ON POSSIBLE GROWTHS OF ARITHMETICAL COMPLEXITY *

Anna E. Frid[1]

Abstract. The arithmetical complexity of infinite words, defined by Avgustinovich, Fon-Der-Flaass and the author in 2000, is the number of words of length n which occur in the arithmetical subsequences of the infinite word. This is one of the modifications of the classical function of subword complexity, which is equal to the number of factors of the infinite word of length n. In this paper, we show that the orders of growth of the arithmetical complexity can behave as many sub-polynomial functions. More precisely, for each sequence u of subword complexity $f_u(n)$ and for each prime $p \geq 3$ we build a Toeplitz word on the same alphabet whose arithmetical complexity is $a(n) = \Theta(n f_u(\lceil \log_p n \rceil))$.

Mathematics Subject Classification. 68R15, 68Q45.

Introduction

Arithmetical complexity of an infinite word is a function relative to its subword complexity. It is larger than or equal to the subword complexity since it counts not only factors of the word but all words occurring in its arithmetic progressions. By the definitions, if the infinite word is $u = u_1 u_2 \cdots u_k \cdots$, where u_i are symbols, then for all n its subword complexity $f_u(n)$ is the number of distinct words of the form $u_i u_{i+1} \cdots u_{i+n-1}$ for arbitrary i, and its arithmetical complexity $a_u(n)$ is the number of all words of the form $u_k u_{k+d} \cdots u_{k+(n-1)d}$ for arbitrary initial symbols k and differences d. Note that by the classical Van der Waerden theorem [23, 25], there are arbitrarily large powers of some symbol among these words.

Arithmetical complexity was introduced in 2000 by Avgustinovich, Fon-Der-Flaass and Frid [5]; its idea belongs to S.V. Avgustinovich. It has become one of the most well-explored modified complexity functions relative to the subword complexity. Other definitions include *d-complexity* introduced in 1987 by Iványi [11]; *palindrome complexity* (see Allouche *et al.* [1] for a survey); modified complexity by Nakashima, Tamura, and Yasutomi [21, 22]; *pattern complexity* introduced in 2002 by Restivo and Salemi [24]; and *maximal pattern complexity* by Kamae and Zamboni [16] which is also dated 2002 and well-explored [16–19]. For a survey on subword complexity, see Ferenczi [12].

All published results on arithmetical complexity concern words of linear subword complexity. Arithmetical complexity of such words can behave variously: it can be exponential, as for the Thue-Morse word, or linear, as for the paperfolding word [5]. In subsequent papers, uniformly recurrent words of linear arithmetical complexity were characterized [15], and a family of words having lowest complexity was found among them [4]. On the other hand, the arithmetical complexity of Sturmian words has been estimated as $\Theta(n^3)$ and found for Sturmian words of many slopes [8, 13].

Here we find a variety of rates of growth of arithmetical complexity and show that it is not less than the variety of possible subword complexity rates of growth: each function of the subword complexity of a word can be included to a formula for the arithmetical complexity of another word. More precisely, recall that the notation $f(n) = \Theta(g(n))$ for functions f and g of positive integers means that $C_1 g(n) \leq f(n) \leq C_2 g(n)$ for all sufficiently large n and some positive constants C_1 and C_2. We prove the following

Theorem 0.1. *For each infinite word u on a finite alphabet Σ, of subword complexity $f_u(n)$, and for each prime $p \geq 3$ there exists an infinite word on Σ whose arithmetical complexity is $\Theta(n f_u(\lceil \log_p n \rceil))$.*

In the next section, we describe the (well-known) technique of bispecial words which is used below for estimates of arithmetical complexity. Then in Section 2 the construction of the required word is introduced; this word is a Toeplitz word of a special kind, it depends on u, and its subword complexity is always linear. In the next four sections, we analyse the set of arithmetical factors of the constructed word, find its bispecial words and relate them to factors of u; compute their parameters and at last obtain estimates for the arithmetical complexity, thus proving Theorem 0.1. After that in Section 7 we discuss particular cases of Theorem 0.1.

1. COMPLEXITIES AND BISPECIAL WORDS

Let us define notions of complexity of words and languages used in this paper and discuss ways to compute them.

Let Σ be a finite alphabet of cardinality q; as usual, the set of all finite words on Σ including the empty word λ will be denoted by Σ^*, and the set of all (right) infinite words on Σ will be denoted by Σ^ω. A finite word v is called a *factor*, or *subword*, of a finite or infinite word u if $u = s_1 v s_2$ for some (possibly empty)
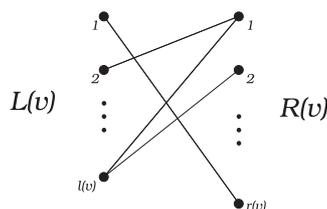
FIGURE 1. A bispeciality graph.

words s_1 and s_2. A language $\mathcal{F} \subseteq \Sigma^*$ is called *factorial* if it is closed under taking factors. Clearly, the language $\mathcal{F}(u)$ of factors of a finite or infinite word u is always factorial.

A word v is called an *arithmetical subword* of a word $u = u_1 u_2 \cdots u_m \cdots$, where $u_i \in \Sigma$ for all i, if $v = u_k u_{k+d} \cdots u_{k+(n-1)d}$ for some k and d. The set of all arithmetical subwords of u (or of words from a language $\mathcal{F}$) is called *arithmetical closure* of u (or $\mathcal{F}$) and denoted by $\mathcal{A}(u)$ ($\mathcal{A}(\mathcal{F})$). The term "closure" is chosen since $\mathcal{A}(\mathcal{A}(\mathcal{F})) = \mathcal{A}(\mathcal{F})$ for all $\mathcal{F}$; clearly, for any word u we have $\mathcal{A}(u) \supseteq \mathcal{F}(u)$ since the difference $d = 1$ in the definition of an arithmetical subword gives all factors of u.

The *subword complexity* of a factorial language $\mathcal{F}$, or of a word u, is the function equal to the number of distinct elements of $\mathcal{F}$ (factors of u) of length n and denoted by $f_{\mathcal{F}}(n)$ ($f_u(n)$). The *arithmetical complexity* of u is defined as $a_u(n) = f_{\mathcal{A}(u)}(n)$. Clearly, $a_u(n) \geq f_u(n)$ for any word u and length n. We see that both complexities count factors of some factorial language, so, it is not surprising that some techniques useful for computing subword complexity can be used for computing the arithmetical one. In this paper, we shall use one of them which involves bispecial words. For proofs and details of this method, see Cassaigne [6].

Let $\mathcal{F}$ be a factorial language. The set of all elements of $\mathcal{F}$ of length n will be denoted by $\mathcal{F}(n)$. For a word $v \in \mathcal{F}$, let us define the set of symbols $R(v)$ as $R(v) = \{a \in \Sigma | va \in \mathcal{F}\}$. Its cardinality $r(v) = \#R(v)$ is called the (right) *speciality degree* of v.

