

On k -abelian equivalence and generalized Lagrange spectra

by

JARKKO PELTOMÄKI and MARKUS A. WHITELAND (Turku)

1. Introduction. The critical exponent of an infinite word \mathbf{w} is the supremum of exponents of fractional powers occurring in \mathbf{w} . Famously, Thue [30] showed in 1906 that the fixed point of the substitution $0 \mapsto 01, 1 \mapsto 10$, now known as the Thue–Morse word due to Morse’s independent contribution [21], has critical exponent 2, meaning that it avoids powers with exponent at least 3. The notion of critical exponent is central in the study of powers and their avoidance, which have since Thue been a central theme in combinatorics on words.

Another important subject in combinatorics on words is the theory of Sturmian words. These are a large class of extensively studied words with strong connections to number theory, particularly to continued fractions (see, e.g., [2], [17, Ch. 2], [25, Ch. 6] and the references therein). The powers occurring in Sturmian words are well-understood, and a formula for the critical exponent of a Sturmian word was determined by Damanik and Lenz [7] and Justin and Pirillo [13]. For example, the critical exponent of the Fibonacci word, the fixed point of the substitution $0 \mapsto 01, 1 \mapsto 0$, is $(5 + \sqrt{5})/2$, as was already derived in [20]. The critical exponent of the Fibonacci word is minimal among all Sturmian words.

In recent years, there has been a substantial amount of research in generalizations of the concept of a power. A popular generalization is that of an abelian power; other generalizations are k -abelian powers (see below) and those based on k -binomial equivalence [27]. Two words u and v are *abelian equivalent*, written $u \sim_1 v$, if one is obtained from the other by permuting letters. If u_0, u_1, \dots, u_{n-1} are abelian equivalent words of length m , then their concatenation $u_0u_1 \cdots u_{n-1}$ is an *abelian power of exponent n and period m*

2020 *Mathematics Subject Classification*: Primary 11J06; Secondary 68R15.

Key words and phrases: Lagrange spectrum, Sturmian word, k -abelian equivalence, continued fraction.

Received 27 September 2018; revised 10 May 2019.

Published online *.

(only integer exponents are considered). Thus an abelian power is a generalization of the usual notion of a power: the abelian equality relation is used in place of the usual equality relation. Questions regarding abelian powers were already raised by Erdős [8] in 1957. More recently there has been a burst of activity on the subject starting, perhaps, with the 2011 paper [26] by Richomme, Saari, and Zamboni. See, e.g., the references of [9] and especially the papers [24, 28, 9, 10] related to Sturmian words.

Fici et al. [9] studied the abelian critical exponents of Sturmian words and showed that there are abelian powers of arbitrarily high exponent starting at each position of a Sturmian word, a result also obtained in [26]. This means that directly generalizing the notion of a critical exponent to the abelian setting only in terms of the exponent does not produce a quantity of interest (at least for Sturmian words). Thus an alternative definition was adopted in [9]. The *abelian critical exponent* of an infinite word \mathbf{w} is defined as the quantity

$$(1) \quad \limsup_{m \rightarrow \infty} \{n/m : u \text{ is an abelian power of exponent } n \text{ and period } m \text{ occurring in } \mathbf{w}\},$$

measuring the maximal ratio between the exponents and periods of abelian powers in \mathbf{w} . This alternative definition does lead to an interesting quantity. The abelian critical exponent of a Sturmian word can be finite or infinite, and again the Fibonacci word has minimal exponent, this time the value being $\sqrt{5}$.

Surprisingly, the set of abelian critical exponents of all Sturmian words turns out to coincide with the Lagrange spectrum. 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 constant $\lambda(\alpha)$ of α is computed as follows:

$$\lambda(\alpha) = \limsup_{t \rightarrow \infty} (q_t \|q_t \alpha\|)^{-1} = \limsup_{t \rightarrow \infty} ([a_{t+1}; a_{t+2}, \dots] + [0; a_t, a_{t-1}, \dots, a_1]),$$

where $[a_0; a_1, a_2, \dots]$ is the continued fraction expansion of α , and (q_k) is the sequence of denominators of its convergents ($\|x\|$ measures the distance of x to the nearest integer). The connection here is that for a fixed Sturmian word, the number n in (1), when maximal, equals the integer part of $1/\|m\alpha\|$ for a certain irrational α (for details, see Section 2 and Lemma 3.8).

