#A7 INTEGERS 11B (2011)

SUBWORD COMPLEXITY AND LAURENT SERIES

Alina Firicel
Institut Camille Jordan, Université Lyon 1, Villeurbanne, France
firicel@math.univ-lyon1.fr

Received: 10/18/10 , Revised: 4/4/11, Accepted: 7/14/11, Published: 12/2/11

Abstract

Decimal expansions of classical constants such as $\sqrt{2}$, π and $\zeta(3)$ have long been a source of difficult questions. In the case of Laurent series with coefficients in a finite field, where "no carries appear," the situation seems to be simplified and drastically different. In 1935 Carlitz introduced analogs of real numbers such as π, e or $\zeta(3)$ and it became reasonable to enquire how "complex" the Laurent representation of these "numbers" is. In this paper we prove that the inverse of Carlitz's analog of π, Π_q, has, in general, linear subword complexity, except in the case $q = 2$, when the complexity is quadratic. In particular, this gives a new proof of the transcendence of Π_2 over $\mathbb{F}_2(T)$. In the second part, we consider the classes of Laurent series of at most polynomial complexity and of zero entropy. We show that these satisfy some nice closure properties.

1. Introduction and Motivations

The sequence of digits of the real number $\pi = 3.14159\cdots$ has baffled mathematicians for a long time. Though the decimal expansion of π has been calculated to billions of digits, we do not even know, for example, if the digit 1 appears infinitely often. Actually, it is expected that, for any $b \geq 2$, the b-ary expansion of π should share some of the properties of a random sequence (see, for instance, [10]). More concretely, it is widely believed that π is normal, meaning that all blocks of digits of equal length occur in the b-ary representation of π with the same frequency, but we are very far from having a proof of this claim. A common way to describe the disorder of an infinite sequence $\mathbf{a}$ is to compute its subword complexity, which is the function that associates with each positive integer m the number of distinct blocks of length m occurring in the sequence $\mathbf{a}$. Let α be a positive real number and let $a_{-k}a_{-k+1}\cdots a_0.a_1a_2\cdots$ be the representation of α in an integer base $b \geq 2$. The complexity function of α is defined as follows

$$p(\alpha, b, m) = \mathrm{Card}\{(a_j, a_{j+1}, \ldots, a_{j+m-1}),\ j \in \mathbb{N}^*\},$$

for any positive integer m. Here $\mathbb{N}$ (respectively $\mathbb{N}^*$) stands for the set $\{0, 1, 2, 3, \ldots\}$ (resp. $\{1, 2, 3, \ldots\}$). Notice that if π were normal, then its complexity would be maximal, that is $p(\pi, b, m) = b^m$, for every $b \geq 2$ and $m \geq 1$. In this direction, similar questions have been asked about other well-known constants and it is for instance widely believed that for every $\alpha \in \{e, \log 2, \zeta(3), \sqrt{2}\}$, one should have

$$p(\alpha, b, m) = b^m,$$

for any $m \geq 1$ and $b \geq 2$.

In this paper we use the familiar asymptotic notation of Landau. We write $f = O(g)$ if there exist two positive real numbers k and n_0 such that, for every $n \geq n_0$ we have $|f(n)| < k\,|g(n)|$. We also write $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$.

If α is a rational real number then for every integer $b \geq 2$ we have $p(\alpha, b, m) = O(1)$, where the constant depends on b and α. Moreover, there is a classical theorem of Morse and Hedlund [31] which implies that for every irrational real number α, we have

$$p(\alpha, b, m) \geq m + 1,$$

for all integers $m \geq 1$ and $b \geq 2$.

Concerning irrational algebraic numbers, the main result known to date in this direction is due to Adamczewski and Bugeaud [3]. These authors proved that the complexity of an irrational algebraic real number α satisfies

$$\lim_{m \to \infty} \frac{p(\alpha, b, m)}{m} = +\infty, \tag{1}$$

for any base $b \geq 2$. For more details about complexity of real numbers, see [1, 3, 4].

The present paper is motivated by this type of question, but in the setting of Laurent series with coefficients in a finite field. In the sequel we let $\mathbb{F}_q(T)$, $\mathbb{F}_q[[T^{-1}]]$ and $\mathbb{F}_q((T^{-1}))$ denote respectively the field of rational functions, the ring of formal power series and the field of formal Laurent power series over the finite field $\mathbb{F}_q$, q being a power of a prime number p.

By analogy with the real numbers, the complexity of a Laurent series is defined as the subword complexity of its sequence of coefficients. Again, the theorem of Morse and Hedlund gives a complete description of the rational Laurent series; more precisely, they are the Laurent series of bounded complexity. Furthermore, there is a remarkable theorem of Christol [19] (see also [21]) that precisely describes the algebraic Laurent series over $\mathbb{F}_q(T)$ as follows. Let $f(T) = \sum_{n \geq -n_0} a_n T^{-n}$ be a Laurent series with coefficients in $\mathbb{F}_q$. Then f is algebraic over $\mathbb{F}_q(T)$ if, and only if, the sequence of coefficients $(a_n)_{n \geq 0}$ is p-automatic. More references on automatic sequences can be found in [8]. Furthermore, Cobham proved that the subword complexity of an automatic sequence is at most linear [23]. Hence, a straightforward consequence of those two results is the following.

Theorem 1. *Let $f \in \mathbb{F}_q((T^{-1}))$ be algebraic over $\mathbb{F}_q(T)$. Then we have*

$$p(f, m) = O(m).$$

The converse is obviously not true, since there are uncountably many Laurent series with linear complexity.[1]

In contrast with real numbers, the situation is thus clarified in the case of algebraic Laurent series. Also, notice that (1) and Theorem 1 point out the fact that the subword complexities of algebraic elements in $\mathbb{F}_q((T^{-1}))$ and in $\mathbb{R}$ are quite different.

On the other hand, Carlitz introduced in [17] functions in positive characteristic by analogy with the Riemann ζ function, the usual exponential and the logarithm function. Many values of these functions, including an analog of the real number π, were shown to be transcendental over $\mathbb{F}_q(T)$ (see [24, 29, 35, 36, 37]). In the first part of this paper we focus on the analog of π, denoted, for each q, by Π_q, and we prove that its inverse has a "low" complexity. More precisely, we will prove the following result in Section 3.

Theorem 2. *Let q be a power of a prime number p. The complexity of the inverse of Π_q satisfies*

(a) *if $q = 2$ then*

$$p\left(\frac{1}{\Pi_2}, m\right) = \Theta(m^2);$$

(b) *if $q \geq 3$ then*

$$p\left(\frac{1}{\Pi_q}, m\right) = \Theta(m).$$

Since any algebraic series has a linear complexity (by Theorem 1), the following corollary yields.

Corollary 3. *Π_2 is transcendental over $\mathbb{F}_2(T)$.*

The transcendence of Π_q over $\mathbb{F}_q(T)$ was first proved by Wade in 1941 (see [36]) using an analog of a classical method of transcendence in zero characteristic. Another proof was given by Yu in 1991 (see [37]), using the theory of Drinfeld modules. De Mathan and Chérif, in 1993 (see [24]), using tools from Diophantine approximation, proved a more general result, but in particular their result implied the transcendence of Π_q. Christol's theorem has also been used as a combinatorial criterion in order to prove the transcendence of Π_q. This is what is usually called an "automatic proof." The non-automaticity, and also the transcendence, were first

[1] Consider for example the uncountable set of Laurent series of the form $f(T) = \sum_{n\geq 0} a_n T^{-n!}$, where $(a_n)_{n\geq 0}$ is an arbitrary sequence with values in $\{0, 1\}$. Furthermore, it follows for instance from [28] that f has sublinear complexity since its sequence of coefficients is a lacunary sequence.

obtained by Allouche, in [6], via the so-called q-kernel. Notice that our proof here is based also on Christol's theorem, but we obtain the non-automaticity of Π_2 over $\mathbb{F}_2(T)$ as a consequence of the subword complexity.

Another motivation for this work comes from a paper of Beals and Thakur [11]. These authors proposed a classification of Laurent series by their space or time complexity. They showed that some classes of Laurent series have good algebraic properties (for instance, the class of Laurent series corresponding to any deterministic space class at least linear form a field). The authors also place some of Carlitz's analogs in this computational hierarchy. Furthermore, motivated by Theorems 1 and 2, we consider the classes of Laurent series of at most polynomial complexity $\mathcal{P}$ and of zero entropy $\mathcal{Z}$ (see Section 4), which seem to be good candidates to enjoy some nice closure properties. In particular, we prove the following theorem.

Theorem 4. *$\mathcal{P}$ and $\mathcal{Z}$ are vector spaces over $\mathbb{F}_q(T)$.*

We will also show that both classes are closed under some usual operations such as the Hadamard product, the formal derivative and the Cartier operators. In particular, Theorem 4 provides a criterion of linear independence over $\mathbb{F}_q(T)$ (see Proposition 28).

This paper is organized as follows. Some definitions and basic notions from combinatorics on words are recalled in Section 2. Section 3 is devoted to the study of the Carlitz's analog of π; we prove Theorem 2. In Section 4 we study some closure properties of Laurent series of "low" complexity and we prove Theorem 4. Finally, we conclude in Section 5 with some remarks concerning the complexity of the Cauchy product of two Laurent series, which seems to be a more difficult problem.

2. Terminology and Basic Notions

In this section, we briefly recall some definitions and well-known results from combinatorics on words.

A *word* is a finite, as well as infinite, sequence of symbols (or letters) belonging to a nonempty finite set $\mathcal{A}$, called *alphabet*. We usually denote words by juxtaposition of their symbols.

Given an alphabet $\mathcal{A}$, we let $\mathcal{A}^* := \cup_{k=0}^{\infty}\mathcal{A}^k$ denote the set of finite words over $\mathcal{A}$. Let $V := a_0a_1\cdots a_{m-1} \in \mathcal{A}^*$. Then the integer m is the length of V and is denoted by $|V|$. The word of length 0 is the empty word, usually denoted by ε. We also let $\mathcal{A}^m$ denote the set of all finite words of length m and by $\mathcal{A}^{\mathbb{N}}$ the set of all infinite words over $\mathcal{A}$. We typically use the uppercase italic letters X, Y, Z, U, V, W to represent elements of $\mathcal{A}^*$. We also use bold lowercase letters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}$ to represent infinite words. The elements of $\mathcal{A}$ are usually denoted by lowercase letters $a, b, c, \cdots$.

We say that V is a *subword* (or *factor*) of a finite word U if there exist some finite words A, B, possibly empty, such that $U = AVB$, and we let $V \lhd U$ denote this property. Otherwise, $V \ntriangleleft U$. We say that X is a *prefix* of U, and we let $X \prec_p U$ denote this property, if there exists Y such that $U = XY$. We say that Y is a *suffix* of U, and we let $Y \prec_s U$ denote this property, if there exists X such that $U = XY$.

Also, we say that a finite word V is a subword (or factor) of an infinite word $\mathbf{a} = (a_n)_{n\geq 0}$ if there exists a nonnegative integer j such that $V = a_j a_{j+1} \cdots a_{j+m-1}$. The integer j is called an occurrence of V.