Analogously, let us define $L(v) = \{a \in \Sigma | av \in \mathcal{F}\}$ and $l(v) = \#L(v)$. A word v is *bispecial* in $\mathcal{F}$ if $l(v) \neq 1$ and $r(v) \neq 1$. Its *bilateral order* $b(v)$ is defined as $b(v) = \#\{(a,b)|a,b \in \Sigma, avb \in \mathcal{F}\} - l(v) - r(v) + 1$. It is not difficult to check that $b(v)$ can be not equal to zero only if v is bispecial in $\mathcal{F}$; so, if $b(v) \neq 0$, we say that v is an *essential bispecial* word in $\mathcal{F}$. The set of all essential bispecial words of length n in $\mathcal{F}$ is denoted by $\mathcal{B}_{\mathcal{F}}(n)$.

It is convenient to depict words of the form $avb \in \mathcal{F}$, where $v \in \mathcal{F}$, $a, b \in \Sigma$, as edges of the bipartite *bispeciality graph* of the word v whose parts are $L(v)$ and $R(v)$ (see Fig. 1). The bilateral order of v is thus the number of edges of this graph, minus number of its vertices, plus 1.

Let us denote by $f'(n)$ ($f''(n)$) the first (second) difference of a function $f(n)$: $f'(n) = f(n+1) - f(n)$, $f''(n) = f(n+2) - 2f(n+1) - f(n)$. We know [6] the

following relations for complexity of a factorial language $\mathcal{F}$:

$$f'_{\mathcal{F}}(n) = \sum_{v \in \mathcal{F}(n)} (r(v) - 1), \tag{1}$$

$$f''_{\mathcal{F}}(n) = \sum_{v \in \mathcal{F}(n)} b(v).$$

By the definition of the set of essential bispecial words, in the latter formula we can ignore all words but them:

$$f''_{\mathcal{F}}(n) = \sum_{v \in \mathcal{B}_{\mathcal{F}}(n)} b(v). \tag{2}$$

At last, we mention another pair of easy relations valid for any infinite word u and any $n > 0$ (recall that $q = \#\Sigma$):

$$\frac{f_u(n)}{q} \leq f_u(n-1) \leq f_u(n). \tag{3}$$

2. Patterns and Toeplitz words

Let $? \notin \Sigma$ be a new symbol called *gap*. A *pattern* is a finite word on $\Sigma \cup \{?\}$. For a pattern T and an infinite word $u = u_1 u_2 \cdots u_n \cdots$, $u_i \in \Sigma \cup \{?\}$, we shall denote by $T \cdot u$ the result of substituting symbols of u to successive gaps of $T^\omega (= TT \cdots T \cdots)$. If all patterns $T_1, T_2, \ldots, T_n, \ldots$ start with a symbol of Σ, then there exists a limit $t = T_1 \cdot T_2 \cdot \cdots \cdot T_n \cdot \cdots \in \Sigma^\omega$ of the sequence $T_1 \cdot (T_2 \cdot \cdots \cdot (T_n \cdot ?^\omega) \cdots)$. The sequence t is called the *Toeplitz word* generated by the sequence of patterns $\{T_i\}_{i=1}^\infty$. For studies of Toeplitz words of specific forms, see [9, 20].

Note that each of the words $T_1 \cdot (T_2 \cdot \cdots \cdot (T_n \cdot ?^\omega) \cdots)$ is periodic, and if we naturally denote the pattern corresponding to its minimal period by $T_1 \cdot T_2 \cdot \cdots \cdot T_n$, we obtain an associative operation $(\cdot)$ on patterns.

Example 2.1. Let $T = a?b?$, then $T \cdot ?^\omega = a?b?a?b?a?b?a?b? \cdots$, $T \cdot (T \cdot ?^\omega) = aab?abb?aab?abb? \cdots$, and thus $T \cdot T = aab?abb?$. Continuing the process, we get the famous *paperfolding word* $t = T \cdot T \cdot T \cdot \cdots = aabaabbaaabbabba \cdots$. Both its subword [2] and arithmetical [5] complexities are linear.

Let us fix a prime p and a letter $a \in \Sigma$ and define the pattern $T_p(a) = \underbrace{a \cdots a}_{p-1}?$.

The pattern $T_p(a_1) \cdot T_p(a_2) \cdot \cdots \cdot T_p(a_n)$ will be denoted by $T_p(a_1 a_2 \cdots a_n)$ for short; its length is p^n and the only gap in it is the last symbol. The word on Σ obtained from $T_p(a_1 a_2 \cdots a_n)$ by erasing the last gap will be denoted by $t_p(a_1 a_2 \cdots a_n)$. It is natural to define $T_p(\lambda) = ?$ and $t_p(\lambda) = \lambda$.

Example 2.2. Since $T_2(a) = a?$ and $T_2(b) = b?$ (and thus $t_2(a) = a$ and $t_2(b) = b$), we have $T_2(aa) = aaa?$ and $t_2(aa) = aaa$, then $T_2(aab) = aaabaaa?$ and $t_2(aab) = aaabaaa$, etc.

Now let us fix an infinite word $u = u_1 u_2 \cdots u_n \cdots$, where $u_i \in \Sigma$, and define the Toeplitz word

$$t_p(u) = T_p(u_1) \cdot T_p(u_2) \cdot \cdots \cdot T_p(u_n) \cdots :$$

each symbol of this infinite word is well-defined since the first $p^n - 1$ symbols are determined already in $t_p(u_1 \cdots u_n)$. Equivalently, $t_p(u)$ can be defined as the limit of the sequence $\{t_p(u_1 \cdots u_n)\}_n$; it always exists.

Example 2.3. The famous *period doubling word* (see, for example, [10]) is defined as

$$t_2(ababab \cdots) = abaaababababaaabaaabaaabab \cdots .$$

The subword complexity of $t_p(u)$ is linear for all u [20]. As for the arithmetical one, the word $t_p(u)$ is what we need to prove Theorem 0.1: in what follows, we show that $a_{t_p(u)}(n) = \Theta(n f_u(\lceil \log_p(n) \rceil))$.

3. ARITHMETICAL CLOSURE OF $t_p(u)$

Consider $u = u_0 u_1 u_2 \cdots u_n \cdots$, where $u_i \in \Sigma$. Without loss of generality, we assume that each symbol of Σ occurs in u at least once. In what follows, we denote by Su the shift $Su = u_1 u_2 u_3 \cdots u_n \cdots$; correspondingly, $S^k u = u_k u_{k+1} u_{k+2} \cdots$ for all k.

Let us count symbols of $t_p(u)$, unlike those of u, starting with 1 not 0, *i.e.*, $t_p(u) = w_1 w_2 \cdots w_m \cdots$ for $w_i \in \Sigma$. We say that a position in a Toeplitz word $t_p(u)$, and the symbol occupying it, are of *nth order* if the index of this position in $t_p(u)$ is divisible by p^n, or, equivalently, if in $T_p(u_0 u_1 \cdots u_{n-1}) \cdot ?^\omega$ there is still a gap at this position. In particular, all symbols of $t_p(u)$ are of order 0. We shall say that the *exact order* of a position, or a symbol occupying it, is equal to n if it is of order n but not of order $n + 1$. By the definition of $t_p(u)$, for all $n \geq 0$ all symbols of exact order n in it are equal to u_n.

