On the asymptotic behaviour of the abelian complexity of pure morphic binary words (extended abstract)

Markus Whiteland*
mawhit@utu.fi

June 2014

Abstract

Two words $u, v \in \Sigma^*$ are abelian equivalent, if the number of occurrences of a in u is equal to that of v for all letters $a \in \Sigma$. The abelian complexity function $\mathcal{P}_w^{\mathrm{ab}}$ of an infinite word w assigns to each $n \in \mathbb{N}$ the number of abelian equivalence classes of factors of length n. In this paper, we study the asymptotic growth of the abelian complexity of fixed points of binary morphisms. From a result of B. Adamczewski in 2003 concerning the balance function of words, we extract a full description of upper bounds of the asymptotic abelian complexity of primitive pure morphic words. We obtain a complete characterization of the tight upper and lower bounds of the asymptotic abelian complexity for non-primitive morphic binary words, completing the description of tight upper bounds for binary morphisms.

1 Introduction

The *factor complexity function* $\mathcal{P}_w$ of an infinite word $w \in \Sigma^{\mathbb{N}}$ counts, for each $n \in \mathbb{N}$, the number of distinct factors of w of length n. The notion has turned out to be a fundamental one, as shown by the theorem of M. Morse and G. A. Hedlund in [10] which characterises ultimately periodic words to be exactly the words admitting $\mathcal{P}(n_0) \leq n_0$ for some $n_0 \in \mathbb{N}$. For a survey on factor complexity we refer the reader to [3]. This notion has spawned other complexity functions of infinite words, see for instance [12, 8, 13]. The topic discussed in this work is the *abelian complexity* $\mathcal{P}^{\mathrm{ab}}$ of infinite words. The notion is close to that of the *balance* of infinite words and, in the case of binary words, they are essentially the same. Using this relation, the work of E. M. Coven and G. A. Hedlund in [4] can be translated into a characterisation of periodic words to be exactly the words for which $\mathcal{P}^{\mathrm{ab}}(n_0) = 1$ for some $n_0 \in \mathbb{N}$. Though the abelian complexity is a natural one, the study was formally initiated only recently by G. Richomme, K. Saari, and L. Q. Zamboni in [12].

The topic of this paper is the asymptotic abelian complexity of pure morphic words. As the asymptotic factor complexity of pure morphic infinite words is completely known by the result of J. J. Pansiot in [11] (see also [3]), it is natural to turn to other complexity classifications of such an important class of words. For example, B. Adamczewski in [1] gives a full description of the asymptotic upper bound behaviour of the balance function of primitive pure morphic words over any alphabet. The work on describing the abelian complexity for pure morphic words was initiated in [2], where F. Blanchet-Sadri and N. Fox, among other things, fully describe the asymptotic abelian complexities of uniform pure morphic binary words, apart from

*Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland, supported by grant no. ?

one case of lower bound behaviour. As the case of the balance function of primitive binary morphisms is essentially the same as the abelian complexity, we can extract from the result of [1], a classification of the upper abelian complexity of primitive pure morphic words. Note that the classification only holds for the lim sup behaviour. We study the case of non-primitive pure morphic words, which also was briefly and partially covered in [2], and we obtain a complete description of the asymptotic upper and lower bound abelian complexity. In the most interesting case of uniformly recurrent fixed points of non-primitive binary morphisms, we use the notion of a *descendant* or a *derived word* of a uniformly recurrent word in order to apply the result of [1] concerning the balance function of infinite words.

2 Preliminaries and notation

An alphabet is denoted by Σ, the set of finite words over Σ by Σ^*, and the set of infinite words over Σ by $\Sigma^{\mathbb{N}}$. The empty word is denoted by ε. We fix the binary alphabet to be $\{a, b\}$. The length of a word w is denoted by $|w|$. For a non-empty $u \in \Sigma^*$ and $w \in \Sigma^*$ we denote by $|w|_u$ the number of occurrences of u in w. The set of factors of an infinite word $w \in \Sigma^{\mathbb{N}}$ is denoted by $F(w)$ and factors of length n by $F_n(w)$. For a non-empty factor u of $w \in \Sigma^{\mathbb{N}}$ we denote the set of *complete first returns to u in w* by $\Re_u(w)$:

$$\Re_u(w) = \{x \in F(w) \mid x = ux_1 = x_2u,\ x_1, x_2 \in \Sigma^*, |x|_u = 2\}.$$

The factors of w in $\Re_u(w)u^{-1}$ are called *first returns* to u in w. See [14] for more on the notion.

