RAIRO-Theor. Inf. Appl. **42** (2008) 599–613
DOI: 10.1051/ita:2008012

Available online at:
www.rairo-ita.org

DRUNKEN MAN INFINITE WORDS COMPLEXITY

Marion Le Gonidec[1]

Abstract. In this article, we study the complexity of drunken man infinite words. We show that these infinite words, generated by a deterministic and complete countable automaton, or equivalently generated by a substitution over a countable alphabet of constant length, have complexity functions equivalent to $n(\log_2 n)^2$ when n goes to infinity.

Mathematics Subject Classification. 11B85, 68R15.

Introduction

The structure of an infinite word, and especially the diversity of the subwords appearing in an infinite word $m = m_0m_1m_2\ldots$ over a finite alphabet, can be described by a function measuring the subword complexity. This function, usually denoted p_m, is the function from $\mathbb{N}$ to $\mathbb{N}$ which maps an integer n to the number of different subwords of length n appearing in m. Namely, if we denote by $F_n(m)$ the set of subwords of length $n \geq 0$ appearing in m, the subword language of m is $L(m) = \cup_{n\geq 0}F_n(m)$ and $p_m(n) = \mathrm{Card}\,(F_n(m))$.

Even if a complexity function must fit many simple properties, its computation often remain awkward and their possible behaviours are various, from constant functions to exponential growth functions and from simple to irregular non-decreasing functions. Consequently, there are several questions related to the computation of subword complexity: can the complexity of a given word be computed exactly? Can we, at least, give an equivalent or the growth order of the complexity? Which function can be realized as a complexity function? Which properties holds for words with same complexity? All these questions have been the aims of many works. See for example [1] or [6] for surveys about these questions.

As further motivations for computing complexity functions, the sequence $\frac{\log p_m(n)}{n}$ tends to the entropy of the dynamical system associated with the word (see for example [9]) and the function $p_m(n)$ is also linked to the Kolmogorov complexity of the infinite word m (see [15]).

[1] Institut de Mathématiques, Université de Liège, Grande traverse 12 B.37, 4000 Liège, Belgium; M.LeGonidec@ulg.ac.be

Article published by EDP Sciences

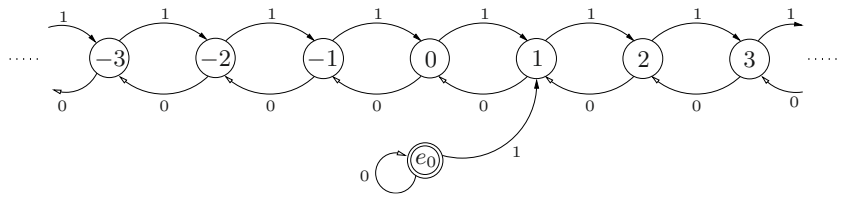

FIGURE 1. Automaton generating the drunken man infinite words.

If we cannot, in the general case, compute exactly all the complexity functions of large classes of infinite words, we can determine for some classes of words, the possible growth orders of the complexity functions. This is the case for classes of words generated by simple algorithms as automatic words, generated by finite automata. A result of Cobham [5] ensures that an automatic word which is not ultimately periodic (so that its complexity is at least $n \mapsto n+1$) have a sublinear complexity function (see [2] for an overview on automatic words). This theorem has been extended to fixed points of substitutions over finite alphabet by Pansiot [14].

One can ask, in order to extend the result of Cobham, if we can obtain a similar result for infinite words generated by deterministic countable automata. As a first result in this direction, we have shown in [10] that the complexity functions of a large class of words over finite alphabets, generated by complete, deterministic and countable automata of uniformly bounded degree, are at most polynomial, that is, $p_m(n) \leq Cn^\alpha$ for a constant α. Moreover, the constant α only depends on the number of transition labels and on the uniform bound of the in-degree.

The purpose of this article is to study the complexity functions of a family of infinite binary words, generated by the first basic example of countable automaton $\mathcal{A}$ represented in Figure 1. These words have their complexity functions equivalent to $n(\log_2 n)^2$ when n goes to infinity.

We construct the family of drunken man infinite words from a unique word, denoted $\overline{m}$, over the infinite alphabet $\mathbb{Z} \cup \{e_0\}$, which is the output word of the automaton. The n-th letter of $\overline{m}$ is given by the label of the output state obtained by feeding the automaton with the proper binary representation of n, read from the most to the least significant digit. That is, if we denote by the word $\rho_2(n) = n_l \dots n_1 n_0$ the proper binary representation of a positive integer $n = \sum_{i=0}^{l} n_i 2^i$, with $n_i \in \{0,1\}$ and the convention $\rho_2(0) = \varepsilon$, the output state obtained by feeding the automaton with $\rho_2(n)$ is labeled by $\overline{m}_n = |\rho_2(n)|_1 - |\rho_2(n)|_0$, where $|\rho_2(n)|_i$ represents the number of occurrences of the letter i in the word $\rho_2(n)$.

The initial state e_0 allows to feed the automaton with non proper binary representations without changing the output state.

The family of the *drunken man infinite words* $\left\{m^{(f)} \in \{0,1\}^{\mathbb{N}},\ f \in \mathbb{Z}\right\}$, where $m^{(f)}$ is generated by the countable automaton $\mathcal{A}$ and the following output function Π_f,

$$\begin{array}{rccc} \Pi_f: & \mathbb{Z} \cup \{e_0\} & \to & \{0,1\} \\ & f & \mapsto & 1 \\ & e_0 & \mapsto & 1 \\ & e \neq f, e_0 & \mapsto & 0 \end{array}$$

extended by concatenation to $(\mathbb{Z} \cup \{e_0\})^{\mathbb{N}}$, so that, for example,

$$m^{(-1)} = \Pi_{-1}(\overline{m}) = 0000100000000000011010001000000000000000000000000\ldots$$

$$m^{(0)} = \Pi_0(\overline{m}) = 1010000001101000000000000000000000010110011010000110 1\ldots$$
$$m^{(1)} = \Pi_1(\overline{m}) = 0100011000000000000101100110100000000000000000000000\ldots$$
$$m^{(2)} = \Pi_2(\overline{m}) = 0001000000010110000000000000000000000001000101100001 0\ldots$$

These words are 2^{∞}-automatic in the sense of [11] and [12] and are closely related to some context-free languages of binary representations. Indeed, the infinite word $m^{(f)}$ is the characteristic word of the following set:

$$\mathcal{S}_f = \{n \in \mathbb{N},\ |\rho_2(n)|_1 = |\rho_2(n)|_0 + f\}.$$

That is, by definition of $m^{(f)}$,

$$\forall n \in \mathbb{N},\ m_n^{(f)} = \begin{cases} 1 & \text{if } n \in \mathcal{S}_f, \\ 0 & \text{otherwise.} \end{cases}$$

We can find in [8,13] and [12] a study of arithmetic and statistical properties of the set $\mathcal{S}_0$.

