Università degli Studi di Napoli Federico II

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Dottorato di Ricerca in Scienze Matematiche XX Ciclo

On the repetitivity index of infinite words

Valerio D'Alonzo

Tutore:
Ch.mo Prof.
Aldo de Luca

Relatore:
Ch.mo Prof.
Arturo Carpi

Coordinatore:
Ch.mo Prof.
Salvatore Rionero

Contents

1 Introduction **3**

2 Preliminaries **7**

2.1 Words 7

2.2 Automata and automatic sequences 10

3 Repetitivity index **15**

3.1 Main properties 15

4 Beck's Theorem **22**

4.1 Preliminaries 23

4.2 Beck's Theorem: Lovász Local Lemma 23

4.3 Beck's Theorem: an effective construction 27

5 Thue-Morse word **34**

5.1 Definitions and main properties 35

5.2 Computing the repetitivity index 37

6 C^∞-words **42**

6.1 Preliminaries 43

6.2 Main properties of smooth words 45

6.3 Repetitivity index of smooth words 49

6.4 Further remarks 52

7 Synchronized sequences **53**

7.1 Preliminaries 53

7.2 Repetitivity index of k-synchronized sequences 55

8 Final remarks **58**

Bibliography **60**

Acknowledgments **65**

Chapter 1

Introduction

The existence of infinite words on a finite alphabet without adjacent repeats of a same factor is one of the oldest problems in Combinatorics on words, [4]. In [40] and [41], the norwegian mathematician Axel Thue (1863-1922) noted that any binary sequence of lenght larger than 4 must contain a *square* i.e. two consecutive identical factors. He then asked whether it was possible to find an infinite binary sequence that neither should contain any *cube* i.e. three consecutive identical factors nor *overlaps* i.e. a factor of the form $awawa$, where $a \in \{0, 1\}$ and w is a factor. The answer to all these questions was positive. Thue used a sequence

$$\mathbf{t} = 0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1 \cdots$$

which construction is given in Chapter 5. More general repetitions can be considered as well. Thue himself called a word on n letters *irreducible* if two distinct occurrences of a non empty factor are always separated by at least $n-2$ letters. Thus, irreducible means overlap-free if $n = 2$ and square-free if $n = 3$. A more general concept, first considered by F. Dejean [16] is what we call *fractional repetition*: a word xyx, where x is nonempty, is

a repetition of *exponent* l, where $l = |xyx|/|xy|$. Thue proved that every binary word of lenght 4 contains a square and that there exist infinite binary overlap-free words (such as the Thue-Morse word). A similar property holds for ternary words: Dejean proved that every ternary word of lenght 39 contains a repetition of exponent $7/4$ and exists a word generated by a suitable endomorphism that has no repetitions of exponent larger than $7/4$. We call *repetition threshold* the smallest number $\mathrm{RT}(k)$ such that there exists an infinite word over k letters that has only repetitions of exponent less than or equal to $\mathrm{RT}(k)$. We know that $\mathrm{RT}(2) = 2$, $\mathrm{RT}(3) = 7/4$, [16]. It was conjectured, in the same paper, that $\mathrm{RT}(4) = 7/5$ and $\mathrm{RT}(k) = k/(k-1)$, for $k \geq 5$. This conjecture has been proved to be true for $k = 4$ by Pansiot [34] and, with extensive use of computer, for $5 \leq k \leq 11$ by Moulin-Ollagnier [29] and more recently, for $12 \leq k \leq 14$ by Mohammad-Noori and Currie [28] and for $k \geq 33$, by Carpi [8]. Recently, Shallit et al. [24] has introduced the notion of *generalized repetition threshold* $\mathrm{RT}(k, l)$ which takes into account not only the exponent but also the lenght of the factors to be avoided. Dejean's conjecture is equivalent to say that, for any $k \geq 5$, there exists a word on k letters such that the distance among any two occurrences of a same factor of lenght n is at least $n(k-1)$ and this bound is tight. To formalize this notion of minimal distance among two occurrences of a same factor of lenght n in a fixed word w, A. Carpi and V. D'Alonzo introduced in [10] a particular function $\mathrm{I}_w(n)$, called *repetitivity index.* In this thesis we will study some of its interesting properties. In particular, we will study the repetitivity index of some classes of words. In [3], J. Beck proved that given an arbitrary small $\epsilon > 0$, there exists a binary infinite word w such that, for all sufficiently large n, $\mathrm{I}_w(n) \geq (2-\epsilon)^n$. This proof is not constructive and it uses a powerful tool in combinatorics, namely the Lovász Local Lemma [22].

In Chapter 4, we effectively construct an infinite word w having a similar (though weaker) property: for any $\nu > 0$, there exists an integer $n(\nu)$ such that, for all $n > n(\nu)$, $\mathrm{I}_w(n) > n^\nu$. In Chapter 5, we study the repetitivity index $\mathrm{I}_w(n)$ when w is the Thue-Morse word. We effectively obtain all the values of this function, for all $n > 0$. In particular, we obtain these values:

$$\mathrm{I}_{\mathbf{t}}(n) = n,$$

if $n = 1, 2, 3, 4$ and

$$\mathrm{I}_{\mathbf{t}}(n) = 2\,\mathrm{I}_{\mathbf{t}}(\lceil n/2 \rceil),$$

for any $n > 4$.

In Chapter 6, we study this function when w is the Kolakoski word. This word,

$$\mathbf{k} = \underbrace{22}_{2}\underbrace{11}_{2}\underbrace{2}_{1}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{11}_{2}\cdots$$

introduced by Kolakoski himself in [33], is a word over the alphabet $\{1, 2\}$ defined by the property that the word of its *runlenghts* is equal to $\mathbf{k}$ itself. Here a run is a maximal sub-word of consecutive identical letters. This sequence is representative by the class of infinite C^∞-words. In this Chapter, we prove that the repetitivity index, for any C^∞-word, and in particular for the Kolakoski word, is ultimately bounded from below by $n + rn^{1/q}$ where r is a suitable constant and $q = \log 1.5006 / \log 1.4994$. This research is motivated by the fact that A. Carpi in [9] conjectured that, for any rational $e > 1$, the lenght of the factors of the Kolakoski word with exponent larger than e is bounded. This is equivalent to say, as proved in [10], that repetitivity index of Kolakoski word is not linearly bounded. In last Chapter, we study the repetitivity index of k-synchronized sequences. These are integer sequences whose graph is represented, in a fixed base k, by a right synchronized rational relation. The notion of a k-synchronized sequence was introduced in [23] as an

intermediate notion between those of k-automatic and k-regular sequences, introduced respectively by Cobham [14] and by Allouche and Shallit [2]. The main result of this section is that repetitivity index of a k-synchronized sequence is a k-synchronized sequence itself.

Chapter 2

Preliminaries

2.1 Words

All through this paper, we will denote by A^* the *free monoid* generated by an *alphabet* A. Its elements are called *words*. The neutral element of A^*, or *empty word*, will be denoted by ϵ. The *length* of a word $w \in A^*$ will be denoted by $|w|$, while $|w|_a$ will denote the number of occurrences of the letter $a \in A$ in w. A word $v \in A^*$ is a *factor* of the word w if there exist $r, s \in A^*$ such that $w = rvs$. The set of the factors of a word w is denoted by $\mathrm{Fact}(w)$. If $r = \epsilon$, then v is a *prefix* of w; if $s = \epsilon$, then v is a *suffix* of w. The sets of all prefixes and suffixes of a word w is denoted with $\mathrm{Pref}(w)$ and $\mathrm{Suff}(w)$, respectively. The *reversal operation* is the unary operation $\sim$ in A^* recursively defined as $\epsilon^\sim = \epsilon$ and $(ua)^\sim = (au^\sim)$ for all $u \in A^*$ and $a \in A$. A word w which coincides with its reversal is called *palindrome*. Let $A = \{1, 2\}$. The *mirror image* of a word $w \in A^*$ is the word obtained by the interchange of $1's$ and $2's$. For example, the word 112212 is the mirror image of the word 221121. Let t be a non-negative integer. A *repetition with gap* t is any word of the form uvu with $u, v \in A^*$ and $|v| = t$. A non-empty repetition with

gap 0 is said to be a *square*. A *cube* is any word of the form uuu with u a non-empty word. An *infinite word* on A is any unending sequence of letters

$$w = w_1 w_2 \cdots w_n \cdots,\ w_i \in A,\ i \geq 1.$$

Its *factors* are the words $w_i w_{i+1} w_{i+2} \cdots w_j$ $(1 \leq i \leq j)$, as well as the empty word. In particular, the factors $w_1 w_2 \cdots w_j$ are the *prefixes* of w. The set of all infinite words is denoted with A^ω.

A word $w \in A^\omega$ is *uniformly recurrent* if for each finite factor of symbols r occurring in w there exists an integer n such that for all i, the factor $w_{i+1} \cdots w_{i+n}$ contains an occurrence of r. A word $w \in A^\omega$ is *ultimately periodic* if there exist integers $p \geq 1,\ N \geq 0$ such that $w_i = w_{i+p}$, for all $i \geq N$.

A sequence $(u_n)_{n \geq 0}$ of finite words over an alphabet A *converges* to an infinite word u if every prefix of u is a prefix of all but a finite number of the word u_n. This word u is unique and is denoted by

$$u = \lim_n u_n.$$

As an example, the sequence $a^n b^n$ converges to $a^\omega = aaaaa \cdots$.

Let $w \in A^\omega$. The *subword complexity* λ_w of w is the map $\lambda_w : \mathbb{N} \to \mathbb{N}$ defined by

$$\lambda_w(n) = \mathrm{Card}(\mathrm{Fact}(w) \cap A^n),$$

for all $n \in \mathbb{N}$, where A^n is the set of all words of lenght n on the alphabet A.

Example 2.1.1 Let $A = \{0, 1\}$. The Fibonacci word is the infinite word inductively defined as follows

$$f_0 = 0,\ \ f_1 = 1,\ \ f_{n+1} = f_n f_{n-1},\ n \geq 1$$

and

$$f = \lim_n f_n = 010010100\cdots$$

It results that $\lambda_f(n) = n + 1$. □

The *recurrency index* of an infinite word w is the function $\rho_w : \mathbb{N} \to \mathbb{N} \cup \{\infty\}$ defined as follows: for any $n \geq 0$, $\rho_w(n)$ is the least integer, if any exists, such that each factor of w of lenght $\rho_w(n)$ contains every factor of w of lenght n. If such an integer does not exist, then $\rho_w(n) = \infty$. If $\rho_w(n)$ is finite for all $n \geq 0$, then w is *uniformly recurrent*.

A *morphism* is a map ϕ satisfying $\phi(xy) = \phi(x)\phi(y)$, for all $x, y \in A^*$. Let $k \geq 2$ an integer. A morphism ϕ is k-uniform if $|\phi(a)| = k$, for all $a \in A$.

Definition 2.1.1 *A positive integer p is a period of $w = w_1 w_2 \cdots w_n$ if whenever $1 \leq i, j \leq |w|$, one has that*

$$i \equiv j \pmod{p} \;\Rightarrow\; w_i = w_j$$

As is well known [25], a word w has a period $p \leq |w|$ if and only if $w = ur = su$, with $|s| = |r| = p$. We recall the famous theorem of Fine and Wilf stating that if a word w has two periods p and q, and $|w| \geq p + q - gcd(p, q)$, then w has also the period $gcd(p, q)$. The minimal period of a word w is denoted by π_w. For example, the periods of the word $w = 101101$ are $3, 5$ and any integer $p \geq 6$. Therefore, its minimal period is $\pi_w = 3$.

Definition 2.1.2 *The exponent of a word w, denoted by $e(w)$ is the ratio among its lenght and its minimal period.*

For example, the word $w = 1011011$ has minimal period $\pi_w = 3$ and lenght $|w| = 7$; therefore, $e(w) = 7/3$.

Definition 2.1.3 *The critical exponent of an infinite word w, is defined in this way*

$$\mathrm{ce}(w) = \sup\{e(x) \mid x \in \mathrm{Fact}(w)\}.$$

For any integer $l \geq 1$, the *generalized critical exponent* of an infinite word w is defined as

$$\mathrm{ce}(w, l) = \sup\{e(x) \mid x \in \mathrm{Fact}(w), \text{ with } |x| \geq l\}.$$

We observe that

- $\mathrm{ce}(w) \geq \mathrm{ce}(w, l)$, for all $l \geq 1$.
- $\mathrm{ce}(w) = \mathrm{ce}(w, 1)$

2.2 Automata and automatic sequences

A *deterministic finite automaton*, or DFA, is one of the simplest possible models of computation, [1]. It is an *acceptor*; that is, strings are given as input and are either accepted or rejected. A DFA starts in an *initial state* and after reading the input can be in one of a finite number of states. The DFA takes as input a string w and-based on the symbols of w, read in order from left to right-moves from state to state. If after reading all the symbols of w the DFA is in a distinguished state called an *accepting state* (or *final state*), then the string is accepted; otherwise, it is rejected. The language accepted by the DFA is the set of all accepted strings. A DFA can be represented by a directed graph called a *transition diagram*. A directed edge labeled with a letter indicates the new state of the machine if the given letter is read. More formally, a DFA M is defined to be a 5-tuple

$$M = (Q, A, \delta, q_0, F),$$

where

- Q is a finite set of states
- A is the finite input alphabet
- $\delta : Q \times A \to Q$ is the *transition function*
- $q_0 \in Q$ is the *initial state* and
- $F \subset Q$ is the set of accepting states.