Let U, V, W be three finite words over $\mathcal{A}$, V possibly empty. We let

$$i(U, V, W) := \{AVB,\ A \prec_s U,\ B \prec_p W,\ A, B \text{ possibly empty}\},$$

and

$$i(U, V, W)^+ := \{AVB,\ A \prec_s U,\ B \prec_p W,\ A, B \text{ nonempty}\}.$$

If n is a nonnegative integer, we let U^n denote $\underbrace{UU\cdots U}_{n \text{ times}}$. We also let U^∞ denote $UU\cdots$, that is U concatenated (with itself) infinitely many times. An infinite word $\mathbf{a}$ is *periodic* if there exists a finite word V such that $\mathbf{a} = V^\infty$. An infinite word is *ultimately periodic* if there exist two finite words U and V such that $\mathbf{a} = UV^\infty$.

2.1. Subword Complexity and Topological Entropy

Let $\mathbf{a} = (a_n)_{n\geq 0}$ be an infinite word over $\mathcal{A}$. As already mentioned in the introduction, the *subword complexity* of $\mathbf{a}$ is the function that associates with each $m \in \mathbb{N}$ the number $p(\mathbf{a}, m)$ defined as follows

$$p(\mathbf{a}, m) = \mathrm{Card}\{(a_j, a_{j+1}, \ldots, a_{j+m-1}),\ j \in \mathbb{N}\}.$$

We now give an useful tool in order to obtain bounds on the subword complexity function (for a proof see, for example, [8], p. 300–302).

Lemma 5. *Let $\boldsymbol{a}$ be an infinite word over an alphabet $\mathcal{A}$. We have the following properties:*

- $p(\boldsymbol{a}, m) \leq p(\boldsymbol{a}, m+1) \leq \mathrm{card}\,\mathcal{A} \cdot p(\boldsymbol{a}, m)$, *for every integer* $m \geq 0$*;*
- $p(\boldsymbol{a}, m+n) \leq p(\boldsymbol{a}, m)p(\boldsymbol{a}, n)$, *for all integers* $m, n \geq 0$.

The *(topological) entropy* of $\mathbf{a}$ is defined as follows

$$h(\mathbf{a}) = \lim_{m\to\infty} \frac{\log p(\mathbf{a}, m)}{m}.$$

Notice that this definition makes sense because the function $\log p(\mathbf{a}, m)$ is subadditive.

Let $\mathbb{F}_q$ be the finite field with q elements, where q is a power of a prime number p. In this paper, we are interested in Laurent series with coefficients in $\mathbb{F}_q$. Let $n_0 \in \mathbb{N}$ and consider

$$f(T) = \sum_{n=-n_0}^{+\infty} a_n T^{-n} \in \mathbb{F}_q((T^{-1})).$$

Let m be a nonnegative integer. We define *the complexity* of f, denoted by $p(f, m)$, as being equal to the complexity of the infinite word $\mathbf{a} = (a_n)_{n\geq 0}$. We also define *the entropy* of f, denoted by $h(f)$, as being equal to the entropy of the infinite word $\mathbf{a} = (a_n)_{n\geq 0}$.

3. An Analog of Π

In 1935, Carlitz [17] introduced for function fields in positive characteristic an analog of the exponential function defined over $\mathcal{C}_\infty$, which is the completion of the algebraic closure of $\mathbb{F}_q((T^{-1}))$ (this is the natural analog of the field of complex numbers). In order to get good properties analogous to the complex exponential, the resulting analog, $z \to e_C(z)$, satisfies

$$e_C(0) = 0,\ d/dz(e_C(z)) = 1 \text{ and } e_C(Tz) = Te_C(z) + e_C(z)^q.$$

This is what is called the Carlitz exponential and the action $u \to Tu + u^q$ leads to the definition of the Carlitz $\mathbb{F}_q[T]$-module, which is in fact a particular case of a Drinfeld module. The Carlitz exponential, $e_C(z)$, can be defined by the following infinite product

$$e_C(z) = z \prod_{a\in\mathbb{F}_q[T],\, a\neq 0} (1 - \frac{z}{a\widetilde{\Pi}_q})$$

where

$$\widetilde{\Pi}_q = (-T)^{\frac{q}{q-1}} \prod_{j=1}^{\infty} \left(1 - \frac{1}{T^{q^j - 1}}\right)^{-1}.$$

Since $e^z = 1$ if, and only if, $z \in 2\pi i\mathbb{Z}$ and since $e_C(z)$ was constructed by analogy such that $e_C(z) = 0$ if, and only if, $z \in \widetilde{\Pi}_q\mathbb{F}_q[T]$ (in other words the kernel of $e_C(z)$ is $\widetilde{\Pi}_q\mathbb{F}_q[T]$), we get a good analog $\widetilde{\Pi}_q$ of $2\pi i$. Hence, it seems that a good analog of π may be considered

$$\Pi_q = \prod_{j=1}^{\infty} \left(1 - \frac{1}{T^{q^j - 1}}\right)^{-1}.$$

We mention that we use the notation that appears in [35] (page 32 and 47) and we consider the analog of π defined by the formula above (as explained in [35], page 48 and page 365). For more details about analogs given by the theory of Carlitz modules, and in particular about the exponential function or its fundamental period $\widetilde{\Pi}_q$, we refer the reader to the monographs [29] (pages 51, 362) and [35] (pages 32, 47, 365).

If we look for the Laurent series expansion of Π_q, we obtain that

$$\Pi_q = \prod_{j=1}^{\infty}\left(1 - \frac{1}{T^{q^j-1}}\right)^{-1} = \sum_{n\geq 0} a_n T^{-n},$$

where a_n is defined as the number of partitions of n whose parts take values in $I = \{q^j - 1, j \geq 1\}$, taken modulo p. To compute the complexity of Π_q, we would like to find a closed formula or some recurrence relations for the sequence of partitions $(a_n)_{n\geq 0}$. This question seems quite difficult and, unfortunately, we are not able to solve it at this moment.

However, it was shown in [6] that the inverse of Π_q has the following simple Laurent series expansion

$$\frac{1}{\Pi_q} = \prod_{j=1}^{\infty}\left(1 - \frac{1}{T^{q^j-1}}\right) = \sum_{n=0}^{\infty} p_n T^{-n}$$

where the sequence $\mathfrak{p}_q = (p(n))_{n\geq 0}$ is defined as follows

$$p_n = \begin{cases} 1 \text{ if } n = 0; \\ (-1)^{\operatorname{card} J} \text{ if there exists a nonempty set } J \subset \mathbb{N}^* \text{ such that } n = \sum_{j\in J}(q^j - 1); \\ 0 \text{ if there is no set } J \subset \mathbb{N}^* \text{ such that } n = \sum_{j\in J}(q^j-1). \end{cases} \tag{2}$$

Remark 6. In [16], the authors show, using a well-known result of Fraenkel [27], that a nonnegative integer n can be written of the form $n = \sum_{i\geq 0} a_i(q^i - 1)$, where $a_i \in \{0, 1, \ldots, k-1\}$, if, and only if, n is a multiple of $q - 1$. If it is the case, then n has a unique representation of this form.

3.1. Proof of Part (a) of Theorem 2

In this subsection we study the sequence $\mathfrak{p}_2 = (p_n^{(2)})_{n\geq 0}$, defined by the formula (2) in the case where $q = 2$. More precisely:

$$p_n^{(2)} = \begin{cases} 1 \text{ if } n = 0 \text{ or if there exists } J \subset \mathbb{N}^* \text{ such that } n = \sum_{j\in J}(2^j - 1); \\ 0 \text{ otherwise.} \end{cases} \tag{3}$$

In order to simplify the notation, in the rest of this subsection we set $p_n := p_n^{(2)}$ so that $\mathfrak{p}_2 = p_0 p_1 p_2 \cdots$. For every $n \geq 1$, we let W_n denote the factor of $\mathfrak{p}_2$ that occurs between positions $2^n - 1$ and $2^{n+1} - 2$, that is:

$$W_n := p_{2^n-1} \cdots p_{2^{n+1}-2}.$$

We also set $W_0 := 1$. Observe that $|W_n| = 2^n$. With this notation the infinite word $\mathfrak{p}_2$ can be factorized as:

$$\mathfrak{p}_2 = \underbrace{1}_{W_0}\underbrace{10}_{W_1}\underbrace{1100}_{W_2}\underbrace{11011000}_{W_3}\cdots = W_0W_1W_2\cdots .$$

Remark 7. The sequence $\mathfrak{p}_2$ is related to von Neumann's sequence (see [7], Theorem 2). In [7], the authors proved that $\mathfrak{p}_2$ is the infinite fixed point beginning with 1 of the morphism σ defined by $\sigma(1) = 110$ and $\sigma(0) = 0$.

There is a classical theorem of Pansiot that describes the asymptotic behavior of the subword complexity of pure morphic sequences in function of the order of growth of letters (see [32]). Note that Cassaigne and Nicolas [18] recently gave a very clear and detailed exposition of the proof of this theorem. An important step towards establishing Pansiot's theorem is the following result which corresponds to Theorem 4.7.66 in [18].

Theorem 8. *Let $\boldsymbol{a} \in \mathcal{A}^{\mathbb{N}}$ be a purely morphic sequence and let σ be the morphism that generates $\boldsymbol{a}$. If $\boldsymbol{a}$ is not ultimately periodic and if infinitely many distinct factors of $\boldsymbol{a}$ are bounded under σ, then $p(\boldsymbol{a}, m) = \Theta(m^2)$.*

We recall that a word $U \in A^*$ is said to be bounded under a morphism σ if $|\sigma^n(U)|$ remains bounded when n tends to ∞. Following Example 4.7.67 of [18], one can use Theorem 8 to easily deduce that $p(\mathfrak{p}_2, m) = \Theta(m^2)$. Indeed, the infinite sequence $\mathfrak{p}_2$ is not ultimately periodic since the word 10^n1 is a factor of $\mathfrak{p}_2$ for every $n \in \mathbb{N}$. Furthermore, for every positive integer n, the word 0^n is a factor of $\mathbf{a}$ that is bounded under σ.

In an unpublished note [5], Allouche showed the weaker result that the complexity of the sequence $\mathfrak{p}_2$ satisfies

$$p(\mathfrak{p}_2, m) \geq Cm \log m,$$

for some positive constant C and every positive integer m. The elementary proof given by the author is based on the morphism σ but does not use Pansiot's theorem.

We provide below an elementary proof in which we use neither Pansiot's theorem nor the morphism σ. One interest for such a proof is that it leads to a more precise result (the hidden constants in Theorem 8 are explicitly given). Furthermore, the approach we use for the case $q = 2$ naturally extends to the case $q > 2$.