Thanks to the correspondence between countable automata and substitutions over countable alphabets (we refer to [12] regarding to this correspondence), there is another way to construct the word $m^{(f)}$. Indeed, the substitution $\overline{\sigma}$ over the alphabet $\mathbb{Z} \cup \{e_0\}$, that is,

$$\begin{array}{rccc} \overline{\sigma}: & \mathbb{Z} \cup \{e_0\} & \to & \mathbb{Z}^* \cup \{e_0 1\} \\ & e & \mapsto & (e-1)(e+1) \\ & e_0 & \mapsto & e_0 1 \end{array}$$

admits $\overline{m}$ as unique fixed point:

$$\overline{m} = \overline{\sigma}(\overline{m}) = e_0 102(-1)113(-2)0020224(-3)(-1)(-1)1(-1)113(-1)1131335(-4)$$
$$(-2)(-2)0(-2)002(-2)0020224(-2)002022402242446(-5)(-3)(-3)(-1)(-3)$$
$$(-1)(-1)1(-3)(-1)(-1)1(-1)113(-3)(-1)(-1)1(-1)113(-1)1131335(-1)1131$$
$$33513353557\ldots$$

and the drunken man infinite words $m^{(f)}$ are obtained from $\overline{m}$ by the projection Π_f.

Ferenczi have studied in [7] dynamical systems associated with substitutions over countable alphabets. He shows that, under conditions of aperiodicity and irreducibility, subshifts associated with substitutions over countable alphabets admit invariant measures, which can be finite or infinite. These invariant measures are also ergodic for some examples of substitutions.

In this article, we will focus on combinatorial aspects of the words $m^{(f)}$ to show that their complexity functions are equivalent to $n(\log_2 n)^2$ when n goes to infinity. It seems that the construction of $m^{(f)}$ as projection of the fixed point of $\overline{\sigma}$ is the best point of view to obtain this result.

1. Drunken man infinite word complexity

Theorem 1.1. *The complexity function of all drunken man infinite word is equivalent to* $n\log_2^2 n$.

To prove this theorem, we search an upper bounding function and a lower bounding function for $p_{m^{(f)}}$, both equivalent to $n\log_2^2 n$. Thus, we will show the following two results.

Claim 1.2. For all integers f in $\mathbb{Z}$, the complexity function of $m^{(f)}$ admits the following upper bound:

$$\forall n \geq 1, \quad p_{m^{(f)}}(n) \leq n(\log_2 n + 9)(\log_2 n + 2).$$

Claim 1.3. For all integers f in $\mathbb{Z}$, the complexity function of $m^{(f)}$ admits the following lower bound:

$$\forall n \in \mathbb{N}, n \geq 3, \ p_{m^{(f)}}(n) \geq n\left((\log_2 n)^2 - 2\right) + 2.$$

The upper bound of Claim 1.2 follows from a thin bounding of the number of subwords of length 2 appearing in m, based on methods from [10]. The lower bound of Claim 1.3 results from computing special subwords of fixed length and need a deep study of the words $\overline{\sigma}^k(e)$'s structure, for all e in $\mathbb{Z} \cup \{e_0\}$.

Note 1.4. To lighten the following proofs, we will fix for the end of the section, an integer f of $\mathbb{Z}$ and we denote, for all positive integers k, $\sigma^k = \Pi_f \circ \overline{\sigma}^k$.

We will also denote by $[\![x, y]\!]$ the set $\{z \in \mathbb{Z} \mid x \leq z \leq y\}$.

Remark 1.5. Even if the words $m^{(f)}$ share the same kind of complexity function and also share special subwords and their constructions, we must notice that, they do not necessarily share the same language. For example, the word $w = 100011$ belongs to $L(m^{(1)})$ but does not belong to $L(m^{(0)})$.

Indeed, if we assume $w = 100011$, w belongs to $L(m^{(1)})$ and also belongs to $L(m^{(0)})$, there is, by definition of $m^{(0)}$, a word $\overline{w} = 0x_1x_2x_300$ in $\overline{m}$ which projects on w by Π_0. As $\overline{m}$ is the fixed point of $\overline{\sigma}$ and as 00 is not the image of a letter by $\overline{\sigma}$, x_1x_2 and x_30 have to be the images of letters so $x_3 = -2$ and there exists a letter x

such that $\overline{w} = 0(x-1)(x+1)(-2)00$. Moreover, $\overline{w}$ always appears in $\overline{m}$ followed by a 2 and preceded by a -2. As $(-2)\overline{w}2$ is a subword of $\overline{m}$, its pre-image by $\overline{\sigma}$ is also a subword of $\overline{\sigma}$ as $\overline{m}$ is the fixed point of $\overline{\sigma}$ so the word $\overline{u} = (-1)x(-1)1$ must be a subword of $\overline{m}$. According to the images of letters by $\overline{\sigma}$, x could be a letter 1 or -3. The case $x = 1$ is excluded because it would imply $x_1x_2 = 02$ so the second letter of w should be a 1 which is not the case. Thereby, the word $\overline{u} = (-1)(-3)(-1)1$ is the only possibility, but it also lead to a contradiction. Indeed, as $(-3)(-1)$ is not the image of a letter by $\overline{\sigma}$, $\overline{u}$ always appears in $\overline{m}$ followed by a 3 and preceded by a -3. The pre-image $(-2)(-2)2$ of $(-3)\overline{u}3$ by $\overline{\sigma}$ must also be a subword of $\overline{\sigma}$, which is not the case as neither $(-2)(-2)$ nor $(-2)2$ are images of letters by $\overline{\sigma}$.