Note that a DFA is said to be *complete* if δ is defined for all pairs in its range. In order to formally define acceptance by DFA, we need to extend the domain of δ to $Q \times A^*$, where A^* is the free monoid generated by the alphabet A, respect to the operation of concatenation. We do this as follows: first, we define $\delta(q, \epsilon) = q$, for all $q \in Q$, and define $\delta(q, xa) = \delta(\delta(q, x), a)$ for all $q \in Q,\ x \in A^*$ and $a \in A$. Then $L(M)$, the language accepted by M, is defined to be:

$$L(M) = \{w \in A^* \ : \ \delta(q_0, w) \in F\}.$$

We call a state q of a DFA *reachable* if there exists $x \in A^*$ such that $\delta(q_0, x) = q$, and *unreachable*, otherwise. A *deterministic finite automaton* with output, DFAO, is defined to be a 6-tuple:

$$M = (Q, A, \delta, q_0, \Delta, \tau),$$

where Q, A, δ, q_0 are as in the definition of DFA, Δ is the output alphabet and $\tau \ : \ Q \to \Delta$ is the output function. We define a function from A^* to Δ, which we denote as $f_M(w)$, as follows

$$f_M(w) = \tau(\delta(q_0, w)).$$

This function is called a *finite-state function*. We can represent a DFAO with a transition diagram, much the same way we did for DFAs; the only difference is that a state labeled q/a indicates that the output associated with the state q is the symbol a. When the input alphabet $A = A_k = \{0, 1, 2 \ldots, k-1\}$ for an integer $k \geq 2$, we call DFAO as k-DFAO. Now, we are ready to define the notion of a k-automatic sequence. Informally, a sequence $(a_n)_{n\geq 0}$ is k-automatic if a_n is a finite-state function of the base-k digits of n. More precisely, we compute a_n by feeding a finite automaton with the base-k representation of n, starting with the most significant digit, and then applying an output mapping τ to the last state reached. More formally, we say that a sequence $(a_n)_{n\geq 0}$ over a finite alphabet Δ is *k-automatic* if there exists a k-DFAO

$$M = (Q, A_k, \delta, q_0, \Delta, \tau),$$

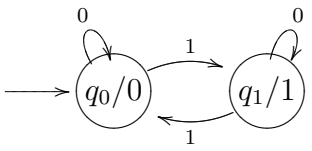
such that $a_n = \tau(\delta(q_0, w_n))$, for all $n \geq 0$, where w_n is the expansion of n in base k. If M is as above, we say M *generates* the sequence $(a_n)_{n\geq 0}$.

Example 2.2.1 (**Thue-Morse sequence**)

Let $A = \{0, 1\}$ and $\mu : A^* \to A^*$ the 2-uniform morphism such that $\mu(0) = 01$ and $\mu(1) = 10$. The Thue-Morse sequence is defined in this way:

$$\mathbf{t} = \lim_n \mu^n(0) = 01101001 \cdots,$$

where $\mu^n(0) = \mu(\mu^{n-1}(0))$. The Thue-Morse sequence is 2-automatic, since it can be generated by the 2-DFAO in figure below



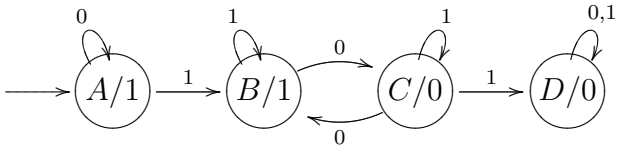
n	0	1	2	3	4	5	6	7
t_n	0	1	1	0	1	0	0	1

Example 2.2.2 (**The Baum-Sweet sequence**)

This sequence $\mathbf{b} = (b_n)_{n\geq 0}$ takes the value 1 if the binary representation of n contains no block of consecutive 0's of odd lenght and 0 otherwise. Here are the first few therms of this sequence:

n	0	1	2	3	4	5	6	7
b_n	1	1	0	1	1	0	0	1

The 2-DFAO in figure below generates this sequence:



Here the meaning of the states is as follows:

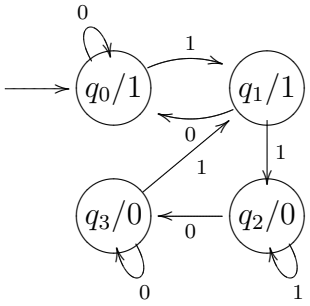
- A: reading the leading zeros of the input;
- B: all blocks of zeros (including current one) are even lenght;
- C: the last block of zeros seen has odd lenght so far, but all previous ones have even lenght;
- D: we' ve seen a block of zeros of odd lenght.

Example 2.2.3 (**The regular paperfolding sequence**) First, take a rectangular piece of paper and fold it in half lenghtwise, then fold the result in

half again, etc., ad infinitum, taking care to make the folds in the same direction each time. Next, unfold the paper. The sequence $(R_i)_{i\geq 1}$ of 'hills' (1) and 'valleys' (0) that results is called the *regular paperfolding sequence*. For example, after fold and unfolding to 90° we obtain this sequence:

n	1	2	3	4	5	6	7	8
R_n	1	1	0	1	1	0	0	1

The regular paperfolding sequence $\mathbf{R} = (R_n)_{n\geq 1}$ is generated by the 2-DFAO in figure below:



Chapter 3

Repetitivity index

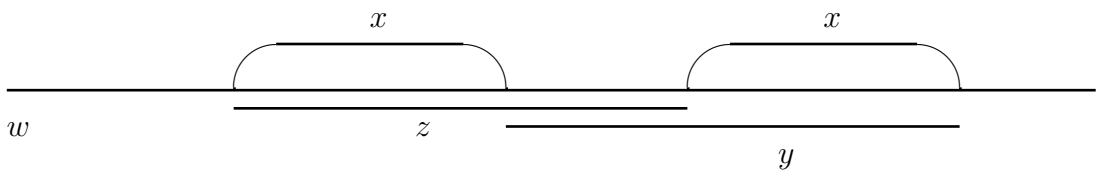
3.1 Main properties

Let w an infinite word on finite alphabet A.

Definition 3.1.1 *The repetitivity index of w is the function $\mathrm{I}_w : \mathbb{N} \to \mathbb{N}$ defined as follows:*

$$\mathrm{I}_w(n) = \min\{k > 0 \mid \exists x, y, z \ s.t. \ |x| = n, \ |y| = |z| = k, \ xy = zx \in \mathrm{Fact}(w)\}\,.$$

In other terms, $\mathrm{I}_w(n)$ gives the minimal distance among any two occurences of a same factor of length n in w. Grafically,



$$I_w(n) = \min_{|x|=n} |z| = \min_{|x|=n} |y|$$

In this section, we study some properties of the repetitivity index.

Proposition 3.1.1 *The repetitivity index of an infinite word w is a non-decreasing function.*

Proof. For any $n > 0$ we can find $x, y, z \in A^*$ such that $|x| = n$, $|y| = |z| = \mathrm{I}_w(n)$, $xy = zx \in \mathrm{Fact}(w)$. Setting $x = x'a$, $ay = y'b$, $a, b \in A$, one has $x'y'b = xy = zx = zx'a$ and therefore $x'y' = zx'$, $a = b$, with $|x'| = n - 1$, $|y'| = |z| = \mathrm{I}_w(n)$. This implies $\mathrm{I}_w(n-1) \leq \mathrm{I}_w(n)$. □

There is a strictly relation between critical exponent, generalized critical exponent and repetitivity index of an infinite word w. In fact, we obtain

Proposition 3.1.2 *Let w be an infinite word. Then one has*

$$\mathrm{ce}(w) = 1 + \sup_{n \geq 0} \frac{n}{\mathrm{I}_w(n)}.$$

$$\mathrm{ce}(w, l) \geq 1 + \sup_{n \geq l} \frac{n}{\mathrm{I}_w(n)},$$

for any $l \geq 1$.

Proof. Let $w \in A^\omega$ and $v = ur = su \in \mathrm{Fact}(w)$, with $|u| = n$, $|r| = |s| = \mathrm{I}_w(n)$. Then $\mathrm{I}_w(n)$ is a period of v and, therefore

$$e(v) = \frac{|v|}{\pi_v} \geq 1 + \frac{n}{\mathrm{I}_w(n)}.$$

As $|v| > n$, we obtain that $\mathrm{ce}(w, n) \geq e(v)$. On the other hand

$$\mathrm{ce}(w) \geq \mathrm{ce}(w, n) \geq 1 + \frac{n}{\mathrm{I}_w(n)},$$

for all $n \geq 0$. Therefore

$$\mathrm{ce}(w) \geq 1 + \sup_{n \geq 0} \frac{n}{\mathrm{I}_w(n)}.$$

On the other hand, let v be a factor of w with exponent $t > \mathrm{ce}(w) - \varepsilon$. Then $v = ur = su$, with $|r| = |s| = \pi_v$ and, therefore, $|r| \geq \mathrm{I}_w(|u|)$. It follows

$$t = 1 + \frac{|u|}{|r|} \leq 1 + \frac{|u|}{\mathrm{I}_w(|u|)} \leq 1 + \sup_{n \geq 0} \frac{n}{\mathrm{I}_w(n)}$$

and by the arbitrariness of ε,

$$\mathrm{ce}(w) \leq 1 + \sup_{n \geq 0} \frac{n}{\mathrm{I}_w(n)};$$

then the first relation is proved. On the other hand, by relation

$$\mathrm{ce}(w, n) \geq 1 + \frac{n}{\mathrm{I}_w(n)},$$

one has that

$$\sup_{n \geq 0} \mathrm{ce}(w, n) \geq 1 + \sup_{n \geq 0} \frac{n}{\mathrm{I}_w(n)};$$

by definition of generalized critical exponent, one has

$$\mathrm{ce}(w, l) \geq \mathrm{ce}(w, n),$$

for all $n \geq l > 0$. Therefore the conclusion follows. □

Proposition 3.1.3 *Let w be an infinite word on a d-letter alphabet A. Then for all $n > 0$,*

$$\mathrm{I}_w(n) \leq d^n.$$

Proof. Since there are d^n distinct words of lenght n on the alphabet A, any factor of w of lenght $d^n + n$ necessarily contains two occurrences of a same factor of lenght n. Let v be a factor of w of minimal length with two occurrences of a same factor of length n. Then v has the form $v = xy = zx$ with $x, y, z \in A^*$, $|x| = n$, $|v| \leq d^n + n$. Thus, $|y| = |z| = |v| - |x| \leq d^n$. Since $\mathrm{I}_w(n) \leq |y|$, the conclusion follows. □

Proposition 3.1.4 *Let $w \in A^\omega$. Then for all $n \geq 0$,*

$$\mathrm{I}_w(n) \leq \rho_w(n) - n + 1\,. \tag{3.1}$$

Moreover, one has

$$\mathrm{I}_w(n) = \rho_w(n) - n + 1\,,$$

for some $n > 0$ if and only if w is ultimately periodic.

Proof. Equation (3.1) holds trivially if $\rho_w(n) = \infty$. Thus we assume $\rho_w(n) \in \mathbb{N}$. As any factor of w of length $\rho_w(n)$ contains all factors of w of length n, a factor of w of length $\rho_w(n)+1$ necessarily contains two occurrences of a same factor of length n. Thus, Equation (3.1) can be easily obtained proceeding as in the proof of the Proposition 3.1.3.

Now, suppose that $\mathrm{I}_w(n) = \rho_w(n) - n + 1$ for some $n > 0$. From the definition of repetitivity index, for any $i \geq 1$, the words

$$w[i, i+\mathrm{I}_w(n)-1],\ w[i+1, i+\mathrm{I}_w(n)],\ \ldots,\ w[i+\mathrm{I}_w(n)-1, i+\mathrm{I}_w(n)+n-2],$$

are pairwise distinct. Moreover, as $\rho_w(n) = \mathrm{I}_w(n) + n - 1$, these are all the factors of w of length n. For the same reason, also the words

$$w[i+1, i+\mathrm{I}_w(n)],\ w[i+2, i+\mathrm{I}_w(n)+1],\ \ldots,\ w[i+\mathrm{I}_w(n), i+\mathrm{I}_w(n)+n-1],$$

are all the factors of w of length n. Hence, necessarily,

$$w[i, i+\mathrm{I}_w(n)-1] = w[i+\mathrm{I}_w(n), i+\mathrm{I}_w(n)+n-1]$$

and, consequently, $w_i = w_{i+\mathrm{I}_w(n)}$. Thus, w has period $\mathrm{I}_w(n)$.