In [7], the authors also showed that, for every $n \geq 2$, we have

$$W_n = 1W_1W_2 \cdots W_{n-1}0. \tag{4}$$

Since the subword W_n ends with 0, we can define U_n by $W_n := U_n0$, for every $n \geq 1$. Thus, $U_1 = 1$, $U_2 = 110$. For every $n \geq 1$, we have

$$U_{n+1} = U_nU_n0. \tag{5}$$

Lemma 9. *For every $n \geq 2$, there exists a word Z_n such that $W_n = 1Z_n10^n$ and $0^{n-1} \ntrianglelefteq Z_n$ (in other words W_n ends with exactly n zeros and Z_n does not contain blocks of 0 of length larger than $n-2$). This is equivalent to saying that $U_n = 1Z_n10^{n-1}$ and $0^{n-1} \ntrianglelefteq Z_n$.*

Proof. We argue by induction on n. For $n = 2$, $W_2 = 1100$ ends with two zeros and obviously there are no other zeros. We assume that W_n ends with n zeros and does not contain any other block of zeros of length greater than $n-2$. We show that this statement holds for $n+1$. By (5),

$$W_{n+1} = U_{n+1}0 = U_nU_n00.$$

As U_n ends exactly with $n-1$ zeros (by the induction hypothesis), then W_{n+1} also ends with $n+1$ zeros. Now, $U_n = 1Z_n10^{n-1}$ and $0^{n-1} \ntrianglelefteq Z_n$; so we have $W_{n+1} = 1Z_n10^{n-1}1Z_n10^{n-1}00 = 1Z_{n+1}10^{n+1}$, where $Z_{n+1} := Z_n10^{n-1}1Z_n$. Since $0^{n-1} \ntrianglelefteq Z_n$, then $0^n \ntrianglelefteq Z_{n+1}$. This completes the proof. □

Lemma 10. *For every $n \geq 1$, let $A_n := \{U_n^2 0^k,\ k \geq 1\}$. Then $\mathfrak{p}_2 \in A_n^{\mathbb{N}}$.*

Proof. Let $n \geq 1$. By definition of W_n and U_n and using the relation (4), the infinite word $\mathfrak{p}_2$ can be factorized as:

$$\mathfrak{p}_2 = \underbrace{1W_1W_2 \cdots W_{n-1}}_{U_n}\underbrace{U_n0}_{W_n}\underbrace{U_{n+1}0}_{W_{n+1}}\underbrace{U_{n+2}0}_{W_{n+2}} \cdots . \tag{6}$$

We prove that for every positive integer k, there exist a positive integer r and $k_1, k_2, \ldots, k_r \in \mathbb{N}^*$ such that:

$$U_{n+k} = U_n^2 0^{k_1} U_n^2 0^{k_2} \cdots U_n^2 0^{k_r}. \tag{7}$$

We argue by induction on k. For $k = 1$, we have $U_{n+1} = U_nU_n0 = U_n^2 0$. We suppose that the relation (7) is true for k and we show it for $k+1$. By (5)

$$U_{n+k+1} = U_{n+k}U_{n+k}0 = U_n^2 0^{k_1} U_n^2 0^{k_2} \cdots U_n^2 0^{k_r} U_n^2 0^{k_1} U_n^2 0^{k_2} \cdots U_n^2 0^{k_r+1}.$$

By Eq. (6), this ends the proof. □

Remark 11. In fact, in Eq. (7), one can easily see that $k_r = k$ and $k_i < k$, for $i < r$. Hence, we also have the following expression for W_{n+k}:

$$W_{n+k} = U_n^2 0^{k_1} U_n^2 0^{k_2} \cdots U_n^2 0^{k+1}.$$

Fix $m \in \mathbb{N}, m \geq 2$. Then, there is a unique integer n such that:

$$2^{n-1} < m \leq 2^n. \tag{8}$$

Lemma 12. *Let $m \in \mathbb{N}$. All distinct words of length m of $\mathfrak{p}_2$ occur in the prefix:*

$$P_m = W_0 W_1 \cdots W_m.$$

Proof. Let m and n be two positive integers satisfying (8).

We show that all distinct words of length m occur in the prefix

$$P_m = W_0 W_1 W_2 \cdots W_{n-1} W_n \cdots W_m = U_n \underbrace{U_n 0}_{W_n} \underbrace{U_n U_n 00}_{W_{n+1}} \underbrace{U_n U_n 0 U_n U_n 000}_{W_{n+2}} \cdots W_m;$$

the second identity following from (4) and (5).

Notice that we cannot choose a shorter prefix because, for instance, the word 0^m first occurs in W_m.

By Remark 11, W_i ends with $U_n U_n 0^{i-n+1}$, for every $i \geq n+1$ and there are no other block of zeros of length greater than $i - n + 1$. Hence, all the words $U_n U_n 0^k$, with $0 \leq k \leq m - n + 1$, are factors of P_m. More precisely, we can easily prove that

$$\text{If } B_n := \{U_n U_n 0^k, 0 \leq k \leq m - n + 1\}, \text{then } P_m \in B_n^*.$$

After the prefixe P_m, it is not possible to see new different subwords of length m. Indeed, suppose that there exists a word F of length m such that $F \triangleleft W_{m+1} W_{m+2} W_{m+3} \cdots$ and $F \not\triangleleft P_m$.

Since $1 + |U_n| = 2^n \geq m$ and using Lemma 10, F must occur in the words $U_n 0^k U_n$, with $k \geq m - n + 1$. But since U_n ends with $n - 1$ zeros (by Lemma 9), F must be equal to 0^m or $0^i R_i$, where $i \geq m - n + 2$ and $R_i \prec_p U_n$, or $F \triangleleft U_n$. But all these words already appear in P_m. This contradicts our assumption. $\square$

3.1.1. An Upper Bound for $p(\frac{1}{\Pi_2}, m)$

In this part we prove the following result:

$$p(\mathfrak{p}_2, m) \leq \frac{(m - \log_2 m)(m + \log_2 m + 2)}{2} + 2m. \tag{9}$$

In order to find all different factors of length m that occur in $\mathfrak{p}_2$, it suffices, by Lemmas 10 and 12, to consider factors appearing in the word $U_n U_n$ and in the sets

$i(U_n, 0^k, U_n)$, where $1 \le k \le m-n$. To these, we add the word 0^m, which cannot occur neither in U_nU_n nor $i(U_n, 0^k, U_n)$.

In the word U_nU_n we can find at most $|U_n|$ distinct words of length m. Since $|U_n| = 2^n - 1$ and $2^{n-1} < m \le 2^n$, the number of different factors of length m that occur in U_nU_n is at most $2m-1$.

Also, it is not difficult to see that $\left|i(U_n, 0^k, U_n)\right| \cap \mathcal{A}^m \le m-k+1$.

The number of distinct subwords occurring in all these sets, for $1 \le k \le m-n$, is less than or equal to:

$$\sum_{k=1}^{m-n} (m-k+1) = (m+1)(m-n) - \frac{(m-n)(m-n+1)}{2}.$$

Counting all these words and using the fact that $2^{n-1} < m \le 2^n$, we obtain that:

$$p(\mathfrak{p}_2, m) \le 2m + \frac{(m-n)(m+n+1)}{2} < \frac{(m-\log_2 m)(m+\log_2 m+2)}{2} + 2m$$

as claimed.

3.1.2. A Lower Bound for $p(\frac{1}{\Pi_2}, m)$

In this part we prove the following result:

$$p(\mathfrak{p}_2, m) \ge \frac{(m-\log_2 m)(m-\log_2 m+1)}{2}. \tag{10}$$

By Lemma 12, we have to look for distinct words of length m occuring in $W_nW_{n+1}\cdots W_m$.

In order to prove this proposition, we use the final blocks of 0 from each W_i. These blocks are increasing (as we have shown in Lemma 9). First, in the word W_m we find for the first time the word of length m: 0^m.

In the set $i(W_{m-1}, \varepsilon, W_m)$, we find two distinct words of length m that do not occur previously (10^{m-1} and $0^{m-1}1$) since there are no other words containing blocks of zeros of length $m-1$ in $i(W_k, \varepsilon, W_{k+1})$, for $k < m-1$.

More generally, fix k such that $n \le k \le m-2$. Since

$$W_kW_{k+1} = \underbrace{1Z_k10^k}_{W_k}\underbrace{1Z_{k+1}10^{k+1}}_{W_{k+1}},$$

in $i(W_k, \varepsilon, W_{k+1})$ we find $m-k+1$ words of length m of the form $\alpha_k 0^k \beta_k$. More precisely, the words we count here are the following $S_{m-k-1}10^k$, $S_{m-k-2}10^k1$, $S_{m-k-3}10^k1T_1$,..., $S_10^k1T_{m-k-2}$, 0^k1T_{m-k-1}, where $S_i \prec_s Z_k$ and $T_i \prec_p Z_{k+1}$, $|S_i| = |T_i| = i$, for every integer i, $1 \le i \le m-k-1$.

All these words do not occur before, that is in $i(W_s, \varepsilon, W_{s+1})$, for $s < k$, since there are no blocks of zeros of length k before the word W_k (according to Lemma

9). Also, in $i(W_s, \varepsilon, W_{s+1})$, for $s > k$, we focus on the words $\alpha_s 0^s \beta_s$ and hence they are different from all the words seen before (because $k < s$).

Consequently, the total number of subwords of length m of the form $\alpha_k 0^k \beta_k$ considered before, is equal to

$$1 + 2 + \ldots + (m - n + 1) = \frac{(m - n + 1)(m - n + 2)}{2}.$$

Since $2^{n-1} < m \leq 2^n$ we obtain the desired lower bound.

Proof of Part (a) of Theorem 2. Follows from inequalities (9) and (10). $\square$

We mention that a consequence of Theorem 2 and Theorem 1 is the following result on transcendence.

Corollary 13. *Let $\mathbb{K}$ be a finite field and let $(p_n^{(2)})_{n\geq 0}$ be the sequence defined in (3). Let us consider the associated formal series over $\mathbb{K}$:*

$$f(T) := \sum_{n\geq 0} p_n^{(2)} T^{-n} \in \mathbb{K}[[[T^{-1}]].$$

Then f is transcendental over $\mathbb{K}(T)$.

Notice that, if $\mathbb{K} = \mathbb{F}_2$ then the formal series f coincide with $1/\Pi_2$ and hence Corollary 13 implies Corollary 3.

3.2. Proof of Part (b) of Theorem 2

In this section we study the sequence $\mathfrak{p}_q = (p_n^{(q)})_{n\geq 0}$ defined by the formula (2) in the case where $q \geq 3$. In the following, we will consider the case $q = p^n$, where $p \geq 3$. We will discuss at the end the case $q = 2^n$.

Proposition 14. *Let $q \geq 3$. For every positive integer m, we have*

$$p(\mathfrak{p}_q, m) \leq (2q + 4)m + 2q - 5.$$

In particular, this proves Part (b) of Theorem 2. Indeed, we do not have to find a lower bound for the complexity function, as the sequence $\mathfrak{p}_q$ is not ultimately periodic (see Remark 19) and thus, by Morse and Hedlund's theorem we have

$$p(\mathfrak{p}_q, m) \geq m + 1,$$

for any $m \geq 0$.

In order to simplify the notation, we set in the sequel $p_n := p_n^{(q)}$ so that $\mathfrak{p}_q = p_0 p_1 p_2 \cdots$.

For every $n \geq 1$, we let W_n denote the factor of $\mathfrak{p}_q$ defined in the following manner

$$W_n := p_{q^n-1} \cdots p_{q^{n+1}-2}.$$

Let us fix $W_0 := 0^{q-2} = \underbrace{00\cdots 0}_{q-2}$ and $\alpha_0 := q-2$. Thus $W_0 = 0^{\alpha_0}$.