Lemma 3.1. *For any $u \in \Sigma^\omega$ and prime p the arithmetical closure of $t_p(u)$ is*

$$\mathcal{A}(t_p(u)) = \bigcup_{n=0}^{\infty} \mathcal{F}(t_p(S^n u)) \bigcup \left(\bigcup_{x \in \Sigma} x^* \right).$$

Example 3.2. If $u = abaaab \cdots$, then the arithmetical closure of $t_2(u)$ is the union of sets of factors of $t_2(u) = t_2(abaaab \cdots) = (abaa)^7 abab \cdots$, $t_2(Su) = t_2(baaab \cdots) = (ba)^7 bb \cdots$, $t_2(S^2 u) = t_2(aaab) = a^7 ba^7 \cdots$, etc., plus arbitrarily large powers of a and b.

Proof. Let us use the notation $t_p(u) = w_1 w_2 \cdots w_n \cdots$, where $w_i \in \Sigma$, and denote by $w(d, k)$ the arithmetical subsequence of $t_p(u)$ of difference d starting with w_k: $w(d, k) = w_k w_{k+d} \cdots w_{k+nd} \cdots$.

First suppose that $(d, p) = 1$. Then for each m, exactly one of the first p^m symbols of $w(d, k)$ (say, $w_{k+n_m d}$ for some $n_m \in \{1, \ldots, p^m\}$) is of mth order in $t_p(u)$.

It is not difficult to see that the factor of $w(d,k)$ of length $p^m - 1$ starting from its next symbol, *i.e.*, $w_{k+(n_m+1)d}w_{k+(n_m+2)d}\cdots w_{k+(n_m+p^m-1)d}$, coincides with a prefix of $t_p(u)$. Since m is arbitrary, $t_p(u)$ and $w(d,k)$ thus have arbitrarily long common factors. But $t_p(u)$ is by construction *uniformly recurrent* [9], that is, each of its factors occurs an infinite number of times with bounded gaps. By a (folklore) lemma from [5], its arithmetical subsequence $w(d,k)$ is thus also uniformly recurrent.

It is well-known also that if languages of factors of two uniformly recurrent words have an infinite intersection, then they coincide. Indeed, let $R_x(n)$ denote the recurrence function of an infinite word x, that is the maximal distance between successive occurrences of the same word of length n in x. If x and y are uniformly recurrent and contain a common factor of length $\max(R_x(n), R_y(n)) + n$, then the sets of their factors of length n coincide. So, if they contain arbitrarily long common factors, then the sets of its factors coincide.

We can conclude that $\mathcal{F}(w(d,k)) = \mathcal{F}(t_p(u))$.

Now suppose that $(d,p) \neq 1$; since p is prime, this implies that $p^m | d$ for some $m > 0$. Suppose that m is maximal (*i.e.*, $p^{m+1} \nmid d$) and define $m' \geq 0$ as the maximal exponent of p dividing k: $p^{m'} | k$, $p^{m'+1} \nmid k$. If $m' \geq m$, then all symbols of $w(d,k)$ are of order (at least) m in $t_p(u)$, and $w(d,k)$ is a subsequence of $w(p^m, p^m) = t_p(S^m u)$ of difference d/p^m coprime with p. Analogously to the previous paragraph, $\mathcal{F}(w(d,k)) = \mathcal{F}(t_p(S^m u))$. If $m' < m$, then all symbols of $w(d,k)$ are of exact order m' in $t_p(u)$, and clearly $w(d,k) = u_{m'}^\omega$. For different m' we obtain all symbols of Σ which occur in u.

We listed all possible cases for d and k, and the observation that $\mathcal{A}(t_p(u)) = \cup_{k,d>0}\mathcal{F}w(d,k)$ completes the proof of the lemma. $\square$

In what follows, we are going to characterize essential bispecial words of the language $\mathcal{A} = \mathcal{A}(t_p(u))$. It will help us to estimate its subword complexity, which is the arithmetical complexity of $t_p(u)$: $f_{\mathcal{A}}(n) = a_{t_p(u)}(n)$. In what follows, the notations $f_{\mathcal{A}}(n)$ and $a_{t_p(u)}(n)$ are used equivalently.

4. Bispecial words of $\mathcal{A}$

Let us define the Toeplitz product of a pattern of the form $T = t?$, $t \in \Sigma^*$, and a word $v = v_1 \cdots v_n$, $v_i \in \Sigma$, as the word $T \cdot v = tv_1 tv_2 t \cdots tv_n t$. Clearly, the length of $T \cdot v$ is $|T \cdot v| = (|t| + 1)(|v| + 1) - 1$.

Example 4.1. For each word $v = v_1 \cdots v_n$ with $v_i \in \Sigma$ we have $? \cdot v = v$ and $T_p(v_1 \cdots v_{n-1}) \cdot v_n = t_p(v)$. For each pattern $T = t?$ and for all $x, y \in \Sigma$ we have $T \cdot \lambda = t$, $T \cdot x = txt$, $T \cdot xy = txtyt$.

Now consider a word $v \in \mathcal{F}(u)$ and denote by a_v its last symbol. We shall write $v = v'a_v^{\sigma(v)}$, where v' is the longest prefix of v which does not end with a_v. Thus, $\sigma(v)$ is defined as the length of the longest suffix of v in which only one letter occurs. We shall call the word v *saturated* in $\mathcal{F}(u)$ if $v \in \mathcal{F}(u)$ but $va_v \notin \mathcal{F}(u)$, *i.e.*, if v cannot be extended to an element of $\mathcal{F}(u)$ by repeating its last symbol.

For technical reasons, we for a while assert that

$$\text{all symbols of } \Sigma \text{ occur in } u \text{ infinitely many times.} \tag{4}$$

Later, in the end of Section 6, we will turn this condition down, but at the moment it is convenient to use it.

Lemma 4.2. *Let $p \geq 3$ be a prime integer, and $u \in \Sigma^\omega$ be an infinite word satisfying Condition (4). Then each essential bispecial word of the arithmetical closure $\mathcal{A}$ of the Toeplitz word $t_p(u)$ is of one of the following four forms:*

(1) $t_p(v)$, *where* $v \in \mathcal{F}(u)$;
(2) $T_p(v') \cdot a_v^{2p^{\sigma(v)}-2}$, *where* $v = v'a_v^{\sigma(v)}$ *is a saturated word in* $\mathcal{F}(u)$ *with* $v' \neq \lambda$;
(3) $T_p(v) \cdot x = t_p(v)xt_p(v)$, *where* $x \in \Sigma$, $v \in \mathcal{F}(u)$;
(4) $T_p(v) \cdot xy = t_p(v)xt_p(v)yt_p(v)$, *where* $x, y \in \Sigma$, $x \neq y$, $v \in \mathcal{F}(u)$.

Proof. Let us consider a bispecial word $w \in \mathcal{A}$. Due to Lemma 3.1, it must belong to some $\mathcal{F}(t_p(S^n u))$: indeed, even if w is a power of a symbol x, we must be able to extend it not only by x, so it must occur somewhere else than in x^*.