Conversely, suppose that w is periodic and let p be its minimal period. In such a case [27], for any $n \geq p$, one has $\rho_w(n) = p + n - 1$ and, moreover, w has p distinct factors of length n. This implies that for any $i \geq 1$, the words

$$w[i, i+n-1],\ w[i+1, i+n],\ \ldots,\ w[i+p-1, i+p+n-2],$$

are all the factors of w of length n and they are pairwise distinct. Consequently, $\mathrm{I}_w(n) \geq p$. Since, as we have seen, $p = \rho_w(n) - n + 1 \geq \mathrm{I}_w(n)$, one derives $\rho_w(n) - n + 1 = \mathrm{I}_w(n)$. □

Example 3.1.1 Let $A = \{0, 1\}$ and $\mu : A^* \to A^*$ the 2-uniform morphism such that $\mu(0) = 01$ and $\mu(1) = 10$. The Thue-Morse word is defined in this way

$$\mathbf{t} = \lim_n \mu^n(0) = 01101001\cdots,$$

where $\mu^n(0) = \mu(\mu^{n-1}(0))$, for all $n > 1$. In this infinite word any factor has exponent minus than or equal to 2; on the other hand, any word of lenght 4 on a binary alphabet has a factor of exponent 2; therefore,

$$\mathrm{ce}(\mathbf{t}) = 2.$$

By Proposition 3.1.2, we obtain that the repetitivity index of Thue-Morse word satisfies a relation of this type

$$\mathrm{I}_{\mathbf{t}}(n) \geq n.$$

□

Example 3.1.2 The *Fibonacci word* is the limit of the sequence of words

$$f_1 = b,\ f_2 = a,\ f_{n+1} = f_n f_{n-1},\ n \geq 2.$$

Notice that for all $n \geq 1$, the length of f_n is the n-th term of the Fibonacci numerical series F_n. The critical exponent of the Fibonacci word is $ce(f) = 2 + \phi$, where $\phi = (\sqrt{5} + 1)/2$ is the golden ratio [30]. By Proposition 3.1.2, for all $n \geq 0$, one has $1 + n/\,\mathrm{I}_f(n) \leq \mathrm{ce}(f)$, that is,

$$\mathrm{I}_f(n) \geq \frac{n}{\mathrm{ce}(f) - 1} = \frac{3 - \sqrt{5}}{2}\, n\,.$$

Let $n \geq 3$ and set $v = f_n f_{n+1}$. Then $v \in \mathrm{Fact}(f)$. Indeed, one has $f_{n+3} = f_{n+2}f_{n+1} = f_{n+1}f_n f_{n+1}$. Moreover, one has

$$v = f_n f_n f_{n-1} = f_n f_{n-1} f_{n-2} f_{n-1} = f_{n+1} f_{n-2} f_{n-1}\,,$$

so that $v = xy = zx$, with $x = f_{n+1}$, $y = f_{n-2}f_{n-1}$, $z = f_n$. One derives that $\mathrm{I}_f(F_{n+1}) \leq F_n$.
Now let $k > 2 = F_3$. Then we can find $n \geq 3$ such that $F_n < k \leq F_{n+1}$. Since I_f is non decreasing, one obtains

$$\mathrm{I}_f(k) \leq F_n < k\,.$$

The first few values of $\mathrm{I}_f(k)$ are

$$\mathrm{I}_f(0) = \mathrm{I}_f(1) = 1\,,\ \mathrm{I}_f(2) = \mathrm{I}_f(3) = 2\,,\ \mathrm{I}_f(4) = \mathrm{I}_f(5) = \mathrm{I}_f(6) = 3\,,$$

$$\mathrm{I}_f(7) = \mathrm{I}_f(8) = \mathrm{I}_f(9) = \mathrm{I}_f(10) = \mathrm{I}_f(11) = 5\,.$$

Example 3.1.3 Let $A = \{a, b, c\}$ a ternary alphabet. Dejean word [16], is a ternary infinite word w generated by the 19-uniform morphism h

$$\begin{aligned} a &\to abcacbcabcbacbcacba \\ b &\to bcabacabcacbacabacb \\ c &\to cabcbabcabacbabcbac \end{aligned}$$

Any factor of this word has exponent minus than or equal to 7/4. Moreover, this word has infinitely many factors of exponent 7/4. In fact, for example, the factor $acbcacb$ of the word $h(a) = abcacbcabcb\overbrace{acbcacb}a$ has exponent 7/4 as the factor $h^n(acbcacb)$, for all $n \geq 1$. Therefore,

$$ce(w) = 7/4.$$

By Proposition 3.1.2, the repetitivity index of Dejean word is:

$$\mathrm{I}_w(n) \geq \frac{4}{3}n. \tag{3.2}$$

As we have seen, the equality is verified by infinitely many values of the integer n. □

Remark 3.1.1 For any $n \geq 2$, the minimal critical exponent of infinite words on n letters is called the *repetition threshold on n letters* and it is denoted by $\mathrm{RT}(n)$. As we have seen, Dejean [16] proved that repetition threshold on 3 letters is $7/4$. Dejean has also showed that for $n \geq 5$, the repetition threshold on n letters is not smaller than $n/(n-1)$ while if $n = 4$, then it is not smaller than $7/5$. She conjectured that these are the actual values of the repetition threshold. This conjecture has been proved to be true for $n = 4$ by Pansiot [34] and, with extensive use of a computer, for $5 \leq n \leq 11$ by Moulin-Ollagnier [29] and, more recently, for $12 \leq n \leq 14$, by Mohammad-Noori and Currie [28] and for $n \geq 33$, by Carpi, [8].

Remark 3.1.2 Let $\alpha > 1$ be a rational number, and let $l \geq 1$ be an integer. A word w is a *repetition of order α and lenght l* if we can write it as $w = (xy)^n x$, with $|xy| = l$ and $|w| = \alpha l$. For any integers $k \geq 2$ and $l \geq 1$, the *generalized repetititon threshold* $\mathrm{RT}(k, l)$ was defined in [24] as the real number α such that there exists an infinite word on k letters which avoid repetition of order α', and lenght l', for all $\alpha' > \alpha$ and $l' \geq l$. Notice that $\mathrm{RT}(k, 1) = \mathrm{RT}(k)$, for all $k > 1$. The values of $\mathrm{RT}(k, l)$ for some small values of k and l was computed in [24]. For instance, $\mathrm{RT}(3, 2) = 3/2$, $\mathrm{RT}(2, 5) = 7/5$.

Chapter 4

Beck's Theorem

A well-known theorem in combinatorics states that given an arbitrary natural number n, there exists an infinite binary word w, the de Bruijn cicle, such that $\mathrm{I}_w(n) = 2^n$, (see, e.g., L. Lovász problem book [22], Problem 8); for example, let $A = \{a, b\}$ and we consider the word $w = (bbbabaaa)^\omega$. This word satisfies the equation $\mathrm{I}_w(3) = 8$. In [3], J. Beck proved that given an arbitrarily small $\epsilon > 0$, there is an infinite binary word w such that $\mathrm{I}_w(n) \geq (2 - \epsilon)^n$, for all $n > n(\epsilon)$. This proof is not costructive: it is based on a probabilistic lemma due to L. Lovász [20]. On the other hand, in this chapter, we will construct effectively an infinite binary word w having a similar (though weaker) property. For any $\nu > 0$, there exists an integer $n(\nu)$ such that for all $n > n(\nu)$, $\mathrm{I}_w(n) \geq n^\nu$; the integer $n(\nu)$ can be effectively computed. This result has been obtained with A. Carpi, [10]. An interesting consequence concerns the exponents of the factors of w. Indeed, from the preceding result we derive that for any $\varepsilon > 0$ the number of distinct factors of w with exponent larger than $1 + \varepsilon$ is finite. The chapter is organized as follows: in Section 4.1, we give some preliminary definitions useful in the sequel. In Section 4.2, we recall the proof of Beck; in Section 4.3, we prove

the main result of this chapter.

4.1 Preliminaries

A set of words is *prefix-closed* if it contains the prefixes of all its elements. A set of words is *factor-closed* or *factorial* if it contains the factors of all its elements. The following result is known as König's Lemma, [25].

Proposition 4.1.1 *If X is an infinite prefix-closed set of words over a finite alphabet A, there is an infinite word w having all its prefixes in X.*

In other terms, if X is a set on an alphabet A then this two properties are equivalent

1. There are infinitely many words on A^* that have no factors in X.
2. There is an infinite word on A that has no factors in X.

We can observe that, in general, there is not an effective method to construct an infinite word by a set which verifies 1. In mathematics, a *σ-algebra* over a set X is a non empty collection Σ of subset of X that is closed under complementation and countable uniones of its members. In mathematics, a probability of an event A is represented by a real number in the range 0 to 1 and written as $\Pr(A)$. The complement of the event A, is the event $\overline{A}$ such that $\Pr(\overline{A}) = 1 - \Pr(A)$. Let p, q be integers. We shall write $p \mid q$ (resp., $p \nmid q$) to denote that p divides q (resp., p does not divide q).

4.2 Beck's Theorem: Lovász Local Lemma

In this section, we recall the proof of Beck's Theorem, [3]. We first need the following purely probabilistic theorem:

Theorem 4.2.1 *(Lovász local lemma) [20]*
Let G be a simple graph on the vertex set $V(G) = \{1, 2, \ldots, m\}$ and let an event A_i be associated with each vertex i. Suppose that there are real numbers $x_1, \ldots, x_m$, $(0 \leq x_i \leq 1)$ such that
(a) every A_i is independent of the σ-algebra generated by the set of all A_j's for which j is not adjacent to i;
(b)

$$\Pr(A_i) \leq (1 - x_i) \prod_{\{i,j\} \in G} x_j, \quad (i = 1, \ldots, m).$$

Then

$$\Pr(\overline{A_1} \cap \overline{A_2} \ldots \cap \overline{A_m}) > 0.$$

□

Concerning further applications of Lovász local lemma, see Spencer [36] and Graham - Rotschild - Spencer [35]. Now we can prove the Theorem of Beck.

Theorem 4.2.2 *Given an arbitrary small $\epsilon > 0$, there is an integer $n(\epsilon)$ and an infinite binary word w such that $\mathrm{I}_w(n) \geq (2 - \epsilon)^n$, for all $n > n(\epsilon)$.*

Proof. Let $f(n) = (2 - \epsilon)^n$, for $n > n(\epsilon)$ and 0 for $1 \leq n \leq n(\epsilon)$, where the threshold $n(\epsilon)$ will be specified later. By König lemma, it sufficies to prove the following finite version of the theorem: given an arbitrary natural number N, there exists a binary word w, with $|w| = N$ having the property that $\mathrm{I}_w(n) > f(n)$, for each $1 \leq n \leq N$. Let $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ be independent random variables such that $\Pr(\epsilon_i = 0) = \Pr(\epsilon_i = 1) = \frac{1}{2}$. Let $A(k, l, n)$ denote the event that the intervals $(\epsilon_{k+1}, \ldots, \epsilon_{k+n})$ and $(\epsilon_{l+1}, \epsilon_{l+2}, \ldots, \epsilon_{l+n})$ are identical, i.e. $\epsilon_{k+i} = \epsilon_{l+i}$, for each $1 \leq i \leq n$. By properties of the Pr function, one has

$$\Pr(A(k, l, n)) = 2^{-n},$$

for $k \neq l$. Now we define a graph G. Let the vertex set $V(G)$ be the set of all triplets (k, l, n) for which $0 \leq k < l,\ l + n \leq N,\ l - k \leq f(n)$ and $n > n(\epsilon)$. The vertices (k_1, l_1, n_1) and (k_2, l_2, n_2) are adjacentes in G if and only if the unions of intervals

$$\mathbb{J}(k_i, l_i, n_i) = [k_i + 1, k_i + n_i] \cup [l_i + 1, l_i + n_i],$$

for $i = 1, 2$ have at least one common element. Observe that if $\mathbb{J}(k_i, l_i, n_i),\ (i = 1, 2, \ldots r)$ are disjoint from $\mathbb{J}(k, l, n)$, then the event $A(k, l, n)$ is indipendent of the σ-algebra generated by the set of events $A(k_i, l_i, n_i),\ (i = 1, 2, \ldots, r)$. Thus, in our case, condition (a) of Lovász Local Lemma is satisfied. Let $x_{k,l,n} = 1 - \frac{1}{f(n)n^3}$. In the next step we will verify condition (b) of Lovász Local Lemma, that is the inequality below

$$2^{-n_0} = \Pr(A(k_0, l_0, n_0)) \leq (1 - x_{k_0,l_0,n_0}) \prod_{(k,l,n),(k_0,l_0,n_0) \in G} x_{k,l,n}. \tag{4.1}$$

For notational convenience let

$$P^{(i)} = \prod_{(k,l,n)}{}^{(i)}\ x_{k,l,n}\ \ (i = 1, 2, 3, 4)$$

where the products $\prod^i\ (i = 1, 2, 3, 4)$ are extended over all triplets (k, l, n) for which $0 \leq k < l,\ l + n \leq N,\ l - k \leq f(n),\quad n > n(\epsilon)$ and property $(\pi_i)\ \ (i = 1, 2, 3, 4)$ holds, respectively

$$(\pi_1)\ [k + 1, k + n] \cap [k_0 + 1, k_0 + n_0] \neq \emptyset;$$

$$(\pi_2)\ [k + 1, k + n] \cap [l_0 + 1, l_0 + n_0] \neq \emptyset;$$

$$(\pi_3)\ [l + 1, l + n] \cap [k_0 + 1, k_0 + n_0] \neq \emptyset;$$

$$(\pi_4)\ [l + 1, l + n] \cap [l_0 + 1, l_0 + n_0] \neq \emptyset.$$

By definition,

$$\prod_{\{(k,l,n),(k_0,l_0,n_0)\}\in G} x_{k,l,n} \geq \prod_{i=1}^{4} P^{(i)} \tag{4.2}$$

