different from 0, there exists a maximal left special factor v and $r \in \{\alpha_0, \dots, \alpha_{n-1}\}$ such that $w = f_\theta(v)0^r$. In this particular case $r = \alpha_0$, as $\alpha_0 = \alpha_1 = \dots = \alpha_{n-1}$. □

We know by the definition of the θ-expansion, that $\alpha_0 \geq \max\{\alpha_1, \dots, \alpha_n\}$. If this inequality is strict, we obtain the following theorem.

Theorem 4.2.24. *Let $\alpha_0 > \max\{\alpha_1, \dots, \alpha_{n-1}\}$. Then for every $i \in \mathbb{N}_0$ we have*

$$V_i = f_\theta\left(V_{i-1}\right) 0^{\alpha_k}, \qquad \text{where } k \in \{0, \dots, n-1\}, k = i \bmod (n). \tag{4.2}$$

Furthermore, the only right extensions of V_i in X_θ for $i \in \mathbb{N}$ are 0 and $k+1$ where $k \in \{0, \dots, n-1\}, k = i \bmod (n)$.

Proof. We prove by induction on i, that V_i is of the form (4.2) and that the only right extensions of V_i in X_θ are 0 or $k+1$. Base case: We notice that because $V_0 = 0^{\alpha_0+\alpha_n-1}$ due to Lemma 4.2.13, we know that the only two right extensions of V_0 are 1 and 0. So $V_1 = f_\theta(V_0)0^{\alpha_1}$ because $f_\theta(V_0 0) = f_\theta(V_0)0^{\alpha_0}1$ and $f_\theta(V_0 1) = f_\theta(V_0)0^{\alpha_1}2$. So as V_1 is a maximal left special factor and due to the fact that $\alpha_0 > \max\{\alpha_1 \cdots \alpha_n\}$ we can conclude. The only right extensions for V_1 are 0 and 2. Induction: We assume that $V_i = f_\theta\left(V_{i-1}\right) 0^{\alpha_k}$, where $k \in \{0, \dots, n-1\}$, $k = i(\mod n)$ and that V_i0 and $V_i(k+1)$ are the only factors of X_θ which have V_i as a suffix and are of length $|V_i| + 1$. We will be distinguishing between two cases:

- if $k+1 < n$, then we have

$$f_\theta(V_i 0) = f_\theta(V_i)0^{\alpha_0}1$$

and

$$f_\theta(V_i(k+1)) = f_\theta(V_i)0^{\alpha_{k+1}}(k+2)$$

are factors of X_θ. We obtain $V_{i+1} = f_\theta(V_i)0^{\alpha_{k+1}}$ where

$$k+1 \in \{0, \dots, n-1\}$$
$$k+1 = i+1(modn)$$

because V_{i+1} is a maximal left special factor and due to the fact that $\alpha_0 > \max\{\alpha_1 \cdots \alpha_{n-1}\}$ and 0 and $k+1$ are the only right extensions of V_i. So $f_\theta(V_i)0^{\alpha_0}1$ and $f_\theta(V_i)0^{\alpha_{k+1}}(k+2)$ are factors of X_θ and $\alpha_0 > \max\{\alpha_1, \dots, \alpha_{n-1}\}$, 0 and $k+2$ are the only right extensions of V_{i+1};

- if $k+1 = n$, $V_i n$ and $V_i 0$ are factors of X_θ. We know due to Lemma 4.2.18 that the only right extension of $V_i n$ is 0. So $V_i n0$ and $V_i 1$ are two factors of X_θ and therefore $f_\theta(V_i n0) = f_\theta(V_i)0^{\alpha_0+\alpha_n}1$ and $f_\theta(V_i 0) = f_\theta(V_i)0^{\alpha_0}1$. We obtain $V_{i+1} = f_\theta(V_i)0^{\alpha_0}$ because V_{i+1} is a left special factor and 0 and $k+1$ are the only right extension of V_i. Furthermore $k = n-1$ and $n-1 = i(\text{mod}n)$ so $0 = i+1(\text{mod}n)$. The right extensions of V^{i+1} are 1 and 0.

□

Example 4.2.25. We reconsider Example 4.2.9. The θ-expansion of 1 is $D_\theta(1) = 22$. We notice that $\alpha_0 = \alpha_1$ (and $\alpha_1 > 1$). Using Theorem 4.2.22, we are now able to calculate all the maximal left special factors of the corresponding fixed point. We have

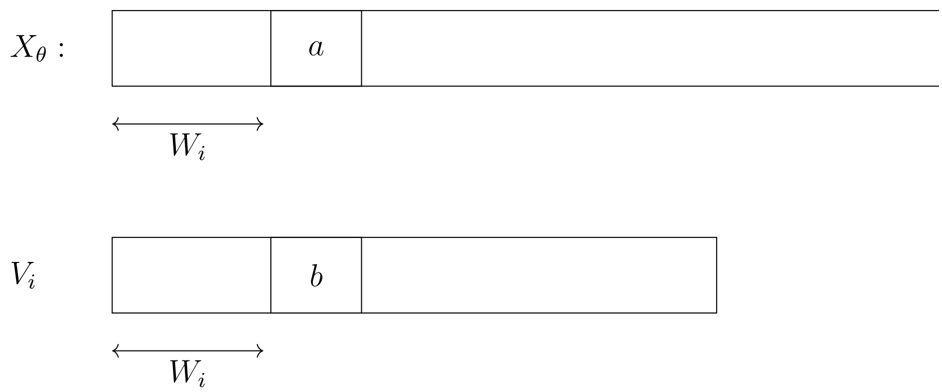
$$\begin{aligned} &V_0 = 0^3 \\ &V_1 = f_\theta\left(0^3\right)0^2 = 00100100100 \\ &V_2 = f_\theta\left(00100100100\right)0^2 = 00100100001001000010010000100100 \\ &\vdots \end{aligned}$$

4.3 Bispecial factors in the fixed point of the Fabre substitution associated with a simple Parry number

We can notice that every left special factor of X_θ is either a prefix of the infinite left special factor X_θ or a prefix of some maximal left special factor V_i, for $i \in \mathbb{N}$. We would like to be able to determine the complexity of X_θ. In order to do this we need the following definition.

Definition 4.3.1. For every $i \in \mathbb{N}$, let W_i be the maximal common prefix of X_θ and V_i.

Remark 4.3.2. For $i \in \mathbb{N}$, due to definition of V_i, V_i is not a prefix of X_θ. So W_i has a least two right extensions.



where $a \neq b$. Furthermore V_i is a maximal left special factor so $V_i \neq W_i$ because $W_i a$, $W_i b$ are factors of X_θ. Furthermore, each prefix of X_θ and each V_i for $i \in \mathbb{N}$ is a left special factor. So W_i is a bispecial factor for each $i \in \mathbb{N}$.

Remark 4.3.3. We know that $0^{\alpha_0}1$ is a prefix of X_θ and by definition of $V_0 = 0^{\alpha_0+\alpha_n-1}$. So the maximal common prefix of X_θ and V_0 is 0^{α_0}. So we have $W_0 = 0^{\alpha_0}$.

Definition 4.3.4. A factor w is called a total bispecial factor of X_θ if there exist two distinct letters a and b such that $a, b \in \{0, \ldots, n\}$ and wa and wb are left special factors of X_θ.

Remark 4.3.5. Each W_i is a total bispecial factor of X_θ for $i \in \mathbb{N}$ because W_i has two distinct left extensions a, b(one in X_θ and one in V_i). Furthermore, as $W_i \neq V_i$ both $W_i a$ and $W_i b$ are left special factors.

Proposition 4.3.6. *Let w be a total bispecial factor of X_θ containing a letter different than 0. Then there exists a total bispecial factor v and $r \in \{\alpha_0, \ldots, \alpha_{n-1}\}$ such that $w = f_\theta(v)0^r$.*

The proof of this theorem is similar to the proof of Proposition 4.2.20.