A mapping $\varphi : \Delta^* \to \Sigma^*$ for two alphabets Δ and Σ is a *morphism*, if $\varphi(uv) = \varphi(u)\varphi(v)$ for all $u, v \in \Delta^*$. A morphism extends naturally to infinite words, and we will not distinguish between the extension and the morphism itself. A morphism is called *primitive*, if there exists a number $n_0 \in \mathbb{N}$ such that for all letters $a, b \in \Sigma$ we have $|\varphi^{n_0}(a)|_b \geq 1$. For an enumeration of the alphabet $\Sigma = \{a_1, a_2, \ldots, a_n\}$ and a morphism $\varphi : \Sigma^* \to \Sigma^*$, the *incidence matrix* of φ is the matrix $A_\varphi = \left(|\varphi(a_j)|_{a_i}\right)_{i,j\in\{1,\ldots n\}}$. In other words, the jth entry of the ith row equals the number of occurrences of a_i in $\varphi(a_j)$. For such a morphism φ we have $A_{\varphi^n} = A_\varphi^n$ for all $n \in \mathbb{N}$.

The abelian complexity function of an infinite word w assigns to each $n \in \mathbb{N}$ the number of abelian equivalence classes among the factors of w of length n. The abelian complexity function of $w \in \Sigma^{\mathbb{N}}$ is denoted by $\mathcal{P}_w^{\text{ab}}$. In general, $\mathcal{P}_w^{\text{ab}}$ can be strongly fluctuating (see [9]), and it might be rather hard to obtain closed formulas for the abelian complexity of a given word. We are thus interested in the asymptotic growth of the upper and lower bounds, so we define the functions $\mathcal{U}_w^{\text{ab}}$ and $\mathcal{L}_w^{\text{ab}}$ to be $\mathcal{U}_w^{\text{ab}}(n) = \max\{\mathcal{P}_w^{\text{ab}}(m) \mid 0 \leq m \leq n\}$ and $\mathcal{L}_w^{\text{ab}}(n) = \min\{\mathcal{P}_w^{\text{ab}}(m) \mid m \geq n\}$. For a word $w \in \Sigma^{\mathbb{N}}$ and a letter $a \in \Sigma$, we define $\max_{w,a}(n) = \max\{|u|_a \mid u \in F_n(w)\}$ and $\min_{w,a}(n) = \min\{|u|_a \mid u \in F_n(w)\}$. We omit w from the subscript when it is clear from context. Note that for any $w \in \{a, b\}^{\mathbb{N}}$, we have $\mathcal{P}_w^{\text{ab}} n = \max_{w,b}(n) - \min_{w,b}(n) + 1$ for all $n \in \mathbb{N}$.

For two functions $f, g : \mathbb{N} \to \mathbb{N}$ we use the notations $f(n) = \mathcal{O}(g(n))$ if there exists an $n_0 \in \mathbb{N}$ and a constant C such that $f(n) \leq Cg(n)$ for all $n \geq n_0$, $f(n) = \Omega(g(n))$ if $\limsup_{n\to\infty} f(n)/g(n) > 0$, and $f(n) = \Theta(g(n))$ if there exists an $n_0 \in \mathbb{N}$ and two constants C_1 and C_2 such that $C_1 g(n) \leq f(n) \leq C_2 g(n)$ for all $n \geq n_0$. If $f(n) = \mathcal{O}(g(n))$ and $f(n) = \Omega(g(n))$ then we denote this by $f(n) = (\mathcal{O} \cap \Omega)(g(n))$. Finally, we denote by $f(n) = o(g(n))$, if $\lim_{n\to\infty} f(n)/g(n) = 0$.