First of all, let us prove that

$$w = T_p(v_1 \cdots v_k) \cdot z \tag{5}$$

for some $k \geq 0$ and z which is either of length not greater than 2 or is a power of a letter x; and each word $S^n u$ such that $w \in \mathcal{F}(t_p(S^n u))$ starts with $v_1 \cdots v_k$. Indeed, if $|w| \leq 2$ or $w = x^m$ for some letter x and integer m, then it is true with $k = 0$ (since $T_p(\lambda) \cdot w =? \cdot w = w$). So, suppose that $|w| \geq 3$. Since w is a factor of some $t_p(S^n u)$, where $S^n u$ starts with a letter v_1, and since w contains other symbols than v_1, it looks like

$$w = v_1^\alpha x_1 v_1^{p-1} x_2 \cdots x_m v_1^\beta,$$

where $0 \leq \alpha, \beta \leq p-1$, $x_1 \cdots x_m \in \mathcal{F}(t_p(S^{n+1}u))$, and at least one of the symbols x_i is not equal to v_1. We shall denote this symbol by x. Since $|w| \geq 3$, w contains at least one of the factors $v_1 v_1 x$, $v_1 x v_1$, or $x v_1 v_1$. Hence w cannot be a factor of any $t_p(S^{n'}u)$ where $S^{n'}u$ starts with a symbol $y \neq v_1$ because in such $S^{n'}u$ occurrences of any symbol but y are separated by $p-1$ symbols y. (Note that here we use the condition $p > 2$: if $p = 2$, arbitrarily long words of the form $v_1 y v_1 y v_1 y \cdots$ occur both in $t_2(v_1 yy \cdots)$ and in $t_2(y v_1 v_1 \cdots)$, complicating the situation.)

Thus, w occurs in $\mathcal{A}$ only as a factor of some words of the form $t_p(v_1 \cdots)$. A symbol not equal to v_1 may occur in it, or in its extension, only at a distance divisible by p from x. In particular, if $\alpha \neq p-1$ ($\beta \neq p-1$), then w can be extended to the left (to the right) in $\mathcal{A}$ only by v_1. But w is by assertion bispecial, so, $\alpha = \beta = p-1$ and thus $w = T_p(v_1) \cdot (x_1 \cdots x_m)$. Here $x_1 \cdots x_m$ is a factor of $t_p(S^{n+1}u \cdots)$ for all n such that w is a factor of $t_p(S^n u)$. If the length $m \leq 2$ or $x_1 = x_2 = \cdots = x_m$, then (5) is proved with $k = 1$, otherwise we can continue the process with $x_1 \cdots x_m$ instead of w: it is also bispecial in $\mathcal{A}$ since it can be extended

at least by the same letters than w. After several steps we shall necessarily come to (5).

If $|z| = 2$ and $z = xy$, $x \neq y$, we get exactly Case 4. In all other cases z is a power of some symbol: $z = a^m$, $a \in \Sigma$, $m \geq 0$.

If $k = 0$ and $m = 0$, we have $w = \lambda = t_p(\lambda)$, so that w falls into Case 1. If $k > 0$, then $m > 0$: we cannot have $z = \lambda$ since $T_p(v) \cdot \lambda = t_p(v) = T_p(v') \cdot a_v^{p^{\sigma(v)}-1}$, and if $w = t_p(v)$ we would stop the process earlier, at the representation $w = T_p(v') \cdot a_v^{p^{\sigma(v)}-1}$ (with $a = a_v$, $m = p^{\sigma(v)} - 1$). So, from now on we consider that $m > 0$.

Also, $v_k \neq a$: otherwise we also would stop the process earlier, at most at the representation $w = T_p(v_1 \cdots v_{k-1}) \cdot a^{p(m+1)-1}$.

Let us mention that the greatest power of a occurring in $\mathcal{F}(t_p(a^i b \cdots))$, where $b \neq a$, is a^{2p^i-1}: indeed, $a^{2p^i-1} = T_p(a^i) \cdot a$; the letter a will somewhere occur between two $t_p(a^i)$ because of Condition (4).

So, suppose first that $m = p^i - 1$ (or $m = 2p^i - 1$) with $i > 0$; then, due to the previous paragraph, w occurs in $\mathcal{A}$ if and only if $v = v_1 \cdots v_k a^i \in \mathcal{F}(u)$; here $v_1 \cdots v_k = v'$, $a = a_v$ and $i = \sigma(v)$. In the first situation, $w = T_p(v_1 \cdots v_k) \cdot a^m = t_p(v_1 \cdots v_k a^i)$, and this gives us exactly Case 1; in the second one, $w = T_p(v_1 \cdots v_k) \cdot a^m = T_p(v_1 \cdots v_k a^i) \cdot a = t_p(v_1 \cdots v_k a^i)\, a\, t_p(v_1 \cdots v_k a^i)$. This situation falls into Case 3; to describe this case completely, we add the possibility of $i = 0$, that is, of $m = 1$ and $z = a$.

It remains to consider the case $z = a^m$, where $m \neq p^i - 1$, $m \neq 2p^i - 1$ for any i, and $m \geq 2$. As we have mentioned, $w = T_p(v_1 \cdots v_k) \cdot a^m$ occurs in $\mathcal{A}$ only as a factor of some $t_p(va^i \cdots)$, where $2p^{i-1} - 1 < m \leq 2p^i - 1$.

Suppose that $bw \in \mathcal{A}$ for some $b \neq a$. Then bw occurs as a factor in words of the form $t_p(v_1 \cdots v_k a^i \cdots)$ and only in them, moreover, the first b of bw always occurs in such a word at a position of order $k + i$, i.e., is always followed by the word $t_p(v_1 \cdots v_k a^i)$ of length $p^{k+i} - 1$, then by a symbol (of order $k + i$ again), then by $t_p(v_1 \cdots v_k a^i)$ again etc. But we assumed that $m \neq p^i - 1$ and $m \neq 2p^i - 1$ for any i, and thus $|bw| = 1 + |T_p(v_1 \cdots v_k) \cdot a^m| = p^k(m + 1)$ is not divisible by p^{k+i}. So, the word bw can be extended to the right by only one of the letters; its exact order in $t_p(v_1 \cdots v_k a^i \cdots)$ lies between k and $k + i - 1$, so it is equal to a. The symbol $b \neq a$ can be chosen arbitrarily by Condition (4), so, for all $b \neq a$ we have $bwa \in \mathcal{A}$, $bwc \notin \mathcal{A}$ for any $c \neq a$.

Analogously, $awb \in \mathcal{A}$ for all $b \neq a$ (and $cwb \notin \mathcal{A}$ for all $b, c \neq a$, but we already know this because of the symmetry of b and c). So, to find the bilateral order of w it remains to find out if $awa \in \mathcal{A}$: this word must be considered individually since we cannot determine the position of its first (or last) symbol modulo p^{k+i}, we know only that it is of order k in $t_p(v_1 \cdots v_k a^i \cdots)$ and thus that awa may occur in $\mathcal{A}$ only as a factor of $T_p(v_1 \cdots v_k) \cdot a^{m+2} = t_p(v_1 \cdots v_k)awat_p(v_1 \cdots v_k)$.