Proof. Let $w = w_0 \cdots w_k$ be a total bispecial factor of X_θ containing a letter different than 0 with $k \in \mathbb{N}$. So in particular, w is a left special factor and each prefix of w is a left special factor. We define $k' = \max\{l \in \mathbb{N} : w_l \neq 0\}$. Due to Lemma 4.2.4, we know that there exists a left special factor $v = v_0 \cdots v_{k''}$ such that $w_0 \cdots w^{k'} = f_\theta(v)$ and therefore, we have

$$w = w_0 \cdots w^{k'} 0^r = f_\theta(v)0^r \qquad \text{where } r = k - k'.$$

If $r \notin \{\alpha_0, \cdots, \alpha_n\}$, then w has a unique right extension according to Corollary 4.2.15 which contradicts the hypothesis that w is a bispecial factor. By definition, there exist

$a, b, c, d, c', d' \in \{0, \cdots, n\}$ such that $a \neq b$, $c \neq d$ and $c' \neq d'$ such that cwa, dwa, $c'wb$ and $d'wb$ are factors of X_θ. Therefore, $cf_\theta(v)0^r a$, $df_\theta(v)0^r a$, $c'f_\theta(v)0^r b$ and $d'f_\theta(v)0^r b$ are factors of X_θ. By definition of the Fabre substitution and the definition of a fixed point, there exists $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, $\tilde{d}$, $\tilde{c'}$, $\tilde{d'} \in \{0, \cdots, n\}$ such that $\tilde{a} \neq \tilde{b}$, $\tilde{c} \neq \tilde{d}$ and $\tilde{c'} \neq \tilde{d'}$ such that $\tilde{c}w\tilde{a}$, $\tilde{d}w\tilde{a}$, $\tilde{c'}w\tilde{b}$ and $\tilde{d'}w\tilde{b}$ are factors of X_θ. So v is a total bispecial factor of X_θ. □

Corollary 4.3.7. *There exists an integer sequence $(t_i)_{i\in\mathbb{N}}$ such that*

$$W_0 = 0^{\alpha_0}$$
$$W_i = f_\theta\left(W_{i-1}\right)0^{t_{i-1}} \qquad \text{for } i \in \mathbb{N}_0\,.$$

Proof. This is a direct consequence of Remark 4.3.3 and Proposition 4.3.6. □

As before, we study the particular case, if we have $\alpha_0 = \alpha_1 = \cdots = \alpha_{n-1}$.

Lemma 4.3.8. *If $\alpha_0 = \alpha_1 = \cdots = \alpha_{n-1}$, then the sequence $(t_m)_{m\in\mathbb{N}}$ is a constant sequence, where each term is equal to α_0.*

Proof. This is a direct result of the definition of $(t_i)_{i\in\mathbb{N}}$ in Corollary 4.3.7 and Proposition 4.3.6, which tells us that for every total bispecial factor w which contains at least one letter different from 0, there exists a total bispecial factor v and $r \in \{\alpha_0, \cdots, \alpha_{n-1}\}$ such that $w = f_\theta(v)0^r$. In this particular case $r = \alpha_0$, as $\alpha_0 = \alpha_1 = \ldots = \alpha_{n-1}$. □

Remark 4.3.9. We notice, that in this particular case ($\alpha_0 = \alpha_1 = \cdots = \alpha_{n-1}$), the sequences $(t_i)_{i\in\mathbb{N}}$ and $(s_i)_{i\in\mathbb{N}}$ are the same.

Again, as before we will study the particular case if we have $\alpha_0 > \max\{\alpha_1 \cdots \alpha_{n-1}\}$.

Theorem 4.3.10. *Let $\alpha_0 > \max\{\alpha_1 \cdots \alpha_{n-1}\}$. Then for every $i \in \mathbb{N}_0$ we have*

$$W_i = f_\theta\left(W_{i-1)}\right)0^{\alpha_k}, \qquad \text{where } k \in \{0,\ldots,n-1\}, k = i \bmod (n). \tag{4.3}$$

Furthermore, the only right extensions of $W_i = f_\theta\left(W_{i-1}\right)0^{\alpha_k}$ in X_θ are 0 and $k+1$.

Proof. We prove by induction on $i \in \mathbb{N}_0$, that W_i is of the form 4.3 and that the only right extension of W_i in X_θ are 0 or $k+1$. Base case: If $i = 0$, we have $W_0 = 0^{\alpha_0}$, we know that the only two right extensions of W_0 are 1 and 0. So $W_1 = f_\theta(W_0)0^{\alpha_1}$ because W_1 is a total bispecial factor and $\alpha_0 > \max\{\alpha_1 \cdots \alpha_{n-1}\}$. Induction: We assume that $W_i = f_\theta\left(W_{i-1}\right)0^{\alpha_k}$, where $k \in \{0,\ldots,n-1\}$, $k = i \bmod (n)$ and that W_i0 and $W_i(k+1)$ are the only factors of X_θ which have W_i as a suffix and are of length $|W_i|+1$. We will be distinguishing between two cases:

- if $k+1 < n-1$, then we have

$$f_\theta(W_i0) = f_\theta(W_i)0^{\alpha_0}1$$

and

$$f_\theta(W_i(k+1)) = f_\theta(W_i)0^{\alpha_{k+1}}(k+2)$$

are factors of X_θ. We obtain $W_{i+1} = f_\theta(W_i)0^{\alpha_{k+1}}$ where

$$k+1 \in \{1, \dots, n-1\}$$
$$k+1 = i+1(\text{mod}n)$$

because W_{i+1} is a total bispecial factor, due to the fact that $\alpha_0 > \max\{\alpha_1 \cdots \alpha_{n-1}\}$ and that the only right extensions of W_i are 0 and $k+1$.

Furthermore, as 0 and $k+1$ are the only right extensions of W_i, we know that 0 and $k+2$ are the only right extensions of W^{i+1};

- if $k+1 = n$,$W_i n$ and $W_m 0$ are have two factors of X_θ. We know due to Lemma 4.2.18 that the right extension of $W_m n$ is 0. So $W_i n0$ and $W_i 0$ are two factors of X_θ and therefore $f_\theta(W_i n0) = f_\theta(W_i)0^{\alpha_0+\alpha_n}1$ and $f_\theta(W_i 0) = f_\theta(W_i)0^{\alpha_0}1$ too. We obtain $W_{i+1} = f_\theta(W_i)0^{\alpha_0}$ because W_{i+1} is a total bispecial factor and the the only right extensions of W_i are 0 and $k+1$. Furthermore $k = n-1$ and $k = n-1 = i(\text{mod}n)$ so $0 = i+1(\text{mod}(n-1))$. The right extensions of W_{i+1} are 1 and 0.

□

Remark 4.3.11. If $\alpha_0 > \max\{\alpha_1 \dots \alpha_{n-1}\}$, the sequences $(t_i)_{i\in\mathbb{N}}$ and $(s_i)_{i\in\mathbb{N}}$ coincide.

Example 4.3.12. Let's reconsider the Example 4.2.9. We are able to determine W_i, the maximal common prefix between X_θ and V_i for $i \in \mathbb{N}$.

$$W_0 = 00$$
$$W_1 = f_\theta(00)00 = 00100100$$
$$W_2 = f_\theta(00100100)00 = 0010010000100100000100100$$
$$\vdots$$

Theorem 4.3.13. *Let $\alpha_0 > \max\{\alpha_1, \dots, \alpha_{n-1}\}$ or $\alpha_0 = \dots = \alpha_n$. Then for $i \in \mathbb{N}$*

1. $|V_i| = (\alpha_n - 1)U_i + |W_i|$;
2. $|W_i| = \sum_{j=0}^{i} c_j U_{i-j}$;

   where $c_i = \alpha_k$ and $k \in \{0, \dots.n-1\}$

   such that $i = k \bmod (n)$

3. $|V_i| < |W_{i+1}|$.

Proof. 1. We will proceed by induction on i to prove that $V_i = (f_\theta^i(0))^{\alpha_n - 1} W_i$. Base case: If $i = 0$ then $V_0 = 0^{\alpha_0 + \alpha_n - 1}$, $f_\theta^0(0) = 0$ and $W_0 = 0^{\alpha_0}$, so we have

$$V_0 = (f_\theta^0(0))^{\alpha_n - 1} W_0.$$

Induction: Let's suppose that this equality is verified for all integer smaller than $i \in \mathbb{N}_0$, and we will prove that the equality is true for i. By Theorem 4.2.24, we have $\forall i \in \mathbb{N}_0$, $V_i = f_\theta(V_{i-1})0^{\alpha_k}$ where $k \in \{0, \dots, n-1\}$, $k = i \bmod (n)$. Using Remark 4.3.11, we have

$$\begin{aligned} V_i &= f_\theta(V_{i-1})0^{\alpha_k} \\ &= f_\theta((f_\theta^{i-1}(0))^{\alpha_n - 1} W_{i-1})0^{\alpha_k} \\ &= (f_\theta^i(0))^{\alpha_n - 1} f_\theta(W_{i-1})0^{\alpha_k} \\ &= (f_\theta^i(0))^{\alpha_n - 1} W_i. \end{aligned}$$

Furthermore, we know that $|f_\theta^i(0)| = U_i$ due to Property 3.3.15 for every $i \in \mathbb{N}$, which allows us to conclude that $|V_i| = (\alpha_n - 1)U_i + |W_i|$.