So there is no subword of $\overline{m}$ projecting on w by Π_0, that is, $w = 100011$ does not belong to $L(m^{(0)})$.

Using similar arguments, one can shows that the word 000100000001011 belongs to $L(m^{(2)})$ but does not belong to $L(m^{(0)})$ and the word 10001100 belongs to $L(m^{(1)})$ but does not belong to $L(m^{(2)})$.

1.1. An upper bound for the complexity functions

To give an upper bound of the complexity function, we need combinatorial results on fixed points of substitutions of constant length over countable alphabets, which are recalled in the following proposition (see [10] for the proof).

Proposition 1.6. *Let m be the fixed point beginning by the letter a_0 of a substitution ς of constant length q over a countable alphabet A, defined by $\varsigma(a) = \varsigma_0(a)\varsigma_1(a)\dots\varsigma_{q-1}(a)$.*

*For a given letter a appearing in m and a given positive integer $k \geq 1$, the word $\varsigma^k(a) = u_0u_1\dots u_{q^k-1}$, called the k-*th iterated of *a by ς, satisfies:*

$$\forall n \in \{0, 1, \dots, q^k - 1\},\ u_n = \varsigma_{n_0} \circ \varsigma_{n_1} \circ \cdots \circ \varsigma_{n_l} \circ \varsigma_0^{k-(l+1)}(a),$$

where $\rho_q(n) = n_l \dots n_1n_0$ is the q-ary representation of n.

This proposition allows to obtain useful facts about the structure of the subwords $\sigma^k(x)$ which appear in the drunken man infinite word:

Lemma 1.7. *Let x be an element of $\mathbb{Z}$ and k be a positive integer, we have:*

1. *If $x - f$ and k have the same parity, all the values in $[\![x - k, x + k]\!]$ with the same parity as f, and only these, appear in $\overline{\sigma}^k(x)$.*
2. *If $x - f$ and k have different parity, all the values in $[\![x - k, x + k]\!]$ with different parity from f, and only these, appear in $\overline{\sigma}^k(x)$.*

In particular, the word $\sigma^k(x) \neq 0^{2^k}$ if and only if there exists an integer $p \in [\![0, k]\!]$ such that $x = f + k - 2p$.

Proof. The first two item of this lemma are consequence of the fist part of Proposition 1.6. Indeed, if we note $\overline{\sigma}^k(x) = w_0w_1\dots w_{2^k-1}$ then, using Proposition 1.6, we have $w_i = x + |\rho_2(i)|_1 - (|\rho_2(i)|_0 + k - |\rho_2(i)|)$, that is $w_i = x + 2|\rho_2(i)|_1 - k$.

Thus all the letters w_i belong to $[\![x-k, x+k]\!]$ and have same parity as x if k is even and have different parity from x if k is odd.

From it, we obtain that letters w_i have the same parity as f if and only if x and f have the same parity and k is even or if x has different parity from f and k is odd, that is letter w_i have the same parity as f when $x-f$ and k have the same parity. □

To obtain Claim 1.2, we are going to give an upper bound to the number of subwords of $m^{(f)}$ of length 2^k. The idea is to extract from the words of $F_2(\overline{m})$ the words leading to subwords of $m^{(f)}$ of type $\sigma^k(x_1)\sigma^k(x_2)$ which are not $0^{2^{k+1}}$. Indeed, if w is a subword of length 2^k of m, w is a subword of some $\sigma^k(x_1)\sigma^k(x_2)$, due to the structure of fixed point of $\overline{m}$. Nevertheless, infinitely many pairs (x_1', x_2') can lead to subwords $\sigma^k(x_1')\sigma^k(x_2')$ containing w as a subword, due to the projection Π_f, but we can find a finite set of subwords x_1x_2 of $\overline{m}$ such that every subword of length 2^k of $m^{(f)}$ is a subword of one of the $\sigma^k(x_1)\sigma^k(x_2)$. The following proposition about subwords of length two of $\overline{m}$, proved in [10], will help us to find this finite set.

Proposition 1.8. *The set of subwords of length two of $\overline{m}$ is:*

$$F_2(\overline{m}) = \{e_0 1\} \cup \{x_1x_2 \mid x_1 \in \mathbb{Z}, x_2 = x_1 - 2p,\ p \geq -1 \text{ or } x_2 = -x_1+1 \text{ if } x_1 > 0\}.$$

Lemma 1.9. *For some integer $k \geq 1$, we denote by $\mathcal{U}_k(f)$ the following set of words of length two:*

$$\{e_0 1\} \cup \Big\{(f+k-2q)(f+k-2p) \mid (q,p) \in [\![-1,k]\!] \times [\![0,k+1]\!] \setminus \{(-1,k+1)\},\ p \geq q-1\Big\}.$$

For every word w of $m^{(f)}$ of length 2^k, there is a word x_1x_2 of $\mathcal{U}_k(f)$ such that w is a subword of $\sigma^k(x_1)\sigma^k(x_2)$.

Proof. There is a suitable word for 0^{2^k} in $\mathcal{U}_k(f)$, for example $\sigma^k(f+k)\sigma^k(f+k) = 10^{2^{k+1}-2}1$.

Let w a subword of $m^{(f)}$ of length 2^k different from 0^{2^k}. If w is not a subword of $\sigma^k(e_0)\sigma^k(1)$, there is at least one pair (x_1, x_2) of $\mathbb{Z}^2$ such that w is a subword of $\sigma^k(x_1)\sigma^k(x_2)$.

As w is not 0^{2^k}, $\sigma^k(x_1)$ or $\sigma^k(x_2)$ contains an occurrence of 1 and then, the letter x_1 or the letter x_2 belongs to $[\![f-k, f+k]\!]$ and can be written $f+k-2q$ for some q in $[\![0,k]\!]$.

