

Every nonnegative real number is an abelian critical exponent

Jarkko Peltomäki^{1,2,3}[0000-0003-3164-1559] and Markus A. Whiteland³

¹ The Turku Collegium for Science and Medicine TCSM, University of Turku, Turku, Finland

² Turku Centre for Computer Science TUCS, Turku, Finland

³ University of Turku, Department of Mathematics and Statistics, Turku, Finland
{jspelt,mawhit}@utu.fi

Abstract. The abelian critical exponent of an infinite word w is defined as the maximum ratio between the exponent and the period of an abelian power occurring in w . It was shown by Fici et al. that the set of finite abelian critical exponents of Sturmian words coincides with the Lagrange spectrum. This spectrum contains every large enough positive real number. We construct words whose abelian critical exponents fill the remaining gaps, that is, we prove that for each nonnegative real number θ there exists an infinite word having abelian critical exponent θ . We also extend this result to the k -abelian setting.

Keywords: abelian equivalence · k -abelian equivalence · critical exponent · Sturmian word

1 Introduction

The study of powers and their avoidance has been one of the central themes in combinatorics on words; see [2, Ch. 4]. The central notion here is that of the *critical exponent* which measures the maximum exponent of a power occurring in a given word. Recently it has been popular to generalize the notion of a power using some equivalence relation in place of the usual equality of words. For example, abelian equivalence (see the references of [6]), and its generalizations k -abelian equivalence [9,3] and binomial equivalence [19,16] have been popular options.

Two words u and v are *abelian equivalent*, written $u \sim v$, if they are permutations of each other. An *abelian power* of exponent e and period m is a word of the form $u_0 \cdots u_{e-1}$ such that $m = |u_0|$ and u_0, \dots, u_{e-1} are nonempty and abelian equivalent. For example, $01 \cdot 10$ (a square) and $abc \cdot bca \cdot cab$ (a cube) are abelian powers. Now it is possible to define the abelian critical exponent of an infinite word as the maximum exponent of an abelian power occurring in it. However, this does not give any interesting information on abelian powers occurring in Sturmian words or, more generally, in words with bounded abelian complexity because such words contain abelian powers of arbitrarily high exponent [18]. In order to capture more information on abelian powers of an infinite

word to a single quantity, it was proposed in [6] to define the *abelian critical exponent* $\mathcal{A}c(\mathbf{w})$ of an infinite word \mathbf{w} as the quantity

$$\limsup_{m \rightarrow \infty} \frac{\mathcal{A}e_{\mathbf{w}}(m)}{m},$$

where $\mathcal{A}e_{\mathbf{w}}(m)$ is the supremum of exponents of abelian powers of period m occurring in \mathbf{w} . This notion turns out to be much more interesting. For example, $\mathcal{A}c(\mathbf{f}) = \sqrt{5}$ for the Fibonacci word \mathbf{f} , the fixed point of the substitution $0 \mapsto 01$, $1 \mapsto 0$ [6, Thm. 5.14]. Furthermore $\sqrt{5}$ is the minimum abelian critical exponent among all Sturmian words [6, Thm. 5.14]. It follows that for each Sturmian word \mathbf{s} and each $\delta > 0$, there exists an increasing sequence (m_i) of integers such that \mathbf{s} contains an abelian power of period m_i and total length greater than $(\sqrt{5} - \delta)m_i^2$. Notice that if \mathbf{w} does not contain abelian powers with arbitrarily large exponent, then $\mathcal{A}c(\mathbf{w}) = 0$. Many examples of such words are known; see, e.g., [2, Ch. 4.6]. It is also possible that $\mathcal{A}c(\mathbf{w}) = \infty$. Take for example the Thue-Morse word \mathbf{t} , the fixed point of the substitution $0 \mapsto 01$, $1 \mapsto 10$. Indeed, it is straightforward to see that \mathbf{t} can be factored as a product of abelian equivalent words of length $2n$ for all $n \geq 0$. This shows that $\mathcal{A}c(\mathbf{t}) = \infty$.

Further study in [6] showed the surprising fact that the set of finite abelian critical exponents of Sturmian words equals the Lagrange spectrum \mathcal{L} . The Lagrange constant of an irrational α is the infimum of the real numbers λ such that for every $c > \lambda$ the inequality $|\alpha - n/m| < 1/cm^2$ has only finitely many rational solutions n/m . The Lagrange spectrum is the set of finite Lagrange constants of irrational numbers. The Lagrange spectrum has been extensively studied in number theory since the works of Markov [12,13] in the 19th century. The famous theorems of Markov show that the initial part of \mathcal{L} inside the interval $[\sqrt{5}, 3)$ is discrete. Later in 1947 Hall proved that \mathcal{L} contains a half-line [8]. After a series of improvements by multiple authors, it was finally determined by Freiman in 1975 [7] that the largest half-line contained in the Lagrange spectrum is $[c_F, \infty)$, where

$$c_F = \frac{2221564096 + 283748\sqrt{462}}{491993569} = 4.5278295661\dots$$

Good sources for information on the Lagrange spectrum are the monograph of Cusick and Flahive [4] and Aigner's book [1]. See also the recent book [17] of Reutenauer for a more word-combinatorial flavor.