2. We will first prove by induction, that $\forall i \in \mathbb{N}$,

$$W_i = (f_\theta^i(0))^{c_0} (f_\theta^{i-1}(0))^{c_1} \cdots (f_\theta^1(0))^{c_{i-1}} 0^{c_i}$$

where c_i is defined as in the statement. Base case: We know that $W_0 = 0^{\alpha_0}$ and $W_1 = f_\theta(W_0)0^{\alpha_1} = f_\theta(0^{\alpha_0})0^{\alpha_1} = (f_\theta(0))^{\alpha_0} 0^{\alpha_1}$ due to Theorem 4.3. Induction: If

$$W_i = (f_\theta^i(0))^{c_0} (f_\theta^{i-1}(0))^{c_1} \cdots (f_\theta^1(0))^{c_{i-1}} 0^{c_i}$$

for $i \in \mathbb{N}$ then

$$\begin{aligned} W_{i+1} &= f_\theta(W_i)0^{\alpha_k} \\ &= f_\theta((f_\theta^i(0))^{c_0} (f_\theta^{i-1}(0))^{c_1} \cdots (f_\theta^1(0))^{c_{i-1}} 0^{c_i})0^{\alpha_k} \\ &= (f_\theta^{i+1}(0))^{c_0} (f_\theta^i(0))^{c_1} \cdots (f_\theta^2(0))^{c_{i-1}} f_\theta(0^{c_i})0^{c_{i+1}} \end{aligned}$$

where $k \in \{0, \dots . n-1\}$, $k = i + 1 \bmod (n)$. Furthermore, we know that $|f_\theta^m(0)| = U_i$ due to Property 3.3.15 for every $i \in \mathbb{N}$, which allows us to conclude.

3. Proving that $|V_i| < |W_{i+1}|$ is equal to proving

$$\begin{aligned} &(\alpha_n - 1)U_i + |W_i| < \sum_{j=0}^{i+1} c_j U_{i-j+1} \\ \Leftrightarrow &(\alpha_n - 1)U_i + \sum_{j=0}^{i} c_j U_{i-j} < \sum_{j=0}^{i+1} c_j U_{i-j+1} \end{aligned}$$

$$\Leftrightarrow (\alpha_n - 1)U_i < \sum_{j=0}^{i} c_j (U_{i-j+1} - U_{i-j}) + c_{i+1}$$

where $c_j = \alpha_k$ and $k \in \{0, ...n-1\}$ such that $j = k \bmod (n)$. The first term of the sum on the right is

$$\begin{aligned} \alpha_0 (U_{i+1} - U_i) &> \alpha_0 (\alpha_0 U_i - U_i) \\ &> \alpha_0 U_i (\alpha_0 - 1) \\ &> U_i (\alpha_n - 1). \end{aligned}$$

Furthermore, all the other terms of the sum on the right are positive as a result of the construction of U_i which implies that $U_{i-j+1} > U_{i-j}$.

□

4.4 Complexity of the fixed point of the Fabre substitution associated with a simple Parry number where $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$ or $\alpha_0 = \ldots = \alpha_{n-1}$

After having studied the factors of X_θ with more than one extension, we are now able to study the complexity of the infinite word X_θ for simple Parry number. Frougny, Masákoá and Pelantová proved the following theorems related to this in [15]. The complexity of an infinite word is often described with the complexity function, which is defined as follows:

Definition 4.4.1. Let w be an infinite word over the alphabet Σ, the complexity function of w is

$$\begin{aligned} \mathcal{C} \colon \; & \mathbb{N} \to \mathbb{N} \\ & m \mapsto |\operatorname{Fact}(w) \cap \Sigma^m|. \end{aligned}$$

To simplify the notion for the rest of this chapter, we define the sequence $(l_i)_{i \in \mathbb{N}}$.

Definition 4.4.2. For θ such that $D_\theta(1) = \alpha_0 \cdots \alpha_n$, we denote

$$\begin{aligned} l_0 &:= 0 \\ l_{i+1} &:= |W_i| = \sum_{j=0}^{i} c_j U_{i-j} \end{aligned}$$

where $i \in N$, and $c_j = \alpha_k$ and $k \in \{0, \ldots, n-1\}$ such that $j = k \bmod (n)$.

Notation 4.4.3. For technical reasons, we suppose that $U_{-1} = 0$.

We are now able to determine the complexity of X_θ under specific conditions on $\alpha_0, \ldots, \alpha_n$. This theorem is one of the main results of [15].

Theorem 4.4.4. *Let* $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$ *or* $\alpha_0 = \ldots = \alpha_{n-1}$.

1. *Suppose that* $\alpha_0 = 1$. *Then for any* $i \in \mathbb{N}_0$, *we have*

$$\mathcal{C}(i+1) - \mathcal{C}(i) = n;$$

2. *then for* $i \in \mathbb{N}$, *there exists a* $k \in \mathbb{N}$ *such that*

$$l_k < i \leq l_{k+1}$$

and

$$\mathcal{C}(i+1) - \mathcal{C}(i) = \begin{cases} n+1 & \text{if } l_k < i < l_k + (\alpha_n - 1)U_{k-1} \\ n & \text{if } l_k + (\alpha_n - 1)U_{k-1} < i < l_{k+1}. \end{cases}$$

Proof. 1. First, we notice that if $n = 0$, there exists exactly one factor of each size so $\mathcal{C}(i+1) - \mathcal{C}(i) = 0 = n$ for all $i \in \mathbb{N}_0$. If $n > 0$, due to Theorem 4.2.7, we know that each prefix of X_θ has $n+1$ left extensions. So for every $i \in \mathbb{N}_0$, we have $\mathcal{C}(i+1) - \mathcal{C}(i) \geq n$. Let's proceed by contradiction that there exist a $i \in \mathbb{N}_0$ such that $\mathcal{C}(i+1) - \mathcal{C}(i) > n$. In this case, there exists a factor w of X_θ of size i, which has at least two left extensions and is not a prefix of X_θ. Furthermore, we know, due Theorem 4.2.22 that for every maximal left special factor v of X_θ there exists a $k \in \mathbb{N}$ such that $v = V_k$. So for every left special factor w of X_θ, which is not a prefix of X_θ, there exists a $k \in \mathbb{N}$ such that v is a prefix of V_k. However, due to Corollary 4.2.21, we know that the fixed point X_θ does not have any maximal left special factors, so it is not possible v is a prefix of V_k for some $k \in \mathbb{N}$, where our contradiction. Therefore, there exists no other left special factor w of size i and we have $\mathcal{C}(i+1) - \mathcal{C}(i) = n$.

2. As before, we first notice that, due to Theorem 4.2.7, we know that each prefix of X_θ has $n+1$ left extensions. So for every $i \in \mathbb{N}$, we have $\mathcal{C}(i+1) - \mathcal{C}(i) \geq n$. Let $(B_i)_{i \in \mathbb{N}}$ be the sequence such that for every $i \in \mathbb{N}$, we have

$$\mathcal{C}(i+1) = \mathcal{C}(i) + n + B_i. \tag{4.4}$$

For every $i \in \mathbb{N}$, B_i is equal to the number of left special factors of length i that are not prefixes of any infinite left special factor. We know, due to Theorem 4.2.22 that for every maximal left special factor w of X_θ there exists a $k \in \mathbb{N}$ such that $w = V_k$. So for every left special factor v of X_θ there exist a $k \in \mathbb{N}$ such that v is a prefix of V_k. Furthermore for some k, W_k describes the longest common prefix between V_k and X_θ, so left special factors of length i that are not prefixes of any infinite left special factor can only exists if there exists $k \in \mathbb{N}_0$ such that $|W_{k-1}| < i \leq |V_{k-1}|$. Moreover, for every $i \in \mathbb{N}$ there exists only one such interval as $|V_{k-1}| < |W_k|$ due to Theorem 4.3.13.

We obtain

$$B_i = \begin{cases} 1 & \text{if } |W_{k-1}| < i \leq |V_{k-1}| \\ 0 & \text{if } |V_{k-1}| < i \leq |W_k|. \end{cases}$$

Using the formula from Theorem 4.3.13, we have

$$\mathcal{C}(i+1) - \mathcal{C}(i) = \begin{cases} n+1 & \text{if } l_k < i < l_k + (\alpha_n - 1)U_{k-1} \\ n & \text{if } l_k + (\alpha_n - 1)U_{k-1} < i < l_{k+1}. \end{cases}$$

□

A direct consequence of the first point of Theorem 4.4.4 is:

Corollary 4.4.5. *Let $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_{n-1}\}$ or $\alpha_0 = \ldots = \alpha_{n-1}$. If $\alpha_n = 1$, we have for every $i \in \mathbb{N}$*

$$\mathcal{C}(i) = ni + 1.$$

Proof. We will proceed by induction on i. Base case: If $i = 0$, the only factor of X_θ of this size is ϵ and $\mathcal{C}(0) = (n-1)0 + 1 = 1$. If $i = 1$, there exist $n+1$ letters in X_θ and $\mathcal{C}(1) = n1 + 1 = n + 1$. Induction: If $\mathcal{C}(i) = ni + 1$ for all integer smaller than or equal to $i \in \mathbb{N}$, we have due to Theorem 4.4.4

$$\begin{aligned} \mathcal{C}(i+1) &= n + \mathcal{C}(i) \\ &= n + ni + 1 = n(i+1) + 1. \end{aligned}$$

□

Example 4.4.6. Let's reconsider the Example 4.2.9. We are now able to determine the complexity function:

$$\mathcal{C}(i) = i + 1.$$

So the fixed point always contains $i + 1$ distinct factors of size i.