Assume $x_1 = f+k-2q$ for a q in $[\![0,k]\!]$. Proposition 1.8 implies $x_2 = e_1 - 2p$ for some $p \geq -1$ or $x_2 = -e_1+1$ if $x_2 \leq 0$. According to the cases, we get different words of $\mathcal{U}_k(f)$:

- if $x_2 = x_1 - 2p$ and x_2 also belongs to $[\![f-k, f+k]\!]$, the word x_1x_2 belongs $\mathcal{U}_k(f)$;
- if $x_2 = -x_1+1$ or $x_2 = x_1 - 2p$ does not belong to $[\![-k,k]\!]$, then $\sigma^k(x_2) = 0^{2^k}$, so $\sigma^k(x_2) = \sigma^k(f-k-2)$.

Proposition 1.8 ensures that $x_1(f-k-2)$ is also a subword of $\overline{m}$ and it follows: w is a subword of $\sigma^k(x_1)\sigma^k(f-k-2)$ and $x_1(f-k-2)$ belongs to $\mathcal{U}_k(f)$.
Assume $x_2 = f+k-2q$ for a q in $[\![0,k]\!]$. Proposition 1.8 implies $x_1 = x_2+2p$ for some $p \geq -1$ or $x_1 = -x_2+1$, only if $x_2 < 0$. According to the cases, we get different words of $\mathcal{U}_k(f)$:

- if $x_1 = x_2+2p$ and x_1 belongs to $[\![f-k, f+k]\!]$, the word x_1x_2 is in $\mathcal{U}_k(f)$;
- if $x_1 = -x_2-1$ or $x_1 = x_2+2p$ does not belong to $[\![f-k, f+k]\!]$, we get $\sigma^k(x_1) = 0^{2^k} = \sigma^k(f+k+2)$. Proposition 1.8 ensures that $(f+k+2)x_2$ is a subword of $\overline{m}$ and it follows: w is a subword of $\sigma^k(f+k+2)\sigma^k(x_2)$ and $(f+k+2)x_2$ belongs to $\mathcal{U}_k(f)$.

Thus, we found, for all words w of length 2^k of $m^{(f)}$, a word x_1x_2 of $\mathcal{U}_k(f)$ such that w is a subword of $\sigma^k(x_1)\sigma^k(x_2)$. □

Proof of Claim 1.2. According to Lemma 1.9, every subword of length 2^k of m is a subword of some $\sigma^k(x_1)\sigma^k(x_2)$ where (x_1,x_2) belongs to $\mathcal{U}_k$. It leads to the following inequality: $p_m(2^k) \leq 2^k \cdot \mathrm{Card}(\mathcal{U}_k(f))$, where 2^k represents the number of possible ways of extracting a subword of length 2^k from a subword $\sigma^k(x_1)\sigma^k(x_2)$ and $Card(\mathcal{U}_k(f))$ is the cardinal of $\mathcal{U}_k(f)$. As

$$\mathrm{Card}(\mathcal{U}_k(f)) = 1+(k+1)+(k+2)+\sum_{i=3}^{k+2} i = \frac{k^2+9k+8}{2},$$

it follows

$$p_{m^{(f)}}(2^k) \leq 2^k\frac{k^2+9k+8}{2} \quad \text{and} \quad p_{m^{(f)}}(2^{k+1}) \leq 2^{k+1}\frac{k^2+11k+18}{2}.$$

As the complexity function is an increasing function, using the fact that, for all integers n in $[\![2^k, 2^{k+1}[\![$, $\log_2(n)$ belongs to $[\![k, k+1[\![$, we obtain:

$$\forall n \geq 1, \quad p_{m^{(f)}}(n) \leq n(\log_2 n+9)(\log_2 n+2),$$

that is the announced upper bound for $p_{m^{(f)}}(n)$. □

1.2. A lower bound for the complexity functions

Among all the subwords of a given infinite word, the *special factors* form the main tool to find a lower bound for complexity function.

The *left special factors* (resp. *right special factors, bispecial factors*) of m are subwords of m appearing in m extended on the left (resp. on the right, on the both sides) by at least two different letters. For binary words, the difference $p_m(n+1)-p_m(n)$ is exactly the number of right special factors or the number of left special factors. Many other deeper formulas hold between special factors and complexity (see [3,4]).

These subwords are really useful to compute complexity, especially when we consider infinite words generated by simple algorithms (so we can organize special

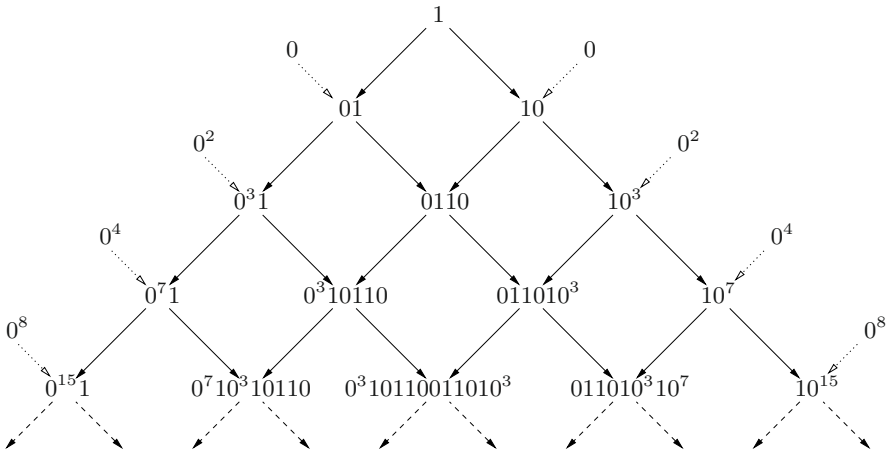

FIGURE 2. Production process of the words $\sigma^k(x) \neq 0^{2^k}$.

factors in families) or when we consider infinite words of low complexity (so the special factors are not too numerous).

To find a lower bound for the complexity function of the drunken man infinite words, we need to know more precisely the structure of the subwords $\sigma^k(f+k-2p)$ for a fixed integer k and some p in $[\![0,k]\!]$, that means the subwords $\sigma^k(x)$ which are different from 0^{2^k}. Indeed, knowing exactly where are the occurrences of 1 in the subwords $\sigma^k(f+k-2p)$ will leads us to localize and extract a family of special factors of $m^{(f)}$. Thanks to this family of subwords, we will be able to give a lower bound for the complexity function of the drunken man infinite word $m^{(f)}$.