The connection between the Lagrange spectrum and abelian critical exponents of Sturmian words shows that each real number larger than c_F is the abelian critical exponent of some infinite word. This raises the obvious question of whether this can be extended to hold for all nonnegative numbers. In this paper, we answer the question in the positive. The main result of this paper is the following theorem.

Theorem 1. *Let θ be a nonnegative real number. Then there exists an infinite word \mathbf{w} such that $\mathcal{A}c(\mathbf{w}) = \theta$. The word \mathbf{w} can be taken over an alphabet of at most three letters.*

This result should be compared with a result of Krieger and Shallit stating that every real number $\theta > 1$ is a critical exponent (in the usual sense) of some infinite word [10]. Notice that here the number of letters required tends to infinity when θ tends to 1 [10], but in our setting we need at most three letters.

We prove an analogue of Theorem 1 for k -abelian critical exponents; see Section 3 for the extension and the necessary definitions.

Our proof method is to exploit the properties of the Lagrange spectrum, that is, the fact that Theorem 1 is already known to be true for all reals greater than c_F . The idea is to find a suitable N -uniform substitution σ such that each abelian power in $\sigma(\mathbf{w})$ can be decoded to an abelian power in \mathbf{w} with the same exponent. This means, in essence, that the abelian powers in $\sigma(\mathbf{w})$ are the abelian powers of \mathbf{w} blown up by a factor of N . Roughly speaking, the ratio of exponents and periods corresponding to $\mathcal{A}(\mathbf{w})$ gets divided by N , that is, $\mathcal{A}(\sigma(\mathbf{w})) = \mathcal{A}(\mathbf{w})/N$. The conclusion is that Theorem 1 is true for each real in the interval $[c_F/N, \infty)$, where $[c_F, \infty)$ is the largest half-line contained in the Lagrange spectrum. We may choose N to be arbitrarily large, so Theorem 1 follows. The extension of Theorem 1 to the k -abelian setting is proved using the same ideas.

We use the usual notions and notation from combinatorics on words. If the reader encounters anything undefined, we refer him or her to [11]. Even though we mention Sturmian words several times in this paper, we do not need any properties of these binary words. For their definition, we refer the reader to [11, Ch. 2] and [14, Ch. 4].

2 Proof of Theorem 1

Let θ be a nonnegative real number. If $\theta = 0$, then θ is the abelian critical exponent of any infinite word that avoids abelian powers with large enough exponent. Such words exist by [5] (abelian fourth powers are avoidable over two letters); see also [2, Ch. 4.6].

Assume then that $\theta > 0$, and let N be an integer such that $N\theta \in [c_F, \infty)$. Let \mathbf{w} be an infinite binary word. Our aim is to find an N -uniform substitution f defined on a two-letter alphabet with the following properties:

- (i) If an abelian power $u_0 \cdots u_{e-1}$ occurs in \mathbf{w} , then $f(u_0) \cdots f(u_{e-1})$ is an abelian power occurring in $f(\mathbf{w})$.
- (ii) If an abelian power $u_0 \cdots u_{e-1}$, $e \geq N$, occurs in $f(\mathbf{w})$, then \mathbf{w} contains an abelian power $v_0 \cdots v_{e-1}$ with $|v_0| = |u_0|/N$.

Let us show how to prove Theorem 1 under the assumption that such f exists.

Let \mathbf{s} be a Sturmian word having $\mathcal{A}(\mathbf{s}) = N\theta$. In fact, any binary word \mathbf{s} with $\mathcal{A}(\mathbf{s}) = N\theta$ will do, we just know that such a Sturmian word exists by the results of [6]. We claim that $\mathcal{A}(f(\mathbf{s})) = \theta$. This proves Theorem 1 when $\theta > 0$ (assuming that $f(\mathbf{w})$ has at most three letters).