In other words, W_n is the factor of $\mathfrak{p}_q$ occurring between positions $q^n - 1$ and $q^{n+1} - 2$. Notice that $|W_n| = q^n(q-1)$.

With this notation the infinite word $\mathfrak{p}_q$ can be factorized as follows

$$\mathfrak{p}_q = 1\underbrace{00\cdots 0}_{W_0}\underbrace{(-1)00\cdots 0}_{W_1}\underbrace{(-1)\cdots 00100\cdots 0}_{W_2}(-1)00\cdots .$$

Now, we prove some lemmas that we use in order to bound from above the complexity function of $\mathfrak{p}_q$.

First, we can deduce from Remark 6 the following properties of $\mathfrak{p}_q$:

$$\text{for any } k,n \in \mathbb{N}^* \text{ such that } k \in \left[2(q^n-1), q^{n+1}-2\right], \text{ we have } p_k = 0; \tag{11}$$

$$\text{for any } k,n \in \mathbb{N}^* \text{ such that } k < q^n - 1, \text{ we have } p_k = -p_{k+(q^n-1)}. \tag{12}$$

For a word $W = a_1 a_2 \cdots a_l \in \{0,1,-1\}^l$ then set $\widehat{W} := (-a_1)(-a_2)\cdots(-a_l)$.

Lemma 15. *For every $n \geq 1$, we have the following*

$$W_n = (-1)\widehat{W_0}\widehat{W_1}\cdots\widehat{W_{n-1}}0^{\alpha_n}$$

with $\alpha_n = (q^{n+1}-1) - 2(q^n-1)$.

Proof. Obviously, the word W_n begins with -1 since $p_{q^n-1} = -1$. In order to prove the relation above, it suffices to split W_n into subwords as follows

$$W_n = \underbrace{p_{q^n-1}}_{-1}\underbrace{0\cdots 0}_{W'_0}\underbrace{p_{(q^n-1)+(q-1)}\cdots p_{(q^n-1)+(q^2-2)}}_{W'_1}\underbrace{p_{(q^n-1)+(q^2-1)}\cdots p_{(q^n-1)+(q^3-2)}}_{W'_2}$$

$$\cdots\underbrace{p_{(q^n-1)+(q^{n-1}-1)}\cdots p_{(q^n-1)+(q^n-2)}}_{W'_{n-1}}\underbrace{p_{2(q^n-1)}\cdots p_{q^{n+1}-2}}_{0^{\alpha_n}}.$$

Since $p_{(q^n-1)+k} = -p_k$, for every $k < q^n - 1$ (by (12)), we obtain that $W'_i = \widehat{W}_i$, for $0 \leq i \leq n-1$. The relation (11) ends the proof. $\square$

Since the subword W_n ends with 0^{α_n}, we can define U_n as prefix of W_n such that $W_n := U_n 0^{\alpha_n}$, for every $n \geq 1$. Notice that $|U_n| = q^n - 1$.

Lemma 16. *For every $n \geq 1$, we have $U_{n+1} = U_n \widehat{U_n} 0^{\alpha_n}$.*

Proof. By Lemma 15, $U_n = (-1)\widehat{W_0}\widehat{W_1} \cdots \widehat{W_{n-1}}$. Consequently:

$$U_{n+1} = \underbrace{(-1)\widehat{W_0}\widehat{W_1} \cdots \widehat{W_{n-1}}}_{U_n} \widehat{W_n} = U_n \widehat{U_n 0^{\alpha_n}} = U_n \widehat{U_n} 0^{\alpha_n}.$$

□

Remark 17. Since $q \geq 3$ we have $\alpha_n \geq |U_n|$ for every $n \geq 1$. Moreover $(\alpha_n)_{n \geq 1}$ is a positive and increasing sequence.

Lemma 18. *For every $n \geq 1$, let $A_n := \{U_n, \widehat{U_n}, 0^{\alpha_i},\ i \geq n\}$. Then $\mathfrak{p}_q \in A_n^{\mathbb{N}}$.*

Proof. Let $n \geq 1$. By definition of W_n and U_n, the infinite word $\mathfrak{p}_q$ can be factorized as:

$$\mathfrak{p}_q = \underbrace{1W_0 W_1 \cdots W_{n-1}}_{V_n} W_n W_{n+1} \cdots .$$

By Lemma 15, since $U_n = (-1)\widehat{W_0}\widehat{W_1} \cdots \widehat{W_{n-1}}$ we have that the prefix $V_n := 1W_0 W_1 \cdots W_{n-1} = \widehat{U_n}$.

Also, by Lemma 16 $W_{n+1} = U_n \widehat{U_n} 0^{\alpha_n} 0^{\alpha_{n+1}}$, $W_{n+2} = U_n \widehat{U_n} 0^{\alpha_n} \widehat{U_n} U_n 0^{\alpha_n + \alpha_{n+1} + \alpha_{n+2}}$. Iterating, W_n can be written as a concatenation of U_n, $\widehat{U_n}$ and 0^{α_i}, $i \geq n$. More precisely, $\mathfrak{p}_q$ can be written in the following manner:

$$\mathfrak{p}_q = \widehat{U_n} \underbrace{U_n 0^{\alpha_n}}_{W_n} \underbrace{U_n \widehat{U_n} 0^{\alpha_n + \alpha_{n+1}}}_{W_{n+1}} \underbrace{U_n \widehat{U_n} 0^{\alpha_n} \widehat{U_n} U_n 0^{\alpha_n + \alpha_{n+1} + \alpha_{n+2}}}_{W_{n+2}} \cdots .$$

□

Proof of Proposition 14. Let $m \in \mathbb{N}$. Then there exists a unique positive integer n, such that:

$$q^{n-1} - 1 \leq m < q^n - 1.$$

By Lemma 18 and Remark 17, between the words U_n and $\widehat{U_n}$ (when they do not occur consecutively), there are only blocks of zeros of length greater than $\alpha_n \geq |U_n| = q^n - 1$ and, thus, greater than m. Hence, all distinct factors of length m appear in the following words: $U_n \widehat{U_n}$, $\widehat{U_n} U_n$, $0^{\alpha_n} U_n$, $0^{\alpha_n} \widehat{U_n}$, $U_n 0^{\alpha_n}$ and $\widehat{U_n} 0^{\alpha_n}$.

In $U_n \widehat{U_n}$ we can find at most $|U_n \widehat{U_n}| - m + 1 = 2|U_n| - m + 1$ factors at length m. In $\widehat{U_n} U_n$ we can find at most $(m-1)$ new different factors of length m: they form the set $i(\widehat{U_n}, \varepsilon, U_n)^+$.

In $0^{\alpha_n} U_n$ (respectively $0^{\alpha_n} \widehat{U_n}$, $U_n 0^{\alpha_n}$, $\widehat{U_n} 0^{\alpha_n}$) we can find at most m (respectively $m-1$) new different factors (they belong to $i(0^{\alpha_n}, \varepsilon, U_n)^+ \cup \{0^m\}$, respectively $i(0^{\alpha_n}, \varepsilon, \widehat{U_n})^+$, $i(U_n, \varepsilon, 0^{\alpha_n})^+$ and $i(\widehat{U_n}, \varepsilon, 0^{\alpha_n})^+$).

Consequently, the number of such subwords is at most $2|U_n| + 4m - 3$. Since $U_n = q^n - 1 = q(q^{n-1} - 1) + q - 1 \leq qm + q - 1$ we obtain that:

$$p(\mathfrak{p}_q, m) \leq 2(qm + q - 1) + 4m - 3 \leq (2q + 4)m + 2q - 5.$$

□

Remark 19. It is not difficult to prove that $\mathfrak{p}_q$ is not ultimately periodic. Indeed, recall that $\mathfrak{p}_q = W_0 W_1 W_2 \cdots$. Using Lemma 15 and the Remark 17,

$$\mathfrak{p}_q = A_1 0^{l_1} A_2 0^{l_2} \cdots A_i 0^{l_i} \cdots, \tag{13}$$

where A_i, $i \geq 1$, are finite words such that $A_i \neq 0^{|A_i|}$ and $(l_i)_{i \geq 1}$ is a strictly increasing sequence.

Remark 20. This part concerns the case where $q \geq 3$. If the characteristic of the field is 2, that is, if $q = 2^n$, where $n \geq 2$, then, in the proof we have $-1 = 1$, but the structure of $\mathfrak{p}_q$ remains the same (hence, the proof follows as previously). We will certainly have a lower complexity, but $\mathfrak{p}_q$ is still of the form (13), and thus $m + 1 \leq p(\mathfrak{p}_q, m) \leq (2q + 4)m + 2q - 5$.

4. Closure Properties of Two Classes of Laurent Series

The subword complexity offers a natural way to classify Laurent series with coefficients in a finite field. In this section we study some closure properties of the following classes:

$$\mathcal{P} = \{f \in \mathbb{F}_q((T^{-1})), \text{there exists } K \text{ such that } p(f, m) = O(m^K)\},$$

and

$$\mathcal{Z} = \{f \in \mathbb{F}_q((T^{-1})), \text{ such that } \ h(f) = 0\}.$$

Clearly, $\mathcal{P} \subset \mathcal{Z}$. We recall that h denotes the topological entropy, as defined in Section 2. We have already seen, in Theorem 1, that all algebraic Laurent series belong to $\mathcal{P}$. Also, by Theorem 2, $1/\Pi_q$ belongs to $\mathcal{P}$. Hence, $\mathcal{P}$, and more generally $\mathcal{Z}$, seem to be two relevant sets for this classification.

The main result we will prove in this section is Theorem 4. We will also prove that $\mathcal{P}$ and $\mathcal{Z}$ are closed under usual operations such as the Hadamard product, the formal derivative and the Cartier operators.

4.1. Proof of Theorem 4

The proof of Theorem 4 is a straightforward consequence of Propositions 21 and 27 below.

Proposition 21. *Let f and g be two Laurent series belonging to $\mathbb{F}_q((T^{-1}))$. Then, for every integer $m \geq 1$, we have:*

$$\frac{p(f,m)}{p(g,m)} \leq p(f+g,m) \leq p(f,m)p(g,m).$$

Proof. Let $f(T) := \sum_{i \geq -i_1} a_i T^{-i}$ and $g(T) := \sum_{i \geq -i_2} b_i T^{-i}$, $i_1, i_2 \in \mathbb{N}$.

By definition of the complexity of Laurent series (see Section (2.1)), for every $m \in \mathbb{N}$:

$$p(f(T)+g(T),m) = p(\sum_{i \geq 0} c_i T^{-i}, m),$$

where $c_i := (a_i + b_i) \in \mathbb{F}_q$. Thus, we may suppose that

$$f(T) := \sum_{i \geq 0} a_i T^{-i} \text{ and } g(T) := \sum_{i \geq 0} b_i T^{-i}.$$

We let $\mathbf{a} := (a_i)_{i \geq 0}$, $\mathbf{b} := (b_i)_{i \geq 0}$ and $\mathbf{c} := (c_i)_{i \geq 0}$.