First of all, there is a simple way to compute the subwords $\sigma^k(x)$ different from 0^{2^k}. We can construct a triangle (see Fig. 2), using the substitutive relation $\sigma^k(x) = \sigma^{k-1}(x-1)\sigma^{k-1}(x+1)$, where the p-th element of the k-th row is $\sigma^k(f+2p-k)$.

We now focus on the properties of words $\sigma^k(f+k-2p)$ for $p \in [\![0,k]\!]$.

Note 1.10. Let $k>0$ and $q>0$ be two positive integers and x be an integer.

We denote by $R_k(x,q)$ the rank, from 0 to 2^k-1, of the q-th occurrence of 1 in $\sigma^k(x)$, that is, if we set $\sigma^k(x) = u_0 u_1 \dots u_{2^k-1}$, then $R_k(x,q)$ is characterized by:

$$\sum_{i=0}^{R_k(x,q)-1} u_i = q-1 \quad \text{and} \quad \sum_{i=0}^{R_k(x,q)} u_i = q.$$

Proposition 1.11. *Let $k>0$ be a positive integer and p be in $[\![0,k]\!]$. We have the following properties:*

(1) *For all integers* n, $0 \leq n \leq 2^k - 1$,

$$(\sigma^k(f-(k-2p)))_n = (\sigma^k(f+k-2p))_{2^k-n-1},$$

that is, the words $\sigma^k(f-(k-2p))$ *and* $\sigma^k(f+k-2p)$ *are mirror image one of each other. In particular,* $\sigma^{2k}(f)$ *is a palindrome different from* 0^{2^k} *if* k *is even.*

(2) *The letter* 1 *occurs in* $\sigma^k(f+k-2p)$ *exactly* $\binom{k}{p}$ *times.*

(3) *For all integer* p *in* $[\![0,k]\!]$ *and all integers* q *in* $[\![1, \binom{k}{p}]\!]$*:*

$$R_k(f+k-2p,q) = \begin{cases} R_{k-1}(f+k-2p-1,q) & \text{if } q \leq \binom{k-1}{p}, \\ 2^{k-1} + R_{k-1}(f+k-2p+1, q-\binom{k-1}{p}) & \text{otherwise.} \end{cases}$$

In particular, thanks to this formula, we can obtain, for $x = f+k-2p$*:*

- $R_k(x,1) = 2^p - 1$,
- $R_k(x,2) = 2^p + 2^{p-1} - 1$,
- $R_k(x, \binom{k}{p}) = 2^k - 2^{k-p}$,
- $R_k(x, \binom{k}{p} - 1) = 2^k - 2^{k-p} - 2^{k-p-1}$.

Proof. Let $k > 0$ be a positive integer and p be in $[\![0,k]\!]$.

(1) First, we must notice that $\overline{\sigma}^k(f) = \overline{u}_0 \dots \overline{u}_{2^k-1}$ satisfies:

$$\forall n \in [\![0, 2^k-1]\!], \forall b \in [\![-k, k]\!],\ \overline{u}_n = f + b \Longleftrightarrow \overline{u}_{2^k-1-n} = f - b.$$

It can be shown by induction on $k \geq 1$.

If we denote $\sigma^k(f) = u_0 \dots u_{2^k-1}$, we have directly:

$$\forall n \in [\![0, 2^k-1]\!],\ u_n = 1 \iff u_{2^k-1-n} = 1,$$

and so $\sigma^k(f)$ is a palindrome of length 2^k, different from 0^{2^k} when k is even.

To show that $\sigma(f-(k-2p))$ and $\sigma^k(f+k-2p)$ are mirror images, we denote

$$\overline{\sigma}^k(f+k-2p) = \overline{y}_0 \dots \overline{y}_{2^k-1}, \quad \sigma^k(f+k-2p) = y_0 \dots y_{2^k-1},$$

$$\overline{\sigma}^k(f-(k-2p)) = \overline{w}_0 \dots \overline{w}_{2^k-1}, \quad \sigma^k(f-(k-2p)) = w_0 \dots w_{2^k-1}.$$

As the coordinates applications of $\overline{\sigma}$: $\overline{\sigma}_0 : n \mapsto n-1$ and $\overline{\sigma}_1 : n \mapsto n+1$ are linear, we get: $\forall n \in [\![0, 2^k-1]\!]$, $\overline{y}_n = \overline{u}_n + (k-2p)$ and $\overline{w}_n = \overline{u}_n - (k-2p)$. Thus, we obtain the following equivalences for n in $[\![0, 2^k-1]\!]$:

$$y_n = 1 \iff \overline{y}_n = f \iff \overline{u}_n = f - (k-2p),$$

and, on the other hand,

$$w_{2^k-1-n} = 1 \iff \overline{w}_{2^k-1-n} = f \iff \overline{u}_{2^k-1-n} = f + (k-2p).$$

This is sufficient to show from those equivalences that $\sigma^k(f-(k-2p))$ and $\sigma^k(f+k-2p)$ are mirror images.

(2) For p an element of $\mathbb{N}$, we note $N_k(p)$ the number of occurrences of the letter 1 in $\sigma^k(f+k-2p)$. The set $\{N_k(p) \mid k \geq 1, p \in \mathbb{N}\}$ have the following properties:

- $\forall p \notin [\![0,k]\!],\ N_k(p) = 0$.
- As $\sigma^k(k) = 10^{2^k-1}$ and $\sigma^k(-k) = 0^{2^k-1}1$, we get

$$\forall k \geq 1,\ N_k(0) = N_k(k) = 1.$$

- The relation $\sigma^k(k-2p) = \sigma^{k-1}(k-1-2p)\sigma^{k-1}(k-1-2(p-1))$ implies

$$N_k(p) = N_{k-1}(p) + N^{k-1}(p-1).$$

Thus, the function $(k,p) \mapsto N_k(p)$ satisfies the same functional properties as the function $(k,p) \mapsto \binom{k}{p}$ so:

$$\forall k \geq 1,\ \forall p \in [\![0,k]\!],\ N_k(p) = \binom{k}{p}.$$

Notice that the construction in Figure 2, expounded at the beginning of the section, allows to see directly this property.