Now we estimate the factors $P^{(i)}$ $(1 \leq i \leq 4)$. We have $P^{(1)} = P^{(11)}P^{(12)}$, where

$$P^{(11)} = \prod_{k=k_0}^{k_0+n_0-1} \prod_{n=n(\epsilon)+1}^{N-k} \prod_{l=k+1}^{\min\{k+f(n),N-n} x_{k,l,n}$$

and

$$P^{(12)} = \prod_{k=0}^{k_0-1} \prod_{n=\max\{k_0+1-k,n(\epsilon)+1\}}^{N-k} \prod_{l=k+1}^{\min\{k+f(n),N-n\}} x_{k,l,n}.$$

Clearly

$$P^{(11)} \geq \prod_{n=n(\epsilon)+1}^{\infty} \Big(1 - \frac{1}{f(n)n^3}\Big)^{n_0 f(n)} > \Big(1 - \sum_{n=n(\epsilon)+1}^{\infty} \frac{1}{n^3}\Big)^{n_0} > \Big(1 - \frac{1}{n(\epsilon)}\Big)^{n_0} \tag{4.3}$$

On the other hand,

$$P^{(12)} \geq \prod_{n=n(\epsilon)+1}^{\infty} \Big(1 - \frac{1}{f(n)n^3}\Big)^{nf(n)} > 1 - \sum_{n=n(\epsilon)+1}^{\infty} \frac{1}{n^2} > 1 - \frac{1}{n(\epsilon)}. \tag{4.4}$$

Thus, by (4.2) and (4.3), one has

$$P^{(1)} \geq \Big(1 - \frac{1}{n(\epsilon)}\Big)^{n_0+1}. \tag{4.5}$$

The same computation gives

$$P^{(i)} \geq \Big(1 - \frac{1}{n(\epsilon)}\Big)^{n_0+1}, \text{ for } 2 \leq i \leq 4. \tag{4.6}$$

Summarizing, by equations (4.2), (4.3), (4.4) and by definition of $x_{k,l,n}$, one has

$$\begin{aligned}(1 - x_{k_0,l_0,n_0}) \prod_{\{(k,l,n),(k_0,l_0,n_0)\}\in G} x_{k,l,n} &\geq \frac{1}{f(n_0)(n_0)^3} \prod_{i=1}^{4} P^{(i)} \\ &\geq \frac{\Big(1 - \frac{1}{n(\epsilon)}\Big)^{4n_0+4}}{f(n_0)(n_0)^3} = \Big(1 - \frac{1}{n(\epsilon)}\Big)^{4n_0+4} (2-\epsilon)^{-n_0} (n_0)^{-3}.\end{aligned}$$

Simple computation shows that

$$\left(1-\frac{1}{n(\epsilon)}\right)^{4n_0+4}(2-\epsilon)^{-n_0}(n_0)^{-3} \geq 2^{-n_0} \tag{4.7}$$

for all $n_0 > n(\epsilon)$, if $n(\epsilon)$ is sufficiently large depending only on ϵ. Thus, (4.6) and (4.7) complete the proof of the relation (4.1). By the application of Lovász Local Lemma we obtain the existence of a binary word w of lenght N such that $\mathrm{I}_w(n) \geq (2-\epsilon)^n$, for each $n(\epsilon) < n \leq N$. □

4.3 Beck's Theorem: an effective construction

In this section, we construct the infinite binary word w with the above mentioned properties. The following lemma is proved in [8], in the particular case $p = 3$.

Lemma 4.3.1 *For any $p \geq 3$, let $w_p = (b_{p,i})_{i\geq 1}$ be the infinite word on the alphabet $\{0,1\}$ defined by*

$$b_{p,i} = \begin{cases} 0 & \text{if } i \equiv 1 \pmod p \\ 1 & \text{if } i \not\equiv 0,1 \pmod p \\ b_{p,i/p} & \text{if } i \equiv 0 \pmod p \end{cases}$$

If a factor x of w_p has a period r and lenght $|x| \geq r + p^k$ with $k \geq 0$, then p^k divides r.

Proof. Let $x = b_{p,i}b_{p,i+1}\cdots b_{p,i+|x|-1}$ with $i \geq 1$. By the periodicity of x one has

$$b_{p,i}b_{p,i+1}\cdots b_{p,i+p^k-1} = b_{p,i+r}b_{p,i+r+1}\cdots b_{p,i+r+p^k-1}\,. \tag{4.8}$$

Let p^q be the maximal power of p dividing r and assume by contradiction $q < k$. We can write $r = p^q t$ with $p \nmid t$. First we consider the case that

$p \nmid t+1$. In this case, $1+t \not\equiv 0,1 \pmod{p}$. Let j be the integer such that $i \le j < i+p^k$ and $j \equiv p^q \pmod{p^k}$. Then one has $j = p^q(1+t'p^{k-q})$ for a suitable integer $t' \ge 0$ and $j+r = p^q(1+t+t'p^{k-q})$. From the definition of w_p, one derives

$$b_{p,j} = b_{p,1+t'p^{k-q}} = 0 \quad \text{and} \quad b_{p,j+r} = b_{p,1+t+t'p^{k-q}} = 1\,.$$

This yields a contradiction because by (4.8), $b_{p,j} = b_{p,j+r}$. Now we consider the case that $p \mid t+1$. In this case, $2+t \equiv 1 \pmod{p}$. We can take j such that $i \le j < i+p^k$ and $j \equiv 2p^q \pmod{p^k}$. Then one has $j = p^q(2+t'p^{k-q})$ and $j+r = p^q(2+t+t'p^{k-q})$ for a suitable integer $t' \ge 0$. From the definition of w_p, one derives

$$b_{p,j} = b_{p,2+t'p^{k-q}} = 1 \quad \text{and} \quad b_{p,j+r} = b_{p,2+t+t'p^{k-q}} = 0\,,$$

so that one has again a contradiction. □

We obtain a similar result also when $p = 2$. In fact, one has:

Lemma 4.3.2 *Let $w_2 = (b_{2,i})_{i\ge 1}$ be the infinite word on alphabet $\{0,1\}$ defined by*

$$b_{2,i} = \begin{cases} 0 & \text{if } i \equiv 1 \pmod 4 \\ 1 & \text{if } i \equiv 3 \pmod 4 \\ b_{2,i/2} & \text{if } i \equiv 0,2 \pmod 4 \end{cases}$$

If a factor x of w_2 has a period r and lenght $|x| \ge r + 2^k$ with $k \ge 3$, then 2^k divides r.

Proof. Let $x = b_{2,i}b_{2,i+1}\cdots b_{2,i+|x|-1}$ with $i \ge 1$. By the periodicity of x one has

$$b_{2,i}b_{2,i+1}\cdots b_{2,i+2^k-1} = b_{2,i+r}b_{2,i+r+1}\cdots b_{2,i+r+2^k-1}\,. \tag{4.9}$$

First we prove that r is even. Indeed assume r odd. Let j be the integer such that $i \leq j \leq i+3$ and $j \equiv 2 \pmod 4$. Then one has $j = 2j'$ for some odd integer $j' \geq 1$. Since j is even, r is odd and $j + r \equiv j + r + 4 \pmod 4$, from the definition of w_2 one obtains

$$b_{2,j+r} = b_{2,j+r+4}\,.$$

Moreover, from (4.9) one has $b_{2,j+r} = b_{2,j}$ and $b_{2,j+r+4} = b_{2,j+4}$ and from the definition of w_2, $b_{2,j} = b_{2,j'}$ and $b_{2,j+4} = b_{2,j'+2}$. Hence,

$$b_{2,j'} = b_{2,j'+2}\,.$$

As j' is odd, from the definition of w_2 one derives $j' \equiv j'+2 \pmod 4$, which is a contradiction. Thus we can write $r = 2^q t$ with $q \geq 1$ and t odd. Let us assume, by contradiction, $q < k$. Let j be the integer such that $i \leq j < i+2^k$ and $j \equiv 2^{q-1} \pmod{2^k}$. Then one has $j = 2^{q-1}(1 + t'2^{k-q+1})$ for a suitable integer $t' \geq 0$ and $j + r = 2^{q-1}(1 + 2t + t'2^{k-q+1})$. Since t is odd and $k > q$ one has $1 + t'2^{k-q+1} \equiv 1 \pmod 4$ and $1 + 2t + t'2^{k-q+1} \equiv 3 \pmod 4$. One easily derives

$$b_{2,j} = 0 \quad \text{and} \quad b_{2,j+r} = 1$$

which contradicts (4.9). We conclude that $q \geq k$. □

We notice that the word w_2 defined above is the so-called *regular paperfolding sequence* (see Preliminaries, [1] and [26]). Now we construct the infinite binary word

$$w = a_1 a_2 \cdots a_n \cdots$$

with the announced properties. In the sequel, we shall denote by p_r the r-th prime number. Let $i > 0$ and r be the least positive integer such that $2^r \nmid i$. We set

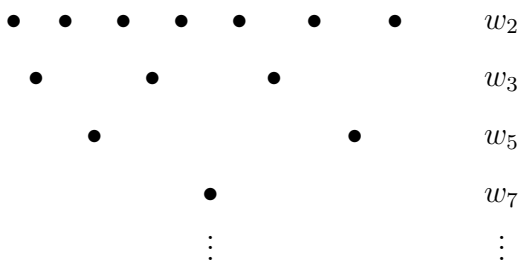
$$a_i = b_{p_r,(i+2^{r-1})/2^r}\,.$$

In other words, for all $k \geq 0$, the word

$$a_{2^k} a_{3 \cdot 2^k} a_{5 \cdot 2^k} \cdots a_{(2n+1) \cdot 2^k} \cdots$$

is equal to $w_{p_{k+1}}$.
Grafically,

$$a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} a_{12} a_{13} \cdots$$



$$w = 0\,0\;\;0\;\;0\;\;1\;\;1\;\;0\;\;0\;\;0\;\;\;0\;\;\;1\;\;\;1\;\;\;1\cdots$$

The word w verifies the following property.

Lemma 4.3.3 *Let $x \in \mathrm{Fact}(w)$ and p be a period of x. If $|x| \geq p + 40$, then p is even.*

Proof. Let $x = a_i a_{i+1} \cdots a_{i+l}$. By the periodicity of x,

$$a_i a_{i+1} \cdots a_{i+39} = a_{i+p} a_{i+p+1} \cdots a_{i+p+39} \tag{4.10}$$

Let r be the least odd integer such that $r \geq i$. By definition of w, we obtain

$$a_r a_{r+2} \cdots a_{r+38} = b_{2,r'} b_{2,r'+1} \cdots b_{2,r'+19}, \tag{4.11}$$

for a suitable $r' \geq 1$. Now, let r'' be the least odd integer such that $r'' \geq r'$. From the definition of w_2 one derives that the word

$$v = b_{2,r''} b_{2,r''+2} \cdots b_{2,r''+18}$$

is a factor of $(01)^n$ for some $n > 0$ and therefore v has period 2. By equations (4.10) and (4.11) one derives that

$$v = a_s a_{s+4} \cdots a_{s+36}, \tag{4.12}$$

with $s \in \{r + p, r + p + 2\}$. Suppose by contradiction that p is odd. In this case, s is even, as r is odd. First, we consider the case that $4 \nmid s$. By (4.12) and the definition of w one has that v is a factor of w_3. As v has period 2 and lenght $|v| = 10$, this yields a contradiction by Lemma 4.3.2. Now, we consider the case that $4 \mid s$. Select $t \in \{s, s + 4\}$ such that $8 \nmid t$. From the definition of w, the word

$$v' = a_t a_{t+8} a_{t+16} a_{t+24} a_{t+32}$$

is a factor of w_5. On the other hand, v' is obtained from v by taking only the letters of odd places or those of even places. Hence, as v has period 2, the word v' has period 1, i.e. $v' \in \{00000, 11111\}$. This is a contradiction since, as one easily verifies, neither 0^5 nor 1^5 are factors of w_5. We conclude that p has to be even. □

Lemma 4.3.4 *Let ν be a positive integer and set*

$$n(\nu) = \max\{2^\nu, 40\}\,, \quad c(\nu) = 2^{\nu(\nu+1)/2} \prod_{i=2}^{\nu} p_i\,.$$

If $x \in \mathrm{Fact}(w)$ has period p and length $|x| \geq n(\nu) + p$, then

$$p > \frac{(|x| - p)^\nu}{c(\nu)}\,. \tag{4.13}$$

Proof. Let $x \in \mathrm{Fact}(w)$ be a word of period p and length $|x| \geq n(\nu) + p$. For any $r \leq \nu$ one has

$$p + 2^r \cdot p_r^{k_r} \leq |x| < p + 2^r \cdot p_r^{k_r+1}, \tag{4.14}$$

for a suitable integer $k_r \geq 0$. In particular, for $r = 1$ one has $p + 40 \leq |x| < p + 2^{k_1+2}$ which implies $k_1 \geq 3$ so that by Lemma 4.3.3 the period p is even. Hence, the word

$$w_2 = a_1 a_3 \cdots a_{2n+1} \cdots$$

will contain a factor $\widetilde{x}$ with period $q = p/2$ and lenght $|\widetilde{x}| \geq \lfloor |x|/2 \rfloor$. As $\lfloor |x|/2 \rfloor \geq \lfloor (p + 2^{1+k_1})/2 \rfloor = q + 2^{k_1}$, by Lemma 4.3.2 one has that 2^{k_1} divides q, i.e., 2^{k_1+1} divides p. We notice that $p + 2^\nu \leq |x| < p + 2^{k_1+2}$ and therefore $\nu \leq k_1 + 1$. Thus, $2^\nu \mid p$. Now, let $3 \leq r \leq \nu$. Since $2^r \mid p$ there is a factor $\widehat{x}$ of the word

$$w_{p_r} = a_{2^r} a_{3 \cdot 2^r} \cdots a_{(2n+1)2^r} \cdots$$

with period $q' = p/2^r$ and lenght $|\widehat{x}| \geq \lfloor |x|/2^r \rfloor$. As $|x| \geq p + 2^r p_r^{k_r}$, we obtain $|\widehat{x}| \geq q' + p_r^{k_r}$ and therefore, by Lemma 4.3.1, $p_r^{k_r} \mid q'$. This proves that p is divided by $p_1^{k_1},\ p_2^{k_2},\ \ldots p_\nu^{k_\nu}$s and therefore

$$p \geq p_1^{k_1} p_2^{k_2} \cdots p_\nu^{k_\nu} .$$

In view of (4.14) one has $p_r^{k_r} > (|x| - p)/(2^r p_r)$, $1 \leq r \leq \nu$. Thus, from the previous equation,

$$p > \frac{(|x| - p)^\nu}{2^{\nu(\nu+1)/2} p_2 \cdots p_\nu}. \tag{4.15}$$

□

Theorem 4.3.1 *For all $\nu > 0$ the repetitivity index I_w of the word w verifies eventually the following inequality*

$$\mathrm{I}_w(n) \geq n^\nu .$$

Proof. From the definition of repetivity index there exist words x, y, z such that $xy = zx$ is a factor of w, $|x| = n$ and $|y| = \mathrm{I}_w(n)$. Thus the word