For the sake of simplicity, throughout this part, we set $x(m) := p(f,m)$ and $y(m) := p(g,m)$. Let $\mathcal{L}_{f,m} := \{U_1, U_2, \ldots, U_{x(m)}\}$ (resp. $\mathcal{L}_{g,m} := \{V_1, V_2, \ldots, V_{y(m)}\}$) be the set of different factors of length m of the sequence of coefficients of f (resp. of g). As the sequence of coefficients of the Laurent series $f+g$ is obtained by the termwise addition of the sequence of coefficients of f and the sequence of coefficients of g, we deduce that:

$$\mathcal{L}_{f+g,m} \subseteq \{U_i + V_j,\ 1 \leq i \leq x(m),\ 1 \leq j \leq y(m)\}$$

where $\mathcal{L}_{f+g,m}$ is the set of all distinct factors of length m occurring in $\mathbf{c}$, and where the sum of two words with the same length $A = a_1 \cdots a_m$ and $B = b_1 \cdots b_m$ is defined as

$$A + B = (a_1 + b_1) \cdots (a_m + b_m)$$

(each sum being considered over $\mathbb{F}_q$). Consequently, $p(f+g,m) \leq p(f,m)p(g,m)$.

We shall now prove the first inequality using Dirichlet's principle.

Notice that if $x(m) < y(m)$ the inequality is obvious.

Now assume that $x(m) \geq y(m)$. Notice that if we extract $x(m)$ subwords of length m from $\mathbf{b}$, there is at least one word which appears at least $\left\lceil \frac{x(m)}{y(m)} \right\rceil$ times.

For every fixed m, there exist exactly $x(m)$ different factors of $\mathbf{a}$. The subwords of $\mathbf{c}$ will be obtained adding factors of length m of $\mathbf{a}$ with factors of length m of $\mathbf{b}$.

Consider all distinct factors of length m of $\mathbf{a}$: $U_1, U_2, \ldots, U_{x(m)}$, that occur in positions $i_1, i_2, \ldots, i_{x_m}$. Looking at the same positions in $\mathbf{b}$, we have $x(m)$ factors of length m belonging to $\mathcal{L}_{g,m}$. Since $x(m) \geq y(m)$, by the previous remark there is one word W which occurs at least $\left\lceil \frac{x(m)}{y(m)} \right\rceil$ times in $\mathbf{b}$.

Since we have $U_i + W \neq U_j + W$ if $U_i \neq U_j$, the conclusion follows immediately. $\square$

Remark 22. In fact, the first inequality can also be easily obtained from the second one, but we chose here to give a more intuitive proof. Indeed, if we let $f := h_1 + h_2$, $g := -h_2$, where $h_1, h_2 \in \mathbb{F}_q((T^{-1}))$, the first relation follows immediately, since $p(h_2, m) = p(-h_2, m)$, for any $m \in \mathbb{N}$.

Remark 23. If $f \in \mathbb{F}_q((T^{-1}))$ and $A \in \mathbb{F}_q[T]$ then, obviously, there exists a constant C (depending on the degree of the polynomial A) such that, for any $m \in \mathbb{N}$,

$$p(f + A, m) \leq p(f, m) + C.$$

Remark 24. Related to Proposition 21, one can naturally ask if it is possible to attain the bounds in Proposition 21. By Remark 22, it suffices to show that this is possible for one inequality. In the sequel, we give an example of two Laurent series of linear complexity whose sum has quadratic complexity.

Let α and β be two irrational numbers such that 1, α and β are linearly independent over $\mathbb{Q}$. For any $i \in \{\alpha, \beta\}$ we consider the following rotations:

$$R_i : \mathbb{T}^1 \to \mathbb{T}^1 \qquad x \to \{x + i\},$$

where $\mathbb{T}^1$ is the circle $\mathbb{R}/\mathbb{Z}$, identified to the interval $[0, 1)$.

We can partition $\mathbb{T}^1$ in two intervals I_i^0 and I_i^1, delimited by 0 and $1 - i$. We let ν_i denote the coding function:

$$\nu_i(x) = \begin{cases} 0 & \text{if } x \in I_i^0; \\ 1 & \text{if } x \in I_i^1. \end{cases}$$

We define $\mathbf{a} := (a_n)_{n \geq 0}$ such that, for any $n \geq 0$,

$$a_n = \nu_\alpha(R_\alpha^n(0)) = \nu_\alpha(\{n\alpha\})$$

and $\mathbf{b} := (a_n)_{n \geq 0}$ such that, for any $n \geq 0$,

$$b_n = \nu_\beta(R_\beta^n(0)) = \nu_\beta(\{n\beta\}).$$

Let us consider $f(T) = \sum_{n \geq 0} a_n T^{-n}$ and $g(T) = \sum_{n \geq 0} b_n T^{-n}$ be two elements of $\mathbb{F}_3((T^{-1}))$. We will prove that, for any $m \in \mathbb{N}$, we have

$$p(f + g, m) = p(f, m)p(g, m). \tag{14}$$

We thus provide an example of two infinite words whose sum has a maximal complexity, in view of Proposition 21.

A sequence of the form $(\nu(R_\alpha^n(x)))_{n \geq 0}$ is a particular case of a rotation sequence. It is not difficult to see that the complexity of the sequence $\mathbf{a}$ satisfies $p(\mathbf{a}, m) = m+1$ for any $m \in \mathbb{N}$ and hence $\mathbf{a}$ is Sturmian. For a complete proof, the reader may consult the monograph [33], but also the original paper of Morse and Hedlund [31], where they prove that every Sturmian sequence is a rotation sequence.

Let $m \in \mathbb{N}$. Let

$$\mathcal{L}_{\mathbf{a},m} := \{U_1, U_2, \ldots, U_{m+1}\}$$

and

$$\mathcal{L}_{\mathbf{b},m} := \{V_1, V_2, \ldots, V_{m+1}\}$$

be the set of distinct factors of length m that occur in **a**, resp. in **b**.

In order to prove the relation (14), we show that

$$\mathcal{L}_{\mathbf{a}+\mathbf{b},m} = \{U_i + V_j, 1 \le i, j \le m+1\}. \tag{15}$$

Let $I := [0, 1)$. It is well-known (see, for example, Proposition 6.1.7 in [33]) that, using the definition of the sequence **a** and **b**, respectively, we can split I in $m+1$ intervals of positive length $J_1, J_2, \ldots, J_{m+1}$ (resp., $L_1, L_2, \ldots, L_{m+1}$) corresponding to $U_1, U_2, \ldots, U_{m+1}$ (resp., $V_1, V_2, \ldots, V_{m+1}$) such that:

$$\{n\alpha\} \in J_k \text{ if, and only if, } a_n a_{n+1} \cdots a_{n+m-1} = U_k$$

$$(\text{resp., } \{n\beta\} \in L_k \text{ if, and only if, } b_n b_{n+1} \cdots b_{n+m-1} = V_k).$$

In other words, $\{n\alpha\} \in J_k$ (resp. $\{n\beta\} \in L_k$) if, and only if, the factor U_k (resp., V_k) occurs in **a** (resp., **b**) at the position n.

Now we use the well-known Kronecker theorem (see for example [30]) which asserts that the sequence of fractional parts $(\{n\alpha\}, \{n\beta\})_{n\ge0}$ is dense in the square $[0, 1)^2$ since by assumption 1, α and β are linearly independent over $\mathbb{Q}$.

In particular, this implies that, for any pair $(i, j) \in \{0, 1, \ldots, m+1\}^2$, there exists a positive integer n such that $(\{n\alpha\}, \{n\beta\}) \in J_i \times L_k$. This is equivalently to saying that, for any pair of factors $(U_i, V_j) \in \mathcal{L}_{\mathbf{a},m} \times \mathcal{L}_{\mathbf{b},m}$, there exists n such that $U_i = a_n a_{n+1} \cdots a_{n+m-1}$ and $V_k = b_n b_{n+1} \cdots b_{n+m-1}$. This proves Equality (15) and more precisely, since we are in characteristic 3 (i.e., $1 + 1 = 2 \ne 0$), we have the following equality

$$\operatorname{Card} \mathcal{L}_{\mathbf{a}+\mathbf{b},m} = \operatorname{Card} \mathcal{L}_{\mathbf{a},m} \cdot \operatorname{Card} \mathcal{L}_{\mathbf{b},m} = (m+1)^2.$$

We mention that the idea of our construction here appears, briefly, in Theorem 7.6.6 in [8].

We point out the following consequence of Proposition 21.

Corollary 25. *Let $f_1, f_2, \ldots, f_l \in \mathbb{F}_q((T^{-1}))$. Then for every $m \in \mathbb{N}$ and for every integer $i \in [1; l]$ we have the following*

$$\frac{p(f_i, m)}{\prod_{j\ne i, 1\le j\le l} p(f_j, m)} \le p(f_1 + f_2 + \cdots + f_l, m) \le \prod_{1\le j\le l} p(f_j, m).$$

Notice that the bounds in these inequalities can be attained, just generalizing the construction above (choose l Sturmian sequences of irrational slopes $\alpha_1, \alpha_2, \ldots, \alpha_l$, such that $1, \alpha_1, \alpha_2, \ldots, \alpha_l$ are linearly independent over $\mathbb{Q}$).

We shall prove next that the sets $\mathcal{P}$ and $\mathcal{Z}$ are closed under multiplication by rational functions. Let us begin with a particular case, that is, the multiplication by a polynomial.

Proposition 26. *Let $b(T) \in \mathbb{F}_q[T]$ and $f(T) \in \mathbb{F}_q((T^{-1}))$. Then there is a positive constant M (depending only on $b(T)$), such that for all $m \in \mathbb{N}$:*

$$p(bf, m) \le M\, p(f, m).$$

Proof. Let

$$b(T) := b_0T^r + b_1T^{r-1} + \cdots + b_r \in \mathbb{F}_q[T]$$

and

$$f(T) := \sum_{i \ge -i_0} a_iT^{-i} \in \mathbb{F}_q((T^{-1})),\ i_0 \in \mathbb{N}.$$

Then

$$\begin{aligned} b(T)f(T) &= b(T)\left(\sum_{i=-i_0}^{-1} a_iT^{-i} + \sum_{i \ge 0} a_iT^{-i}\right) \\ &= b(T)\left(\sum_{i=-i_0}^{-1} a_iT^{-i}\right) + b(T)\left(\sum_{i \ge 0} a_iT^{-i}\right). \end{aligned} \tag{16}$$

Now, the product

$$\begin{aligned} b(T)(\sum_{i \ge 0} a_iT^{-i}) &= T^r(b_0 + b_1T^{-1} + \cdots + b_rT^{-r})(\sum_{i \ge 0} a_iT^{-i}) \\ &:= T^r(\sum_{j \ge 0} c_jT^{-j}) \end{aligned} \tag{17}$$

where the sequence $\mathbf{c} := (c_j)_{j \ge 0}$ is defined as follows

$$c_j = \begin{cases} b_0a_j + b_1a_{j-1} + \cdots + b_ja_0 \text{ if } j < r \\ b_0a_j + b_1a_{j-1} + \cdots + b_ra_{j-r} \text{ if } j \ge r. \end{cases}$$

According to the definition of complexity (see Section 2.1) and using (16) and (17), for every $m \in \mathbb{N}$, we have

$$p\left(b(t)f(T), m\right) = p\left(b(T)(\sum_{i \ge 0} a_iT^{-i}), m\right) = p\left((\sum_{j \ge r} c_jT^{-j}), m\right).$$