The *Lagrange spectrum* is the set of finite Lagrange constants of irrational numbers. The Lagrange spectrum has been studied extensively, but many of its properties still remain a mystery. The spectrum has a curious structure: its initial part inside the interval $[\sqrt{5}, 3)$ is discrete as shown by Markov already in late 19th century [18, 19], but it contains a half-line as was famously proven by Hall [12] in 1947. Good sources of information on

the Lagrange spectrum are the monograph of Cusick and Flahive [6] and Aigner's book [1].

Another relatively recent development in combinatorics on words is the systematic study of a generalization of abelian equivalence called *k-abelian equivalence* initiated by Karhumäki, Saarela, and Zamboni [15]. This generalization originally appears in a 1980 paper of Karhumäki [14]. Two words u and v are said to be *k-abelian equivalent*, written $u \sim_k v$, if $|u|_w = |v|_w$ for each nonempty word w of length at most k (here $|u|_w$ stands for the number of occurrences of w as a factor of u). Thus 1-abelian equivalence is simply the abelian equivalence discussed above. The *k-abelian equivalence* relation is clearly an equivalence relation, but it is also a congruence relation. For $k = 1, 2, \dots$, the corresponding *k-abelian equivalence* relations can be seen as refinements of the abelian equivalence relation approaching the usual equality relation. The *k-abelian equivalence* has been studied especially from the points of view of factor complexity and power avoidance; for more information, see the recent paper [4] and its references.

The purpose of the current paper is to generalize the research of [9] on abelian critical exponents of Sturmian words to the *k-abelian* setting. That is, we use the general *k-abelian equivalence* in place of abelian equivalence to obtain the notion of *k-abelian critical exponent*, and study the set \mathcal{L}_k of *k-abelian critical exponents* of Sturmian words. As \mathcal{L}_1 is the Lagrange spectrum, the sets \mathcal{L}_k for $k > 1$ can be seen as combinatorial generalizations of the Lagrange spectrum.

Our main contribution is the characterization of the *k-Lagrange spectrum* \mathcal{L}_k in terms of \mathcal{L}_1 . Our result, Theorem 3.12, states that the *k-abelian critical exponent* of a Sturmian word \mathbf{s} with abelian critical exponent K equals cK for a particular constant c , $0 < c < 1$, which depends on k and \mathbf{s} . The relation between \mathcal{L}_1 and \mathcal{L}_k is thus quite simple. However, the sets \mathcal{L}_k inherit the complicated structure of the Lagrange spectrum \mathcal{L}_1 . We show that for $k > 1$ we have $\mathcal{L}_k \subseteq (\sqrt{5}/(2k - 1), \infty)$, the number $\sqrt{5}/(2k - 1)$ being the least accumulation point of \mathcal{L}_k (Theorem 3.14). Moreover, we prove that the set \mathcal{L}_k is dense in $(\sqrt{5}/(2k - 1), \infty)$ (Theorem 3.15). This contrasts the case $k = 1$ where the initial part of \mathcal{L}_1 is discrete. The set \mathcal{L}_1 is known to contain a half-line. We do not know if \mathcal{L}_k contains an analogous half-line for $k > 1$; we leave this problem open.

Our approach is to first give an arithmetical and geometric interpretation for what it means for two factors of a Sturmian word to be *k-abelian equivalent*, and then to employ continued fractions to derive our results. This approach is similar to that of [9] where the usage of continued fractions was crucial. The arithmetical interpretation complements the combinatorial methods of [15]: we make some results of [15] on Sturmian words more precise.