If $awa \in \mathcal{A}$, then the bispeciality graph of w looks as it is depicted in Figure 2; it has $2q - 1$ edges and $2q$ vertices, so, $b(w) = 0$ and w is not essential bispecial. If $awa \notin \mathcal{A}$, then the bispeciality graph of w looks as it is depicted in Figure 3, and $b(w) = -1$, so, w *is* essential bispecial. But this is the case only if, first, i is
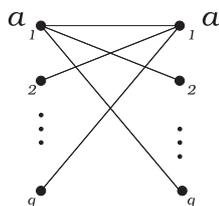
FIGURE 2. $b(w) = 0$.
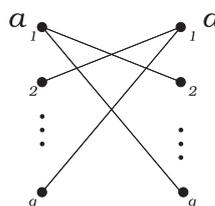
FIGURE 3. $b(w) = -1$ (Case 2).

the maximal power of a which can follow $v_1 \cdots v_k$ in u (*i.e.*, if the word $v_1 \cdots v_k a^i$ is saturated), and second, if $m = 2p^i - 2$, that is, if the maximal number of consecutive a's in all $t_p(S^n u)$ of the form $t_p(S^k v_1 \cdots v_k a^i \cdots)$ is exactly $m + 1$, so that $T_p(v_1 \cdots v_k) \cdot a^{m+2} \notin \mathcal{A}$. Moreover, the case $k = 0$ (*i.e.*, of $v_1 \cdots v_k = \lambda$ and $w = a^{2p^i - 2}$) is excluded since any power of a including a^{2p^i} always belongs to $\mathcal{A}$ due to Lemma 3.1. We have obtained exactly Case 2 of Lemma 4.2 and completed its proof. $\qquad \square$

5. BILATERAL ORDERS AND DIAGRAM OF FIRST DIFFERENCES

As announced at the end of the introduction, we continue the proof of Theorem 0.1 finding bilateral orders of all essential bispecial words listed in Lemma 4.2. We shall consider all the four cases and will find second differences of the subword complexity function. Let us start from

Case 1. Consider a word of the form $w = t_p(v)$, where $v \in \mathcal{F}(u)$, $|v| = m$. Note that w occurs in $\mathcal{A}$ only as a factor of words of the form $t_p(v \cdots)$. Suppose that $a_v = a$, that is, $w = T_p(v') \cdot a^{p^k - 1}$ for some $k > 0$, where v' does not end with a. By the construction, all symbols of this $a^{p^k - 1}$ are always of order $|v'|$ in any infinite word $t_p(v \cdots)$. Let us consider possible extensions bwc of w in $\mathcal{A}$.

If we choose an arbitrary $b \neq a$, we see that the first b of bwc must be of order $|v| = m$ (in any occurrence of bw in $t_p(v \cdots)$). If $vb \in \mathcal{F}(u)$, then the order of b can be equal exactly to m, and thus the order of the last c of $bwc \in \mathcal{A}$ can be arbitrarily large. Due to (4), c can be arbitrary, so, $bwc \in \mathcal{A}$ for all $b \in R(v) \backslash \{a\}$, $c \in \Sigma$ (and, symmetrically, for all $b \in \Sigma$, $c \in R(v) \backslash \{a\}$). On the other hand, if $b \notin R(v)$, then the order of the first b of bwc must be greater than m, the order of the last c must be equal exactly to m, and thus vc must belong to $\mathcal{F}(u)$. So, if $b \neq a$ (or, symmetrically, $c \neq a$), and $b, c \notin R(v)$, then $bwc \notin \mathcal{A}$.

It remains to find out if $awa \in \mathcal{A}$. If $a \in R(v)$, this is the case by the same arguments as above: the exact order of the first and the last a can be equal to m. But this is the case even if $a \notin R(v)$, *i.e.*, if v is a saturated word in $\mathcal{F}(u)$. Indeed, in this case all $p^k + 1$ symbols of order $|v'|$ in awa are equal to a. Any of them except the first and the last one may have order $|v|$ (and arbitrarily large exact
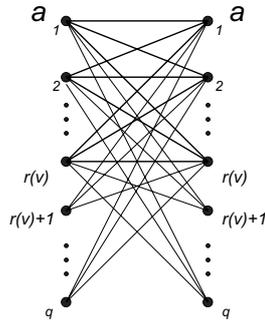
FIGURE 4. Case 1, v is not saturated.
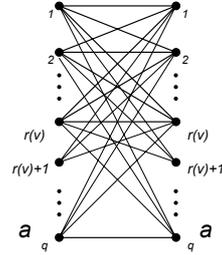
FIGURE 5. Case 1, v is saturated.

order, so, it can be equal to a due to (4)). All other symbols are of exact order less than $|v|$. So, $awa \in \mathcal{A}$ anyway.

We see that the form of the bispeciality graph of $t_p(v)$ depends whether v is saturated in $\mathcal{F}(u)$ or not (see Figs. 4, 5). The number of edges in the graph is respectively $qr(v) + (q - r(v))r(v) = r(v)(2q - r(v))$ and $r(v)(2q - r(v)) + 1$, the number of vertices is always $2q$, and thus

$$b(w) = \begin{cases} (r(v) - 1)(2q - r(v) - 1) + 1, & \text{if } v \text{ is saturated,} \\ (r(v) - 1)(2q - r(v) - 1) & \text{otherwise.} \end{cases}$$

As it follows from Lemma 4.2, all essential bispecial words of $\mathcal{A}$ of length $p^m - 1$ correspond to words of $\mathcal{F}(u)$ of length m. If we denote the set of factors of u of length m by $\mathcal{F}_u(m)$ and the number of saturated words in $\mathcal{F}_u(m)$ by $s_u(m)$, we have, due to (2),

$$f''_{\mathcal{A}}(p^m - 1) = s_u(m) + \sum_{v \in \mathcal{F}_u(m)} (r(v) - 1)(2q - r(v) - 1).$$

Case 2. As we have shown in the proof of Lemma 4.2, each saturated word v of length m with $v' \neq \lambda$ in u gives an essential bispecial word $T_p(v') \cdot a_v^{2p^{\sigma(v)} - 2}$ in $\mathcal{A}$. Its length is $(2p^{\sigma(v)} - 1)p^{|v'|} - 1 = 2p^m - p^{|v'|} - 1$ and its bilateral order is -1. So, lengths of all such words lie between $2p^m - p^{m-1} - 1$ and $2p^m - 2$ (in fact, the latter value can be specified to $2p^m - p - 1$ since $v' \neq \lambda$, but we neglect it). Their combined number is $s_u(m)$ minus the number of saturated words with $v' = \lambda$, that is, of symbols whose maximal power in u is m. If we denote the latter number by $q'_u(m)$, we get

$$f'_{\mathcal{A}}(2p^m - 1) - f'_{\mathcal{A}}(2p^m - p^{m-1} - 1) = \sum_{n=2p^m - p^{m-1} - 1}^{2p^m - 2} f''_{\mathcal{A}}(n) = -s_u(m) + q'_u(m).$$
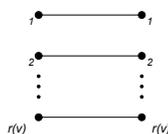
FIGURE 6. Case 3, $x \notin L(v)$.
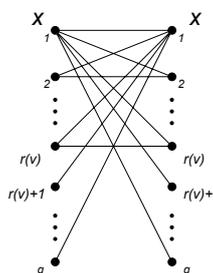
FIGURE 7. Case 3, $x \in L(v)$.