By Property (i), we have $\mathcal{A}_{f(\mathbf{s})}(tN) \geq \mathcal{A}_{\mathbf{s}}(t)$ for all positive integers t . Since $\mathcal{A}(\mathbf{s}) > 0$, the word \mathbf{s} contains abelian powers of arbitrarily high exponent, and thus by Property (i) the word $f(\mathbf{s})$ contains abelian powers of arbitrarily

high exponent and period divisible by N . If $\mathcal{A}e_{f(\mathbf{s})}(tN) \geq N$, then $\mathcal{A}e_{f(\mathbf{s})}(tN) \leq \mathcal{A}e_{\mathbf{s}}(t)$ by Property (ii). Therefore there exists a sequence (t_i) such that $\mathcal{A}e_{\mathbf{s}}(t_i) = \mathcal{A}e_{f(\mathbf{s})}(t_i N)$ for all i . Hence

$$\limsup_{i \rightarrow \infty} \frac{\mathcal{A}e_{f(\mathbf{s})}(t_i N)}{t_i N} = \limsup_{i \rightarrow \infty} \frac{\mathcal{A}e_{\mathbf{s}}(t_i)}{t_i N} = \frac{1}{N} \limsup_{i \rightarrow \infty} \frac{\mathcal{A}e_{\mathbf{s}}(t_i)}{t_i} = \frac{1}{N} \mathcal{A}c(\mathbf{s}) = \theta,$$

so $\mathcal{A}c(f(\mathbf{s})) \geq \theta$. If $\mathcal{A}c(f(\mathbf{s})) > \theta$, then there exists an increasing sequence (ℓ_i) such that

$$\frac{\mathcal{A}e_{f(\mathbf{s})}(\ell_i)}{\ell_i} > \theta > 0$$

for all i . By the preceding, only finitely many of the numbers in the sequence (ℓ_i) are divisible by N . By Property (ii), we thus have $\mathcal{A}e_{f(\mathbf{s})}(\ell_i) \geq N$ only for finitely many i meaning that

$$\frac{\mathcal{A}e_{f(\mathbf{s})}(\ell_i)}{\ell_i} < \frac{N}{\ell_i}$$

for i large enough. This is impossible as $N/\ell_i \rightarrow 0$ as $i \rightarrow \infty$. The conclusion is that $\mathcal{A}c(f(\mathbf{s})) = \theta$. This concludes the proof of Theorem 1.

Let us then show how to choose a suitable substitution f . Let N be a fixed positive and *even* integer, and define the N -uniform substitution $\sigma: \{0, 1\}^* \rightarrow \{0, 1, \#\}^*$ by

$$\begin{aligned} 0 &\mapsto \#0^{N-1}, \\ 1 &\mapsto \#1^{N-1}. \end{aligned}$$

Lemma 2. *The substitution σ satisfies Property (i).*

Proof. Property (i) trivially holds for any nonerasing substitution. \square

Before showing that the substitution σ satisfies Property (ii), we show that the period of an abelian power with large enough exponent is divisible by N , the length of the substitution σ .

Lemma 3. *Let \mathbf{w} be an infinite binary word. If an abelian power $u_0 \cdots u_{e-1}$, with $e \geq N$, occurs in $\sigma(\mathbf{w})$, then N divides $|u_0|$.*

Proof. Let $m = |u_0|$, and write $m = tN + r$ for some $t \geq 0$ and $0 \leq r < N$. The claim is thus that $r = 0$. Assume, for a contradiction, that $r > 0$. Observe that for $\sigma(\mathbf{w}) = a_0 a_1 \cdots$, where $a_n \in \{0, 1, \#\}$ for each $n \geq 0$, we have $a_n = \#$ if and only if $n \equiv 0 \pmod{N}$. Let us denote the position of the occurrence of u_j in $\sigma(\mathbf{w})$ by i_j , that is,

$$u_j = a_{i_j} a_{i_j+1} \cdots a_{i_j+m-1}.$$

Observe that $i_j = i_0 + jm$, and $i_j \equiv i_0 + jr \pmod{N}$ for each $j = 0, \dots, e-1$. Notice also that the number of occurrences of the letter $\#$ in u_j equals the number of indices k in the set $\{i_j, i_j + 1, \dots, i_j + m - 1\}$ for which $k \equiv 0$

(mod N). Let $n_j = i_j \pmod N$. If $n_j = 0$, then we may compute the value $|u_j|_{\#}$ as follows:

$$|u_j|_{\#} = \left\lceil \frac{m}{N} \right\rceil = \left\lceil \frac{tN + r}{N} \right\rceil = t + \left\lceil \frac{r}{N} \right\rceil = t + 1$$

since $0 < r < N$ by assumption. If $n_j > 0$, then none of the first $N - n_j$ letters of u_j equals $\#$. The value $|u_j|_{\#}$ is thus computed as follows:

$$|u_j|_{\#} = \left\lceil \frac{m - (N - n_j)}{N} \right\rceil = \left\lceil \frac{tN + r - (N - n_j)}{N} \right\rceil = t - 1 + \left\lceil \frac{r + n_j}{N} \right\rceil.$$

We conclude that $|u_j|_{\#} = t + 1$ if and only if $n_j = 0$ or $n_j > N - r$, and otherwise $|u_j|_{\#} = t$.