$v = xy = zx$ has period $\mathrm{I}_w(n)$ and lenght $n + \mathrm{I}_w(n)$. With the notations of Lemma 4.3.4 for all $n \geq n(\nu)$ one has

$$\mathrm{I}_w(n) > \frac{n^\nu}{c(\nu)}\,.$$

By the arbitrariness of ν, the conclusion follows easily. □

From the previous theorem, one derives the following

Corollary 4.3.1 *For any $\varepsilon > 0$, the number of distinct factors of w with exponent larger than $1 + \varepsilon$ is finite.*

Proof. Let v be a factor of w. One can write

$$v = xy = zx \tag{4.16}$$

with $|y| = |z| = \pi_v$. Set $r = |x|$. Then one has necessarily

$$e(v) = \frac{|x|}{\pi_x} = 1 + \frac{r}{\pi_v}\,.$$

If $r < \varepsilon\pi_v$, one derives $e(v) < 1+\varepsilon$. From (4.16) one easily derives $|z| \geq \mathrm{I}_w(r)$. By Theorem 4.3.1, there exists an integer k such that $\mathrm{I}_w(n) \geq n^2$ for all $n > k$. Thus, if $r > k_0 = \max\{k, 1/\varepsilon\}$, then $\pi_v = |z| \geq I_w(r) > r^2 > r/\varepsilon$ and therefore

$$e(v) = 1 + \frac{r}{\pi_v} < 1 + \varepsilon\,.$$

Thus, if $e(x) \geq 1 + \varepsilon$ one has

$$\varepsilon\pi_x \leq r \leq k_0\,.$$

This implies that

$$|v| = \pi_v + r \leq \frac{r}{\varepsilon} + r \leq k_0\left(1 + \frac{1}{\varepsilon}\right).$$

This proves that the factors of w with exponent larger than $1+\varepsilon$ have bounded length. □

Chapter 5

Thue-Morse word

As we have seen in the introduction, in [40] and [41] the norvegian mathematician Axel Thue noted that any binary word of lenght ≥ 4 must contain a square, i.e. two consecutive identical factors. He then asked whether it was possible to find an infinite binary word that neither should contain any cube i.e. three consecutive identical factors nor overlaps i.e. factors of the form $awawa$, where $a \in \{0, 1\}$ and $w \in A^*$. The answer to all two questions was positive. This work of Thue was the starting point of an important branch of combinatorics, now called *combinatorics on words*. It is worth noting that Thue explained he had no particular application in mind, but he thought the problem was interesting enough in itself to deserve attention. Thue's papers were rediscovered by several different authors, including Marton Morse. Although there are uncontably many overlap-free sequences on two symbols [6], the Thue-Morse sequence is, roughly speaking, the 'canonical example'. In this chapter, we study the repetitivity index of this word. We prove that the function $\mathrm{I}_{\mathbf{t}}(n)$ satisfies a relation of this type

$$\mathrm{I}_{\mathbf{t}}(n) = 2\,\mathrm{I}_{\mathbf{t}}(\lceil n/2 \rceil),$$

for any $n > 4$ and

$$\mathrm{I}_{\mathbf{t}}(n) = n,$$

for any $1 \leq n \leq 4$. The chapter is organized as follows: in Section 5.1, we give a formal definition of the Thue-Morse word and we recall some useful properties. In Section 5.2, we prove the main result of this chapter.

5.1 Definitions and main properties

We first give a formal definition of the Thue-Morse word. We denote by $\mathbf{t} = (t_n)_{n\geq 0}$ the Thue-Morse word over $\{0, 1\}$, defined recursively by $t_0 = 0$ and $t_{2n} = t_n$, $t_{2n+1} = \overline{t_n}$ for all $n \geq 0$, where, for $x \in \{0, 1\}$, we define $\overline{x} = 1-x$. Denote by $s_k(n)$ the sum of the digits in the base-k representation of the integer n. Since we clearly have $s_2(2n) = s_2(n)$ and $s_2(2n+1) = s_2(n)+1$ for every integer $n \geq 0$, we easily obtain the following equivalent definition

Proposition 5.1.1 *The Thue-Morse word* $\mathbf{t}$ *is equal to the word* $(s_2(n) \bmod 2)_{n\geq 0}$.

Yet another definition, which is easily seen to be equivalent to the previous two, is the following

Proposition 5.1.2 *[1]*
Let X *be an indeterminate. Then we have*

$$\prod_{i\geq 0}(1 - X^{2^i}) = (1 - X)(1 - X^2)(1 - X^4)\ldots = \sum_{j\geq 0}(-1)^{t_j}X^j.$$

Proposition 5.1.3 *Define the morphism* μ *on the alphabet* $\{0, 1\}$ *by* $\mu(0) = 01$, $\mu(1) = 10$. *Then the Thue-Morse word* $\mathbf{t}$ *is the unique fixed point of* μ *that begins with* 0.

Our first theorem is the one we mentioned in the introduction. It is due to Thue [40].

Theorem 5.1.1 *The Thue-Morse word* $\mathbf{t}$ *is overlap-free and does not contain a cube.*

A natural arising question is whether possible to build another binary word that is both overlap-free and generated by a morphism. The next theorem due to Berstel and Séébold [5] answers this question negatively.

Theorem 5.1.2 *If an overlap-free binary word is a fixed point of a non trivial morphism, then it is equal either to the Thue-Morse word* $\mathbf{t}$ *or to its complement* $\overline{\mathbf{t}} = (t_n)_{n\geq 0} = 1001011001101001\cdots$.

The Thue-Morse sequence has the nice property that it exhibits regularity without being ultimately periodic.

Proposition 5.1.4 *(Morse) There exists an infinite sequence over* $\{0,1\}$ *which is uniformly recurrent but not ultimately periodic.*

The sequence that Morse gives is exactly $\mathbf{t}$. The Thue-Morse word is the prototype of a class of sequences called 2-*automatic sequences*. Roughly speaking, a sequence is k-automatic if its n-th term is generated by a finite-state machine which takes as input the base-k expansion of n, cfr. Section 2.2. For more about this class of sequences, see for example [14]. Now, we define the notion of *fractional power*. We say that a (finite or infinite) word w contains an α-power (real $\alpha > 1$) if w has a factor of the form $x^{\lfloor\alpha\rfloor}x'$, where x' is a prefix of x and $|x^{\lfloor\alpha\rfloor}x'| \geq \alpha|x|$. For example, the word:

$$2301\overbrace{01234567}\overbrace{01234567}\overbrace{0123}310,$$

has a $\frac{5}{2}$-power. A word is α-power-free if it contains no α-power. Given an infinite word w it is an interesting and challenging task to determine its

critical exponent $\mathrm{ce}(w)$, such that w contains α-powers for all $\alpha < \mathrm{ce}(w)$ but has no α-powers, for $\alpha > \mathrm{ce}(w)$.

Theorem 5.1.3 *The critical exponent of the Thue-Morse word* $\mathbf{t}$ *is 2.*

Proof. The word $\mathbf{t}$ begins $011\cdots$ and hence contains a square. If $\mathbf{t}$ would contain a $(2+\epsilon)$-powers for some $\epsilon > 0$, then it would contains an overlap. But $\mathbf{t}$ is overlap-free by Theorem 5.1.1.

5.2 Computing the repetitivity index

Let w an infinite word on finite alphabet A. We recall that the repetitivity index of w is the function $\mathrm{I}_w(n)$ defined as follows

$$\mathrm{I}_w(n) := \min\{k > 0 \,|\, \exists\, x, y, z \in A^*,\, |x| = n, |y| = |z| = k,\, xy = zx \in \mathrm{Fact}(w)\}.$$

In this section, we prove that this function satisfies the following relations: for $k > 0,\ 2^k < n \leq 2^k + 2^{k-1}$, one has:

$$\mathrm{I}_{\mathbf{t}}(n) = 2^k + 2^{k-1};$$

for $k > 0,\ 2^k + 2^{k-1} < n \leq 2^{k+1}$, one has:

$$\mathrm{I}_{\mathbf{t}}(n) = 2^{k+1}.$$

Lemma 5.2.1 *Let* $\mathbf{t}$ *be the Thue-Morse word. Then*

$$\mathrm{I}_{\mathbf{t}}(n) \geq n,$$

for all $n > 0$.

Proof. We recall by Preliminaries that

$$\mathrm{ce}(\mathbf{t}) = 1 + \sup_{n>0} \frac{n}{\mathrm{I}_{\mathbf{t}}(n)};$$

therefore, as $\mathrm{ce}(\mathbf{t}) = 2$, we obtain that

$$\mathrm{I}_{\mathbf{t}}(n) \geq n.$$

□

Lemma 5.2.2 *Let $w \in \{0110, 1001\}^\omega$. For all $n \geq 4$, $\mathrm{I}_w(n)$ is even.*

Proof. Write $w = w_0 w_1 \cdots w_i \cdots, w_i \in A,\ i \geq 0$. As $w \in \{0110, 1001\}^\omega$, for any odd i one has

$$w_i w_{i+1} w_{i+2} w_{i+3} \in \{110, 001\}A \cup A\{011, 100\}.$$

In other therms, for any odd i, the word $w_i w_{i+1} w_{i+2} w_{i+3}$ either starts or ends with 00 or 11. On the contrary, for any even i,

$$w_i w_{i+1} w_{i+2} w_{i+3} \in \{01, 10\}^2,$$

i.e. $w_i w_{i+1} w_{i+2} w_{i+3}$ neither starts nor ends with 00 or 11. We conclude that the parity of i is uniquely determined by the word $w_i w_{i+1} w_{i+2} w_{i+3}$. Now, for any $n \geq 4$ there exists i s.t.

$$w_i w_{i+1} \cdots w_{i+n-1} = w_{i+\mathrm{I}_w(n)} \cdots w_{i+\mathrm{I}_w(n)+n-1}.$$

Thus, $w_i w_{i+1} w_{i+2} w_{i+3} = w_{i+\mathrm{I}_w(n)} \cdots w_{i+\mathrm{I}_w(n)+3}$. Hence i and $i + \mathrm{I}_w(n)$ have the same parity, so that $\mathrm{I}_w(n)$ is even.

□

Lemma 5.2.3 *Let w be a fixpoint of a uniform morphism of lenght k. Then for all $n > 0$,*

$$\mathrm{I}_w(nk) \leq k\, \mathrm{I}_w(n).$$

Proof. We can find words r, s, t, such that

$$rs = st \in \mathrm{Fact}(w),\ |s| = n,\ |r| = |t| = \mathrm{I}_w(n).$$

Since w is a fixedpoint of a uniform morphism f of lenght k, setting $r' = f(r),\ s' = f(s),\ t' = f(t),$ one has that:

$$r's' = s't' \in \mathrm{Fact}(w),\ |s'| = nk,\ |r'| = |t'| = k\,\mathrm{I}_w(n),$$

and the conclusion follows.