Our approach also makes it possible to efficiently find the possible exponents and locations of k -abelian powers occurring in a given Sturmian word.

The paper is organized as follows. In Section 2, we give the necessary definitions and background information on Sturmian words and number theory. After this we present the main results and their proofs in Section 3. Section 4 provides further discussion on some matters raised in Section 3. Finally, Section 5 concludes the paper with open problems.

2. Preliminaries. We use standard terminology from combinatorics on words; we refer the reader to the book [17] for any undefined terms. The words considered in this paper are finite or infinite binary words over the alphabet $\{0, 1\}$. We distinguish infinite words from finite words by referring to the former with boldface symbols. By $|w|$ we mean the length of the finite word w . The n th power of a finite word w is the word obtained by repeating it consecutively n times, and it is denoted by w^n . For the infinite repetition of w , we use the notation w^ω . An infinite word is *ultimately periodic* if it can be written in the form uv^ω for some finite words u and v ; otherwise it is *aperiodic*.

We denote by $|w|_u$ the number of occurrences of the nonempty word u as a factor of w . If u and v are finite words over an alphabet A , then u and v are *abelian equivalent*, written $u \sim_1 v$, if $|u|_a = |v|_a$ for each letter a of A . Let then k be a fixed positive integer. We say that u and v are *k -abelian equivalent*, written $u \sim_k v$, if $|u|_w = |v|_w$ for each word w of length at most k . Notice that if $k = 1$, then k -abelian equivalence is simply abelian equivalence. For words of length at least $k - 1$ we can alternatively say that $u \sim_k v$ if and only if u and v have a common prefix and a common suffix of length $k - 1$ and $|u|_w = |v|_w$ for each word w of length k [15, Lemma 2.3]. Thus, for words of length at most $2k - 1$, the k -abelian equivalence is in fact the equality relation [15, Lemma 2.4]. The k -abelian equivalence relation is a congruence relation. If u_0, u_1, \dots, u_{n-1} are k -abelian equivalent words of length m , then their concatenation $u_0u_1 \cdots u_{n-1}$ is a *k -abelian power of exponent n and period m* . In this paper, we consider only nondegenerate powers, that is, we assume that $n \geq 2$.

Recall that every irrational real number α has a unique infinite continued fraction expansion:

$$(2) \quad \alpha = [a_0; a_1, a_2, a_3, \dots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}$$

with $a_0 \in \mathbb{Z}$ and $a_t \in \mathbb{Z}_+$ for $t \geq 1$. The numbers a_i are called the *partial quotients* of α . By cutting the expansion after $t+1$ terms, we obtain a rational

number $[a_0; a_1, a_2, a_3, \dots, a_t]$, which we denote by p_t/q_t . These rationals p_t/q_t are the *convergents* of α . The convergents of α satisfy the best approximation property, that is,

$$\|q_t \alpha\| = \min_{0 < m < q_{t+1}} \|m \alpha\|$$

for all $t \geq 1$. Here $\|x\|$ measures the distance from x to the nearest integer. In other words, $\|x\| = \min\{\{x\}, 1 - \{x\}\}$, where $\{x\}$ denotes the fractional part of x . Two numbers with continued fraction expansions $[a_0; a_1, \dots]$ and $[b_0; b_1, \dots]$ are *equivalent* if there exist integers N and M such that $a_{N+i} = b_{M+i}$ for all $i \geq 0$. As we shall see later, continued fractions are useful in studying Sturmian words (defined below). More details on the connection with Sturmian words can be found, e.g., in [22, Ch. 4].