Case 3. Consider $w = t_p(v)xt_p(v)$, where $v \in \mathcal{F}(u)$, $|v| = m$ (and thus $|w| = 2p^m - 1$), $x \in \Sigma$. We have to consider two cases, $a_v \neq x$ and $a_v = x$.

Suppose first that v does not end with x: $a_v \neq x$. Then the order of the central symbol x of w in any of its occurrences in $t_p(v \cdots)$ is (at least) m.

If we assume in addition that $vx \notin \mathcal{F}(u)$, then the *exact* order of the central x of w must always be greater than m. So, the symbols situated at the distance p^m from it must be of exact order m, which means that they both must be equal and must belong to $L(v)$ in $\mathcal{F}(u)$. But these two symbols are exactly symbols by which w can be extended to the left and to the right in $\mathcal{A}$: the word w occurs in each $t_p(va \cdots)$, where $a \in L(v)$, and can be extended in it only as awa. So, the bispeciality graph of w looks as it is depicted in Figure 6; it has $r(v)$ edges and $2r(v)$ vertices, and the bilateral order of w is $b(w) = -r(v) + 1$.

All these $r(v)$ edges by the same reasons occur also when $vx \in \mathcal{F}(u)$, but in this case, the central x of w may have exact order m in some $t_p(vx \cdots)$. So, if $bwc \in \mathcal{A}$ for some $b, c \in \Sigma$, then the exact order of b can be arbitrarily large, and due to (4), b can be arbitrary; but if $b \neq x$, this implies that the exact order of c is m and thus $c = x$. Similarly, c can be arbitrary with $b = x$. So, the bispeciality graph of w looks as it is shown in Figure 7; it has $2q + r(v) - 2$ edges and $2q$ vertices, and $b(w) = r(v) - 1$.

Now let us consider $x = a_v$. In this case, $w = T_p(v') \cdot x^{2p^{\sigma(v)}-1}$, and each of these $2p^{\sigma(v)} - 1$ letters x is of order (at least) $|v'| = m - \sigma(v)$.

Suppose that $vx \notin \mathcal{F}(u)$; then, as it was mentioned earlier, $2p^{\sigma(v)} - 1$ is the maximal number of successive letters x of order $|v'|$ in a word of the form $t_p(v \cdots)$ whose factors belong to $\mathcal{A}$. So, the central one of them, that is the central x of w, always is of order m (and even greater), and we can continue the arguments as in the case of $x \neq a_v$ and obtain the bispeciality graph from Figure 6 again. The only difference occurs if $v = x^m$ and thus $w = x^{2p^m-1}$: as a factor of x^ω, it can be extended also as xwx. Its bispeciality graph is depicted in Figure 8, and its bilateral order is $-r(v)$.

Now let $vx \in \mathcal{F}(u)$; then each of the $2p^{\sigma(v)} - 1$ letters x of order $|v'|$ in w may be of order m. However, if this is not the central one, *i.e.*, if the central x in w is of order less than m, then this occurrence of w can be extended only as xwx. On
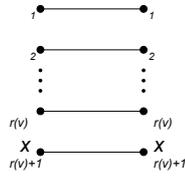
FIGURE 8. Case 3, $v = x^m$, $x^{m+1} \notin \mathcal{F}(v)$.

FIGURE 9. Case 4.

the other hand, if the central x is of order m, all arguments applied for $x \neq a_v$ are valid. We see that the bispeciality graph of w looks again as in Figure 7.

So, for each $v \in \mathcal{F}_u(m)$ there are $r(v)$ words of length $2p^m - 1$ in $\mathcal{A}$ having bilateral order $r(v) - 1$ (namely, words $T_p(v) \cdot x$, $x \in L(v)$), and $q - r(v)$ words of the same length of bilateral order $-r(v) + 1$ or $-r(v)$, where each word of $b(w) = -r(v)$ corresponds to a saturated word of the form $v = x^m$; the number of such words is $q'_u(m)$. Since due to Lemma 4.2 there are no other bispecial words of length $2p^m - 1$, we have

$$f''_{\mathcal{A}}(2p^m - 1) = -q'_u(m) + \sum_{v \in \mathcal{F}_u(m)} (2r(v) - q)(r(v) - 1).$$

Note that this value can be positive, negative, or equal to zero.

Case 4. Consider $w = t_p(v)xt_p(v)yt_p(v)$, where $|v| = m$; then $|w| = 3p^m - 1$. Since $x \neq y$, x and y are always of order m in $t_p(v \cdots)$, and the exact order of one of them is even larger. The other symbol must be of exact order m. Let this symbol be x (so, $vx \in \mathcal{F}(u)$); then w can be extended only as xwx since $p - 1$ symbols of order m surrounding y of order greater than m are of *exact* order m (here we once again use the condition $p \geq 3$). If $vy \notin \mathcal{F}(u)$, then w is not bispecial at all, but otherwise symmetrically $ywy \in \mathcal{A}$. So, for each pair of distinct letters $x, y \in L(v)$ the respective word w has the bispeciality graph depicted in Figure 9; its bilateral order is -1. The number of such pairs x, y for a given v is $r(v)(r(v) - 1)$, so,

$$f''_{\mathcal{A}}(3p^m - 1) = -\sum_{v \in \mathcal{F}_u(m)} r(v)(r(v) - 1).$$

We have found all values of $f''_{\mathcal{A}}(n)$ for $n \in \{p^m - 1, \ldots, p^{m+1} - 1\}$. Figure 10 shows the behaviour of $f'_{\mathcal{A}}(n)$ at this interval.

The rectangle shows the interval from $2p^m - p^{m-1} - 1$ to $2p^m$ where $f'_{\mathcal{A}}(n)$ somehow like stairs decreases by $s_u(m)$. (Recall that $s_u(m)$ is the number of saturated words of length m in u; the notation is introduced in the end of the consideration of Case 1.) All other non-zero values of $f''_{\mathcal{A}}(n)$ are explicitly found above.
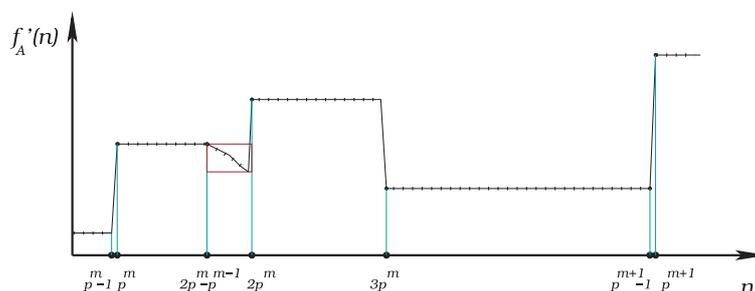
FIGURE 10. The behaviour of $f'_{\mathcal{A}}(n)$.