We exhibit two words u_{j_1} and u_{j_2} from the abelian power for which the number of occurrences of the letter $\#$ differ. This contradiction proves our claim. Since $e \geq N$, we see that the numbers n_j , $n_j \equiv n_0 + jr \pmod N$, $j = 0, \dots, e-1$, form the coset $n_0 + \langle r \rangle$ of the subgroup $\langle r \rangle$ of $\mathbb{Z}/N\mathbb{Z}$. Let now $d = \gcd(r, N)$, so that $\langle r \rangle = \{0, d, 2d, \dots, (N/d - 1)d\}$. For example, if $\gcd(r, N) = 1$, then $\langle r \rangle = \mathbb{Z}/N\mathbb{Z}$. There thus exists an index j_1 such that the letter $\#$ occurs among the first d letters of u_{j_1} . This means that either $n_{j_1} = 0$ or

$$n_{j_1} > N - d \geq N - r.$$

Thus $|u_{j_1}|_{\#} = t + 1$ as was concluded previously. Similarly, there exists an index j_2 such that the letter $\#$ occurs among the d letters immediately preceding u_{j_2} . This means that

$$0 < n_{j_2} \leq d.$$

In this case

$$n_{j_2} + r \leq d + r \leq d + N - d = N$$

since $r \leq N - \gcd(r, N) = N - d$. We thus have $n_{j_2} \leq N - r$ implying that $|u_{j_2}|_{\#} = t$ as was concluded previously. This concludes the proof. \square

Remark 4. The above result may be slightly generalized. Indeed, notice that the only structural properties of σ used in the above proof are that σ is uniform, the images of the letters begin with $\#$, and the images of the letters contain no other occurrences of $\#$. In fact, the property that both images of letters begin with $\#$ is not important, it is only required that $\#$ occurs at the same position in both $\sigma(0)$ and $\sigma(1)$. We are thus led to the following generalization of Lemma 3. Let $\varphi: \{0, 1\}^* \rightarrow \{0, 1, \#\}^*$ be a uniform substitution defined by $\varphi(0) = u\#v$, $\varphi(1) = u'\#v'$, where $u, u', v, v' \in \{0, 1\}^*$, $|u| = |u'|$, and $|v| = |v'|$. Let \mathbf{w} be a binary word. If an abelian power $u_0 \cdots u_{e-1}$, $e \geq |u\#v|$, occurs in $\varphi(\mathbf{w})$, then $|u\#v|$ divides $|u_0|$. We shall need this generalization later in Section 3.

Lemma 5. *The substitution σ satisfies Property (ii).*

Proof. Let $u_0 \cdots u_{e-1}$, $e \geq N$, be an abelian power occurring in $\sigma(\mathbf{w})$. It follows by Lemma 3 that N divides the length of u_0 . Our aim is to show that the abelian

power $u_0 \cdots u_{e-1}$ can be shifted (to the left or the right) to obtain another abelian power $u'_0 \cdots u'_{e-1}$ with $|u'_0| = |u_0|$ such that each u'_i begins with the letter $\#$. Before doing so, let us show how the main claim follows from this. Because σ is injective, as is readily verified, there exist unique factors v_0, \dots, v_{e-1} of \mathbf{w} of length $|u_0|/N$ such that $\sigma(v_i) = u'_i$ for $i = 0, \dots, e-1$. Notice that $v_0 \cdots v_{e-1}$ is a factor of \mathbf{w} . Clearly the words v_i are abelian equivalent as $|v_i|_0 = |u'_i|_0/(N-1)$ and $|u'_i|_0 = |u'_j|_0$ for all j . We conclude that the word $v_0 \cdots v_{e-1}$ is an abelian power in \mathbf{w} .