Let α be an irrational number, and define the *Lagrange constant* $\lambda(\alpha)$ of α as the infimum of real numbers λ such that for every $c > \lambda$ the inequality

$$(3) \quad \left| \alpha - \frac{p}{q} \right| < \frac{1}{cq^2}$$

has only finitely many rational solutions p/q . Famously Hurwitz's Theorem states that $\lambda(\alpha) \geq \sqrt{5}$ for any irrational α , and there exist numbers with $\lambda(\alpha) = \sqrt{5}$. The numbers with finite Lagrange constant are often called *badly approximable* in the literature. The Lagrange constant of α with continued fraction expansion as in (2) is computed as follows:

$$(4) \quad \lambda(\alpha) = \limsup_{t \rightarrow \infty} ([a_{t+1}; a_{t+2}, \dots] + [0; a_t, a_{t-1}, \dots, a_1]).$$

From this formula, it is clear that two equivalent numbers have the same Lagrange constant. The *Lagrange spectrum* is the set of finite Lagrange constants. This set has many curious properties, and we shall return to them at the end of Subsection 3.2. For details on the Lagrange spectrum, see [6] or [1].

Sturmian words are defined as the codings of orbits of points in an irrational circle rotation with two intervals. This understanding is sufficient for our purposes, but many other viewpoints exist; see, e.g., [25, 17]. Identify the unit interval $[0, 1)$ with the unit circle \mathbb{T} , and let α be a fixed irrational. The mapping $R: \mathbb{T} \rightarrow \mathbb{T}$, $x \mapsto \{x + \alpha\}$, defines a rotation on \mathbb{T} . Partition the circle \mathbb{T} into two intervals I_0 and I_1 defined by the points 0 and $\{1 - \alpha\}$. Let ν be the coding function defined by setting $\nu(x) = 0$ if $x \in I_0$ and $\nu(x) = 1$ if $x \in I_1$. Define $\mathbf{s}_{x,\alpha}$ as the infinite word obtained by setting its n th letter, $n \geq 0$, to equal $\nu(R^n(x))$. Then $\mathbf{s}_{x,\alpha}$ is called the *Sturmian word of slope α and intercept x* .

The above definition is not complete because we did not specify how ν behaves at the endpoints 0 and $\{1 - \alpha\}$. There is some choice here, and we take either $I_0 = [0, \{1 - \alpha\})$ and $I_1 = [\{1 - \alpha\}, 1)$, or $I_0 = (0, \{1 - \alpha\}]$ and

$I_1 = (\{1 - \alpha\}, 1]$. These options are determined by whether or not $0 \in I_0$. This little detail makes no difference to us: only interior points of intervals are considered. Let $x, y \in \mathbb{T}$ with $x < y$. Then by both $I(x, y)$ and $I(y, x)$ we mean the interval $[x, y)$ if $0 \in I_0$ and the interval $(x, y]$ if $0 \notin I_0$.

One particular example of a Sturmian word is the Fibonacci word \mathbf{f} . Its slope is $1/\varphi^2$, where φ is the golden ratio, and its intercept equals its slope. We have

$$\mathbf{f} = 01001010010010100101001001010010 \cdots .$$

This word is also the fixed point of the substitution $0 \mapsto 01, 1 \mapsto 0$.

The sequence $(\{n\alpha\})_{n \geq 0}$ is dense in $[0, 1)$ by Kronecker's Theorem, so Sturmian words of slope α have a common language \mathcal{L} (the set of factors). Let w denote a word $a_0 a_1 \cdots a_{n-1}$ of length n in \mathcal{L} . Then there exists a unique subinterval $[w]$ of \mathbb{T} such that the Sturmian word $\mathbf{s}_{x, \alpha}$ begins with w if and only if $x \in [w]$. Clearly $[w] = I_{a_0} \cap R^{-1}(I_{a_1}) \cap \cdots \cap R^{-(n-1)}(I_{a_{n-1}})$ (here the choice of endpoints matters, but we only consider interior points of these intervals). The points $0, \{-\alpha\}, \{-2\alpha\}, \dots, \{-n\alpha\}$ partition the circle into $n + 1$ subintervals which are exactly the intervals $[w]$ for factors of length n . We call these $n + 1$ intervals the *level n intervals*, and we denote the set consisting of them by $L(n)$. We abuse notation and write $\max L(n)$ (resp. $\min L(n)$) for the maximum (resp. minimum) length of a level n interval.