□

Proposition 5.2.1 *Let* $\mathbf{t}$ *the Thue-Morse word. For any* $n \geq 4$,

$$\mathrm{I}_{\mathbf{t}}(n) = 2\,\mathrm{I}_{\mathbf{t}}(\lceil n/2 \rceil).$$

Proof. From Lemma 5.2.3, as $\mathrm{I}_{\mathbf{t}}(n)$ is non decreasing, one has

$$\mathrm{I}_{\mathbf{t}}(n) \leq \mathrm{I}_{\mathbf{t}}(2\lceil n/2 \rceil) \leq 2\,\mathrm{I}_{\mathbf{t}}(\lceil n/2 \rceil). \tag{5.1}$$

Let $n \geq 4$. There exists $i \geq 0$ s.t.

$$t_i t_{i+1} \cdots t_{i+n-1} = t_{i+\mathrm{I}_t(n)} t_{i+\mathrm{I}_t(n)+1} \cdots t_{i+\mathrm{I}_t(n)+n-1}. \tag{5.2}$$

By Lemma 5.2.2, $\mathrm{I}_{\mathbf{t}}(n) = 2k$, for some $k > 0$. We set $m = \lceil n/2 \rceil$. First we consider the case that i is even, say $i = 2j$. By definition of Thue-Morse word, one has:

$$t_i t_{i+2} \cdots t_{i+2m-2} = t_j t_{j+1} \cdots t_{j+m-1},$$

$$t_{i+\mathrm{I}_{\mathbf{t}}(n)} t_{i+\mathrm{I}_{\mathbf{t}}(n)+2} \cdots t_{i+\mathrm{I}_{\mathbf{t}}(n)+2m-2} = t_{j+k} t_{j+k+1} \cdots t_{j+k+m-1}.$$

From (5.2) one derives,

$$t_j t_{j+1} \cdots t_{j+m-1} = t_{j+k} t_{j+k+1} \cdots t_{j+k+m-1}.$$

This implies that $\mathrm{I}_{\mathbf{t}}(m) \leq k$ i.e. $\mathrm{I}_{\mathbf{t}}(n) \geq 2\,\mathrm{I}_{\mathbf{t}}(m)$ and conclusion follows from (5.1). Now consider the case that i is odd. As $\mathbf{t} \in \{01, 10\}^{\omega}$, one has $t_{i-1} = t_i$ and $t_{i+\mathrm{I}_{\mathbf{t}}(n)-1} = t_{i+\mathrm{I}_{\mathbf{t}}(n)}$. From (5.2), $t_i = t_{i+\mathrm{I}_{\mathbf{t}}(n)}$ and consequently, $t_{i-1} = t_{i+\mathrm{I}_{\mathbf{t}}(n)-1}$. Thus (5.2) holds with i replaced by $i-1$, and we are reduced to the previous case. □

Proposition 5.2.2 *Let* $\mathbf{t}$ *be the Thue-Morse word. Then* $\mathrm{I}_{\mathbf{t}}(n) = n$, *for* $n = 1, 2, 3, 4$.

Proof. Since the words

$$11, 1010, 010010, 10011001,$$

are factors of the Thue-Morse word,

$$\mathbf{t} = 0\overbrace{11}0\overbrace{10011001}011\overbrace{010010}11001\overbrace{1010}01\cdots$$

one has

$$\mathrm{I}_{\mathbf{t}}(1) \leq 1,\ \mathrm{I}_{\mathbf{t}}(2) \leq 2,\ \mathrm{I}_{\mathbf{t}}(3) \leq 3,\ \mathrm{I}_{\mathbf{t}}(4) \leq 4.$$

On the other hand, by Lemma 5.2.1,

$$\mathrm{I}_{\mathbf{t}}(n) \geq n,$$

for all $n > 0$. The conclusion follows.

□

Proposition 5.2.3 *Let* $\mathbf{t}$ *be the Thue-Morse word. For* $k > 0,\ 2^k < n \leq 2^k + 2^{k-1}$, *one has*

$$\mathrm{I}_{\mathbf{t}}(n) = 2^k + 2^{k-1}.$$

For $k > 0,\ 2^k + 2^{k-1} < n \leq 2^{k+1}$, *one has*

$$\mathrm{I}_{\mathbf{t}}(n) = 2^{k+1}$$

Proof. In the case $k = 1$, the statement is true, by the previous proposition. The proof can then be achieved easily by induction on k, using the Proposition 5.2.1 □

Chapter 6

C^∞-words

The Kolakoski word $\mathbf{k}$, introduced in [33], is a word over the alphabet $\{1, 2\}$ defined by the property that the word of its runlenght is equal to $\mathbf{k}$ itself

$$\mathbf{k} = \underbrace{22}_{2}\underbrace{11}_{2}\underbrace{2}_{1}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{11}_{2}\dots$$

Here a run is a maximal subword of consecutive identical letters. This word is an example of an infinite C^∞-word, as defined in [17]. For a survey on this word we refer to F.M. Dekking [18], [19]. How is this word generated? Culik et al. in [13] proposed the *double substitution rules* $\sigma_1(1 \to 1, 2 \to 11)$ and $\sigma_2(1 \to 2, 2 \to 22)$ which are applied alternatingly to each letter of the word. These substitutions can also be found in Allouche et al in [1]. In the paper [38], N. Ücoluk proved that the Kolakoski word is not ultimately periodic solving a question introduced by Kolakoski himself in [33]. A. Carpi in [9] proved that Kolakoski word is cube-free and contains only finitely many distant squares. These properties are shared by the class of the infinite C^∞-words of which Kolakoski word is a representative. This notwithstanding, several questions on it remain unanswered: among them we recall the asymptotical density of the letters and recurrence, i.e. whether

each factor appears infinitely often. Recently, [7], A. Ladoucer and S. Brleck proved that the existence of arbitrarly long palindromes implies the recurrence of the Kolakoski word. In this chapter, we start the investigation on the repetitivity index of C^∞-words. This research is motivated by the fact that some arguments pointed out in [9] conjectured that, for any rational $e > 1$, Kolakoski word only contains finitely many distinct factors with exponent larger than e. This is equivalent to say, as proved in [10], that the repetitivity index of Kolakoski word is not linearly bounded. In particular, we prove that this function is ultimately bounded from below by $n + rn^{1/q}$, where r is a suitable constant and $q = \log 1.5006/\log 1.4994$. This result has been obtained with A. Carpi in [11]. The chapter is organized as follows: in Section 6.1, we give some preliminary definitions; in Section 6.2, we recall some results on C^∞-words useful for our purposes; in Section 6.3, we prove the main result of this chapter.

6.1 Preliminaries

All through this chapter, $A = \{1, 2\}$. Any word $w \in A^*$ can be uniquely written as

$$w = a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n}, \tag{6.1}$$

with $n \geq 0$, $a_i \in A$, $k_i \geq 1$, $1 \leq i \leq n$, $a_j \neq a_{j+1}$, $1 \leq j \leq n-1$. The word $a_i^{k_i}$ is said to be the i-th *run* of w. If one has $k_i \leq 2$, for $1 \leq i \leq n$, then w is said to be *differentiable.* In this case, one can consider the word $k_1 k_2 \cdots k_n$ on the alphabet A, which is called the *mother* of w and it is denoted by $M(w)$. The *derivative* of w is the mother of the word obtained by w, by deleting the first and/or the last run, if their lenghts are equal to 1. It is denoted by $D(w)$. For example, $D(22112122) = 22112$, $D(12211) = 22$ and $D(212) = 1$.

We put $D(\epsilon) = \epsilon$, where ϵ is the empty word. Also $D(12) = D(21) = D(2) = D(1) = \epsilon$, in accordance with the definition. A word v such that $D(v) = w$ is called *primitive* of w. A word can have 8 primitives, at most; for example, the primitives of 22 are 1122, 2211, 21122, 12211, 11221, 22112, 122112, 211221. The two primitives with minimal lenght are called *principal* primitives. The maximal set of differentiable words (with respect to inclusion) closed for derivation is denoted by C^∞. Its elements are called *C^∞-words* or *smooth words*. We remark that the language C^∞ is recursive; to decide whether a word belongs to C^∞ it is sufficient to compute the sequence of words $D(w)$, $D^2(w)$, $D^3(w), \ldots$, halting when a non differentiable word, or the empty word, is found: one has $w \notin C^\infty$, in the first case, $w \in C^\infty$, in the second case. For instance, the word $w = 11212212 = 1^2 2^1 1^1 2^2 1^1 2^1$ is differentiable; one has $M(w) = 211211$ and $D(w) = 21121$. The word $D(w)$ is still differentiable, as well as its derivative $D^2(w) = 21$. Since $D^3(w) = D(21) = \epsilon$, we conclude that $w \in C^\infty$. The word $v = 2212122$ is differentiable, too, but $D(v) = 21112$, is not differentiable: therefore, $v \notin C^\infty$. By Equation (6.1), if w is differentiable, then one derives

$$|w| = k_1 + k_2 + \cdots + k_n = |M(w)|_1 + 2 \cdot |M(w)|_2 = |M(w)| + |M(w)|_2. \quad (6.2)$$

Since by the definition of derivative one has $|M(w)| - 2 \leq |D(w)| \leq |M(w)|$ and $|M(w)|_2 = |D(w)|_2$, by (6.2) we derive the useful inequalities

$$|D(w)| + |D(w)|_2 \leq |w| \leq |D(w)| + |D(w)|_2 + 2 \quad (6.3)$$

An *infinite C^∞-word* is an infinite word whose factors belong to C^∞. The *Kolakoski word* is the only infinite C^∞-word $\mathbf{k}$, beginning with the symbol 2, whose set of prefixes is closed for derivation. Thus,

$$\mathbf{k} = 2211212212211211 \cdots .$$

6.2 Main properties of smooth words

In this section we introduce some properties of the smooth words proved by Chvátal [12], Weakley [39] and Carpi [9].

Proposition 6.2.1 (Carpi, [9]) *For all $k \geq 0$, C^∞ contains only finitely many repetitions with gap k. In particular, the lengths of squares belonging to C^∞ are 2, 4, 6, 18, and 54.*

Proposition 6.2.2 (Carpi, [9]) *The maximal exponent of a C^∞-word is* $8/3$. *Consequently, no cube is a C^∞-word.*

Remark 6.2.1 In particular, Kolakoski sequence is cube-free and contains only squares of lenght 2, 4, 6, 18, and 54.

Let w be an infinite word on the alphabet A. For any $a \in A$ if the limit

$$\lim_{n\to\infty} \frac{1}{n}|w[1,n]|_a$$

exists, where $w[1,n] = w_1 w_2 \cdots w_n$, it is called the *density* (or *frequency*) of a in w. It has been conjectured by Keane in 1991 [31] that the asymptotic frequency of each symbol in the Kolakoski word is $1/2$. Chvátal has found a limitation close to this value.

Proposition 6.2.3 (Chvátal, [12]) *There exists a positive constant c such that, for any C^∞-word u one has*

$$0.4994|u| - c < |u|_1, |u|_2 < 0.5006|u| + c$$

Proposition 6.2.4 (Weakley, [39]) *Any factor of a C^∞-word is a C^∞-word. Conversely, for any C^∞-word w there exist words r, s of arbitrarily large length such that $rws \in C^\infty$.*

Let $\gamma(n)$ denote the number of C^∞-words of lenght n. We denote respectively by γ' and γ'' the first and the second difference of γ, i.e.,

$$\gamma'(n) = \gamma(n+1) - \gamma(n)\,, \quad \gamma''(n) = \gamma'(n+1) - \gamma'(n)\,, \quad n \geq 0\,.$$

We say that a C^∞-word w is *left special* if both $1w$ and $2w$ are C^∞-words. For each nonnegative integer n, let LS_n denote the set of left special C^∞-words of lenght n. In view of Proposition 6.2.4 one has that $\gamma(n+1) = \gamma(n) + \mathrm{Card}(LS_n)$, that is, $\mathrm{Card}(LS_n) = \gamma'(n)$.

Proposition 6.2.5 (Weakley, [39]) *Let $w = a_1a_2\cdots a_n$ be a C^∞-word, $a_i \in A$, $1 \leq i \leq n$, $n \geq 2$. The following conditions are equivalent*

1. *w is a left special C^∞-word;*

2. *$a_1 \neq a_2$ and $D(w)$ is a left special C^∞-word.*

For a nonempty word w, let $T(w)$ denote the word obtained by removing the rightmost letter of w.

Proposition 6.2.6 (Weakley, [39]) *Let $n > 0$ and $w \in A^*$. If $w \in LS_n$, then $T(w) \in LS_{n-1}$ and there is a $s \in LS_{n+1}$ such that $T(s) = w$.*

We say that a C^∞-word w is *fully extendable* if $1w1, 1w2, 2w1, 2w2 \in C^\infty$. Let FE_n denote the set of fully extendable words of lenght n. In view of Proposition 6.2.4, one has $|LS_{n+1}| = |FE_n| + |LS_n|$. Since $|LS_n| = \gamma'(n)$, one obtains $\gamma''(n) = |FE_n|$.

Proposition 6.2.7 (Weakley, [39]) *Let $w = a_1a_2\cdots a_n$ be a C^∞-word, $a_i \in A$, $1 \leq i \leq n$, $n \geq 2$. The following conditions are equivalent*

1. *$w \in FE$;*

2. $D(w) \in FE$ *word,* $a_1 \neq a_2$, *and* $a_{n-1} \neq a_n$;

3. $D(w) \in FE$ *and* $w = |D(w)| + |D(w)|_2 + 2$.

The *height* of a C^∞-word w is the least integer k such that $D^k(w) = \epsilon$. We write $ht(w)$ for the height of w. For example, the word 221121 has height 3; in fact, $D(221121) = 221$, $D(221) = 2$, $D(2) = \epsilon$. For each $k \geq 0$, let $A(k)$ denote the minimum and $B(k)$ the maximum lenght of fully extendable words of height k. W. Weakly showed that $B(k-1) < A(k)$ for $k = 0, 1, 2, \ldots, 17$ and for each n satisfying $B(k-1) + 1 \leq n \leq A(k) + 1$, he proved that

1. $\gamma(n) = (n+3)2^k - 2 \cdot 3^k$

2. There are positive constant c_1, c_2 such that $c_1 n^p \leq \gamma(n) \leq c_2 n^p$, where $p = \log 3 / \log 1.5 \approx 2.71$

On the other hand, F. M. Dekking in [18], proved that $cn^{2.15} \leq \gamma(n) \leq n^{7.2}$, where c is a suitable positive constant. Recently, Y. B. Huang in [21] proved that $c_1 n^{p_1} \leq \gamma(n) \leq c_2 n^{p_2}$ for any positive integer n, where $p_1 = \log 4.448 / \log 1.50084 > 3.6757$, $p_2 = \log 4.5063 / \log 1.49916 < 3.749$. Therefore, there are only finitely many positive integer k such that $B(k-1) < A(k)$. In the same paper, the author conjectures that there exist a suitable positive constants c_1 and c_2 such that $c_1 n^p \leq \gamma(n) \leq c_2 n^p$ for any positive integer n, where $p = \log 4.5 / \log 1.5 \approx 3.7095$.