3 Main results

We invoke a result from [1] concerning a concept similar to that of the abelian complexity. We define *balance function* B_w of an infinite word $w \in \Sigma^{\mathbb{N}}$ as follows. For all $n \in \mathbb{N}$ define

$B_w(n) = \max\{max_{w,a}(n) - \min_{w,a}(n) \mid a \in \Sigma\}$. For a binary word, the functions are almost the same, that is, $B_u(n) = \mathcal{P}_u^{\mathrm{ab}}(n) - 1$ for all $n \in \mathbb{N}$.

We are now ready to state the main result used in this work. In the following, let $\sigma : \Sigma \to \Sigma^*$ be a primitive morphism such that $\sigma(a) = ax$ for some $x \in \Sigma^*$. It then admits a fixed point $u = \sigma^\omega(a)$. In [1] B. Adamczewski classifies the tight upper bounds of the asymptotic balance function of primitive pure morphic words. The upper bound of a primitive pure morphic word depends on the eigenvalues of the incidence matrix and, in some cases, *weights of admissible paths* in the *prefix automaton* corresponding to the morphism and word itself. Summarizing this result, a primitive pure morphic word $u \in \Sigma^{\mathbb{N}}$ can have $B_u(n) = (\mathcal{O} \cap \Omega)(1)$, $B_u(n) = (\mathcal{O} \cap \Omega)((\log n)^\alpha n^{\log_\theta \theta_2})$, or $B_u(n) = (\mathcal{O} \cap \Omega)((\log n)^\alpha)$ for some $\theta, \theta_2 \in \mathbb{R}$, $\theta > 1, \theta > \theta_2 > 0$ and $\alpha \in \mathbb{N}$, $0 \leq \alpha \leq |\Sigma| - 1$. For more details, we refer the reader to [1].

The asymptotic growth of $\mathcal{U}_u^{\mathrm{ab}}$ for primitive morphic binary words $u \in \{a, b\}^{\mathbb{N}}$ are easy to extract from the result. For example, in the above, $\alpha = 0$ always. Summarizing this, u can have $\mathcal{P}_u^{\mathrm{ab}}(n) = \Theta(1)$, $\mathcal{U}_u^{\mathrm{ab}}(n) = \Theta(n^{\log_\theta |\theta_2|})$, or $\mathcal{U}_u^{\mathrm{ab}}(n) = \Theta(\log n)$.

We expand this result to all binary morphisms. Let φ be a non-primitive morphism admitting an iterated fixed point $\varphi^\omega(a) = y$. Note that if $|\varphi(a)|_b \geq 1$, then $\varphi(b) \in b^*$ for φ to be non-primitive. Moreover, for y to be aperiodic, we have to have $|\varphi(a)|_b \geq 1$. Indeed, it can be shown that $y \in \Sigma^{\mathbb{N}}$ is ultimately periodic if and only if one of the following holds: $\varphi(a) \in \{a^+, ab^+\}$, $\varphi(b) = \varepsilon$, or $|\varphi(b)|_b = 1$, $\varphi(a)$ ends with a and $|\Re_a(\varphi(a))| = 1$.

By the theorem of E. Coven and G. A. Hedlund [4], an infinite word w is periodic if and only if $\mathcal{P}_w^{\mathrm{ab}}(n) = 1$ for some $n \in \mathbb{N}$. It is a straightforward consequence, that $\mathcal{P}_w^{\mathrm{ab}}(n) = \Theta(1)$ for ultimately periodic words w. We are thus only interested in aperiodic words.