Our aim is to count the number of words of the form $c_j c_{j+1} \cdots c_{j+m-1}$, when $j \geq r$. By definition of $\mathbf{c}$, we notice that for $j \geq r$ these words depend only on $a_{j-r} a_{j-r+1} \cdots a_{j+m-1}$ and on $b_0, b_1, \cdots, b_r$, which are fixed. The number of words $a_{j-r} a_{j-r+1} \cdots a_{j+m-1}$ is exactly $p(f, m+r)$. By Lemma 5 we obtain

$$p(f, m+r) < p(f, r) p(f, m) = M p(f, m),$$

where $M = p(f, r)$. More precisely, we can bound M from above by q^r, since this is the number of all possible words of length r over an alphabet of q letters. □

Proposition 27. *Let $r(T) \in \mathbb{F}_q(T)$ and $f(T) = \sum_{n \geq -n_0} a_n T^{-n} \in \mathbb{F}_q((T^{-1}))$. Then for every $m \in \mathbb{N}$, there is a positive constant M, depending only on $r(T)$ and n_0, such that:*

$$p(rf, m) \leq M p(f, m).$$

Proof. Let $f(T) := \sum_{i \geq -i_0} a_i T^{-i} \in \mathbb{F}_q((T^{-1}))$, $i_0 \in \mathbb{N}$ and $m \in \mathbb{N}$. By Proposition 21, we have

$$p(r(T)f(T), m) \leq p\left(r(T)(\sum_{i=-i_0}^{-1} a_i T^{-i}), m\right) \cdot p\left(r(T)(\sum_{i \geq 0} a_i T^{-i}), m\right).$$

Proposition 26 implies that

$$p\left(r(T)(\sum_{i=-i_0}^{-1} a_i T^{-i}), m\right) \leq R$$

where R does not depend on m. Thus, we may assume that $f(T) = \sum_{i \geq 0} a_i T^{-i}$.

We divide the proof of Proposition 27 into five steps.

`Step 1.` Since $r(T) \in \mathbb{F}_q(T)$, the sequence of coefficients of r is ultimately periodic. Thus, there exist two positive integers $S, L \in \mathbb{N}^*$ and $p_1 \in \mathbb{F}_q[T]$ (with degree equal to $S-1$) and $p_2 \in \mathbb{F}_q[T]$ (with degree equal to $L-1$) such that

$$r(T) = \frac{P(T)}{Q(T)} = \frac{p_1(T)}{T^{S-1}} + \frac{p_2(T)}{T^{S+L-1}}(1 + T^{-L} + T^{-2L} + \cdots).$$

Hence

$$\begin{aligned} r(T)f(T) &= \underbrace{\frac{1}{T^{S-1}} p_1(T) f(T)}_{g(T)} + \underbrace{p_2(T) \frac{1}{T^{S+L-1}} f(T)(1 + T^{-L} + T^{-2L} \cdots)}_{h(T)} \\ &:= \sum_{n \geq 0} f_n T^{-n}. \end{aligned} \tag{18}$$

We let $\mathbf{d} = (d(n))_{n \geq 0}$ denote the sequence of coefficients of $g(T)$ and by $\mathbf{e} = (e_n)_{n \geq 0}$ the sequence of coefficients of $h(T)$. Clearly $\mathbf{f} := (f_n)_{n \geq 0}$ is such that $f_n = d_n + e_n$, for every $n \in \mathbb{N}$.

Fix $m \in \mathbb{N}$. Our aim is to bound from above $p(\mathbf{f}, m)$. First, assume that m is a multiple of L and set $m = kL$, where $k \in \mathbb{N}$.

In order to bound the complexity of $\mathbf{f}$, we will consider separately the sequences $\mathbf{e}$ and $\mathbf{d}$.

`Step 2.` We now study the sequence $\mathbf{e}$, defined in (18).

In order to describe the sequence $\mathbf{e}$, we shall first study the product

$$f(T)(1 + T^{-L} + T^{-2L} + \cdots) = (\sum_{i \geq 0} a_i T^{-i})(1 + T^{-L} + T^{-2L} + \cdots) := \sum_{j \geq 0} c_j T^{-j}.$$

Expanding this product, it is not difficult to see that $c_l = a_l$ if $l < L$ and $c_{kL+l} = a_l + a_{l+L} + \cdots + a_{kL+l}$, for $k \geq 1$ and $0 \leq l \leq L - 1$.

By definition of c_n, $n \in \mathbb{N}$, we can easily obtain

$$c_{n+L} - c_n = a_{n+L}.$$

Consequently, for all $s \in \mathbb{N}$, we have

$$c_{n+sL} - c_n = a_{n+sL} + a_{n+(s-1)L} + \cdots + a_{n+L}. \tag{19}$$

Our goal is now to study the subwords of $\mathbf{c}$ with length $m = kL$.

Let $j \geq 0$ and let $c_j c_{j+1} c_{j+2} \cdots c_{j+kL-1}$ be a finite factor of length $m = kL$. Using identity (19), we can split the factor above in k words of length L as follows

$$\begin{aligned} c_j c_{j+1} c_{j+2} \cdots c_{j+kL-1} = {} & \underbrace{c_j c_{j+1} \cdots c_{j+L-1}}_{D_1} \underbrace{c_{j+L} c_{j+L+1} \cdots c_{j+2L-1}}_{D_2} \cdots \\ & \cdots \underbrace{c_{j+(k-1)L} c_{j+(k-1)L+1} \cdots c_{j+kL-1}}_{D_k} \end{aligned}$$

where the words D_i, $2 \leq i \leq k$ depend only on D_1 and $\mathbf{a}$. More precisely, we have

$$\begin{aligned} D_2 &= (c_j + a_{j+L})(c_{j+1} + a_{j+L+1}) \cdots (c_{j+L-1} + a_{j+2L-1}) \\ &\vdots \\ D_k &= (c_j + a_{j+L} + \cdots + a_{j+(k-1)L})(c_{j+1} + a_{j+L+1} + \cdots + a_{j+(k-1)L+1}) \cdots \\ &\quad (c_{j+L-1} + a_{j+2L-1} + \cdots + a_{j+kL-1}). \end{aligned}$$

Consequently, the word $c_j c_{j+1} c_{j+2} \cdots c_{j+m-1}$ only depends on D_1, which is a factor of length L, determined by $r(T)$, and on the word $a_{j+L} \cdots a_{j+kL-1}$, factor of length $kL - L = m - L$ of $\mathbf{a}$.

Now, let us return to the sequence $\mathbf{e}$. We recall that

$$\sum_{n \geq 0} e_n T^{-n} = \frac{p_2(T)}{T^{S+L-1}} \sum_{j \geq 0} c_j T^{-j}. \tag{20}$$

Using a similar argument to the one used in the proof of Proposition 26 and using Identity (20), a factor of the form $e_j e_{j+1} \cdots e_{j+m-1}$, $j \in \mathbb{N}^*$, only depends on the coefficients of p_2, which are fixed, and on $c_{j-L+1} \cdots c_{j-1} c_j \cdots c_{j+m-1}$. Hence, the number of distinct factors of the form $e_j e_{j+1} \cdots e_{j+m-1}$ only depends on the number of distinct factors of the form $a_{j+1} a_{j+2} \cdots a_{j+(k-1)L}$ and on the number of factors of length L that occur in $\mathbf{c}$.

`Step 3.` We now describe the sequence $\mathbf{d}$, defined in (18).

Doing the same proof as for Proposition 26, we obtain that the number of words $d_j \cdots d_{j+m-1}$, when $j \in \mathbb{N}$, only depends on the coefficients of p_1, which are fixed, and on the number of distinct factors $a_{j-S+1} \cdots a_j \cdots a_{j+m-1}$.

`Step 4.` We now give an upper bound for the complexity of $\mathbf{f}$, when m is a multiple of L.

According to steps 2 and 3, the number of distinct factors of the form $f_j f_{j+1} \cdots f_{j+m-1}$, $j \in \mathbb{N}$, depends on the number of distinct factors of the form $a_{j-S+1} a_{j+2} \cdots a_{j+m-1}$ and on the number of factors of length L that occur in $\mathbf{c}$.

Consequently,

$$p(rf, m) \leq p(f, m+S-1) q^L,$$

and by Lemma 5

$$p(f, m+S-1) \leq p(f, m) p(f, S-1) \leq q^{S-1} p(f, m).$$

Finally,

$$p(rf, m) \leq q^{L+S-1} p(f, m).$$

`Step 5.` We now give an upper bound for the complexity of $\mathbf{f}$, when m is not a multiple of L.

In this case, let us suppose that $m = kL + l$, $1 \leq l \leq L-1$. Using Lemma 5 and according to Step 4:

$$\begin{aligned} p(rf, m) &= p(rf, kL+l) \leq p(rf, kL) p(rf, l) \leq p(rf, kL) p(rf, L-1) \\ &\leq q^{L-1} p(rf, kL) \leq q^{S+2L-2} p(f, m). \end{aligned}$$

□

As a straightforward consequence of Theorem 4, we give a criterion of linear independence over $\mathbb{F}_q(T)$ for two Laurent series in terms of theirs complexity.

Proposition 28. *Let $f, g \in \mathbb{F}_q((T^{-1}))$ be two irrational Laurent series such that:*

$$\lim_{m \to \infty} \frac{p(f, m)}{p(g, m)} = \infty.$$

Then f and g are linearly independent over the field $\mathbb{F}_q(T)$.

4.2. Other Closure Properties

In this section we prove that both classes $\mathcal{P}$ and $\mathcal{Z}$ are closed under various natural operations such as the Hadamard product, the formal derivative and the Cartier operators.

4.2.1. Hadamard Product

Let $f(T) := \sum_{n\geq -n_1} a_n T^{-n}$, $g(T) := \sum_{n\geq -n_2} b_n T^{-n}$ be two Laurent series in $\mathbb{F}_q((T^{-1}))$. The Hadamard product of f and g is defined as follows

$$f \odot g = \sum_{n\geq -\min(n_1,n_2)} a_n b_n T^{-n}.$$

As in the case of addition of two Laurent series (see Proposition 21) one can easily obtain the following.

Proposition 29. *Let f and g be two Laurent series belonging to $\mathbb{F}_q((T^{-1}))$. Then, for every $m \in \mathbb{N}$, we have*

$$\frac{p(f,m)}{p(g,m)} \leq p(f \odot g, m) \leq p(f,m)p(g,m).$$

The proof is similar to the one of Proposition 21. The details are left to the reader.

4.2.2. Formal Derivative

As an easy application of Proposition 29, we present here the following result. First, let us recall the definition of the formal derivative.

Definition 30. Let $n_0 \in \mathbb{N}$ and consider $f(T) = \sum_{n=-n_0}^{+\infty} a_n T^{-n} \in \mathbb{F}_q((T^{-1}))$. The formal derivative of f is defined as follows

$$f'(T) = \sum_{n=-n_0}^{+\infty} (-n \mod p) a_n T^{-n+1} \in \mathbb{F}_q((T^{-1})).$$

We now prove the following result.