The C^∞-words occurring in Kolakoski word $\mathbf{k}$ are called *admissible*. In [32], C. Kimberling asked to prove or disprove the following properties:
Mirror invariance: a word is admissible if and only if its mirror image is admissible.
Recurrence: every admissible word occurs infinitely often in $\mathbf{k}$.
In [18], F. M. Dekking proved that

Proposition 6.2.8 *Mirror invariance implies recurrence.*

Proposition 6.2.9 *Mirror invariance holds if and only if each C^∞-word is admissible.*

At the I.C.M. of Berlin in 1998, J. Cassaigne proved that the above mentioned properties are strictly related to the subword complexity of the Kolakoski word. More specifically, one has

Proposition 6.2.10 *Let* $\mathbf{k}$ *be the Kolakoski word.*
1) Keane problem implies that $\lambda_{\mathbf{k}}(n) = O(n^{\alpha+\epsilon})$,
2) Mirror invariance implies $\lambda_{\mathbf{k}}(n) \geq Cn^{\alpha}$,
with $\alpha = \log(3)/\log(3/2)$, C *a suitable constant and for all* $\epsilon > 0$.

In [7], A. Ladouceur and S. Brlek give a characterization of the palindromes related to the Kolakoski word $\mathbf{k}$. They proved that the set of all smooth palindrome words $\mathcal{P}$ is obtained in this way

$$\mathcal{P} = \{\widetilde{q}\cdot 1\cdot q,\ \overline{\widetilde{q}\cdot 1\cdot q} \mid q \in \operatorname{Pref}\mathbf{k}\} \cup \{\widetilde{q'}\cdot 11\cdot q', \overline{\widetilde{q'}\cdot 11\cdot q'} \mid q' \in \operatorname{Pref}(\Delta_2^{-1}(\overline{\mathbf{k}})\},$$

where $\overline{1} = 2$ and $\overline{2} = 1$ and $\Delta_2^{-1},\ \Delta_1^{-1}\ : A^* \to A^*$are so defined

$$\Delta_2^{-1}(u) = 2^{u[1]}1^{u[2]}2^{u[3]}\dots$$

$$\Delta_1^{-1}(u) = 1^{u[1]}2^{u[2]}1^{u[3]}\dots$$

For example, if $u = 221121$ then $\Delta_2^{-1}(u) = 221121221$.

This characterization of all C^∞-palindromes, based on the left palindromic closure of all prefixes of $\mathbf{k}$ obtained by using a bijection between the class of right infinite words over A and a class of words over the same alphabet, reveals the first link between the existence of some palindromes and the recurrence of $\mathbf{k}$. In fact, they proved that

Proposition 6.2.11 *If* $|\mathcal{P}'' \cap \operatorname{Fact}(\mathbf{k})| = \infty$, *then* $\mathbf{k}$ *is recurrent, with* $\mathcal{P}'' = \{p \in \mathcal{P} \mid p = \widetilde{q}\cdot 1\cdot q$ *and* $q \in \operatorname{Pref}(\mathbf{k})\}$.

6.3 Repetitivity index of smooth words

In [9], A. Carpi introduced the following relation on A^*:

$$\mathcal{R} = \{(u, v) \in A^* \times A^* \mid \exists z \in A^*,\ u = M(z),\ v \approx z,\ \text{ and } |u| \text{ is even}\}$$

and proved that

Lemma 6.3.1 *Let $u, v \in A^*$ be words such that $|u| \geq 2$ and uvu is differentiable. Then there are words $u', v' \in A^*$ such that*

$$D(u) = u',\ D(uvu) = u'v'u',\ (u'v', uv) \in \mathcal{R},\ |u'| < |u|,\ |v'| + |v'|_2 \leq |v| + 2.$$

□

In the sequel, by $\mathcal{R}$-closure of a subset C of A^*, we mean the minimal subset C' of A^* such that $C \subseteq C'$ and $\mathcal{R} \cap (C' \times A^*) \subseteq C' \times C'$. In the same paper the author denotes by $h(n)$ the maximal absolute values of $|x|_2 - |x|_1$, with x a factor of Kolakoski word of lenght n. Supported by machine computations, Keane in [31] conjectured that the density of the symbol 2 in Kolakoski word is asymptotically equal to $1/2$; a strengthening of this conjecture, proved by Chvátal [12], implies the sublinearity of $h(n)$. A direct analysis seems to suggest that $h(n)$ grows like $\sqrt{n}$. Moreover, Chvátal checked that $||x|_2 - |x|_1| \leq 5147$, for any x occurring the prefix of lenght 10^9 of the Kolakoski

$\widetilde{\Delta_2^{-1}}(\overline{\mathbf{k}})$	$\leftarrow$	$\cdots 112122121\overbrace{12112212}^{\widetilde{q'}}$	11	$\overbrace{21221121}^{q'}121221211\cdots$	$\rightarrow$	$\Delta_2^{-1}(\overline{\mathbf{k}})$
$\widetilde{\overline{\mathbf{k}}}$	$\leftarrow$	$\cdots 12212211211212211$	2	$11221211211221221\cdots$	$\rightarrow$	$\overline{\mathbf{k}} = \Delta_1^{-1}(\mathbf{k})$
$\widetilde{\mathbf{k}}$	$\leftarrow$	$\cdots 211211221\underbrace{22121122}_{\widetilde{q}}$	1	$\underbrace{22112122}_{q}122112112\cdots$	$\rightarrow$	$\mathbf{k}$

Table 6.1: Construction of palindromes in C^∞-words

word. Carpi conjectured that $h(n) = O(n^{1-\delta})$, for some $\delta > 0$. On the other hand, by machine computation Carpi checked that the $\mathcal{R}$-closure of any C^∞-word of lenght smaller than 16 is finite. One could think that the same property holds for all C^∞-words. The next proposition proved by Carpi in [9], shows an interesting implication of the conjectures above.

Proposition 6.3.1 *If $h(n) = O(n^{1-\delta})$ and the $\mathcal{R}-$closure of any C^∞ word is finite, then, for all $e > 1$, the lenght of the factors of the Kolakoski word with exponent larger than e is bounded.*

Therefore, if it would prove that the repetitivity index $\mathrm{I}_w(n)$, with w a C^∞-word, is upper bounded by a linear function then, by Proposition 3.1.2 and Proposition 6.3.1 at least one of the two problems above mentioned, $\mathcal{R}$-closure of any differentiable word and the conjecture on the density of the letter 2, is false. In this section we will establish a lower bound for this function. As proved in [9], for all $n \geq 0$, there are only finitely many C^∞-words w of the form $w = uvu$ with $u, v \in A^*$ and $|v| \leq n$. Thus, we can introduce the function $f : \mathbb{N} \to \mathbb{N}$ defined by

$$f(n) = \max\{k \geq 0 \mid \exists u, v \in A^*, \ \ |u| = k, |v| \leq n, \ \ uvu \in C^\infty\}. \tag{6.4}$$

The following holds.

Lemma 6.3.2 *One has $f(n) = O(n^q)$, where*

$$q = \frac{\log 1.5006}{\log 1.4994}.$$

Proof. For any non-negative integer n, we can find two words $u, v \in A^*$ such that

$$|u| = f(n), \ |v| \leq n, \ uvu \in C^\infty.$$

By Lemma 6.3.1, one has

$$D(u) = u',\ D(uvu) = u'v'u' \in C^\infty,\ |v'| + |v'|_2 \leq |v| + 2,$$

for suitable $u', v' \in A^*$, and also from (6.3),

$$|u'| + |u'|_2 \geq |u| - 2.$$

By Proposition 6.3.1 one derives

$$n \geq |v| \geq |v'| + |v'|_2 - 2 > 1.4994|v'| - c - 2$$

and

$$f(n) = |u| < |u'| + |u'|_2 + 2 < 1.5006|u'| + c + 2.$$

As $u'v'u' \in C^\infty$, one has $|u'| \leq f(|v'|)$, so that from the previous equations, taking in account the fact that f is non decreasing, one obtains

$$f(n) < 1.5006\ f\left(\left\lfloor \frac{n + c + 2}{1.4994} \right\rfloor\right) + c + 2.$$

From this inequality, the statement follows easily using the Master Theorem (see, e.g., [15]). □

Theorem 6.3.1 *Let* $\mathbf{w}$ *be an infinite* C^∞*-word. Then there is a constant* $r > 0$ *such that eventually,*

$$\mathrm{I}_{\mathbf{w}}(n) \geq n + rn^{1/q},$$

where q *has been defined in Lemma 6.3.2.*

Proof. Let $n > 54$ and set $\mathrm{I}_{\mathbf{w}}(n) = k$. Then there exist $x, y, z \in A^*$ such that

$$xy = zx, \quad |x| = n, \quad |y| = |z| = k.$$

Let us verify that $k \geq n$. Indeed, suppose $k < n$. In such a case, one has $x = zv = vy$ for some $v \in A^*$, so that $xy = vyy$. By Proposition 6.2.1 one

derives $|y| \leq 27$. If moreover, $|v| \geq |y|$, from $x = zv = vy$ one derives $v = uy$ for some $u \in A^*$, so that $xy = vyy = uyyy$; this yields a contradiction since by Proposition 6.2.1, C^∞-words are cube free. Thus, $|v| < |y|$ and, consequently, $n = |x| = |vy| < 2|y| \leq 54$. Thus, we may assume $k \geq n$. Then one has $z = xv$, $y = vx$ for some $v \in A^*$ and xvx is a factor of $\mathbf{w}$. Hence, $|x| \leq f(|v|)$, where f is the function defined by Equation 6.4. Hence, one has

$$|xv| \leq f(|v|) + |v| \leq c|v|^q + |v|,$$

for a suitable $c > 0$ and consequently, $n = |x| \leq c|v|^q$. One derives that $|v| \geq c^{-1/q} x^{1/q}$ and therefore

$$f(n) = k = |xv| \geq n + c^{-1/q} n^{1/q},$$

which proves the statement by taking $r = c^{-1/q}$. □

6.4 Further remarks

In the paper [33], W. Kolakoski introduced the homonymous infinite word and proposed two questions: what is the n-th term of this word? Is this word periodic? If the second question was solved by Necdet Ücoluk in [38] the following year, the first question was solved by B. Steinsky only in 2006, [37]. The recursive formula for n-term of the Kolakoski word is:

$$K_n = K_{n-1} + (3 - 2K_{n-1})\Big(1 - \frac{1}{2}\frac{K_{n-1} - K_{n-2}}{3 - 2K_{n-2}}\Big(1 + (-1)^{K_{1+\sum_{j=2}^{n-1} \frac{K_j - K_{j-1}}{3-2K_{j-1}}}}\Big)\Big).$$

Chapter 7

Synchronized sequences

In [23], the authors introduce the notion of a k-synchronized sequence, where k is an integer larger than 1. Roughly speaking, a sequence of natural numbers is said to be k-synchronized if its graph is represented, in base k, by a right synchronized rational relation. This is an intermediate notion between k-automatic, cfr. Section 2.2 and [14], and k-regular sequences [2]. In this chapter, we prove that the repetitivity index of a k-synchronized sequence is a k-synchronized sequence itself. This result has been obtained with A. Carpi in [11].

7.1 Preliminaries

Let A_i,, $1 \leq i \leq r$, be r alphabets, $r \geq 1$. By *relation* on the alphabets A_i, $1 \leq i \leq r$, we mean any subset of the direct product

$$M = A_1^* \times A_2^* \times \cdots \times A_r^*,$$

i.e., any element of the monoid $\wp(M)$ of the subsets of M. A relation is *rational* if it belongs to the smallest submonoid of $\wp(M)$ containing the finite

parts and closed for the operations of finite union and submonoid generation. A relation $\rho \subset M$ is *lenght-preserving* if for all element $(w_1, \ldots, w_r) \in \rho$ one has $|w_1| = |w_2| = \cdots = |w_r|$.

Let $\$ \notin \bigcup_{i=1}^{r} A_i$ be a new symbol. A relation $\rho \subset M$ is said to be *right synchronized rational* if the relation

$$\{(\$^{t-|w_1|}w_1, \ldots, \$^{t-|w_r|}w_r) \mid (w_1, \ldots, w_r) \in \rho,\ t = \max_{1 \le i \le r} |w_i|\}\,,$$

is a lenght-preserving rational relation.

Let $k \geq 2$ be an integer. For any $n \in \mathbb{N}$, we shall denote by $[n]_k$ the standard expansion of n in base k. Thus, $[n]_k$ is a word on the *digit alphabet* $D_k = \{0, 1, \ldots k-1\}$. We shall say that a subset σ of $\mathbb{N}^r$ is a *k-synchronized relation* if the relation

$$\{([n_1]_k, \ldots, [n_r]_k) \mid (n_1, \ldots, n_r) \in \sigma\}$$

is a right syncronized rational relation in $D_k^* \times \cdots \times D_k^*$. A sequence of natural numbers $u = (u_n)_{n=0}^{\infty}$ will be called a *k-synchronized sequence* if its graph $G_u = \{(n, u_n) \mid n \in \mathbb{N}\}$ is a k-synchronized relation.