Theorem 3.1. *Let φ and y be as above, and suppose that y is aperiodic. Then the abelian complexity is one of the following.*

i) If $|\varphi(b)|_b = 1$ and $\varphi(a)$ ends in a, then $\mathcal{L}_y^{ab}(n) = \Theta(1)$ and $\mathcal{U}_y^{ab}(n) = \Theta(\log n)$,

ii) if $|\varphi(b)|_b = 1$ and $\varphi(a)$ ends in b, then $\mathcal{P}_y^{ab}(n) = \Theta(n)$,

iii) if $|\varphi(b)|_b > 1$, $|\varphi(b)|_b > |\varphi(a)|_a$, then $\mathcal{P}_y^{ab}(n) = \Theta\left(n^{\log_{|\varphi(b)|_b} |\varphi(a)|_a}\right)$

iv) if $|\varphi(b)|_b > 1$, $|\varphi(b)|_b = |\varphi(a)|_a$, then $\mathcal{P}_y^{ab}(n) = \Theta(n/\log n)$,

v) if $|\varphi(b)|_b > 1$, $|\varphi(b)|_b < |\varphi(a)|_a$, then $\mathcal{P}_y^{ab}(n) = \Theta(n)$.

Remark 3.2. In [2], F. Blanchet-Sadri and N. Fox give a brief sketch of proof for the items *ii)*–*v)*. Nonetheless, we inspect these cases as well.

4 Sketch of the proof of Theorem 3.1

In this section we give a sketch of the proof of Theorem 3.1. The most interesting case of Theorem 3.1 is item *i)* and we focus on the proof of it more than the others. First we give a very brief sketch of items *ii)*–*v)* as the proofs are fairly straightforward.

We first we observe that in each of these cases $b^n \in F(y)$ for all $n \in \mathbb{N}$. This implies that $\mathcal{P}_y^{\mathrm{ab}}(n) = n - \min_b(n) + 1$ for all $n \in \mathbb{N}$ and, in particular, $\mathcal{P}_y^{\mathrm{ab}}(n)$ is monotonically increasing.

The second observation is that $\min_{y,b}(|\varphi^n(a)|) = |\varphi^n(a)|_b$. This gives a direct method of computing $\mathcal{P}_y^{\mathrm{ab}}$ for certain lengths. The growth of the sequence of numbers $|\varphi^n(a)|$ is compared to that of $\mathcal{P}_y^{\mathrm{ab}}(|\varphi^n(a)|)$ in each of the items *ii)*–*v)*: In item *iii)*, the growth of the sequences are $|\varphi(b)|_b^n$ versus $|\varphi(a)|_a^n$, in item *iv)* the ratio is $n\rho^n$ versus ρ^n for a $\rho > 1$, and in item *v)* we have

$|\varphi^n(a)|^n$ versus $|\varphi^n(a)|^n$. The property of $\mathcal{P}_y^{\mathrm{ab}}$ being monotonically increasing then gives valid bounds for $\mathcal{P}_y^{\mathrm{ab}}$ rather than to the upper or lower bound functions $\mathcal{U}_y^{\mathrm{ab}}$ and $\mathcal{L}_y^{\mathrm{ab}}$.

We then proceed to the case of item i). It is more challenging and interesting than the other items. First of all, the abelian complexity function is fluctuating, so it is not enough to find a sub-sequence for which we know the behaviour. This is why we need derived words, for which the result of Adamczewski comes into play.

In the following, we suppose that φ is a morphism and y the fixed point as in item i) of Theorem 3.1. The aim of this section is to prove the claim of item i), that is, $\mathcal{L}_y^{\mathrm{ab}}(n) = \Theta(1)$ and $\mathcal{U}_y^{\mathrm{ab}}(n) = \Theta(\log n)$. For the first, it suffices to find a monotone sequence $(m_n)_{n\in\mathbb{N}}$ such that the values $\mathcal{P}_y^{\mathrm{ab}}(m_n)$ are bounded.

Lemma 4.1. *Let φ and y be as above. Then y is linearly recurrent.*

Remark 4.2. Linearly recurrent words admit uniform frequencies, in particular, the frequencies λ_a, λ_b of the letters a and b, resp., exist. In the case of item i) it is easy to calculate λ_b: $\lambda_b = \frac{|\varphi(a)|_b}{|\varphi(a)|-1}$. Note that $1 \neq \lambda_b \neq 0$. See [6] for references on linearly recurrent words.