(3) From the relation $\sigma^k(f+k-2p) = \sigma^{k-1}(f+k-1-2p)\sigma^{k-1}(f+k-1-2(p-1))$ and the equality $N_k(p) = \binom{k}{p}$, we obtain:

$$\forall q \in [\![1, \tbinom{k-1}{p}]\!],\ R_k(f+k-2p,q) = R_{k-1}(f+k-1-2p,q),$$

as the first letters 1 of an iterate $\sigma^k(x)$ are the letters 1 of $\sigma^{k-1}(x-1)$, but we also obtain

$$\forall q \in [\![\tbinom{k-1}{p}+1, \tbinom{k}{p}]\!],\ R_k(f+k-2p,q) = 2^{k-1} + R_{k-1}(k-1-2(p-1), q-\tbinom{k-1}{p})$$

because the last letters 1 of an iterate $\sigma^k(x)$ are those of $\sigma^{k-1}(x+1)$.

The values of $R_k(f+k-2p,1)$ and $R_k(f+k-2p,2)$ for p in $[\![0,k]\!]$ can be easily computed by induction on k. We obtain the values of $R_k(f+k-2p, \binom{k}{p})$ and $R_k(f+k-2p, \binom{k}{p}-1)$ using the fact that $\sigma(f-(k-2p))$ and $\sigma^k(f+k-2p)$ are mirror images and thus $R_k(f+k-2p, \binom{k}{p}) = R_k(f-(k-2p),1)$ and $R_k(f+k-2p, \binom{k}{p}-1) = R_k(f-(k-2p),2)$. □

Lemma 1.12. *Let (x,q) be an element in $\mathbb{Z} \times \mathbb{N}$.*

The word $(x)(x-2q)$ is a subword of $\overline{m}$ and $(x-2)$ is the unique letter extending $(x)(x-2q)$ on the left to form a subword of m. That is

$$\forall (x,q) \in \mathbb{Z} \times \mathbb{N}, \forall n \in \mathbb{N},\quad \overline{m}_n\overline{m}_{n+1} = (x)(x-2q) \Longrightarrow \overline{m}_{n-1} = (x-2).$$

Proof. For all pairs (x,q) of $\mathbb{Z} \times \mathbb{N}$, according to Proposition 1.8, the word $(x)(x-2q)$ is a subword of $\overline{m}$. Furthermore, as any element of $\mathbb{Z}$ gives, a word of this type, by σ, the subword $(x)(x-2q)$ appears at a junction $\overline{\sigma}(x_1)\overline{\sigma}(x_2)$, with $x_1 = x-1$ and $x_2 = x-2q+1$. □

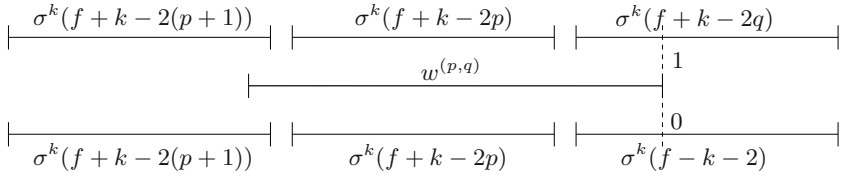

FIGURE 3. Construction of the special factor $w^{(p,q)}$.

Lemma 1.13. *Let k be a positive integer and n be an integer satisfying $2^k \leq n \leq 2^{k+1}-1$. Let (p,q) be a pair of integers such that $0 \leq p \leq q \leq k-1$.*

The subword $v = \sigma^k(f+k-2(p+1))\sigma^k(f+k-2p)\sigma^k(f+k-2q)$ contains a special factor of length n of $m^{(f)}$, denoted $w^{(p,q)}$. More precisely, if we set $v = v_0 v_1 \dots v_{3\cdot 2^k-1}$, we have:

$$w^{(p,q)} = y_{2^{k+1}+2^q-n-2} y_{2^{k+1}+2^q-n-1} \cdots y_{2^{k+1}+2^q-2}.$$

Moreover, if we set

$$\mathcal{W} = \{(1,q)|\ q \in [\![1, k-3]\!]\} \cup \{(p,q) \in \mathbb{N}^2|\ 2 \leq p \leq q \leq k-1\},$$

the function $(p,q) \mapsto w^{(p,q)}$ is one-to-one on $\mathcal{W}$ and every word $w^{(p,q)}$ with $(p,q) \in \mathcal{W}$ contains at least two occurrences of the letter 1.

Proof. Let k be a positive integer, n be an integer satisfying $2^k \leq n \leq 2^{k+1}-1$ and a pair (p,q) of integers such that $0 \leq p \leq q \leq k-1$. The word $w^{(p,q)}$ is extracted from the subword $\sigma^k(f+k-2(p+1))\sigma^k(f+k-2p)\sigma^k(f+k-2q)$ such that the last letter of $w^{(p,q)}$ is the $R_k(f+k-2q,1)$-th letter of $\sigma^k(f+k-2q)$ (see Note 1.10), that is the (2^q-1)-th letter of $\sigma^k(f+k-2q)$ (see Fig. 3).

The words $w^{(p,q)}$ are right special factors of $m^{(f)}$. Indeed, as $(f+k-2(p+1))(f+k-2p)(f+k-2q)$ and $(f+k-2(p+1))(f+k-2p)(f+k-2(q+1))$ are subwords of $\overline{m}$, the word $w^{(p,q)}$ can be extended on the right by the letter 1 if we see it as a subword of $\sigma^k(f+k-2(p+1))\sigma^k(f+k-2p)\sigma^k(f+k-2q)$ or by the letter 0 if we see it as a subword of $\sigma^k(f+k-2(p+1))\sigma^k(f+k-2p)\sigma^k(f+k-2(q+1))$, because the first occurrence of the letter 1 in $\sigma^k(f+k-2(q+1))$ only appears at the 2^{q+1} letter (rank $2^{q+1}-1$).

Nevertheless, we have to ensure that those special factors are different, and this is unfortunately not the case. But if we assume that p belongs to $[\![1, k-1]\!]$ (this ensures that $\sigma^k(f+k-2p)$ contains more than two occurrences of the letter 1) and $w^{(p,q)}$ contains the last two occurrences of 1 of $\sigma^k(f+k-2p)$, all the words $w^{(p,q)}$ are different.