A direct consequence of the second point of Theorem 4.4.4 is

Corollary 4.4.7. *If $\alpha_n > 1$ and $\alpha_0 > \max\{\alpha_1, \ldots, \alpha_n\}$ or $\alpha_0 = \ldots = \alpha_n$. Then for every $i \in \mathbb{N}_0$, we have*

$$ni \leq \mathcal{C}(i) \leq (n+1)i.$$

Furthermore, for every $i \in \mathbb{N}_0$ we have,

$$\mathcal{C}(i) = ni + 1 + (\alpha_n - 1)\left(\sum_{j=0}^{k-2} U_j\right) + \min\{i - 1 - l_k, (\alpha_n - 1)U_{k-1}\}$$

where $l_k < i - 1 \leq l_{k+1}$.

Proof. We first notice that for each $i \in \mathbb{N}_0$ there exists $k \in \mathbb{N}$ such that $l_k < i-1 \leq l_{k+1}$ as for every $j \in \mathbb{N}$, $|W_j| < |W_{j+1}|$. In the demonstration of Theorem 4.4.4, we have defined a sequence $(B_i)_{i\in\mathbb{N}}$ such that 4.4 is verified. We obtain using 4.4, that for every $i \in \mathbb{N}$, we have

$$\begin{aligned}
\mathcal{C}(i) = & \underbrace{\mathcal{C}(i-1)}_{=\mathcal{C}(i-2)+n+B_{i-2}} + n + B_{i-1} \\
& \vdots \\
& = \underbrace{\mathcal{C}(1)}_{=n+1} + (i-1)n + \sum_{j=1}^{i-1} B_j \\
& = (n)i + 1 + \sum_{j=1}^{i-1} B_j.
\end{aligned}$$

Furthermore, we notice that $0 \leq \sum_{j=1}^{i-1} B_j = \sum_{\substack{B_j=1 \\ 1\leq j\leq i-1}}^{i-1} 1 < i-1$. So we have

$$(n)i \leq \mathcal{C}(i) \leq (n+1)i.$$

Moreover, let $k \in \mathbb{N}$ such that $l_k < i-1 \leq l_{k+1}$. Then using Theorem 4.4.4 and Remark 4.4.3, we have

$$\begin{aligned}
\sum_{j=1}^{i-1} B_j &= \sum_{\substack{B_j=1 \\ 1\leq j\leq i-1}}^{i-1} 1 \\
&= \sum_{l=0}^{k-1} (\alpha_n - 1)U_{l-1} + \min\{i-1-l_k, (\alpha_n-1)U_{k-1}\}.
\end{aligned}$$

□

Chapter 5

Application of previous results in a new field: string attractors

To conclude this work, we will look at some applications of some of the results proven in the previous chapter. More precisely, we will look at their use in a new field, namely string attractors.

5.1 Some definitions and lemma

In 2019, Dominik Kempa and Nicola Prezza introduced in [18] the concept of string attractors. A string attractor is defined as follows:

Definition 5.1.1. Let $w = w_0 \cdots w_n$ be an finite word over an alphabet Σ. A string attractor of w is a set of positions $\Gamma \subseteq \{1, ..., n+1\}$ such that every factor of w has an occurrence containing an element of Γ.

Remark 5.1.2. Let $w = w_0 \cdots w_n$ be a finite word. Remark that if a string attractor references the letter w_0 then the string attractor contains the position 1.

Example 5.1.3. If we consider the following word:

$$w = abcdaabcdaac,$$

a string attractor of w is $\Gamma = \{4, 7, 11, 12\}$ (the referenced letters are in blue) because each factor of w has at least one occurrence in w where at least one of its letters is referenced by the string attractor.

Remark 5.1.4. A given finite word w can have different string attractor. For example, if we take the word w of Example 5.1.3, another string attractor of w would be $\Gamma' = \{1, 4, 7, 11, 12\}$. So a string attractor is not unique for a given word.

The fact that a string attractor is not unique for a word, creates the question, which string attractor is the most efficient for a given word. Logically, we can create a string attractor by referencing each position in a word. However this is not optimal. Furthermore, we notice that a string attractor has to contain at least as many positions as there are distinct letters in a word. However, the problem of finding the smallest string attractor is NP-hard. This has been proven in [18]. This however, does not mean that string attractor do not have any applications. string attractors are for example used in in combinatorial pattern matching. Furthermore in [2], the authors Gheeraert, Romana and Stipulanti asked the following question. Given a morphic sequence, does there exist a numeration system S such that the morphic sequence is S-automatic and such that a (the smallest) string attractors of the prefixes of this morphic sequence can be easily described using S? In their paper, they showed a first step to answer this question. They studied the fixed pointed of a morphism f associated with $\alpha_0 \cdots \alpha_n$ defined as follows:

$$f(0) = 0^{\alpha_0}1$$

$$f(1) = 0^{\alpha_1}2$$

$$\vdots$$

$$f(n-1) = 0^{\alpha_{n-1}}n$$

$$f(n) = 0^{\alpha_n}.$$

Again we denote the only fixed point of this substitution X_f. For the rest of this chapter, we will work under the working hypothesis (WH), that $\alpha_0 \cdots \alpha_n \in \mathbb{N}^{n+1}$ and $\alpha_0,\alpha_n \geq 1$. The working hypothesis ensures us , that f is a substitution. The condition, $\alpha_n \geq 1$ ensures us that the corresponding substitution f is non-erasing and that the recurrence relation is of order $n+1$. The condition, that $\alpha_0 \geq 1$ ensures us that the corresponding substitution f is prolongable on 0. We notice that this substitution is a generalisation of the Fabre substitution introduced in section 3.3 as $\alpha_0 \cdots \alpha_n$ does not need to verify the Parry condition. Moreover, for the rest of this chapter, we will consider $(U_i)_{i\in\mathbb{N}}$ the base of the positional numeration system defined as in section 3.1 for $\alpha_0 \cdots \alpha_n$.

As in chapter 4 we will be interested in the study of prefixes of infinite words. Firstly, we notice that:

Lemma 5.1.5. *Let w be a non-empty word and $v = w^r$ and $u = w^s$ with $1 \leq r \leq s$. If Γ is a string attractor of v, then $\Gamma \cup \{|w|\}$ is a string attractor of u.*

Proof. If Γ is a string attractor of v, then Γ covers all the factors appearing in u that also appear in v. Due to the structure of u, the only factors of u, which might not be covered by Γ are factors starting in a position smaller than $|v|$ and ending in a position bigger than $|v|$. However, for those factors, we can find an occurrence of those factors staring in a position smaller than or equal to $|w|$ in u, due to the structure of u. This occurrence will cross the position $|w|$ and therefore $\Gamma \cup \{|w|\}$ is a string attractor of u. □

Example 5.1.6. Let's consider the words $v = aaaba$ and $u = (aaaba)^3$. The set $\Gamma = \{3, 4\}$ is a string attractor of v. However, Γ is not a string attractor of u because it does not cover the factor $aaaa$. Adding the position $|v|$ to Γ, creates a string attractor for u because now the factor $aaaa$ is covered.

Before we are able to study string attractor in more detail, we need to prove some results on particular words, namely Lyndon words.

5.2 Some properties of Lyndon words

Definition 5.2.1. Let w be a non-empty word over an alphabet Σ, w is primitive if for all $v \in \Sigma^* \backslash \{\epsilon\}$ such that $w = v^i$ implies that $i = 1$.

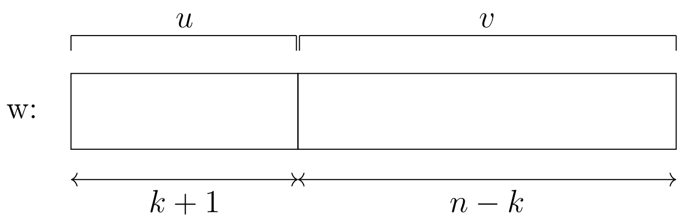
Example 5.2.2. Let $\Sigma = \{a, b, c\}$. The word $w = aabc$ is primitive as it is not possible to represent w as a power of another word. The word $w' = abcabc$ is not primitive as $w' = (abc)^2$.

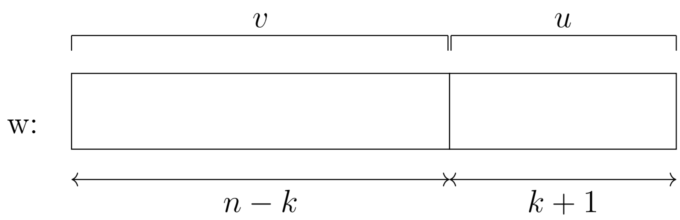
Remark 5.2.3. Two words w and v are conjugates if they are a cyclic permutation of each other.

Example 5.2.4. The word $w'' = bcabca$ and the word w' of Example 5.2.2 are conjugates.