By *projection* of a relation $\rho \subset \mathbb{N}^r$ we mean any of the r relations,

$$\{(x_1, \ldots, x_{i-1}, x_i, \ldots, x_r) \mid \exists x_i \in \mathbb{N} \text{ such that } (x_1, \ldots, x_r) \in \rho\}\,, \quad 1 \le i \le r\,.$$

Example 7.1.1 The 'sum' and the 'order' relation

$$\{(m, n, m+n) \mid m, n \in \mathbb{N}\}\,, \quad \{(m, n) \mid m, n \in \mathbb{N}, m < n\}$$

are k-syncronized relation, for all $k \geq 2$. For all $a, b \geq 0$, the sequence $an + b$ is k-syncronized, for all $k \geq 2$ [23].

From the analogous properties of right syncronized rational relations, one derives the following closure properties of the family of k-synchronized relation (see [23]).

Proposition 7.1.1 *For any $k \geq 1$, the class of k-synchronized relations is closed for Boolean operations, Cartesian product, projection, and permutation of coordinates.*

7.2 Repetitivity index of k-synchronized sequences

In this section, we study the repetitivity index of k-synchronized sequences. In order to study this function the following proposition proved in [23] is useful.

Proposition 7.2.1 *Let $w = (w_n)_{n=0}^{\infty}$ be a k-synchronized sequence. Then the relation*

$$\gamma_w = \{(i, j, h) \in \mathbb{N} \mid h > 0,\ w[i, i+h-1] = w[j, j+h-1]\},$$

is k-synchronized.

Now we can prove the main result of this section.

Proposition 7.2.2 *Let $w = (w_n)_{n=0}^{\infty}$ be a k-synchronized sequence. Then its repetitivity index I_w is a k-synchronized sequence.*

Proof. By definition, the repetitivity index $\mathrm{I}_w(n)$ is the least positive integer h such that $w[i, i+n-1] = w[i+h, i+h+n-1]$ for some $i \geq 0$. In other terms,

$$\mathrm{I}_w(n) = \min\{h > 0 \mid \exists i \in \mathbb{N},\ (i, i+h, n) \in \gamma_w\}\,, \tag{7.1}$$

where γ_w is the relation introduced in Proposition 7.2.1. Consider the relations

$$\rho_1 = \{(i, h, i+h, n) \mid i, h, n \in \mathbb{N}\}\,, \quad \rho_2 = \{(i, h, j, n) \mid h \in \mathbb{N}, (i, j, n) \in \gamma_w\}\,.$$

In view of Propositions 7.1.1 and 7.2.1 they are both k-synchronized, since they can be obtained respectively from the 'sum' and from γ_w by Cartesian product with $\mathbb{N}$ and permutation of coordinates. Still by Proposition 7.1.1, the relation

$$\rho_3 = \rho_1 \cap \rho_2 = \{(i, h, i+h, n) \mid i, h, n \in \mathbb{N},\ (i, i+h, n) \in \gamma_w\}.$$

is k-synchronized. Projecting ρ_3 on the second and fourth coordinates we obtain another k-synchronized relation,

$$\rho_4 = \{(h, n) \in \mathbb{N} \times \mathbb{N} \mid \exists i \in \mathbb{N},\ (i, i+h, n) \in \gamma_w\}.$$

Also the relation

$$\rho_5 = \rho_4 \setminus (\{0\} \times \mathbb{N}) = \{(h, n) \in \mathbb{N} \times \mathbb{N} \mid h > 0 \text{ and } \exists i \in \mathbb{N},\ (i, i+h, n) \in \gamma_w\}$$

is k-synchronized and from (7.1),

$$\mathrm{I}_w(n) = \min\{h \in \mathbb{N} \mid (h, n) \in \rho_5\}\,. \tag{7.2}$$

Now, consider the relation

$$\rho_6 = \{(h, h', n) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N} \mid (h', n) \in \rho_5,\ h > h'\}.$$

Using Propositions 7.1.1 one easily verifies that ρ_6 is k-synchronized, since it can be obtained by intersecting the k-synchronized relations $\mathbb{N} \times \rho_5$ and $> \times \mathbb{N}$. Projecting ρ_6 on the first and third coordinates, we obtain the k-synchronized relation

$$\rho_7 = \{(h, n) \in \mathbb{N} \times \mathbb{N} \mid \exists h' < h,\ (h', n) \in \rho_5\}.$$

One easily verifies that a pair $(h, n) \in \mathbb{N} \times \mathbb{N}$ belongs to the relation $\rho_5 \setminus \rho_7$ if and only if h is the least positive integer such that $(h, n) \in \rho_5$. Thus, in view of (7.2)

$$\rho_5 \setminus \rho_7 = \{(\mathrm{I}_w(n), n) \mid n \in \mathbb{N}\}\,.$$

Since the relation $\rho_5 \setminus \rho_7$ is k-synchronized, from Propositions 7.1.1 we conclude that $\mathrm{I}_w(n)$ is a k-synchronized sequence. □

As proved in [23], any k-synchronized sequence is linearly bounded. One derives immediately the following corollary of Proposition 7.2.2

Corollary 7.2.1 *The repetitivity index of a k-synchronized sequence has a linear upper bound.*

Remark 7.2.1 As the Thue-Morse word is generated by a 2-DFAO, it is 2-automatic, cfr. Section 2.2. As any k-automatic sequence is also k-synchronized, cfr. [23], then the Thue-Morse word is 2-synchronized. Thus, by Proposition 7.2.2 the repetitivity index of the Thue-Morse word has to be 2-synchronized. This is confirmed by Proposition 5.2.3.

Chapter 8

Final remarks

Some questions considered in this thesis remain unsolved. In Chapter 4, we have construct effectively an infinite binary word w having this property: for any $\nu > 0$, there is an integer $n(\nu)$, effectively computed, such that, for all $n > n(\nu)$, $\mathrm{I}_w(n) > n^\nu$. The next objective is to construct an infinite binary word w such that, given an arbitrary small $\epsilon > 0$, $\mathrm{I}_w(n) \geq (2-\epsilon)^n$, for all $n > n(\epsilon)$ and therefore to solve the Beck's theorem in a costructive form. In Chapter 5, we have completely solved the study of the repetitivity index of the Thue-Morse word $\mathbf{t}$, giving a formula to compute $\mathrm{I}_{\mathbf{t}}(n)$. The more general problem of computing the repetitivity index of synchronized sequences has been solved in Chapter 7.

In Chapter 6, we have started the investigation on the repetitivity index of C^∞-words and in particular for the Kolakoski word. We have proved that this function is ultimately bounded from below by $n + rn^{1/q}$, where r is a suitable constant and q is a constant strictly related to the frequency of the symbols. As we have seen, this research is motivated by the fact that A. Carpi has conjectured in [9] that, for any rational $e > 1$, the lenght of the factors of the Kolakoski word with exponent larger than e is bounded; this

result is equivalent to say, as proved in [10], that the repetitivity index of Kolakoski word is not linearly bounded. Therefore, the next objective is to search for an upper bound of this function.

Bibliography

[1] J. P. Allouche and J. Shallit, Automatic Sequences, Cambrige University Press.

[2] J. P. Allouche and J. Shallit, The ring of k-regular sequences, *Theoret. Comput. Sci.* **98** (1992), 163–197.

[3] J. Beck, An application of Lovász lemma: there exists an infinite 01-sequence containing no near identical intervals, *Finite and Infinite Sets, Eger (Hungary)* **37** (1981), 103–107.

[4] J. Berstel, J. Karhumäki, Combinatorics on words: a tutorial, *Bull. EATCS* **79** (2003), 178–228

[5] J. Berstel and P. Seebold, A characterization of overlap-free morphism, *Discrete Appl. Math.* **46** (1993), 275–281.

[6] J. Berstel, A Rewriting of Fife's Theorem about Overlap-Free Words, *Results and Trends in Theoretical Computer Science* (1994), 19–29.

[7] S. Brlek and A. Ladoucer, A note on differentiable palindromes, *Theoret. Comput. Sci.* **302** (2003), 167–178.

[8] A. Carpi, On the repetition threshold for large alphabets, *Theoret. Comput. Sci.* **385** (2007), 137–151.

[9] A. Carpi, On repeated factors in C^∞-words, *Inform. Process. Lett.* **52** (1994), 289–294.

[10] A. Carpi and V. D'Alonzo, On a word avoiding near repeats, Proc.s WORDS 2007, Marseille, France, 17–21 September, pp. 72–78.

[11] A. Carpi and V. D'Alonzo, On the repetitivity index of infinite words, *Int. J. Algebra and Computation*, to appear.

[12] V. Chvatal, Notes on Kolakoski sequence, DIMACS Tech. Rept. 93–84, (1994).

[13] K. Culik, J. Karhumäki, Iterative devices generating infinite words, *Lec. Notes in Comp. Sc.* **577** (1992), 531–544.

[14] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 401–419.

[15] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd Edition, MIT Press and McGraw-Hill, 2001.

[16] F. Dejean, Sur un théorème de Thue, *J. of Combin. Theory, Ser. A* **13** (1972), 90–99

[17] F. M. Dekking, Regularity and irregularity of sequences generated by automata, *Seminarie de Theorie des Nombres de Bordeaux,* (1979–80), exposé n.9

[18] F. M. Dekking, On the structure of selfgenerating sequences, *Seminarie de Theorie des nombres de Bordeaux,* (1980–81), exposé n.31

[19] F. M. Dekking, What is the long range order in the Kolakoski sequence?, In: R. V. Moody (Ed.) *Proc.s NATO Advanced Study Institute,* Water-

loo, ON, August 21-September 1, 1995, Kluwer (Dordrecht 1997), pp. 115–125.

[20] P. Erdös - L. Lovász, Problems and results on 3–chromatic hypergraphs and some related questions, in *Infinite and Finite Sets,* Colloquia Math. Soc. János Bolyai **10** (1973).

[21] Yun Bao Huang, A note on the complexity of C^{∞}-words, submitted

[22] L. Lovász, *Combinatorial problems and exercises,* Akadémiai Kiadó and North-Holland, Amsterdam-New York-London, (1979).

[23] A. Carpi and C. Maggi, On synchronized sequences and their separators, *Theor. Inform. Appl.* **35** (2001), 513–524.

[24] L. Ilie, P. Ochem and J. Shallit, A generalization of Repetition Threshold, in: J. Fiala et al. (eds.), *Proc.s of MFCS 2004*, Lecture Notes in Computer Science, 3153, Springer (Berlin, 2004), pp. 818-826.

[25] Lothaire M., Combinatorics on Words. Addison-Wesley, Reading MA (1983).

[26] M. Mendès France, A. J. van der Poorten, Arithmetic and analytic properties of paperfolding sequences, *Bull. Austr. Math. Soc.* **24**, (1981).

[27] M. Morse and G. Hedlund, Symbolic dynamics, *Amer. J. Math.* **60** (1938), 815–866.

[28] M. Mohammad-Noori, J. D. Currie, Dejean's conjecture and Sturmian words, *European J. Combin.* **28** (2007), 876–890.

[29] J. Moulin-Ollagnier, Proof of Dejean's conjecture for alphabets with 5,6,7,8,9,10 and 11 letters, *Theoret. Comput. Sci.* **95** (1992), 187–205

[30] F. Mignosi and G. Pirillo, Repetitions in the Fibonacci infinite word, *Theor. Imform. Appl.* **26** (1992), 199–204.

[31] Keane, M. S. Ergodic Theory and Subshifts of finite type, in *Ergodic Theory Symbolic Dynamics and Hyperbolic spaces.* T.Belford, M. Keane, C. Series (Eds.), Oxford University Press, Oxford.

[32] C. Kimberling, Problem 6281, *Amer. Math. Montly* **86** (1979), 793.

[33] Kolakoski, W. Self generating runs, Problem 5304, *Amer. Math. Montly* **72** (1965), 674.

[34] J.-J. Pansiot, A propos d'une conjecture de F.Dejean sur les répétitons dans les mots, *Discr. Appl. Math* **7** (1984), 297–311.

[35] R. L. Graham - B. Rotschild - J. Spencer, *Ramsey Theory*, John Wiley, New York, (1980).

[36] J. Spencer, Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20** (1976), 69–76.

[37] B. Steinsky, A Recursive Formula for the Kolakoski Sequence, *J. Integer Seq.* **9** (2006).

[38] N. Üçoluk, *Amer. Math. Montly* **73** (1966), 681–682.

[39] William D.Weakley, On the number of C^{∞}-words of each lenght, *J. Combin. Theory, Ser. A* **51** (1989), 55–62.

[40] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. Mat. Nat. Kl.* **7** (1906), 1–22.

[41] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912) 1–67. Reprinted in:

T. Nagell (Ed.), *Selected Mathematical Papers of Axel Thue*, Universitetsforlaget (Oslo, 1977) pp. 413–478.

Acknowledgments

A conclusione di questo lavoro, penso sia doveroso esprimere i miei piú sinceri ringraziamenti ai professori Aldo de Luca e Arturo Carpi che con pazienza, disponibilitá e comprensione mi hanno accolto e sostenuto.