6. Estimates for arithmetical complexity

In this section, we end the proof of Theorem 0.1 estimating the arithmetical complexity. First, still under Condition (4), we shall sum up the second differences. According to the results of the previous section,

$$
\begin{aligned}
f'_{\mathcal{A}}(p^{m+1} - 1) - f'_{\mathcal{A}}(p^m - 1) &= \sum_{n=p^m-1}^{p^{m+1}-2} f''_{\mathcal{A}}(n) \\
&= f''_{\mathcal{A}}(p^m - 1) - s_u(m) + q'_u(m) + f''_{\mathcal{A}}(2p^m - 1) + f''_{\mathcal{A}}(3p^m - 1) \\
&= (q-1) \sum_{v \in \mathcal{F}_u(m)} (r(v) - 1) = (q-1)f'_u(m).
\end{aligned}
$$

This is valid starting from $m = 0$: indeed, there is one word λ in $\mathcal{F}_u(0)$, it is not saturated since it can be extended to the right by any symbol, and can be considered as well as any other word of $\mathcal{F}(u)$. Clearly, $f'_{\mathcal{A}}(0) = q - 1$, so,

$$
\begin{aligned}
f'_{\mathcal{A}}(p^{m+1} - 1) &= \sum_{n=0}^{m}(q-1)f'_u(n) + f'_{\mathcal{A}}(0) = (q-1)(f_u(m+1) - f_u(0) + 1) \\
&= (q-1)f_u(m+1).
\end{aligned}
$$

This value is valid also for all $n \in \{3p^m, \ldots, p^{m+1}-1\}$; for $n \in \{2p^m, \ldots, 3p^m-1\}$ we have $f'_{\mathcal{A}}(n) := B_2 = (q-1)f_u(m+1) + \sum_{v \in \mathcal{F}_u(m)} r(v)(r(v) - 1)$, and for $n \in \{p^m, \ldots, 2p^m - 1\}$, the value of $f'_{\mathcal{A}}(n)$ lies between $B_1 := (q-1)f_u(m) + \sum_{v \in \mathcal{F}_u(m)}(r(v) - 1)(2q - r(v) - 1)$ and $B_1 + f_u(m)$ (since $0 \leq s_u(m) \leq f_u(m)$). So,

$$
\begin{aligned}
f_{\mathcal{A}}(p^{m+1}) - f_{\mathcal{A}}(p^m) &= \sum_{n=p^m}^{p^{m+1}-1} f'_{\mathcal{A}}(n) \geq p^m B_1 + p^m B_2 + (p^{m+1} - 3p^m)(q-1)f_u(m+1) \\
&= p^m[(pq - p + 1)f_u(m+1) - qf_u(m)] \geq p^m(p-1)(q-1)f_u(m+1).
\end{aligned}
$$

Similarly,

$$f_{\mathcal{A}}(p^{m+1}) - f_{\mathcal{A}}(p^m) \leq p^m(B_1 + f_u(m)) + p^m B_2 + (p^{m+1} - 3p^m)(q-1)f_u(m+1)$$

$$\leq p^m\left(pq - p + 1 - \frac{q-1}{q}\right)f_u(m+1).$$

Summing up, we see that

$$f_{\mathcal{A}}(p^{m+1}) = \sum_{n=0}^{m}[f_{\mathcal{A}}(p^{n+1}) - f_{\mathcal{A}}(p^n)] + q,$$

and using the inequalities above and again (3), we obtain that

$$\frac{(p-1)(q-1)q}{pq-1}f_u(m+1)(p^{m+1}-1) < f_{\mathcal{A}}(p^{m+1}) \leq \frac{pq^2 - pq + 1}{q(p-1)}f_u(m+1)(p^{m+1}-1)+q.$$

Note that for all $n \in \{p^m, \dots, p^{m+1} - 1\}$ the inequality $f'_{\mathcal{A}}(n) \geq (q-1)f_u(m+1)$ holds. So, since $p^m \leq n$, we see that

$$f_{\mathcal{A}}(n+1) \geq f_{\mathcal{A}}(p^m) + (n - p^m)(q-1)f_u(m+1) > f_u(m+1)\left[\frac{(q-1)(p-1)}{pq-1}n - 1\right]$$

and

$$\begin{aligned} f_{\mathcal{A}}(n+1) &\leq f_{\mathcal{A}}(p^{m+1}) - (p^{m+1} - n - 1)(q-1)f_u(m+1) \\ &\leq f_u(m+1)\left[\frac{q(q-1)(2p-1)+p}{q(p-1)}n - \frac{q^2 - q + 1}{q(p-1)}\right] + q. \end{aligned}$$

Since $m + 1 = \lceil \log_p(n+1) \rceil$, the latter two inequalities give us the statement of Theorem 0.1 under Condition (4):

$$(C_1 n - 1)f_u(\lceil \log_p(n+1) \rceil) < f_{\mathcal{A}}(n+1) \leq (C_2 n - C_3)f_u(\lceil \log_p(n+1) \rceil) + q, \quad (6)$$

where $C_1 = \frac{(q-1)(p-1)}{pq-1}$, $C_2 = \frac{q(q-1)(2p-1)+p}{q(p-1)}$, $C_3 = \frac{q^2-q+1}{q(p-1)}$. $\qquad \square$

It remains to turn down Condition (4). Let us consider an infinite word u not satisfying it and denote by Σ' the set of letters which occur in it an infinite number of times. Let the last occurrence of a letter from $\Sigma \backslash \Sigma'$ be the Mth letter of u; then $u' = S^M u$ satisfies (4) and thus Theorem 0.1 is valid for it: $a_{t_p(u')}(n) = \Theta(nf_{u'}(\lceil \log_p n \rceil))$. But for all $n \geq M$, we clearly have $f_u(n) = f_{u'}(n) + M$, so, $\Theta(nf_u(\lceil \log_p n \rceil)) = \Theta(n(f_{u'}(\lceil \log_p n \rceil) + M)) = \Theta(nf_{u'}(\lceil \log_p n \rceil))$ since $f_{u'}$ is a positive non-decreasing function. Then, due to Lemma 3.1, $\mathcal{A}(t_p(u)) = \cup_{n=0}^{M-1}\mathcal{F}(t_p(S^n u)) \cup \mathcal{A}(t_p(u')) \cup_{x \in \Sigma \backslash \Sigma'} x^*$, and thus $a_{t_p(u')}(n) \leq a_{t_p(u)}(n) \leq a_{t_p(u')}(n) + \sum_{n=0}^{M-1} f_{S^n u}(n) + \#(\Sigma \backslash \Sigma')$. Since each of $f_{S^n u}(n)$ grows linearly [20], we have $a_{t_p(u)}(n) \leq a_{t_p(u')}(n) + Dn + q'$ for some D and q'. Since $a_{t_p(u')}$ grows at least as $\Theta(n)$, this implies $a_{t_p(u)} = \Theta(a_{t_p(u')})$. So, since the theorem holds for u', it holds also for u. The proof of Theorem 0.1 is complete. $\qquad \square$

7. DISCUSSION

Let us look what Theorem 0.1 implies for specific cases.