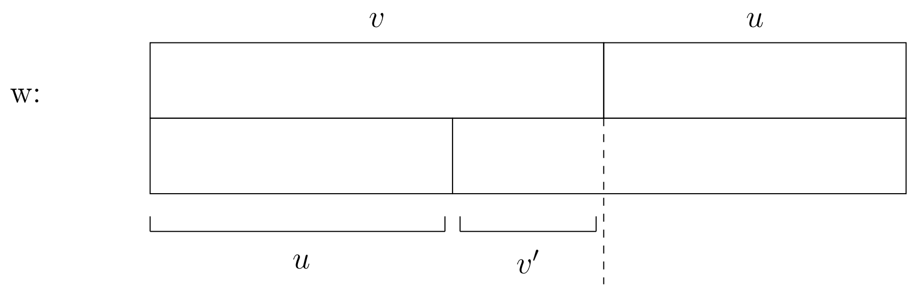
Theorem 5.2.5. *Let w be a primitive word, then w is different from each of its proper conjugates. Furthermore, each of its conjugates is also primitive.*

Proof. We will proceed it by contradiction. Let $w = w_0 \cdots w_n$ be a primitive word such that $w = w_{k+1} \dots w_n w_0 \dots w_k$ with $k \in \mathbb{N}$ and $k < n$. We will denote $u = w_0 \cdots w_k$ and $v = w_{k+1} \cdots w_n$. And we have:

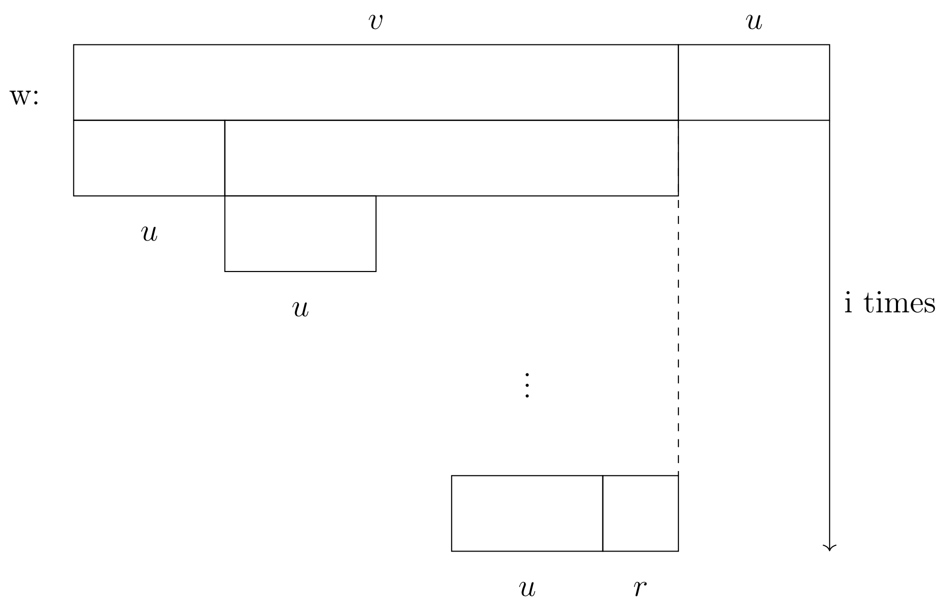




By symmetry, let's suppose that $|v| \geq |u|$. We notice that $v = uv'$ where v' is a prefix of v. We have



We can iterate this procedure on v' until the newly created suffix is smaller then $k+1$.



We obtain $w = u^i r u$ where r is the last suffix with a length smaller then $k+1$. If $r = \epsilon$, than $w = u^i$, which contradicts the primitiveness of w. If $r \neq \epsilon$, we notice that the same iteration can be done on $w = uv$ and we obtain $w = u^{i+1}r$. So we have $ur = ru$, which means that there exists a factor u' such that $u = u'^k$ and $r = u'^{k'}$ with $k, k' \in \mathbb{N}_0$. We have $w = (u'^k)^{i+1}u'^{k'}$, which contradicts the primitiveness of w. Let's suppose that w has a conjugate, which is not primitive. Let's suppose that $w' = w_{k+1} \cdots w_n w_0 \cdots w_k$ with $k \in \mathbb{N}$ and $k < n$ is not primitive *i.e.* there exists a non-empty word v such that $w' = v^i$ for $i \in \mathbb{N}$ and $i \geq 1$. If $w_{k+1} \cdots w_n = v^j$ and $w_0 \cdots w_k = v^{j'}$ for $j, j' \in \mathbb{N}_0$, then $w = v^{j+j'}$ which contradicts the primitiveness of w. If $v = w_l \cdots w_n w_0 \cdots w_{l'}$ where $k+1 \leq l \leq n$ and $0 \leq l' \leq k$. We denote $v_1 = w_0 \cdots w_{l'}$ $v_2 = w_l \ldots w_n$. And we have $w = v_1 v_2 \cdots v_1 v_2$, so w is a power of the word $v_1 v_2$. Furthermore, w contains $v_1 v_2$ at least two times, which contradicts the primitiveness of w.

□

Definition 5.2.6. A word w is Lyndon if it is primitive and lexicographically minimal among its conjugates for some given order.

Example 5.2.7. Let $\Sigma = \{0, 1, 2\}$. The word $w = 0012$ is Lyndon as it is primitive and lexicographically minimal among 0120, 1200 and 2001.

Definition 5.2.8. A word w is anti-Lyndon if it is primitive and lexicographically maximal among its conjugates for some given order.

Proposition 5.2.9. *Let w, u and v be words over an alphabet Σ. If w is not a prefix of u and $w \leq_{lex} u$ in the lexicographical order, than $wv <_{lex} u$ for all $v \in \Sigma^*$.*

Proof. This is a direct result of the definition of the lexicographical order. □

However, there are many different equivalent definitions of a Lyndon word.

Proposition 5.2.10. *A word w is Lyndon if and only if one of the following equivalent properties is satisfied*

1. *w is strictly inferior to each of its proper conjugates for some given order;*

2. *w is strictly inferior to each of its proper suffixes for some given order.*

Proof. Using Theorem 5.2.5, we know that a primitive word is different from each of its proper conjugates. So we have directly that a Lyndon word verifies the first property. Furthermore, if a word w verifies the first property, then we are sure that w is a Lyndon word because w is primitive, otherwise one of its proper conjugate would be equal to w. To prove the equivalence between a Lyndon word and the second property, we can observe that if the second statement is true *i.e.* for all non-empty words u and v such that $w = uv$, we have $w <_{lex} v$. We also have $v <_{lex} vu$, so $w <_{lex} vu$. So the second statement implies the first. Let's suppose that the first statement is true and the second is not. We have a word w such that u and v are two non-empty factors of w such that $w = uv$, $v \leq_{lex} w$ and $w <_{lex} vu$. Using Proposition 5.2.9, we know that v is a prefix of w because otherwise we would have $vu <_{lex} w$. Let's suppose that $w = vu'$ and therefore $vu' <_{lex} vu$. So we have $u' <_{lex} u$ and therefore, $u'v <_{lex} uv = w$, which is a contradiction.

□

As a direct consequence of Proposition 5.2.10, we have the following Lemma.

Lemma 5.2.11. *Let w be a Lyndon word:*

1. *none of the prefixes of w is also a suffix of w;*

2. *if v is a factor of w one of the following statements is true*

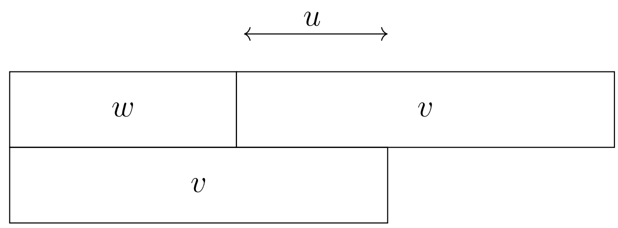
   (a) *v is a prefix of w;*

   (b) *$w <_{lex} v$.*

Proof. The first statement is a direct consequence of the second equivalence in Proposition 5.2.10 and the second statement is a direct consequence of the first statement in Proposition 5.2.10. □

Lemma 5.2.12. *Let w et v be two Lyndon words. The word wv is a Lyndon word if and only if $w <_{lex} v$.*

Proof. If w,v and wv are Lyndon words we have $w <_{lex} wv <_{lex} v$. If $w <_{lex} v$, then v can not be a prefix of w. Furthermore v can not be a prefix of wv



because then, there would exist a suffix u of v which would also be a prefix of v which contradicts Lemma 5.2.11. So we have $wv <_{lex} v$ due to the second statement in Proposition 5.2.10. Let u be a proper suffix of wv, u is either

- a suffix of v and we have $wv <_{lex} v \leq_{lex} u$ due to Proposition 5.2.10; or
- u is a factor of the form $w'v$, where w' is a proper suffix of w. Using Proposition 5.2.10, we have $w <_{lex} w'$ and because w' is a suffix of w it is not a prefix. Using Proposition 5.2.9 we have $wv <_{lex} w'v = u$.