Proposition 4.3. *Let φ and y be as stated above. Then $\mathcal{L}_y^{ab}(n) = \Theta(1)$.*

Proof. We have $a \mapsto ab^{k_1}ab^{k_2}\cdots ab^{k_m}a$ and $b \mapsto b$ where $k_i \geq 0$ for all $i = 1, \dots, m$. Let us fix $n \geq 1$ and denote by $u_n = \varphi^n(a)$. We shall study the factors of length $|u_n|$. First of all, as $\varphi(b) = b$, it is easy to see that $y = \prod_{i=0}^{\infty} ab^{s_i}$ where $s_i \in \{k_1, \dots, k_m\}$ for all $i \in \mathbb{N}$. Furthermore, $y = \varphi^n(y) = \prod_{i=0}^{\infty} u_n b^{s_i}$.

Consider then any factor v of length $|u_n|$. It is then a factor of the word $u_n b^s u_n$ for some $s \in \{k_1, \dots, k_m\}$. Then v is of form qb^t, $qb^s q$, or $b^t q$ for some prefix p and suffix q of w and some $t \in \{0, \dots, s\}$. Then u_n has the minimal number of b's among factors of the same length. Furthermore $|v|_b \leq |u|_b + k_{\max}$, where $k_{\max} = \max\{k_1, \dots, k_m\}$, for any $v \in F_{|u_n|}$. The claim follows, as $\max_b(|u_n|) - \min_b(|u_n|) + 1 \leq k_{\max} + 1$ for any $n \in \mathbb{N}$. □

Then for the asymptotics of $\mathcal{U}_y^{\mathrm{ab}}(n)$. The method is to show first that $\mathcal{U}_y^{\mathrm{ab}}(n) = \mathcal{O}(\log n)$ and then giving a sequence of indices $(m_n)_{n\in\mathbb{N}}$ that grow exponentially, while the sequence $\left(\mathcal{P}_y^{\mathrm{ab}}(m_n)\right)_{n\in\mathbb{N}}$ grows linearly.

4.1 On descendants of uniformly recurrent words

First we discuss the notion of a *descendant* or a *derived word* of a uniformly recurrent word. The notion was introduced by F. Durand in [5] and by C. Holton and L. Q. Zamboni in [7], independently.

Let $\varphi : \Sigma \to \Sigma^*$ be a morphism admitting a fixed point $\varphi^\omega(a) = y$. Furthermore, suppose that y is uniformly recurrent, or, equivalently, $\Re_v(y)$ is finite for each $v \in F(y)$. For any prefix u of y, we define a finite alphabet Δ with $|\Delta| = |\Re_u(y)|$. Let then $\pi : \Delta \to \Re_u(y)$ be a bijection. The mapping π then extends to a morphism in a natural way. A *descendant* of y with respect to u is the infinite word $D_u(y)$ obtained by coding the first returns into the corresponding letters (there is no problem with this, we identify the occurrences of u and we can then see the returns between the occurrences and code them, etc.). Since y is uniformly recurrent, α is also uniformly recurrent. We then have $\pi(D_u(y)) = y$.

The following is a modification of Proposition 5.1 in [5]. The proof of this is essentially the same, as suggested by J. Peltomäki (personal communication).

Proposition 4.4. *Let $y = \gamma^\omega(a)$ be a uniformly recurrent pure morphic word and u a finite prefix of y. Let $D_u(y)$ be the descendant of y with respect to u. Then $D_u(y)$ is a pure morphic word fixed by a primitive morphism.*

We are actually more interested in the construction of the primitive morphism, denoted by μ, mentioned above. The morphism μ is defined for each letter $a \in \Delta$ to be $\mu(a) = \pi^{-1}\gamma\pi(a)$. The reason we are interested in such descendants is that the balancedness of them is known by Adamczewski's classification in [1]. It will then help us establish an upper bound for the lim sup growth of the abelian complexity of y.