First of all, note that (6) is valid for all u satisfying (4), and constants C_1, C_2 and C_3 depend only on the cardinality q of the alphabet and the prime length p of the Toeplitz patterns. In particular, if the alphabet is binary, then $C_1 = \frac{p-1}{2p-1}$ and $C_2 = \frac{5p-2}{2p-2}$. If q is arbitrary and $p \to \infty$, then $C_2/C_1 \to (2q^2 + 1)/(q - 1)$.

As we have mentioned, for all u, the Toeplitz word $t_p(u)$ is uniformly recurrent. Its subword complexity is $\Theta(n)$ due to one of the results of [20].

If u is ultimately periodic, *i.e.*, if $f_u(n) = P$ for some P and for all sufficiently large n, then $a_{t_p(u)}(n) = \Theta(n)$, in agreement with the results of [5,15]. In this case $t_p(n)$ is also p-automatic (see [3] for the definition and discussion of automatic words). Moreover, if u is k-periodic, then $t_p(u)$ is a fixed point of a p^k-uniform morphism.

Example 7.1. If $u = 01010\cdots$, then $t_p(u)$ is the fixed point of the morphism which maps $0 \to (0^{p-1}1)^{p-1}0^p$, $1 \to (0^{p-1}1)^p$. The lower and upper bounds for $\lim_{n\to\infty} a_{t_p(u)}/n$ due to (6) converge to 1 and 5 with $p \to \infty$. The precise upper and lower limits are 2 and 3, and they are minimal among all uniformly recurrent words [4].

If $f_u(n) = \Theta(n)$ (for example, if u is a Sturmian word, or the Thue-Morse word, or $0100110^41^40^81^8\cdots$), then $a_{t_p(u)}(n) = \Theta(n \log n)$, regardless of p. If $f_u(n) = \Theta(n \log n)$, then $a_{t_p(u)}(n) = \Theta(n \log n \log \log n)$. If $f_u(n) = \Theta(n^\alpha)$ for some α, then $a_{t_p(u)}(n) = \Theta(n(\log n)^\alpha)$.

If $f_u(n) = \Theta(\alpha^n)$ for some α, then $a_{t_p(u)}(n) = \Theta(n^{1+\log_p \alpha})$. For example, if $q = p$ and u is a sequence of maximal subword complexity $f_u(n) = p^n$ on the alphabet of cardinality p, then $a_{t_p(u)}(n) = \Theta(n^2)$. The maximal arithmetical complexity of a binary sequence in this family of examples is $\Theta(n^{1+\log_3 2})$, but by increasing q, we can obtain arbitrarily large powers of n.

At last, if the subword complexity of u grows intermediately between polynomials and exponentials, for instance, if u is one of Cassaigne's examples [7], then the arithmetical complexity of $t_p(u)$ also grows intermediately. For example, if $2^{\lceil \sqrt{n} \rceil} \leq f_u(n) \leq n^2 2^{\lceil \sqrt{n} \rceil}$, then $D_1 n 2^{\lceil \sqrt{\log_p n} \rceil} \leq a_{t_p(u)}(n) \leq D_2 n (\log_p n)^2 2^{\lceil \sqrt{\log_p n} \rceil}$ for some positive D_1 and D_2.

REFERENCES

[1] J.-P. Allouche, M. Baake, J. Cassaigne and D. Damanik, Palindrome complexity. *Theoret. Comput. Sci.* **292** (2003) 9–31.

[2] J.-P. Allouche and M. Bousquet-Mélou, Canonical positions for the factors in paperfolding sequences. *Theoret. Comput. Sci.* **129** (1994) 263–278.

[3] J.-P. Allouche and J. Shallit, *Automatic sequences: theory, applications, generalizations.* Cambridge Univ. Press (2003).

[4] S.V. Avgustinovich, J. Cassaigne and A.E. Frid, *Sequences of low arithmetical complexity.* submitted.

[5] S.V. Avgustinovich, D.G. Fon-Der-Flaass and A.E. Frid, Arithmetical complexity of infinite words, in *Words, Languages & Combinatorics III*, Words, Languages & Combinatorics III, Singapore (2003), 51–62 World Scientific Publishing. ICWLC 2000, Kyoto, Japan, March (2000) 14–18.

[6] J. Cassaigne, Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997) 67–88.

[7] J. Cassaigne, Constructing infinite words of intermediate complexity, in *Developments in Language Theory VI*, edited by M. Ito and M. Toyama. *Lect. Notes Comput. Sci.* **2450** (2003) 173–184.

[8] J. Cassaigne and A. Frid, *On arithmetical complexity of Sturmian words*, accepted to WORDS'05.

[9] J. Cassaigne and J. Karhumäki, Toeplitz words, generalized periodicity and periodically iterated morphisms. *Eur. J. Combin.* **18** (1997) 497–510.

[10] D. Damanik, Local symmetries in the period doubling sequence. *Discrete Appl. Math.* **100** (2000) 115–121.

[11] A. Iványi, On the *d*-complexity of words, Ann. Univ. Sci. Budapest. *Sect. Comput.* **8** (1987) 69–90.

[12] S. Ferenczi, Complexity of sequences and dynamical systems. *Discrete Math.* **206** (1999) 145–154.

[13] A. Frid, A lower bound for arithmetical complexity of Sturmian words. *Siberian Electronic Math. Reports* **2** (2005) 14–22.

[14] A. Frid, Arithmetical complexity of symmetric D0L words. *Theoret. Comput. Sci.* **306** (2003) 535–542.

[15] A. Frid, Sequences of linear arithmetical complexity. *Theoret. Comput. Sci.* **339** (2005) 68–87.

[16] T. Kamae and L. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words. *Ergodic Theory Dynam. Syst.* **22** (2002) 1191–1199.

[17] T. Kamae and L. Zamboni, Maximal pattern complexity for discrete systems. *Ergodic Theory Dynam. Syst.* **22** (2002), 1201–1214.

[18] T. Kamae and H. Rao, Maximal pattern complexity over ℓ letters. *Eur. J. Combin.*, to appear.

[19] T. Kamae and Y.-M. Xue, Two dimensional word with 2k maximal pattern complexity. *Osaka J. Math.* **41** (2004) 257–265.

[20] M. Koskas, Complexités de suites de Toeplitz. *Discrete Math.* **183** (1998) 161–183.

[21] I. Nakashima, J. Tamura and S. Yasutomi, Modified complexity and *-Sturmian words. *Proc. Japan Acad. Ser. A* **75** (1999) 26–28.

[22] I. Nakashima, J.-I. Tamura and S.-I. Yasutomi, *-Sturmian words and complexity. *J. Théorie des Nombres de Bordeaux* **15** (2003) 767–804.

[23] J.-E. Pin, *Van der Waerden's theorem*, in *Combinatorics on words*, edited by M. Lothaire. Addison-Wesley (1983) 39–54.

[24] A. Restivo and S. Salemi, Binary patterns in infinite binary words, in *Formal and Natural Computing*, edited by W. Brauer *et al. Lect. Notes Comput. Sci.* **2300** (2002) 107–116.

[25] B.L. Van der Waerden, Beweis einer Baudet'schen Vermutung. *Nieuw. Arch. Wisk.* **15** (1927) 212–216.