Using the second statement in Proposition 5.2.10 we can conclude that wv is a Lyndon word. □

Lemma 5.2.13. *Let w be a Lyndon word over the alphabet Σ and v be a proper prefix of w. If $a \in \Sigma$*

1. *if $w <_{lex} va$, then for all $k \in \mathbb{N}$, $w^k va$ is a Lyndon word;*
2. *if va is not a prefix of w and $va <_{lex} w$, then for all $k \in \mathbb{N}_0$ and $u \in \Sigma^*$, the longest prefix of $w^k vau$, which is a Lyndon word is w.*

Proof. We will prove the first statement. Firstly we want to prove that va is a Lyndon word. We define u as a suffix of w such that $w = vu$ and $v'a$ a proper suffix of va. And we notice that $v'u$ is a proper suffix of w, so we have $vu <_{lex} v'u <_{lex} v'a$. Therefore, we also have $v <_{lex} v'a$. However, $|v| \geq |v'a|$, so we have $va <_{lex} v'a$ and using the second statement of Proposition 5.2.10, we know that va is a Lyndon word. As $w <_{lex} va$ and w and va are Lyndon words, using Lemma 5.2.12 we know that wva is also a Lyndon word. We can iterate this reasoning to obtain that $w^k va$ is a Lyndon word. We will now prove the second statement. If va is not a prefix of w and $va <_{lex} w$, then for all $k \in \mathbb{N}_0$ and $u \in \Sigma^*$ we have $vau <_{lex} w^k vau$. The longest Lyndon prefix of $w^k vau$ is of the form $w^k v$ and the longest Lyndon prefix of $w^k v$ is w due the second statement in Proposition 5.2.10. □

Remark 5.2.14. We notice that all the results which have been proven for Lyndon words are also true for anti-Lyndon words because anti-Lyndon words are Lyndon words, where we reversed the order of the alphabet.

Using those results on anti-Lyndon words, we are now able to show a link between string attractors of prefixes of the fixed points of the Fabre substitution and the base of the positional numeration system.

5.3 Link between string attractors of prefixes of the fixed points of the Fabre substitution and the base of the positional numeration system

To simplify the notion we introduce:

Notation 5.3.1. We will denote $\mathbf{c}$ as the infinite concatenation of the longest anti-Lyndon prefix of the word $\alpha_0 \cdots \alpha_{n-1}$. And for all $i \in \mathbb{N}$, we denote $\mathbf{c}_i$ the $(i+1)^{th}$ letter of $\mathbf{c}$.

Lemma 5.3.2. *Let $\alpha_0 \cdots \alpha_n$ verify WH. Then $\alpha_0 \cdots \alpha_{n-1} \geq_{lex} \mathbf{c}[0, n-1]$.*

Proof. We will proceed by contradiction. Let's suppose that w is the longest anti-Lyndon prefix of $\alpha_0 \cdots \alpha_{n-1}$. Let i be the smallest integer such that $|w| \leq i \leq n-1$ and $\alpha_0 \dots \alpha_i <_{lex} \mathbf{c}[0, i]$. Then $\alpha_0 \cdots \alpha_i = w^k va$ where $k \in \mathbb{N}_0$, v a proper prefix of w and a letter a such that if w' is a prefix of w of length $|va|$, then we have $w' >_{lex} va$. However, Lemma 5.2.13 in the case of anti-Lyndon words implies that $\alpha_0 \cdots \alpha_i$ is an anti-Lyndon word, which contradicts the maximality of w. □

Definition 5.3.3. Let $\alpha_0 \cdots \alpha_n$ verify the WH. For all $i \in \mathbb{N}$, we denote q_i the longest common prefix of X_f and $(f^i(0))^\omega$. Furthermore, we define $Q_i = |q_i|$.

Notation 5.3.4. We will denote $\mathbf{d}$ the periodisation of $\alpha_0 \cdots \alpha_n$. We have

$$\mathbf{d} = (\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1))^\omega.$$

Proposition 5.3.5. *Let $\alpha_0 \cdots \alpha_n$ verify WH. If $\alpha_0 \dots \alpha_{n-1}(\alpha_n - 1)$ is lexicographically maximal among its conjugates, then $\mathbf{d}[0, i] \leq_{lex} \mathbf{c}[0, i]$ for all $i \in \mathbb{N}$.*

Proof. Let's suppose that w is the longest anti-Lyndon prefix of $\alpha_0 \cdots \alpha_{n-1}$. The first part of this demonstration consists off proving that $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1) \leq_{lex} \mathbf{c}[0, n]$. Let's proceed by contradiction and suppose that there exists $k \in \mathbb{N}_0$, a word v, a proper prefix u of w and a letter b such that $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1) = w^k ubv$ and $ub >_{lex} w$. As $ub >_{lex} w$, we also have $ubvw^k >_{lex} \alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ because w is a proper prefix of $\alpha_0 \cdots \alpha_{n-1}$. Therefore $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ is not maximal among its conjugates which contradicts our initial hypothesis and we can conclude that $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1) \leq_{lex} \mathbf{c}[0, n]$. Using Lemma 5.3.2, we can conclude that $\alpha_0 \cdots \alpha_{n-1} = \mathbf{c}[0, n-1]$ and $(\alpha_n - 1) \leq_{lex} \mathbf{c}_n$. We can now prove the actual statement of the Proposition. If $\mathbf{c}_n >_{lex} \alpha_n - 1$, then we can directly conclude. If $\mathbf{c}_n = \alpha_n - 1$, then there exist a proper prefix u of w and $k \in \mathbb{N}_0$ such that $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1) = w^k u$. Let's suppose that v is a proper suffix of w such that $w = uv$. If $u \neq \epsilon$, we have $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1) = w^k u = (uv)^k u = u(vu)^k$. Furthermore

w is anti-Lyndon, so w is maximal among its conjugate. We have $u(vu)^k <_{lex} uw^k$, which contradicts our hypothesis that $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ is lexicographically maximal among its conjugates. If $u = \epsilon$, we have $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1) = w^k$. So we can conclude that we have $\mathbf{d}[0, i] \leq_{lex} \mathbf{c}[0, i]$ for all $i \in \mathbb{N}$. □

Lemma 5.3.6. *Under WH, for all $i \in \mathbb{N}$, we have*

1. $\text{rep}_S(U_i) = 10^i$*;*
2. *the number which have a representation of length $i + 1$ are those in $[U_i, U_{i+1} - 1]$;*
3. $\text{rep}_S(U_{i+1} - 1) = \boldsymbol{d}[0, i]$

where S is the Dumont-Thomas numeration system associated with f. In particular, we have $U_{i+1} - 1 = \sum_{j=0}^{n} \boldsymbol{d}_j U_{n-j}$.

Proof. The first point is a direct result of the construction of a numeration system associated with a fixed point (Theorem 1.3.7) and Property 3.3.15. We have proven that $\text{rep}_S(U_i) = 10^i$ and $\text{rep}_S(U_{i+1}) = 10^{i+1}$, which are both the first words in our language of size $i+1$ (resp. $i+2$). So for all $j \in [U_i, \ldots, U_{i+1} - 1]$, the representation of j is a word of length $i+1$. Lastly, as the representation of $(U_{i+1} - 1)$ is the last word in our alphabet of size $i+1$ in the genealogical order, which is the prefix of $\mathbf{d}$ of size $i+1$, so we can conclude. □

To be able to prove Proposition 5.3.8, we need the following Theorem, which has been proven in [17] and relies on Duval's algorithm computing the Lyndon factorisation of a word in linear time.

Theorem 5.3.7. *Let $\alpha_0 \cdots \alpha_n$ verify WH. For all $i \in \mathbb{N}$, we have $q_i = (f^i(0))^{\mathbf{c}_0} \cdots (f^0(0))^{\mathbf{c}_i}$. In particular, we have $Q_i = \sum_{j=0}^{i} \mathbf{c}_j U_{i-j}$.*

Proposition 5.3.8. *Let $\alpha_0 \cdots \alpha_n$ verify WH. If $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ is lexicographically maximal among its conjugates then $U_{i+1} - 1 \leq Q_i$ for all $i \in \mathbb{N}$.*

Proof. To prove this Proposition, we will proceed by contradiction. Let's suppose that there exists $i \in \mathbb{N}$ such that $U_{i+1} - 1 > Q_i$ so q_i is a proper prefix of $X_f[0, U_{i+1} - 2]$. Using Lemma 5.3.6, we know that $\text{rep}_S(U_{i+1} - 1) = \mathbf{d}[0, i]$, where S is the Dumont-Thomas numeration system associated with f for all $i \in \mathbb{N}_0$. Using Theorem 5.3.7, we know that $q_i = (f^i(0))^{\mathbf{c}_0} \cdots (f^0(0))^{\mathbf{c}_i}$. So $\mathbf{d}_0$ is the largest exponent such that $(f^i(0))^{\mathbf{d}_0}$ is still a prefix of $X_f[0, U_{i+1} - 2]$. We can conclude, that $\mathbf{d}_0 \geq_{lex} \mathbf{c}_0$. We can reiterate the same argument, if $\mathbf{d}_0 = \mathbf{c}_0$ on $\mathbf{d}_1$ and we obtain $\mathbf{d}_0\mathbf{d}_1 \geq_{lex} \mathbf{c}_0\mathbf{c}_1$. Reiterating this reasoning, we obtain $\mathbf{d}[0, i] \geq_{lex} \mathbf{c}[0, i]$. However as q_i is a proper prefix of $X_f[0, U_{i+1} - 2]$, we have $\mathbf{d}[0, i] >_{lex} \mathbf{c}[0, i]$, which contradicts Proposition 5.3.5. □

To simplify our notion, we define P_i for all $i \in \mathbb{N}$.