Indeed, in this configuration, the two last occurrences of 1 of $w^{(p,q)}$ are the last two occurrences of 1 of $\sigma^k(f+k-2p)$. The length of the block of letters 0 between the last two occurrences of 1 of $w^{(p,q)}$ uniquely determines the integer p, because the length of this block is $R_k(f+k-2p, \binom{k}{p}) - R_k(f+k-2p, \binom{k}{p}-1) - 1$, that

is 2^{k-p-1}. Then, the integer q is uniquely determined by the length of the block of letters 0 ending the word $w^{(p,q)}$, because the length of this block is $R_k(f+k-2q,1)-1+2^k-R_k(f+k-2p,\binom{k}{p})$, that is $2^{k-p}+2^q-2$.

According to property 3. of Proposition 1.11, $w^{(p,q)}$ contains the two last occurrences of 1 of $\sigma^k(f+k-2p)$ if and only if:

$$2^k+R_k(f+k-2q,1)-R_k(f+k-2p,\tbinom{k}{p}-1)\leq n.$$

As n belongs to $[\![2^k,2^{k+1}-1]\!]$, no matter if we loose a few special factors, we can restrict this condition to the following one:

$$2^k+R_k(f+k-2q,1)-R_k(f+k-2p,\tbinom{k}{p}-1)\leq 2^k,$$

that is, the word $w^{(p,q)}$ contains the two last occurrences of 1 of $\sigma^k(f+k-2p)$ when $(p,q)\in\mathcal{S}$. So, by construction, the function $(p,q)\mapsto w^{(p,q)}$ is injective on $\mathcal{S}$. □

Lemma 1.14. *Let k be a positive integer and n an integer satisfying $2^k\leq n\leq 2^{k+1}-1$. For an integer q of $[\![1,k-1]\!]$, the word $u^{(q)}=0^{n-2^{q-1}}10^{2^{q-1}-1}$ is a special factor of $m^{(f)}$.*

Proof. First, let q be an integer of $[\![1,k-2]\!]$. The word $u^{(q)}$ can be extracted from $\sigma^k(f+k+2)\sigma^k(f+k+4)\sigma^k(f+k-2q)$ so that the last letter of $u^{(q)}$ is $R_k(f+k-2q,2)$-th letter of $\sigma^k(f+k-2q)$, *i.e.*, the $(2^q+2^{q-1}-1)$-th letter of $\sigma^k(f+k-2q)$, just before the second occurrence of the letter 1 in $\sigma^k(f+k-2q)$ (see Fig. 4). In other words, if we set $\sigma^k(f+k+2)\sigma^k(f+k+4)\sigma^k(f+k-2q)=v_0v_1\dots v_{32^k-1}$, then

$$u^{(q)}=u_{2^{k+1}+2^q+2^{q-1}-n-2}u_{2^{k+1}+2^q+2^{q-1}-n-1}\dots u_{2^{k+1}+2^q+2^{q-1}-2}.$$

But the word $u^{(q)}$ can also be extracted from $\sigma^k(f+k+2)\sigma^k(f+k+4)\sigma^k(f+k-2(q+1))$ so that the last letter of $u^{(q)}$ is the $(R_k(f+k-2(q+1),1)+2^{q-1})$-th letter of $\sigma^k(f+k-2(q+1))$, that is the $(2^{q+1}+2^{q-1}-1)$-th letter of $\sigma^k(f+k-2(q+1))$. In other words, if we set $\sigma^k(f+k+2)\sigma^k(f+k+4)\sigma^k(f+k-2(q+1))=x_0x_1\dots x_{32^k-1}$, then

$$u^{(q)}=x_{2^{k+1}+2^{q+1}+2^{q-1}-n-2}x_{2^{k+1}+2^{q+1}+2^{q-1}-n-1}\dots x_{2^{k+1}+2^{q+1}+2^{q-1}-2}.$$

The words $u^{(q)}$ are right special factors. Indeed, as $(f+k+2)(f+k+4)(f+k-2q)$ and $(f+k+2)(f+k+4)(f+k-2(q+1))$ appear in $\overline{m}$, the word $u^{(q)}$ can be extended on the right by the letter 1 if we see it as a subword of $\sigma^k(f+k+2)\sigma^k(f+k+4)\sigma^k(f+k-2q)$ or by the letter 0 if we see it as a subword of $\sigma^k(f+k+2)\sigma^k(f+k+4)\sigma^k(f+k-2(q+1))$ as the first occurrence of 1 in $\sigma^k(f+k-2(q+1))$ only appears at the $2^{q+1}+2^q$th letter (rank $2^{q+1}+2^q-1$).

The restriction $q\leq k-2$ is needed for $u^{(q)}0$ being a subword of $\sigma^k(f+k+2)\sigma^k(f+k+4)\sigma^k(f+k-2(q+1))$, but the word $u^{(k-1)}=0^{n-2^{k-2}}10^{2^{k-2}-1}$ is also

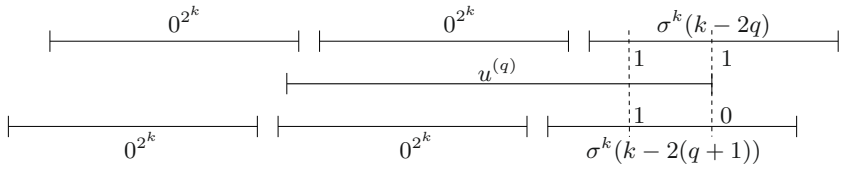

FIGURE 4. Construction of special factor $u^{(p)}$.

special. Using the same construction, the occurrence of the letter 1 of $u^{(k-1)}$ is the last letter and the unique 1 of $\sigma^k(f-k) = 0^{2^k-1}1$. On the other hand, the word $(f+k+2)(f-k)(f-k+2)$ is a subword of $\overline{m}$ and $\sigma^k(f+k+2)\sigma^k(f-k)\sigma^k(f-k+2)$ contains $u^{(k-1)}0$ as a subword because the length of the block of letters 0 between the first two occurrences of 1 in this subword is $2^{k-1}-1$. □

Proof of Claim 1.3. Let k be a positive integer and n be an integer satisfying $2^k \leq n \leq 2^{k+1}-1$.