In the rest of this paper, we keep the slope α with continued fraction expansion $[a_0; a_1, a_2, \dots]$ fixed. Whenever we talk about the convergents q_t , the level n intervals $L(n)$, the rotation R , etc., we implicitly understand that they relate to this fixed α .

3. Main results

3.1. k -abelian equivalence in Sturmian words. Our first aim is to show that the k -abelian equivalence classes of factors of a Sturmian word correspond to certain intervals on the circle \mathbb{T} and to characterize the endpoints of these intervals. We begin by recalling the following result of [15] (specialized to Sturmian words).

PROPOSITION 3.1 ([15, Proposition 2.8]). *Let u and v be two factors of the same length occurring in some Sturmian word. Then $u \sim_k v$ if and only if they share a common prefix and a common suffix of length $\min\{|u|, k - 1\}$ and $u \sim_1 v$.*

This result is interesting as it shows that rather weak conditions are enough for k -abelian equivalence in Sturmian words. This is not unique to Sturmian words: it holds for episturmian words [15, Proposition 2.8], and in [5, Theorem 1] it is shown that Proposition 3.1 also holds for factors of the

Cantor word, the fixed point of the substitution $0 \mapsto 000$, $1 \mapsto 101$. We will return to this matter in Section 4.

Let us also recall the following result which gives an arithmetical characterization of abelian equivalence in Sturmian words.

PROPOSITION 3.2 ([9, Proposition 3.3], [28, Theorem 19]). *Let u and v be two factors of the same length occurring in a Sturmian word of slope α . Then $u \sim_1 v$ if and only if $[u], [v] \subseteq I(0, \{-|u|\alpha\})$ or $[u], [v] \subseteq I(\{-|u|\alpha\}, 1)$.*

In other words, the two possible abelian equivalence classes for factors of length m correspond to two intervals on the circle marked by the points 0 and $\{-m\alpha\}$. Next we generalize Proposition 3.2 to *k*-abelian equivalence.

By Proposition 3.1, we need at least consider the prefixes and suffixes of length up to $k - 1$. Let $m \geq 1$, and define

$$\mathcal{D}_{k,m} = \{0, \{-\alpha\}, \{-2\alpha\}, \dots, \{-\min\{m, k - 1\}\alpha\}\}.$$

These points divide the circle into $\min\{m + 1, k\}$ intervals (which are the level $\min\{m, k - 1\}$ intervals), and if points x and y belong to the same interval, then the prefixes of $\mathbf{s}_{x,\alpha}$ and $\mathbf{s}_{y,\alpha}$ of length $\min\{m, k - 1\}$ are equal. Now if $m \geq k - 1$, then

$$R^{-(m-(k-1))}(\mathcal{D}_{k,m}) = \{\{-(m - (k - 1))\alpha\}, \dots, \{-m\alpha\}\},$$

and these points also divide the circle into k intervals. If x and y belong to the same interval, then the prefixes of $\mathbf{s}_{x,\alpha}$ and $\mathbf{s}_{y,\alpha}$ of length m have a common suffix of length $k - 1$. Set $\mathcal{P}_{k,m} = \mathcal{D}_{k,m} \cup R^{-(m-(k-1))}(\mathcal{D}_{k,m})$ if $m \geq k - 1$; otherwise set $\mathcal{P}_{k,m} = \mathcal{D}_{k,m}$.

DEFINITION 3.3. $\mathcal{I}_{k,m}$ is the set of subintervals of \mathbb{T} determined by the points of $\mathcal{P}_{k,m}$.

What we mean precisely is that to define the intervals I_i of $\mathcal{I}_{k,m}$, we order the points x_i of $\mathcal{P}_{k,m}$: $0 = x_0 < x_1 < \dots < x_{\ell-1} < x_\ell = 1$, $\ell = |\mathcal{P}_{k,m}|$, and set $I_i = [x_i, x_{i+1}]$ if $0 \in I_0$ and $I_i = (x_i, x_{i+1}]$ if $0 \notin I_0$ for $0 \leq i < \ell$. Observe that for $m < k - 1$, the intervals $\mathcal{I}_{k,m}$ coincide with the level m intervals.