Definition 5.3.9. For $i \in \mathbb{N}$, we define

$$P_i = \begin{cases} U_i & \text{if } 0 \leq i \leq n; \\ U_i + U_{i-n} - U_{i-n-1} - 1 & \text{if } i > n. \end{cases}$$

Moreover for the same reason we define the following string attractors.

Definition 5.3.10. For $i \in \mathbb{N}$, we define

$$\Gamma_i = \begin{cases} \{U_0, \ldots, U_i\} & \text{if } 0 \leq i \leq n; \\ \{U_{i-n} \ldots, U_i\} & \text{if } i > n. \end{cases}$$

Lemma 5.3.11. *Let $\alpha_0 \cdots \alpha_n$ verify WH. If $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ is maximal among its conjugates, then $P_i \leq U_{i+1} - 1 \leq Q_i$ for all $i \in \mathbb{N}$.*

Proof. This proof is direct because of the definition of P_i for all $i \in \mathbb{N}$ and because of Theorem 5.3.8. □

To be able to prove the main theorem of this chapter, we need the following Lemma, which gives us a useful characterisation. The proof of this Lemma can be found in [17].

Lemma 5.3.12. *Let $\alpha_0 \cdots \alpha_n$ verify WH. The numeration system $S = (\Sigma, L, <)$ is greedy if and only if, for all $w \in L$ and for all $i \leq |w|$, the suffix of length i of w is smaller than or equal to $\boldsymbol{d}[0, i-1]$. Moreover, we then have*

$$L = \{w = w_0 \cdots w_j \in \mathbb{N}_* \setminus 0\, \mathbb{N}^* \,|\forall 0 \leq i \leq j : w_{j-i} \cdots w_j \leq_{lex} \boldsymbol{d}[0, i]\}$$

where $S = (\Sigma, L, <)$ is the Dumont-Thomas numeration system associated with f as seen in Chapter 1. Furthermore, $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ is lexicographically maximal among its conjugates if and only if S is greedy.

Notation 5.3.13. For technical reasons, we set $\Gamma_{-1} = \emptyset$.

Remark 5.3.14. Even though, the results of section 4.1, were obtained for a simple Parry number. Some of those results, such as Lemma 4.1.2 are still correct in this context because we did not use the hypothesis that $\alpha_0 \cdots \alpha_n$ verifies the Parry condition in the proof.

Theorem 5.3.15. *Let $\alpha_0 \cdots \alpha_n$ verify WH and $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ is maximal among its conjugates. Let $i \geq 0$,*

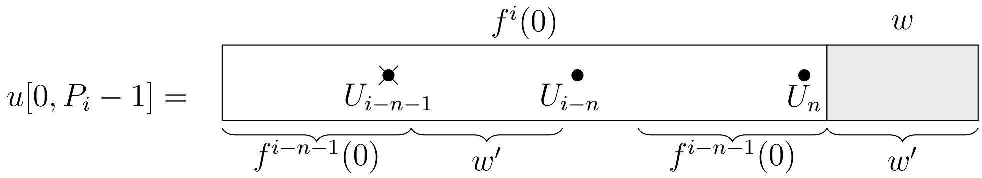
- *if $m \in [U_i, Q_i]$, then $\Gamma_{i-1} \cup U_i$ is a string attractor of $X_f[0, m-1]$;*
- *if $m \in [P_i, Q_i]$, then Γ_i is a string attractor of $X_f[0, m-1]$.*

Proof. We will prove those statements by induction on i. Base case: If $i = 0$, then we have $1 \leq m \leq \alpha_0$ because $Q_0 = \alpha_0$ as $(f^0(0))^\omega$ only contains the letter 0 and X_f has $0^{\alpha_0}1$ as prefix. So we know that $X_f[0, m) = 0^m$ and in booth cases we can directly conclude because a set containing the position $U_0 = 1$ is a string attractor of $X_f[0, m-1]$. Induction:

We assume that the statements hold for all integers smaller than $i \in \mathbb{N}_0$ and we will prove that those statements also hold for i. Firstly, we will prove the first statement. We know that for every $m \in [P_{i-1}, Q_{i-1}]$, Γ_{i-1} is a string attractor of $X_f[0, m-1]$. Using Lemma 5.3.11, we know that $P_{i-1} \leq U_i - 1 \leq Q_{i-1}$, so in particular, we have Γ_{i-1} is a string attractor of $X_f[0, U_i - 2]$. This means that $\Gamma_{i-1} \cup \{U_i\}$ is a string attractor of $(f^i(0))^\omega$ due to Lemma 5.1.5 because using Property 3.3.15, we know that U_i is of length of $f^i(0)$. Furthermore, as Q_i is the size of the longest common prefix between X_f and $(f^i(0))^\omega$, we can conclude that $\Gamma_{i-1} \cup \{U_i\}$ is a string attractor for all $m \in [P_i, Q_i]$. We will now prove the second statement. First, we notice that due to Definition 5.3.9, that $P_i \geq U_i$ for all $i \in \mathbb{N}$. Furthermore, using Lemma 5.1.5 and the condition that Γ_i contains U_i, we know that for all $i \in \mathbb{N}$, if Γ_i is a string attractor for $X_f[0, P_i - 1]$, then Γ_i is also a string attractor for $X_f[0, m-1]$ with $m \in [P_i, Q_i]$. If $0 \leq i \leq n$, then by definition of Γ_i, we have $\Gamma_i = \{U_0, \dots, U_i\} = \Gamma_{i-1} \cup \{U_i\}$ and we can conclude using the first statement. If $i > n$, we know that

$$\begin{aligned}\Gamma_{i-1} \cup \{U_i\} &= \{U_{i-1-n}, \dots, U_{i-1}\} \cup \{U_i\} \\ &= \{U_{i-1-n}\} \cup \{U_{i-n}, \dots, U_i\} \\ &= \Gamma_i \cup \{U_{i-1-n}\}\end{aligned}$$

is a string attractor of $X_f[0, P_i - 1]$ due to the first statement. So we need to prove that the addition of the position U_{i-1-n} is not needed to create a string attractor of $X_f[0, P_i - 1]$. More precisely, we will show that if a factor of $X_f[0, P_i - 1]$ is referenced by U_{i-1-n} and no other position in Γ_i, that this factor has another occurrence in $X_f[0, P_i - 1]$ where one of its positions is referenced by U_i. We have $\Gamma_i = \{U_{i-n}, \dots, U_i\}$, so the smallest position in Γ_i is U_{i-n}, we need to consider the occurrences crossing position U_{i-n-1} in $X_f[0, U_{i-n} - 2]$ but not U_{i-n}. So if we define w as the suffix of $X_f[0, P_i - 1]$ such that $X_f[0, P_i - 1] = f^i(0)w$. We can show that $f^{i-n-1}(0)$ is a suffix of $f^i(0)$ and that $w' := X_f[U_{i-n-1}, U_{i-n} - 2]$ is a prefix of w.



First of all, using Lemma 4.1.2 and because the WH assures us that $\alpha_n \geq 1$, we know that $f^{i-n-1}(0)$ is suffix of $f^i(0)$. Furthermore, we have by definition of P_i

$$|w| = P_i - U_i = U_{i-n} - U_{i-n-1} - 1. \tag{5.1}$$

Using Lemma 4.1.2, we also know that $f^i(0)$ is followed by

$$(f^i(0))^{\alpha_0 - 1} \cdots (f^{i-n}(0))^{\alpha_n}$$

in X_f. Since $f^{i-n}(0)$ is a prefix of $f^{i-n}(0), \ldots, f^{i-1}(0)$, the word $f^i(0)$ is in particular followed by $f^{i-n}(0)$ in X_f. Due to (5.1) we know that $|w| \leq U_{i-n}$, this implies that w is a prefix of $f^{i-n}(0)$. And by construction of X_f, w is also a prefix of X_f. To prove that $w = w'$, we need to distinguish between two cases. If $i - 2(n+1) + 1 \geq 0$, then by definition of w' and Lemma 4.1.2, we know that w' is a prefix of $v := (f^{i-n-1}(0))^{\alpha_0 - 1} \cdots (f^{i-2n-1}(0))^{\alpha_n}$. We define the word $u = (\alpha_0 - 1)\alpha_1 \cdots \alpha_n 0^{i-2n-1}$. If u starts with one or more 0, we consider the longest suffix of u, which does not start with 0. We can notice that the word u corresponds to a factorisation of v into the words $f^{i-n-1}(0), \ldots, f^0(0)$ because we have

$$v = (f^{i-n-1}(0))^{\alpha_0 - 1} \cdots (f^{i-2n-1}(0))^{\alpha_n} \underbrace{(f^{i-2n-2}(0))^0 \cdots (f^0(0))^0}_{i-2n-1 \text{ words to the power of } 0}.$$

Due to the hypothesis that $\alpha_0 \cdots \alpha_{n-1}(\alpha_n - 1)$ is maximal among its conjugates and Lemma 5.3.12, we know that u is in the numeration language of the Dumont-Thomas numeration system associated f, $S = (\Sigma, L, <)$ as seen in Chapter 1. So u is Dumont-Thomas factorisation of v, which implies that the word v is a prefix of X_f.