4.2 The asymptotic growth of $\mathcal{U}_y^{\mathbf{ab}}$

Let now φ and $y = \varphi^\omega(a)$ be as in item *i)* in Theorem 3.1. Let $D_a(y)$ be the descendant of y with respect to the letter a and μ the primitive morphism fixing it. As was noted during the proof of Proposition 4.3, the first returns to a in y can directly be seen from $\varphi(a)$. Also, as in any binary word, the returns to a in $\varphi(a)$ are of form ab^i for some $i \in \mathbb{N}$. We thus take the alphabet Δ to be $\{a_i \mid ab^i \in \Re_a(y)\}$. For ease of notation, we denote by $d = |\Delta|$.

We then have the morphism $\pi : \Delta \to \{a, b\}^*$ defined on the letters by $\pi(a_i) = ab^i$. The construction of μ then gives us $\mu(a_i) = \pi^{-1}\varphi\pi(a_i) = \pi^{-1}\varphi(ab^i) = \pi^{-1}(\varphi(a)b^i) = \pi^{-1}(\varphi(a)a^{-1})a_i$ for all $a_i \in \Delta$. Note that this implies that μ is a *uniform* morphism, i.e. $|\mu(a)| = |\mu(b)|$ for all $a, b \in \Sigma$. The length of μ is $|\mu| = |\pi^{-1}(\varphi(a)a^{-1})| + 1 = |\varphi(a)|_a$. Moreover, the images of letters differ only at the last letter, that is, if we denote $\pi^{-1}(\varphi(a)a^{-1}) = p$, then $\mu(a_i) = pa_i$ for all $a_i \in \Delta$.

The incidence matrix A_μ is of special form $A_\mu = \left(\Psi(p)^T \mid \Psi(p)^T \mid \cdots \mid \Psi(p)^T\right) + I_{d\times d} = A + I_{d\times d}$, where $\Psi(p)$ is the *Parikh vector* of p, $I_{d\times d}$ is the $d \times d$ *identity matrix*, and A is a $d \times d$ matrix, where each column is the same vector $\Psi(p)^T$. Note that since all the first returns to a occur in $\varphi(a)$, all the entries of A_μ are positive.

We are fortunate that enough information of the values θ, θ_2, and α mentioned in the beginning of section 3 can be extracted from A_μ with ease. Indeed, it can be shown that $B_u(n) = \mathcal{O}(\log n)$. This is enough for our considerations.

Lemma 4.5. *Let φ and y be as above. Then $\mathcal{U}_y^{ab}(n) = \mathcal{O}(\log n)$.*

Proof. Let $D_a(y)$, $\mu : \Delta \to \Delta$ and $\pi : \Delta \to \Sigma^*$ be as discussed above. Let $u, v \in F_n(D_a(y))$ be such that $|\pi(u)| - |\pi(v)|$ is maximal. Recall that $\pi(a_i) = ab^i$ for all $a_i \in \Delta$, so that $|\pi(u)|_a = |\pi(v)|_a$. We then have, by the above considerations,

$$\big||\pi(u)| - |\pi(v)|\big| = \big||\pi(u)|_b - |\pi(v)|_b\big| = \left|\sum_{a_i\in\Delta} i(|u|_{a_i} - |v|_{a_i})\right| \leq \sum_{a_i\in\Delta} i\big||u|_{a_i} - |v|_{a_i}\big| = \mathcal{O}(\log n).$$

Let $r = |\pi(u)| - |\pi(v)|$. Then there exists a factor $x \in F_r(y)$ such that $\pi(v)x \in F(y)$. We have $|x|_b = \lambda_b r + o(r)$, where $\lambda_b < 1$ is the uniform frequency of b given in Remark 4.2 in y. Now $|\pi(u)|_b - |\pi(v)x|_b = r - \lambda_b r + o(r) = \Theta(r)$ since $\lambda_b \neq 1 \neq \lambda_a$. Note that $\max_b(|\pi(u)|) - |\pi(u)|_b$ is bounded by a constant and $|\pi(v)x|_b - \min_b(|\pi(u)|)$ is bounded by $o(r)$.