As before, for the level m intervals in $L(m)$, by writing $\max \mathcal{I}_{k,m}$ we mean the maximum length of an interval in $\mathcal{I}_{k,m}$. We claim that the intervals in $\mathcal{I}_{k,m}$ determine the *k*-abelian equivalence classes.

THEOREM 3.4. *Let u and v be two factors of length m occurring in a Sturmian word of slope α . Then $u \sim_k v$ if and only if there exists $J \in \mathcal{I}_{k,m}$ such that $[u], [v] \subseteq J$.*

Proof. Assume that $m < k - 1$. Then $u \sim_k v$ if and only if $u = v$. This means that $[u]$ and $[v]$ equal one of the level m intervals. When $m < k - 1$,

the intervals in $\mathcal{I}_{k,m}$ are precisely the level m intervals, so we are done. We may thus assume that $m \geq k - 1$.

Suppose first that there exists $J \in \mathcal{I}_{k,m}$ such that $[u], [v] \subseteq J$. By the definition of the intervals in $\mathcal{I}_{k,m}$, the words u and v share a common prefix and a common suffix of length $k - 1$. Moreover, they are abelian equivalent by Proposition 3.2 because the point $\{-m\alpha\}$ separating the two abelian equivalence classes is among the points of $\mathcal{P}_{k,m}$. Therefore Proposition 3.1 implies that $u \sim_k v$.

Suppose that $u \not\sim_k v$. Then u and v share a common prefix and a common suffix of length $k - 1$. Assume for a contradiction that $[u]$ and $[v]$ are contained in distinct intervals of $\mathcal{I}_{k,m}$. Without loss of generality, we assume that $\sup [u] \leq \inf [v]$. Let K be the interval containing exactly the points z for which $\sup [u] \leq z \leq \inf [v]$. (If $\sup [u] = \inf [v]$, then we let K be the set containing the common endpoint of $[u]$ and $[v]$.) Since $[u]$ and $[v]$ are contained in distinct intervals of $\mathcal{I}_{k,m}$, there exists a point x in $\mathcal{P}_{k,m}$ such that $x \in K$. Denote the set $R^{-(m-(k-1))}(\mathcal{D}_{k,m})$ by \mathcal{S} . The point x cannot be in $\mathcal{D}_{k,m}$ because u and v share a common prefix of length $k - 1$. Therefore we must have $x \in \mathcal{S}$. Let y be an arbitrary point in \mathcal{S} . If $y \in \mathbb{T} \setminus ([u] \cup [v] \cup K)$, then either $[u] \subseteq I(x, y)$ and $[v] \cap I(x, y) = \emptyset$, or symmetrically $[v] \subseteq I(x, y)$ and $[u] \cap I(x, y) = \emptyset$. Then, by the definition of \mathcal{S} , we see that u and v have distinct suffixes of length $k - 1$, which is impossible. We conclude that $\mathcal{S} \subseteq K$ (see Example 3.5). Since $\{-m\alpha\} \in \mathcal{S}$, it follows by Proposition 3.2 that u and v are not abelian equivalent. This is a contradiction. ■

Notice that $\mathcal{I}_{k,m}$ contains $2k$ intervals when $m \geq 2k - 1$, and $m + 1$ intervals when $0 \leq m \leq 2k - 2$. This number of abelian equivalence classes for factors of length m characterizes Sturmian words [15, Theorem 4.1].