If $i-2(n+1)+1 < 0$, we can reason similarly by using $v := (f^{i-n-1}(0))^{\alpha_0} \cdots (f^0(0))^{\alpha_{i-n-1}}$ and $u = (\alpha_0 - 1)\alpha_1 \cdots \alpha_{i-n-1}$. So w' is also a prefix of X_f and therefore $w = w'$. □

Example 5.3.16. Let's reconsider Example 4.2.9. We have $\theta = 1 + \sqrt{3}$. The θ-expansion of 1 is $D_\theta(1) = 22$. We notice that $D_\theta(1)$ verifies the WH. The corresponding Fabre substitution is $(f_\theta, 0, 1, 0)$ where:

$$\begin{aligned} f :& \{0,1\}^* \to \{0,1\}^* \\ & 0 \mapsto 001 \\ & 1 \mapsto 00. \end{aligned}$$

The fixed point of this substitution is

$$X_\theta = 001001000010010000100100 1001001 \ldots$$

We have $f^2(0) = 00100100$ and therefore $Q_2 = 24$. So $\Gamma_2 = \{3, 8\}$ is a string attractor of the prefix of X_θ of size 24. The letters in orange are the position referenced by the string attractor Γ_2

$$001001000010010000100100.$$

Bibliography

[1] CANTERINI, Vincent and Anne SIEGEL. Automate des préfixes-suffixes associé à une substitution primitive. *Journal de theorie des nombres de bordeaux*. 2001, 13 (2), pp. 353–369.

[2] CASSAIGNE, Julien, France GHEERAERT, Antonio RESTIVO, Giuseppe ROMANA, Marinella SCIORTINO and Manon STIPULANTI. New string attractor-based complexities for infinite words. *Journal of combinatorial theory. Series A*. 2024, 208.

[3] CHARLIER, Emilie, Célia CISTERNINO and Manon STIPULANTI. Regular sequences and synchronized sequences in abstract numeration systems. *European Journal of Combinatorics*. 2022, 101.

[4] CHARLIER, Emilie, Tomi KÄRKI and Michel RIGO. Multidimensional generalized automatic sequences and shape-symmetric morphic words. *Discrete Mathematics*. 2010, 310, pp. 1238–1252.

[5] CHRISTOL, G., T. KAMAE, Michel MENDÈS FRANCE and Gérard RAUZY. Suites algébriques, automates et substitutions. *Bulletin de la Société mathématique de France*. 1980, 108, pp. 401–419.

[6] COBHAM, Alan. On the hartmanis-stearns problem for a class of tag machines. *In*: *9th Annual Symposium on Switching and Automata Theory (swat 1968)*, IEEE, 1968, pp. 51–60. Available from DOI: 10.1109/SWAT.1968.20.

[7] COBHAM, Alan. Uniform tag sequences. *Mathematical Systems Theory*. 1972, 6, pp. 164–192.

[8] DELECROIX, Vincent, Tomáš HEJDA and Wolfgang STEINER. Balancedness of arnoux-rauzy and brun words. *In*: *Combinatorics on Words*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, vol. 8079, pp. 119–131.

[9] DUMONT, Jean Marie and Alain THOMAS. Systemes de numeration et fonctions fractales relatifs aux substitutions. *Theoretical computer science*. 1989, 65 (2), pp. 153–169.

[10] DURAND, Fabien. Decidability of uniform recurrence of morphic sequences. *International journal of foundations of computer science*. 2013, 24 (01), pp. 123–146.

[11] DUVAL, Jean-Pierre. Mots de lyndon et périodicité. *R.A.I.R.O. Informatique théorique*. 1980, 14 (2), pp. 181–191.

[12] FABRE, Stéphane. Substitutions et β-systèmes de numération. *Theoretical computer science*. 1995, 137 (2), pp. 219–236.

[13] FICI, Gabriele and Svetlana PUZYNINA. Abelian combinatorics on words: a survey. *Computer science review*. 2023, 47.

[14] FRAENKEL, Aviezri S. Systems of numeration. *The American Mathematical Monthly*. 1985, 92 (2), pp. 105–114. Available from eprint: `https://doi.org/10.1080/00029890.1985.11971550`.

[15] FROUGNY, Christiane, Zuzana MASÁKOVÁ and Edita PELANTOVÁ. Complexity of infinite words associated with beta-expansions. *RAIRO. Informatique théorique et applications*. 2004, 38 (2), pp. 163–185.

[16] GHEERAERT, France, Giuseppe ROMANA and Manon STIPULANTI. String attractors of fixed points of k-bonacci-like morphisms. *In*: *Combinatorics on Words*, Cham: Springer Nature Switzerland, 2023, (Lecture Notes in Computer Science), vol. 13899, pp. 192–205.

[17] GHEERAERT, France, Giuseppe ROMANA and Manon STIPULANTI. String attractors of some simple-parry automatic sequences. *Theory of Computing Systems*. 2024, 68, pp. 1601–1621.

[18] KEMPA, Dominik and Nicola PREZZA. At the roots of dictionary compression: string attractors. *In*: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, Los Angeles, CA, USA: Association for Computing Machinery, (STOC 2018), pp. 827–840. Available from DOI: `10.1145/3188745.3188814`.

[19] KRECZMAN, Savinien, Sébastien LABBÉ and Manon STIPULANTI. *Positionality of Dumont–Thomas numeration systems for integers*. 2025.

[20] LABBÉ, Sébastien and Jana LEPŠOVÁ. Dumont–thomas numeration systems for $\mathbb{Z}$. *arXiv (Cornell University)*. 2023.

[21] LABBÉ, Sébastien and Jana LEPŠOVÁ. Dumont–thomas complement numeration system for $\mathbb{Z}$. *arXiv (Cornell University)*. 2023.

[22] LECOMTE, P. and M. RIGO. Abstract numeration systems. *In*: *Combinatorics, Automata and Number Theory*, Cambridge University Press, 2010, pp. 108–162.

[23] LECOMTE, Pierre and Michel RIGO. Numeration systems on a regular language. *Theory of Computeing Systems*, 34, pp. 27–44.

[24] LEPŠOVÁ, Jana. *Substitutive structures in combinatorics, number theory, and discrete geometry*. 2024.

[25] MANTACI, Sabrina, Antonio RESTIVO, Giuseppe ROMANA, Giovanna ROSONE and Marinella SCIORTINO. A combinatorial view on string attractors. *Theoretical computer science*. 2021, 850, pp. 236–248.

[26] MARSAULT, Victor. *Énumération et numération*. PhD thesis. Télécom ParisTech. 2016.

[27] MASSUIR, Adeline. *Positional Numeration Systems: Ultimate Periodicity, Complexity and Automatic Sequences*. PhD thesis. ULiège - Université de Liège, Rigo, Michel, 2021.

[28] MITROFANOV, I. V. Periodicity of morphic words. *Journal of mathematical sciences (New York, N.Y.)* 2015, 206 (6), pp. 679–687.

[29] NERODE, A. Linear automaton transformations. *Proceedings of the American Mathematical Society*. 1958, 9 (4), pp. 541–544.

[30] PARRY, W. On the $\beta-$expansions of real numbers. *Acta Mathematica Academiae Scientiarum Hungaricae*. 1960, 11 (3-4), pp. 401–416.

[31] RÉNYI, A. Representations for real numbers and their ergodic properties. *Acta Mathematica Academiae Scientiarum Hungaricae*. 1957, 8 (3-4), pp. 477–493.

[32] RIGO, Michel. *Formal Languages, Automata and Numeration Systems 2: Applications to Recognizability and Decidability*. 1st edition. Vol. 9781848217881. Newark: John Wiley and Sons, Incorporated, 2014.

[33] RIGO, Michel. Generalization of automatic sequences for numeration systems on a regular language. *Theoretical Computer Science*. 2000, 224, pp. 271–281.

[34] RIGO, Michel. Numeration systems on a regular language: arithmetic operations, recognizability and formal power series. *Theoretical computer science*. 2001, 269 (1-2), pp. 469–498.

[35] RIGO, Michel and Arnaud MAES. More on generalized automatic sequences. *Journal of Automata, Languages and Combinatorics*. 2002, 7 (3), pp. 351–376.