Since $r = \mathcal{O}(\log n)$ and $|\pi(u)| = \Theta(n)$, we have $\mathcal{P}_y^{\mathrm{ab}}(n) = \mathcal{O}(\log n)$. The difference $\max\{|\pi(u)| \mid u \in F_{n+1}(y)\} - \max\{|\pi(v)| \mid v \in F_n(y)\}$ is also bounded by a constant, so the estimations hold for all $n \in \mathbb{N}$. □

Proposition 4.6. *Let φ and y be as in item i) of Theorem 3.1. Then $\mathcal{U}_y^{ab}(n) = \Theta(\log n)$.*

Proof. Lemma 4.5 already gives us $\mathcal{U}_y^{\text{ab}}(n) = \mathcal{O}(\log n)$. The proof follows once we establish such growth. Consider thus a factorization $\varphi(a) = xauay$ for some $x, u, y \in \{a, b\}^*$. Let then $u_0 \in F(y)$ be of form $u_0 = au'a$ for some $u' \in F(y)$. Let us then define u_n, $n \geq 0$ recursively as follows: $u_{n+1} = x^{-1}\varphi(u_n)y^{-1}$. Note that u_n is defined for all $n \in \mathbb{N}$, since u_n always begins and ends with a and so $\varphi(u_n)$ always begins with x and ends with y.

Let us now compute $|u_n|$ and $|u_n|_b$ for each n. For this, we have the following. Let A_φ be the incidence matrix of φ. Denote by $\Psi_{x,y}(u_n) = (|u_n|_a, |u_n|_b, -|xy|_a, -|xy|_b)^T$ and consider the 4×4 block matrix

$$A = \begin{pmatrix} A_\varphi & \mathbf{I}_{2\times 2} \\ \mathbf{0} & \mathbf{I}_{2\times 2} \end{pmatrix},$$

where $\mathbf{0} = \mathbf{0}_{2\times 2}$ is the 2×2 zero matrix. It is easy to see that $A^n\Psi_{x,y}(u_0) = \Psi_{x,y}(u_n)$ for all $n \in \mathbb{N}$. By denoting $f_a = |\varphi(a)|_a$ and $f_b = |\varphi(a)|_b$, we have

$$\Psi_{x,y}(u_s) = A^s\Psi_{x,y}(u_0) = \begin{pmatrix} A_\varphi^s & \sum_{i=0}^{s-1} A_\varphi^i \\ \mathbf{0} & \mathbf{I}_{2\times 2} \end{pmatrix} \Psi_{x,y}(u_0) = \begin{pmatrix} f_a^s & 0 & \frac{f_a^s-1}{f_a-1} & 0 \\ f_b\frac{f_a^s-1}{f_a-1} & 1 & f_b\frac{f_a^s-1-(f_a-1)s}{(f_a-1)^2} & s \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Psi_{x,y}(u_0).$$

for all $s \in \mathbb{N}$. We are now able to calculate the desired values of $|u_s|_a = f_a^s|u_0|_a - \frac{f_a^s-1}{f_a-1}|xy|_a$ and $|u_s|_b = f_b\frac{f_a^s-1}{f_a-1}|u_0|_a + |u_0|_b - f_b\frac{f_a^s-1-s(f_a-1)}{(f_a-1)^2}|xy|_a - s|xy|_b$. We then compare the latter to the average number of b's, $\lambda_b|u_s|$, where $\lambda_b = \frac{f_b}{f_a+f_b-1}$ is as in Remark 4.2. Now $|u_s|_b - \frac{f_b}{f_a+f_b-1}|u_s| = c_1 + c_2 s$ where c_1 and c_2 are constants after u_0, x, and y are fixed. If we choose x and y such that $\varphi(a) = xab^{k_{\text{min}}}ay$, where $k_{\text{min}} = \min\{k_i \mid 0 \leq i \leq m\}$, then $c_2 = \frac{f_b|xy|_a-(f_a-1)|xy|_b}{f_a+f_b-1} \neq 0$. This is true, since $c_2 = 0$ if and only if $\frac{f_b}{f_a-1} \neq \frac{|xy|_b}{|xy|_a}$ where the left hand side is the average number of b's in a first return to a in $\varphi(a)$. Since $|\Re_a(\varphi(a))| \geq 2$, the average cannot equal the minimal. Now $(|u_n|)$ grows exponentially while $\left(\mathcal{P}_y^{\text{ab}}(|u_n|)\right)$ has at least linear growth, which gives the claim. □