Let us again write $\sigma(\mathbf{w}) = a_0 a_1 \cdots$ with $a_n \in \{0, 1, \#\}$ for each $n \geq 0$. Let u_0 have the position i in $\sigma(\mathbf{w})$, and let $n = i \bmod N$. If $n = 0$ then we are done since we may choose $u'_i = u_i$ in the above (recall that N divides $|u_0|$). Also, if $n = 1$, each word u_j , $j = 0, \dots, e-1$, is immediately preceded by $\#$ in $\sigma(\mathbf{w})$ and, moreover, each of the words ends with $\#$. By setting $u'_j = \#u_j\#^{-1}$, we see that $\#u_0 \cdots u_{e-1} = u'_0 \cdots u'_{e-1}\#$ occurs in $\sigma(\mathbf{w})$, and clearly $u'_j \sim u_0$ for each $j = 0, \dots, e-1$. Thus $u'_0 \cdots u'_{e-1}$ is an abelian power of the claimed form. Assume now that $n > 1$. Without loss of generality, we assume that u_0 begins with 0 so, in fact, u_0 begins with $0^{N-n}\#$. By the form of the substitution, u_0 is preceded by $\#0^{n-1}$ in $\sigma(\mathbf{w})$. We claim that each of the words u_j , $j = 0, \dots, e-1$, begins with $0^{N-n}\#$ and ends with $\#0^{n-1}$. Let us first show that u_1 begins with $0^{N-n}\#$ (and thus that u_0 ends with $\#0^{n-1}$). Assume for a contradiction that u_1 begins with $1^{N-n}\#$ (whence u_0 ends with $\#1^{n-1}$), and say that u_1 ends with $\#c^{n-1}$ where $c \in \{0, 1\}$. Now the word $\#0^{n-1}u_0(\#1^{n-1})^{-1}$ is the image of a factor x of \mathbf{w} . Similarly, the word $\#1^{n-1}u_1(\#c^{n-1})^{-1}$ is the image of a factor y of \mathbf{w} with $|x| = |y|$. We may write

$$|u_0|_1 = |x|_1(N-1) + n - 1$$

and

$$|u_1|_1 = |y|_1(N-1) - (n-1) + \delta_{c=1} \cdot (n-1),$$

where $\delta_{c=1} = 1$ if $c = 1$, and otherwise $\delta_{c=1} = 0$. Since $u_0 \sim u_1$, by rearranging the terms, we obtain

$$(|y|_1 - |x|_1)(N-1) = (2 - \delta_{c=1})(n-1).$$

Notice here that $1 \leq 2 - \delta_{c=1} \leq 2$ and that $n > 1$. The right side of the inequality is positive, so $|y|_1 - |x|_1 \geq 1$. Since $N > n$, it must be that $|y|_1 - |x|_1 < 2 - \delta_{c=1} \leq 2$. We conclude that $|y|_1 - |x|_1 = 1$ and, furthermore, $\delta_{c=1} = 0$. We now have

$$N-1 = 2(n-1),$$

which is impossible since N was chosen to be even. This contradiction shows that u_1 begins with $0^{N-n}\#$ as well. A symmetric argument shows that u_1 ends with $\#0^{n-1}$. We may repeat the above argument to show that each of the words u_j , $j = 0, \dots, e-1$, begins with $0^{N-n}\#$ and ends with $\#0^{n-1}$.

To finish off the proof, we choose $u'_j = \#0^{n-1}u_j(\#0^{n-1})^{-1}$ for each $j = 0, \dots, e-1$. Observe that $\#0^{n-1}u_0 \cdots u_{e-1} = u'_0 \cdots u'_{e-1}\#0^{n-1}$ and that $u'_0 \sim u'_j$ for each $j = 0, \dots, e-1$. We have thus exhibited an abelian power of the claimed form thus concluding the proof. \square

Since the substitution σ satisfies Properties (i)-(ii) and $\sigma(\mathbf{w})$ has at most three letters, Theorem 1 is proved.

3 Extension to the k -abelian Setting

In this section, we consider a generalization of abelian equivalence. Let k be a positive integer. Two words u and v are *k -abelian equivalent*, written $u \sim_k v$, if $|u|_w = |v|_w$ for all nonempty words w of length at most k [9]. For words of length at least $k - 1$, we can equivalently say that $u \sim_k v$ if and only if u and v share a common prefix and a common suffix of length $k - 1$ and $|u|_w = |v|_w$ for each word w of length k [9, Lemma 2.4]. The k -abelian equivalence relation is a congruence relation. Notice that 1-abelian equivalence is simply abelian equivalence. Moreover, if $u \sim_{k+1} v$, then $u \sim_k v$.

A nonempty word $u_0 \cdots u_{e-1}$ is a *k -abelian power* of exponent e and period m if $|u_0| = m$ and $u_0 \sim_k \cdots \sim_k u_{e-1}$. It was proved in [9, Thm. 5.4] using Szemerédi's theorem that every infinite word having bounded k -abelian complexity contains k -abelian powers of arbitrarily high exponent. Sturmian words are particular examples of such words, so each Sturmian word contains k -abelian powers of arbitrarily high exponent; an alternative proof of this fact is given in [15, Lemma 3.10]

Let \mathbf{w} be an infinite word. Then we set $\mathcal{A}_{k,\mathbf{w}}(m)$ to be the supremum of the exponents of k -abelian powers of period m occurring in \mathbf{w} . We define the *k -abelian critical exponent* of \mathbf{w} to be the quantity

$$\limsup_{m \rightarrow \infty} \frac{\mathcal{A}_{k,\mathbf{w}}(m)}{m},$$

and we denote it by $\mathcal{A}_k(\mathbf{w})$. This generalization of the abelian critical exponent is considered in the preprint [15], where the authors of this paper study the set of finite k -abelian critical exponents of Sturmian words. This set, dubbed as the *k -Lagrange spectrum*, is similarly complicated as the Lagrange spectrum. When $k > 1$, the least accumulation point of the k -Lagrange spectrum is $\sqrt{5}/(2k - 1)$, and the spectrum is dense in the interval $(\sqrt{5}/(2k - 1), \infty)$.