Proposition 31. *Let $f(T) \in \mathbb{F}_q((T^{-1}))$ and k be a positive integer. If $f^{(k)}$ is the derivative of order k of f, then there exists a positive constant M, such that, for all $m \in \mathbb{N}$, we have*

$$p(f^{(k)}, m) \leq M\, p(f,m).$$

Proof. The derivative of order k of f is “almost” the Hadamard product of this series and a rational function. By definition of $p(f,m)$, we may suppose that $f(T) := \sum_{n\geq 0} a_n T^{-n} \in \mathbb{F}_q[[T^{-1}]]$. Then

$$f^{(k)}(T) = \sum_{n\geq k} ((-n)(-n-1)\cdots(-n-k+1)a_n)T^{-n-k} := T^{-k}\sum_{n\geq k} b_n a_n T^{-n},$$

where $b_n := (-n)(-n-1)\cdots(-n-k+1) \mod p$. Since $b_{n+p} = b_n$, the sequence $(b_n)_{n\geq 0}$ is periodic of period p. Hence, let $g(T)$ denote the series whose coefficients are precisely given by $(b_n)_{n\geq 0}$. Thus there exists a positive constant M such that:

$$p(g,m) \leq M.$$

By Proposition 29,

$$p(f^{(k)},m) \leq p(g,m)p(f,m) \leq Mp(f,m),$$

which completes the proof. □

4.2.3. Cartier's Operators

In the fields of positive characteristic, there are natural operators, called Cartier operators, that play an important role in many problems in arithmetic in positive characteristic [19, 20, 25, 34]. In particular, if we consider the field of Laurent series with coefficients in $\mathbb{F}_q$, we have the following definition.

Definition 32. Let $f(T) = \sum_{i\geq 0} a_i T^{-i} \in \mathbb{F}_q[[T^{-1}]]$ and r such that $0 \leq r < q$. The Cartier operator Λ_r is the linear transformation defined by

$$\Lambda_r(\sum_{i\geq 0} a_i T^{-i}) = \sum_{i\geq 0} a_{qi+r} T^{-i}.$$

The classes $\mathcal{P}$ and $\mathcal{Z}$ are closed under this operator. More precisely, we prove the following result.

Proposition 33. *Let $f(T) \in \mathbb{F}_q[[T^{-1}]]$ and $0 \leq r < q$. Then there is M such that, for every $m \in \mathbb{N}$ we have the following*

$$p(\Lambda_r(f),m) \leq qp(f,m)^q.$$

Proof. Let $\mathbf{a} := (a_n)_{n\geq 0}$ be the sequence of coefficients of f and $m \in \mathbb{N}$. In order to compute $p(\Lambda_r(f),m)$, we have to look at factors of the form

$$a_{qj+r}a_{qj+q+r}\cdots a_{qj+(m-1)q+r},$$

for all $j \in \mathbb{N}$. But these only depend on factors of the form

$$a_{qj+r}a_{qj+r+1}\cdots a_{qj+(m-1)q+r}.$$

Using Lemma 5, we obtain that:

$$p(\Lambda_r(f), m) \leq p(f, (m-1)q+1) \leq qp(f, m-1)^q \leq qp(f, m)^q.$$

□

5. Cauchy Product of Laurent Series

In the previous section, we proved that $\mathcal{P}$ and $\mathcal{Z}$ are vector spaces over $\mathbb{F}_q(T)$. This naturally raises the question whether or not these classes form a ring, i.e., whether they are closed under the usual Cauchy product. There are actually some particular cases of Laurent series with low complexity whose product still belongs to $\mathcal{P}$. In this section we discuss the case of automatic Laurent series. However, we are not able to prove whether $\mathcal{P}$ or $\mathcal{Z}$ are or not rings or fields.

5.1. Products of Automatic Laurent Series

A particular case of Laurent series stable by multiplication is the class of k-automatic series, k being a positive integer:

$$\mathrm{Aut}_k = \{f(T) = \sum_{n\geq 0} a_n T^{-n} \in \mathbb{F}_q((T^{-n})), \, \mathbf{a} = (a_n)_{n\geq 0} \text{ is } k\text{-automatic}\}.$$

Since any k-automatic sequence has at most a linear complexity, $\mathrm{Aut}_k \subset \mathcal{P}$. A theorem of Allouche and Shallit [9] states that the set Aut_k is a ring. In particular, this implies that, if f and g belong to Aut_k, then $p(fg, m) = O(m)$. Notice also that, in the case where k is a power of p, the characteristic of the field $\mathbb{F}_q((T^{-1}))$, the result follows from Christol's theorem.

Remark 34. However, we do not know whether or not this property is still true if we replace Aut_k by $\cup_{k\geq 2}\mathrm{Aut}_k$. More precisely, if we consider two Laurent series f and g, which are respectively k-automatic and l-automatic, k and l being multiplicatively independent, we do not know if the product fg still belongs to $\mathcal{P}$. In the sequel, we give a particular example of two such Laurent series for which we prove that their product is still in $\mathcal{P}$.

We now focus on the product of series of the form:

$$f(T) = \sum_{n\geq 0} T^{-d^n} \in \mathbb{F}_q((T^{-1})).$$

It is not difficult to prove that $p(f, m) = O(m)$. The reader may refer to [28] for more general results concerning the complexity of lacunary series. The fact that the complexity of f is linear is also implied by the fact that $f \in \mathrm{Aut}_d$. Notice also that f is transcendental over $\mathbb{F}_q(T)$ if q is not a power of d. This is an easy consequence of Christol's theorem and a theorem of Cobham [22].

In this section we will prove the following result.

Theorem 35. *Let d and e be two multiplicatively independent positive integers (that is $\frac{\log d}{\log e}$ is irrational) and let $f(T) = \sum_{n\geq 0} T^{-d^n}$ and $g(T) = \sum_{n\geq 0} T^{-e^n}$ be two Laurent series in $\mathbb{F}_q((T^{-1}))$. Then we have*

$$p(fg, m) = O(m^4).$$

Let $h(T) := f(T)g(T)$. Then $h(T) = \sum_{n\geq 0} a_n T^{-n}$ where the sequence $\mathbf{a} = (a_n)_{n\geq 0}$ is defined as follows

$$a_n := (\text{the number of pairs } (k,l) \in \mathbb{N}^2 \text{ that verify } n = d^k + e^l) \mod p.$$

The main clue of the proof is the following consequence of the theory of S-unit equations (see [2] for a proof).

Lemma 36. *Let d and e be two multiplicatively independent positive integers. There is a finite number of solutions $(k_1, k_2, l_1, l_2) \in \mathbb{N}^4$, $k_1 \neq k_2$, $l_1 \neq l_2$, that satisfy the equation:*

$$d^{k_1} + e^{l_1} = d^{k_2} + e^{l_2}.$$

Obviously, we have the following consequence concerning the sequence $\mathbf{a} = (a_n)_{n\geq 0}$:

Corollary 37. *There exists a positive integer N such that, for every $n \geq N$ we have $a_n \in \{0, 1\}$. Moreover, $a_n = 1$ if, and only if, there exists one unique pair $(k,l) \in \mathbb{N}^2$ such that $n = d^k + e^l$.*

We now prove Theorem 35. For the sake of simplicity, we consider $d = 2$ and $e = 3$, but the proof is exactly the same in the general case.

Proof. Let $\mathbf{b} := (b_n)_{n\geq 2}$ and $\mathbf{c} := (c_n)_{n\geq 2}$ be the sequences defined as follows

$$b_n = \begin{cases} 1 & \text{if there exists a pair } (k,l) \in \mathbb{N}^2 \text{ such that } n = 2^k + 3^l,\ 2^k > 3^l; \\ 0 & \text{otherwise,} \end{cases}$$

$$c_n = \begin{cases} 1 & \text{if there exists a pair } (k,l) \in \mathbb{N}^2 \text{ such that } n = 2^k + 3^l,\ 2^k < 3^l; \\ 0 & \text{otherwise.} \end{cases}$$

Let $h_1(T) := \sum_{n\geq 2} b_n T^{-n}$ and $h_2(T) := \sum_{n\geq 2} c_n T^{-n}$ be the series associated with $\mathbf{b}$ and $\mathbf{c}$, respectively. Using Corollary 37, there exists a polynomial $P \in \mathbb{F}_q[T]$, with degree less than N, such that h can be written as follows

$$h(T) = h_1(T) + h_2(T) + P(T).$$

By Remark 23, there is $C \in \mathbb{R}$ such that, for any $m \in \mathbb{N}$:

$$p(h, m) \leq p(h_1 + h_2, m) + C.$$

In the sequel, we will show that $p(h_1, m) = p(h_2, m) = O(m^2)$. Theorem 35 will then follow by Proposition 21.

We now study the subword complexity of the sequence of coefficients $\mathbf{b} := (b_n)_{n \geq 2}$. The proof is similar to the proof of Theorem 2. The complexity of the sequence $\mathbf{c}$ can be treated in essentially the same way as for $\mathbf{b}$.

`Step 1.` For all $n \geq 1$, we let W_n denote the factor of $\mathbf{b}$ that occurs between positions $2^n + 1$ and 2^{n+1}, that is

$$W_n := b_{2^n+3^0} b_{2^n+2} b_{2^n+3^1} \cdots b_{2^{n+1}}.$$

We also set $W_0 := 1$.

Observe that $|W_n| = 2^n$.

With this notation the infinite word $\mathbf{b}$ can be factorized as:

$$\mathbf{b} = \underbrace{1}_{W_0}\underbrace{10}_{W_1}\underbrace{1010}_{W_2}\underbrace{10100000}_{W_3} \cdots = W_0 W_1 W_2 \cdots . \tag{21}$$

`Step 2.` Let $n \geq 1$ and m_n be the greatest integer such that $2^n + 3^{m_n} \leq 2^{n+1}$. This is equivalent to saying that m_n is such that

$$2^n + 3^{m_n} < 2^{n+1} < 2^n + 3^{m_n+1}.$$

Notice also that $m_n = n\lfloor \log_3 2 \rfloor$.

With this notation we have (for $n \geq 5$)

$$W_n = 10 10^5 1 \cdots 10^{\alpha_i} \cdots 10^{\alpha_{m_n}} 10^{\beta_n},$$

where $\alpha_i = 2 \cdot 3^{i-1} - 1$, for $1 \leq i \leq m_n$, and $\beta_n = 2^n - 3^{m_n} \geq 0$.

Let U_n denote the prefix of W_n such that $W_n := U_n 0^{\beta_n}$.

Notice that $(m_n)_{n \geq 0}$ is an increasing sequence. Hence $(\alpha_{m_n})_{n \geq 0}$ is increasing. Consequently, $U_n \prec_p U_{n+1}$ and more generally, $U_n \prec_p W_i$, for every $i \geq n + 1$.

`Step 3.` Let $M \in \mathbb{N}$. Our aim is to give an upper bound for the number of distinct factors of length M occurring in $\mathbf{b}$. In order to do this, we will show that there exists an integer N such that all these factors occur either in

$$W_0 W_1 \cdots W_N$$

or

$$A_0 := \{Z \in \mathcal{A}^M;\ Z \text{ is of the form } 0^j P \text{ or } 0^i 10^j P,\ P \prec_p U_N,\ i, j \geq 0, \}.$$

Let $N = \lceil \log_2(M+1) \rceil + 3$. Doing a simple computation we obtain that $\alpha_{m_N} \geq M$. Notice also that, for any $i \geq N$ we have

$$\alpha_{m_i} \geq M.$$

This follows since $(\alpha_{m_n})_{n\geq 0}$ is an increasing sequence.