5 Conclusions

The main focus of the work is on pure morphic binary morphisms. The main reason for this, was to complete the description of the possible asymptotic abelian complexities of such words. Future directions of research include describing the lower asymptotic abelian complexity function for primitive binary morphisms, as well as studying the case of larger alphabets.

The difficulty in the case of lower bound asymptotics for primitive binary words is that there are words for which abelian complexity is fluctuating (e.g. $a \mapsto aba$, $b \mapsto abb$, see [2]) and it is not clear, if B. Adamczewski's technique used in [1], can be modified to obtain lower bound asymptotics. The techniques used in this work cannot be extended to larger alphabets, since the abelian complexity cannot be directly seen from the balance function. On the other hand, the following is proved in [12]. For a word $w \in \Sigma^{\mathbb{N}}$, the function $\mathcal{P}_w^{\text{ab}}$ is bounded, if and only if B_w is bounded. Thus the case of primitive words with bounded balance in the classification of Adamczewski have bounded abelian complexity.

Acknowledgements

I would like to express my gratitude to my advisors Juhani Karhumäki and Svetlana Puzynina for both the mathematical and practical guidance during the preparation of this work. I would also like to thank Jarkko Peltomäki for useful discussions on the topic of combinatorics on words.

References

[1] B. Adamczewski: *Balances for fixed points of primitive substitutions.* Theoretical Computer Science 307 (2003), pp. 47–75.

[2] F. Blanchet-Sadri and N. Fox: *On the Asymptotic Abelian Complexity of Morphic Words.* DLT 2013, LNCS 7907 (2013), pp. 94–105.

[3] J. Cassaigne and F. Nicolas: *Factor complexity* In V. Berthé and M. Rigo (Eds.), *Combinatorics, Automata and Number Theory* (2010), pp. 163–247, Encyclopedia Math. Appl., 135, Cambridge Univ. Press, Cambridge.

[4] E. M. Coven and G. A. Hedlund: *Sequences with minimal block growth.* Math. Systems Theory 7 (1973), pp. 138–153.

[5] F. Durand: *A characterization of substitutive sequences using return words.* Discrete Math. 179 (1998), 89–101.

[6] S. Ferenczi and T. Monteil: *Infinite words with uniform frequencies, and invariant measures* In V. Berthé and M. Rigo (Eds.), *Combinatorics, Automata and Number Theory* (2010), pp. 374–417, Encyclopedia Math. Appl., 135, Cambridge Univ. Press, Cambridge.

[7] C. Holton and L. Q. Zamboni: *Descendants of primitive substitutions*, Theory Comput. Systems 32 (1999), pp. 133–157.

[8] J. Karhumäki, A. Saarela, and L. Q. Zamboni: *On a generalization of Abelian equivalence and complexity of infinite words.* J. Combin. Theory Ser. A 120.8 (2013), pp. 2189–2206.

[9] B. Madill and N. Rampersad: *The abelian complexity of the paperfolding word.* Preprint: arXiv:1208.2856v1.

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

[11] J. J. Pansiot: *Complexité des facteurs des mots infinis engendrés par morphismes itérés.* ICALP 1984, LNCS 172 (1984), pp. 380–389.

[12] G. Richomme, K. Saari, and L. Q. Zamboni: *Abelian Complexity in Minimal Subshifts.* J. London Math. Soc. 83 (2011), 79–95.

[13] M. Rigo, P. Salimov: *Another Generalization of Abelian Equivalence: Binomial Complexity of Infinite Words.* WORDS 2013, LNCS 8079 (2013), pp. 217–228.

[14] L. Vuillon: *A Characterization of Sturmian Words by Return Words.* Europ. J. Combinatorics 22 (2001), pp. 263–275.