Next we prove the following analogue of Theorem 1.

Theorem 6. *Let θ be a nonnegative real number. Then there exists an infinite word \mathbf{w} such that $\mathcal{A}_k(\mathbf{w}) = \theta$. The word \mathbf{w} can be taken over an alphabet of at most three letters.*

Similar to Section 2, we wish to find a substitution f defined on a two-letter alphabet with the following properties:

- (i') If an abelian power $u_0 \cdots u_{e-1}$ occurs in \mathbf{w} , then $f(u_0) \cdots f(u_{e-1})$ is a k -abelian power occurring in $f(\mathbf{w})$.
- (ii') If a k -abelian power $u_0 \cdots u_{e-1}$, $e \geq N$, occurs in $f(\mathbf{w})$, then \mathbf{w} contains an abelian power $v_0 \cdots v_{e-1}$ with $|v_0| = |u_0|/N$.

Given such a substitution f , Theorem 6 is proved exactly as Theorem 1 was proved in Section 2. The case $k = 1$ is handled by Theorem 1, so we may assume that $k > 1$.

Let $N \geq 2k - 1$ be a fixed integer, and define the N -uniform substitution $\tau: \{0, 1\}^* \rightarrow \{0, 1, \#\}^*$ by

$$\begin{aligned} 0 &\mapsto \#0^{k-2}0^{N-2k+2}0^{k-1}, \\ 1 &\mapsto \#0^{k-2}1^{N-2k+2}0^{k-1}. \end{aligned}$$

Let u and v be two words of length greater than $2k - 2$. Suppose that $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$ and $\text{suff}_{k-1}(u) = \text{suff}_{k-1}(v)$, that is, assume that they share a common prefix of length $k - 1$ and a common suffix of length $k - 1$. One easily checks that then $uv \sim_k vu$. Remark then that it follows that $xvvy \sim_k xvyv$ for all words x and y because \sim_k is a congruence.

Lemma 7. *The substitution τ satisfies Property (i').*

Proof. By the form of the substitution τ , we have $\text{pref}_{k-1}(\tau(0)) = \text{pref}_{k-1}(\tau(1))$ and $\text{suff}_{k-1}(\tau(0)) = \text{suff}_{k-1}(\tau(1))$. Therefore $\tau(0)\tau(1) \sim_k \tau(1)\tau(0)$, and hence $\tau(u_i) \sim_k \tau(0)^{|u_i|_0}\tau(1)^{|u_i|_1}$ for $i = 0, \dots, e - 1$. Let $u_0 \cdots u_{e-1}$ be an abelian power in \mathbf{w} . Since the words u_0, \dots, u_{e-1} are abelian equivalent, we have

$$\tau(u_0), \dots, \tau(u_{e-1}) \sim_k \tau(0)^{|u_0|_0}\tau(1)^{|u_0|_1},$$

so $\tau(u_0) \sim_k \cdots \sim_k \tau(u_{e-1})$. □

Lemma 8. *If $u_0 \cdots u_{e-1}$, $e \geq N$, is a k -abelian power occurring in $\tau(\mathbf{w})$, then N divides $|u_0|$.*

Proof. Since $u_0 \cdots u_{e-1}$ is a k -abelian power, it is an abelian power. Observe now that the substitution τ is as in Remark 4. Thus N divides $|u_0|$. □

Lemma 9. *The substitution τ satisfies Property (ii').*

Proof. Suppose that a k -abelian power $u_0 \cdots u_{e-1}$ with $e \geq N$ occurs in $\tau(\mathbf{s})$. By Lemma 8, N divides $|u_0|$. Similar to the proof of Lemma 5, we want to show that the k -abelian power $u_0 \cdots u_{e-1}$ can be shifted (to the left or the right) to obtain another k -abelian power $u'_0 \cdots u'_{e-1}$, $|u'_0| = |u_0|$, such that each u'_i begins with $\#$. Then a slight modification of the argument presented in the first paragraph of the proof of Lemma 5 proves the claim. Indeed, given the preimages v_0, \dots, v_{e-1} of u'_0, \dots, u'_{e-1} , we see that $|v_i|_0 = |u'_i|_{\#0^{k-1}}$ for all i . Since $u'_0 \sim_k \cdots \sim_k u'_{e-1}$, we have $|u'_i|_{\#0^{k-1}} = |u'_j|_{\#0^{k-1}}$ for all i and j , and it follows that $v_0 \sim \cdots \sim v_{e-1}$.