Let V be a factor of length M of $\mathbf{b}$. Suppose that V does not occur in the prefix $W_0W_1 \cdots W_N$. Then, by (21), V must occur in $W_N W_{N+1} \cdots$. Hence, V must appear in some W_i, for $i \geq N+1$, or in $\bigcup_{i\geq N} i(W_i, \varepsilon, W_{i+1})$.

Let us suppose that V occurs in $\bigcup_{i\geq N} i(W_i, \varepsilon, W_{i+1})$. Since W_i ends with $0^{\alpha_{m_i}}10^{\beta_i}$, with $\alpha_{m_i} \geq M$, and since W_{i+1} begins with U_N and $|U_N| = 3^{m_N}+1 \geq M$, we have

$$\mathcal{A}^M \cap (\bigcup_{i\geq N} i(W_i, \varepsilon, W_{i+1})) \subset A_0.$$

Hence, if V occurs in $\bigcup_{i\geq N} i(W_i, \varepsilon, W_{i+1})$ then $V \in A_0$.

Let us suppose that V occurs in some W_i, for $i \geq N+1$. By definition of W_i and α_i, for $i \geq N+1$ and by the fact that we have

$$W_i = 10 10^5 1 \cdots 10^{\alpha_{m_N}} 10^{\alpha_{m_N+1}} \cdots 10^{\alpha_{m_i}} 10^{\beta_i} = U_N 0^{\alpha_{m_N+1}} \cdots 10^{\alpha_{m_i}} 10^{\beta_i}.$$

By assumption, V does not occur in $W_0W_1 \cdots W_N$; hence V cannot occur in U_N which by definition is a prefix of W_N. Consequently, V must be of the form 0^r10^s, $r, s \geq 0$. Indeed, since $\alpha_{m_N} \geq M$, all blocks of zeros that follow after U_N (and before the last digit 1 in W_i) are all longer than M. But the words of the form 0^r10^s, $r, s \geq 0$ belong also to A_0.

Hence, we proved that if V does not occur in the prefix $W_0W_1 \cdots W_N$, then V belongs to A_0, as desired.

`Step 4.` In the previous step we showed that all distinct factors of length M occur in the prefix $W_0W_1 \cdots W_N$ or in the set A_0.

Since

$$|W_0W_1 \cdots W_N| = \sum_{i=0}^{N} 2^i = 2^{N+1} - 1$$

and since $N = \lceil \log_2(M+1) \rceil + 3$ we have

$$2^{N+1} - 1 \leq 2^{\log_2(M+1)+5} - 1 = 32M + 31,$$

and the number of distinct factors that occur in $W_0W_1 \cdots W_N$ is less than or equal to $32M + 31$.

Also, by an easy computation, we obtain that the cardinality of the set A_0 is

$$\mathrm{Card}\, A_0 = \frac{M^2}{2} + \frac{3M}{2}.$$

Finally, $p(\mathbf{b}, m) = p(h_1, m) = O(m^2)$. In the same manner, one could prove that $p(h_2, m) = O(m^2)$. This achieves the proof of Theorem 35, in view of Proposition 21. $\square$

5.2. A More Difficult Case

Set

$$\theta(T) := 1 + 2\sum_{n\geq 1} T^{-n^2} \in \mathbb{F}_q((T^{-1})),\ q \geq 3.$$

The function $\theta(T)$ is related to the classical Jacobi theta function. One can easily prove that:

$$p(\theta, m) = \Theta(m^2).$$

In particular this implies the transcendence of $\theta(T)$ over $\mathbb{F}_q(T)$, for any $q \geq 3$. Notice that this also implies the transcendence over $\mathbb{Q}(T)$ of the same Laurent series, but viewed as an element of $\mathbb{Q}((T^{-1}))$. Since $\theta(T) \in \mathcal{P}$, it would be interesting to know whether or not $\theta(T)^2$ belongs also to $\mathcal{P}$. Notice that

$$\theta(T)^2 = \sum_{n\geq 1} r_2(n) T^{-n}$$

where $r_2(n)$ is the number of representations of n as sum of two squares of integers mod p. In the rich bibliography concerning Jacobi theta function (see, for instance, [26, 30]), there is the following well-known formula

$$r_2(n) = 4(d_1(n) - d_3(n))$$

where $d_i(n)$ denotes the number of divisors of n congruent to i modulo 4, for $i \in \{1, 3\}$, and we can easily deduce that $r_2(n)$ is related to a multiplicative function of n (that is an arithmetic function f which satisfies the property that if m and n are coprime then $f(mn) = f(m)f(n)$). Recall that we would like to study the subword complexity of $r_2(n)_{n\geq 0}$, that is the number of distinct factors of the form $r_2(j)r_2(j+1)\cdots r_2(j+m-1)$, when $j \in \mathbb{N}$. Hence, it would be useful to describe some additive properties of $r_2(n)_{n\geq 0}$; for instance, it would be interesting to find some relations between $r_2(j+N)$ and $r_2(j)$, for some positive integers j and N. This seems to be a rather difficult question about which we are not able to say anything conclusive.

6. Conclusion

It would also be interesting to investigate the following general question: *is it true that Carlitz's analogs of classical constants all have a "low" complexity (i.e., polynomial or subexponential)?*

The first clue in this direction are the examples provided by Theorems 1 and 2. Notice also that a positive answer would reinforce the differences between $\mathbb{R}$ and $\mathbb{F}_q((T^{-1}))$ as hinted at in Section 1. When investigating these problems, we need, in general, the Laurent series expansions of such functions. In this context, one has to mention the work of Berthé [12, 13, 14, 15], where some Laurent series expansions of Carlitz's functions are described.

Acknowledgement I would like to express my gratitude to my advisor Boris Adamczewski for his very valuable comments, suggestions and corrections during the preparation of this article. I also thank Jean-Paul Allouche for sending me his unpublished note concerning the problem studied in Section 3, and Julien Cassaigne for suggesting the idea used in Remark 24.

References

[1] B. Adamczewski, On the expansion of some exponential periods in an integer base, *Math. Ann.* **346** (2010), 107–116.

[2] B. Adamczewski and J. Bell, Function fields in positive characteristic: expansions and Cobham's theorem, *J. Algebra* **319** (2008), 2337–2350.

[3] B. Adamczewski and Y. Bugeaud, On the complexity of algebraic numbers I. Expansions in integer bases, *Ann. of Math.* **165** (2007), 547–565.

[4] B. Adamczewski, Y. Bugeaud and F. Luca, Sur la complexité des nombres algébriques, *C. R. Math. Acad. Sci. Paris* **339** (2004), 11–14.

[5] J.-P. Allouche, private communication.

[6] J.-P. Allouche, Sur la transcendance de la série formelle Π, *J. Théor. Nombres Bordeaux* **2** (1990), 103–117.

[7] J.-P. Allouche, J. Bétréma and J. Shallit, Sur des points fixes de morphismes du monoide libre, *RAIRO Inform. Theor. App.* **23** (1989), 235–249.

[8] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.

[9] J.-P. Allouche and J. Shallit, The ring of k-regular sequences II, *Theoret. Comput. Sci.* **307** (2003), 3–29.

[10] D. H. Bailey, P. B. Borwein and S. Plouffe, On the rapid computation of various polylogarithmic constants, *Math. Comp.* **66** (1997), 903–913.

[11] R. M. Beals and D. S. Thakur, Computational classification of numbers and algebraic properties, *Int. Math. Res. Not.*, **15** (1998), 799–818.

[12] V. Berthé, Automates et valeurs de transcendance du logarithme de Carlitz, *Acta Arith. LXVI* **4** (1994), 369–390.

[13] V. Berthé, Combinaisons linéaires de $\zeta(s)/\Pi^s$ sur $F_q(x)$, pour $1 \leq s \leq q-2$, *J. Number Theory* **53** (1995), 272–299.

[14] V. Berthé, De nouvelles preuves preuves "automatiques" de transcendance pour la fonction zêta de Carlitz, *Journées Arithmétiques de Genève, Astérisque* **209** (1992), 159–168.

[15] V. Berthé, Fonction ζ de Carlitz et automates, *J. Théor. Nombres Bordeaux* **5** (1993), 53–77.

[16] H. Cameron and D. Wood, Pm numbers, ambiguity, and regularity *RAIRO Inform. Theor. App.*, **27** (1993), 261–275.

[17] L. Carlitz, On certain functions connected with polynomials in a Galois field, *Duke Math. J.*, **1** (1935), 137–168.

[18] J. Cassaigne and F. Nicolas, Factor complexity, in: *Combinatorics, Automata and Number Theory. Encyclopedia of Mathematics and its Applications*, vol. **135**, Cambridge University, Press, Cambridge (2010), 179–261.

[19] G. Christol, Ensembles presque périodiques k-reconnaissables, *Theoret. Comput. Sci.* **9** (1979), 141–145.

[20] G. Christol, Opération de Cartier et vecteurs de Witt, *Séminaire Delange-Pisot-Poitou* **12** (1972), 13.1–13.7.

[21] G. Christol, T. Kamae, M. Mendès France and G. Rauzy, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France* **108** (1980), 401–419.

[22] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, *Math. Systems Theory* **3** (1969), 186–192.

[23] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164–192.

[24] H. Chérif and B. de Mathan, Irrationality measures of Carlitz zeta values in characteristic p, *J. Number Theory* **44** (1993), 260–272.

[25] H. Derksen, A Skolem-Mahler-Lech theorem in positive characteristic and finite automata, *Invent. Math.* **168** (2007), 175–224.

[26] D. Duverney, Sommes de deux carrés et irrationalité de valeurs de fonctions thêta, *C. R. Math. Acad. Sci. Paris* **320**, Série I (1995), 1041–1044.

[27] A. Fraenkel, Systems of numeration, *Amer. Math. Mo.* **92** (1985), 10–114.

[28] I. Gheorghiciuc, The subword complexity of finite and infinite binary words, *Adv. in Appl. Math.* **39** (2007), 237–259.

[29] D. Goss, *Basic Structures of Function Field Arithmetic*, Springer-Verlag, Berlin, 1996.

[30] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford Sci. Publ., Oxford Univ. Press, 1989.

[31] M. Morse and G. A. Hedlund, Symbolic dynamics II, *Amer. J. Math.* **62** (1940), 1–42.

[32] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in: *ICALP'84*, Lect. Notes Comput. Sci. **172**, Springer-Verlag (1984), 380–389.

[33] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lecture Notes in Mathematics, vol. **1974**, Springer-Verlag, Berlin, 2002.

[34] H. Sharif and C. F. Woodcock, Algebraic functions over a field of positive characteristic and Hadamard products, *J. Lond. Math. Soc.* **37** (1988), 395–403.

[35] D. S. Thakur, *Function Field Arithmetic*, World Scientific, Singapore, 2004.

[36] L. J. Wade, Certain quantities transcendental over $GF(p_n)(x)$, *Duke Math. J.* **8** (1941), 701–720.

[37] J. Yu, Transcendence and special Zeta values in characteristic p, *Ann. of Math.* **134** (1991), 1–23.