Lemmas 1.13 and 1.14 provide a lower bound for $p_{m^{(f)}}(n+1) - p_{m^{(f)}}(n)$, for a fixed n. Indeed, this lemmas show up two different families of different special factors: words $w^{(p,q)}$ for $(p,q) \in \mathcal{S}$, which are all different and contains at least two occurrences of the letter 1 and words $u^{(q)}$, for $q \in [\![1, k-1]\!]$, which are all different and contain only one occurrence of the letter 1. Moreover, the subword 0^n is a right special factor too. Thereby, there is at least $2k-3+\sum_{i=1}^{k-2} i = (\sum_{i=1}^{k} i) - 2$ right special factors of length n, so

$$\forall k\, 0, \forall n \in [\![2^k, 2^{k+1}-1]\!],\ p_{m^{(f)}}(n+1) - p_{m^{(f)}}(n) \geq \frac{k(k+1)}{2} - 2.$$

Summing those inequalities, we get, for all integers $k > 0$ and all $n \in [\![2^k, 2^{k+1}-1]\!]$, $n \geq 3$:

$$p_{m^{(f)}}(n) - p_{m^{(f)}}(2) \geq (n-2^k)\left(\frac{k(k+1)}{2} - 2\right) + \sum_{i=0}^{k-1} 2^i \left(\frac{i(i+1)}{2} - 2\right).$$

As $p_{m^{(f)}}(2) = 4$, it becomes, for all integers $k > 0$ and all $n \in [\![2^k, 2^{k+1}-1]\!]$, $n \geq 3$:

$$p_{m^{(f)}}(n) \geq (n-2^k)\left(\frac{k(k+1)}{2} - 2\right) + (k+1)(k+2)2^{k-1} - 2^{k+1} - 2 + 4,$$

it follows,

$$p_{m^{(f)}}(n) \geq n\left(\frac{k(k+1)}{2} - 2\right) + (k+1)2^k + 2.$$

Using $k \geq \log_2 n - 1$ and $2^k \geq \frac{n}{2}$, we obtain:

$$\forall n \in \mathbb{N}, n \geq 3,\ p_{m^{(f)}}(n) \geq n\left(\frac{(\log_2 n + 1)\log_2 n}{2} - 2\right) + \frac{n \log_2 n}{2} + 2,$$

and it leads to

$$\forall n \in \mathbb{N}, n \geq 3,\ p_{m^{(f)}}(n) \geq n\left((\log_2 n)^2 - 2\right) + 2. \qquad \square$$

2. Open problems

We found an equivalent for the complexity functions of drunken man infinite words, but their exact computation remain an open problem. Another question, probably linked to this problem, concerns the languages $L(m^{(f)})$ of drunken man infinite words. Indeed, as mentioned in Remark 1.5, the words $m^{(0)}$, $m^{(1)}$ and $m^{(2)}$ have pairwise different languages even if they share many subwords and in particular special factors. One can ask whether this property is also true for any pair of words $m^{(f)}$ and $m^{(f')}$ or not.

As the drunken man infinite words are naturally associated with context-free languages over $\{0, 1\}$, a natural question is, as mentioned in the introduction, to characterize, in the spirit of Cobham's theorem [5], the possible growth orders of complexity functions of infinite words generated in the same way by transition graphs of deterministic pushdown automata. It can be proved from [10], that complexity functions of many of these words are at most polynomial but the upper bound given in the article can probably be improved.

In the other hand, methods displayed in this article can be adapted to more general automata, for example automata supported by lattices of $\mathbb{R}^d$, with similar output functions Π (constant except on a finite number of states). The way we have obtained upper bound for complexity functions can easily be adapted to deterministic and complete automaton with uniformly bounded degree and with a finite number of output states and give thin upper bound for the complexity function (see [10]). The search of lower bounds for complexity functions of words generated by more general automata is also linked to problems of reachability, shortest paths and counting paths in infinite graphs.

References

[1] J.-P. Allouche, Sur la complexité des suites infinies. *Bull. Belg. Math. Soc. Simon Stevin* **1** (1994) 133–143.
[2] J.-P. Allouche and J. Shallit, *Automatic sequences. Theory, applications, generalizations.* Cambridge University Press (2003).
[3] J. Cassaigne, Special factors of sequences with linear subword complexity. In *Developments in language theory (Magdeburg, 1995)*, World Sci. Publishing (1996) 25–34.
[4] J. Cassaigne, Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997) 67–88.
[5] A. Cobham, Uniform-tag sequences. *Math. Syst. Theory* **6** (1972) 164–192.
[6] S. Ferenczi, Complexity of sequences and dynamical systems. *Discrete Math.* **206** (1999) 145–154.
[7] S. Ferenczi, Substitution dynamical systems on infinite alphabets. *Ann. Inst. Fourier* **56** (2006) 2315–2343.
[8] E. Fouvry and C. Mauduit, Sur les entiers dont la somme des chiffres est moyenne. *J. Number Theory* **114** (2005) 135–152.

[9] P. Kůrka, *Topological and symbolic dynamics*, *Cours spécialisés* Vol. 11, SMF (2003).
[10] M. Le Gonidec, Sur la complexité de mots infinis engendrés par des q-automates dénombrables. *Ann. Inst. Fourier* **56** (2006) 2463–2491.
[11] M. Le Gonidec, *Sur la complexité des mots q^∞-automatiques*. Ph.D. thesis, Université Aix-Marseille II (2006).
[12] C. Mauduit, Propriétés arithmétiques des substitutions et automates infinis. *Ann. Inst. Fourier* **56** (2006) 2525–2549.
[13] C. Mauduit and A. Sárközy, On the arithmetic structure of sets characterized by sum of digits properties. *J. Number Theory* **61** (1996) 25–38.
[14] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés. *Lecture Notes Comput. Sci.* **172** (1985) 380–389. Automata, languages and programming (Antwerp, 1984).
[15] L. Staiger, Kolmogorov complexity of infinite words. *CDMTCS Research Report Series* **279** (2006).