Let p be the common prefix of length $k - 1$ of the words u_0, \dots, u_{e-1} and similarly q be the common suffix of length $k - 1$ of these words. Suppose first that $\#$ occurs in p , that is, $p = 0^r \# 0^s$ with $r + s = k - 2$. As each occurrence of $\#$ is preceded by 0^{k-1} and N divides $|u_i|$, the word u_{e-1} is followed by 0^r . Thus we may set $u'_i = (0^r)^{-1}u_i0^r$ for $i = 0, \dots, e - 1$. The same r factors $0^r \# 0^{s+1}$, $0^{r-1} \# 0^{s+2}$, \dots , $0 \# 0^{s+r}$ of length k were removed from each u_i and the same

r factors of length k were added to each u'_i (the final $k - 1$ factors of $q0^r$ of length k) during the shift. Thus $u'_0 \sim_k \cdots \sim_k u'_{e-1}$. If $\#$ occurs in q , that is, say $q = 0^r \# 0^s$ with $r + s = k - 2$ then, like above, we may set $u'_i = \# 0^s u_i (\# 0^s)^{-1}$ for $i = 0, \dots, e - 1$. Suppose then that some word u_i begins with $0^{k-1} \#$. It is straightforward to see that then all of the words u_0, \dots, u_{e-1} begin with $0^{k-1} \#$ and, furthermore, that u_{e-1} is followed by $0^{k-1} \#$. Setting $u'_i = (0^{k-1})^{-1} u_i 0^{k-1}$ for $i = 0, \dots, e - 1$ gives the claim as above.

By the preceding paragraph, we may assume that the occurrence of p as the prefix of u_i is a proper factor of $\tau(c_i)$ for a letter c_i , that is, we may write $\tau(c_i) = x_i p y_i$, with x_i and y_i nonempty, for this occurrence of p . Moreover, the preceding paragraph tells that we may assume that q is a proper suffix of x_i (otherwise $\#$ occurs in q). Since p has length $k - 1$, it is clear from the form of the substitution τ that the letter c_i is uniquely determined by p . Since N divides $|u_i|$, it follows that $c_0 = \dots = c_{e-1}$. This means that each u_i is preceded by x_0 . We still need to know that u_{e-1} ends with x_0 ; the words u_0, \dots, u_{e-2} must end with x_0 . Since N divides $|u_i|$, the suffix q of u_{e-1} occurs in $\tau(d)$, $d \in \{0, 1\}$, in the same position as the occurrence of q preceding $p y_0$ in $\tau(c_0)$. Now q has length $k - 1$, so its occurrence preceding $p y_i$ in $\tau(c_i)$ uniquely determines c_i , and hence its occurrence in $\tau(d)$ in the same position uniquely determines d . Therefore $d = c_0$ and u_{e-1} ends with x_0 . We may now set $u'_i = x_0 u_i x_0^{-1}$ for $i = 0, \dots, e - 1$. The suffix x_0 of u_i is preceded by 0^{k-1} and u_i has prefix p of length $k - 1$, so exactly the same factors of length $k - 1$ are added and removed when shifting each u_i to u'_i . Thus $u'_0 \sim_k \cdots \sim_k u'_{e-1}$. \square

Since τ satisfies Properties (i') and (ii'), Theorem 6 follows.

4 Concluding Remarks

Theorem 1 raises the following question.

Question 10. Given a nonnegative real number θ , does there exist an infinite binary word having k -abelian critical exponent θ ?

We conjecture that the question has a positive answer. To use the presented method, the marker letter $\#$ needs to be replaced by a suitable binary word ensuring that Properties (ii) and (ii') hold. There seems to be no obvious choice, at least no obvious choice leading to reasonable proofs. Perhaps another method is required. It would certainly be very interesting if the answer to the above question turned out to be negative. Nevertheless, we leave the question open.

The k -abelian equivalence is a refinement of abelian equivalence that ‘‘tends’’ to the usual equality of words as $k \rightarrow \infty$. As mentioned in the introduction, it is typical to consider the maximum exponent $\sup_{m \geq 1} \exp(m)$ for the equality relation, not the superior limit of the ratio between the maximum exponent $\exp(m)$ and period m as is done here for abelian equivalence and k -abelian equivalence. What then happens if we consider the unorthodox notion? Does an analogue to Theorem 1 hold? The answer is yes. The following result is proved by the authors in the preprint [15].

Proposition 11. [15, Prop. 3.17] *Given an infinite word \mathbf{w} , let $E(\mathbf{w})$ be the quantity*

$$\limsup_{m \rightarrow \infty} \frac{\exp(m)}{m},$$

where $\exp(m)$ is the supremum of (integral) exponents of powers of period m occurring in \mathbf{w} . For each nonnegative θ , there exists a Sturmian word \mathbf{s} such that $E(\mathbf{s}) = \theta$.

References

1. Aigner, M.: Markov's Theorem and 100 Years of the Uniqueness Conjecture. Springer (2013). <https://doi.org/10.1007/978-3-319-00888-2>
2. Berthé, V., Rigo, M. (eds.): Combinatorics, Words and Symbolic Dynamics. No. 159 in Encyclopedia of Mathematics and Its Applications, Cambridge University Press (2016)
3. Cassaigne, J., Karhumäki, J., Saarela, A.: On growth and fluctuation of k -abelian complexity. *European J. Combin.* **65**, 92–105 (2017). <https://doi.org/10.1016/j.ejc.2017.05.006>
4. Cusick, T.W., Flahive, M.E.: The Markoff and Lagrange Spectra. No. 30 in Mathematical Surveys and Monographs, American Mathematical Society, Providence, Rhode Island (1989)
5. Dekking, F.M.: Strongly non-repetitive sequences and progression-free sets. *J. Combin. Theory Ser. A* **27**(2), 181–185 (1979). [https://doi.org/10.1016/0097-3165\(79\)90044-X](https://doi.org/10.1016/0097-3165(79)90044-X)
6. Fici, G., Langiu, A., Lecroq, T., Lefebvre, A., Mignosi, F., Peltomäki, J., Prieur-Gaston, É.: Abelian powers and repetitions in Sturmian words. *Theoret. Comput. Sci.* **635**, 16–34 (2016). <https://doi.org/10.1016/j.tcs.2016.04.039>
7. Freiman, G.A.: Diophantine approximation and geometry of numbers (Markov's problem). *Kalininskii Gosudarstvennyi Universitet, Kalinin* (1975), (Russian)
8. Hall, Jr., M.: On the sum and products of continued fractions. *Ann. of Math.* **48**(4), 966–993 (1947). <https://doi.org/10.2307/1969389>
9. Karhumäki, J., Saarela, A., Zamboni, L.Q.: On a generalization of Abelian equivalence and complexity of infinite words. *J. Combin. Theory Ser. A* **120**, 2189–2206 (2013). <https://doi.org/10.1016/j.jcta.2013.08.008>
10. Krieger, D., Shallit, J.: Every real number greater than 1 is a critical exponent. *Theoret. Comput. Sci.* **381**, 177–182 (2007). <https://doi.org/10.1016/j.tcs.2007.04.037>
11. Lothaire, M.: Algebraic Combinatorics on Words. No. 90 in Encyclopedia of Mathematics and Its Applications, Cambridge University Press (2002)
12. Markov, A.A.: Sur les formes quadratiques binaires indéfinies. *Math. Ann.* **15**(3–4), 381–406 (1879). <https://doi.org/10.1007/BF02086269>
13. Markov, A.A.: Sur les formes quadratiques binaires indéfinies ii. *Math. Ann.* **17**(3), 379–399 (1880). <https://doi.org/10.1007/BF01446234>
14. Peltomäki, J.: Privileged Words and Sturmian Words. Ph.D. dissertation, Turku Centre for Computer Science, University of Turku, Turku, Finland (2016), <http://urn.fi/URN:ISBN:978-952-12-3422-4>
15. Peltomäki, J., Whiteland, M.A.: On k -abelian equivalence and generalized Lagrange spectra (2018), arXiv:1809.09047 (under review)
16. Rao, M., Rigo, M., Salimov, P.: Avoiding 2-binomial squares and cubes. *Theoret. Comput. Sci.* **572**, 83–91 (2015). <https://doi.org/10.1016/j.tcs.2015.01.029>

17. Reutenauer, C.: From Christoffel Words to Markoff Numbers. Oxford University Press (2019)
18. Richomme, G., Saari, K., Zamboni, L.Q.: Abelian complexity of minimal subshifts. *J. Lond. Math. Soc.* **83**(1), 79–95 (2011). <https://doi.org/10.1112/jlms/jdq063>
19. Rigo, M., Salimov, P.: Another generalization of abelian equivalence: Binomial complexity of infinite words. *Theoret. Comput. Sci.* **601**, 47–57 (2015). <https://doi.org/10.1016/j.tcs.2015.07.025>