EXAMPLE 3.5. Let us consider the 2-abelian equivalence classes of length 5 of the Fibonacci word; its slope α is $1/\varphi^2$. On the left in Figure 1, there

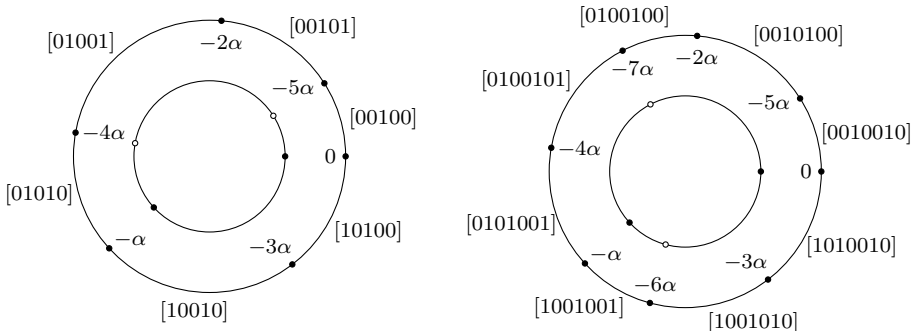


Fig. 1. Factors of length 5 and 7 of the Fibonacci word on the unit circle. The outer circles illustrate the level 5 and 7 intervals, and the inner circles the 2-abelian equivalence classes of length 5 and 7.

are two concentric circles. The outer circle represents the level 5 intervals separated by the points 0 , $\{-\alpha\}$ (≈ 0.62), $\{-2\alpha\}$ (≈ 0.24), $\{-3\alpha\}$ (≈ 0.85), $\{-4\alpha\}$ (≈ 0.47), and $\{-5\alpha\}$ (≈ 0.09). The inner circle shows the endpoints of the 2-abelian equivalence classes. The points 0 and $\{-\alpha\}$ of $\mathcal{D}_{2,5}$ are shown in black, while the points $\{-4\alpha\}$ and $\{-5\alpha\}$ of $R^{-4}(\mathcal{D}_{2,5})$ are represented by circles filled with white. The concentric circles on the right in Figure 1 give the corresponding intervals and points when $m = 7$.

We have four 2-abelian equivalence classes for length 5: $\{00100\}$, $\{00101, 01001\}$, $\{01010\}$, $\{10010, 10100\}$. The singleton classes are special. At the end of the proof of Theorem 3.4, we had to take some extra steps because factors corresponding to two distinct intervals of $\mathcal{I}_{k,m}$ could share prefixes and suffixes of length $k - 1$. Indeed, here 00100 and 01010 have common prefixes and suffixes of length 1, but this does not guarantee abelian equivalence.

We make an observation regarding the part of the proof of Theorem 3.4 showing that if two level m intervals $[u]$ and $[v]$ are included in distinct intervals of $\mathcal{I}_{k,m}$ then $u \approx_k v$. The proof shows that if u and v have common prefixes and suffixes of length $k - 1$, the only way that $u \approx_k v$ is when all the points in \mathcal{S} (this is the set $R^{-(m-(k-1))}(\mathcal{D}_{m,k})$) are contained in one level $k - 1$ interval J . We claim that this phenomenon cannot occur if $k \geq 2$ and $\|\alpha\| > 1/(2(k - 1))$. Notice that in the case $k = 2$ this may happen since $\|\alpha\| < 1/2$ always.

Assume that $k \geq 2$. There exist at least two points at distance $\|\alpha\|$ in \mathcal{S} (e.g., $\{-(m-1)\alpha\}$ and $\{-m\alpha\}$), which implies that the length of J is greater than $\|\alpha\|$. If $k - 1 \geq \lfloor 1/\|\alpha\| \rfloor$, then each interval determined by the points $0, \{-\alpha\}, \dots, \{-(k - 1)\alpha\}$ has length at most $\|\alpha\|$, so we conclude that $k - 1 < \lfloor 1/\|\alpha\| \rfloor$. The intervals determined by the points $0, \{-\alpha\}, \dots, \{-(k - 1)\alpha\}$ are now the same as those determined by the points $0, 1 - \|\alpha\|, 1 - 2\|\alpha\|, \dots, 1 - (k - 1)\|\alpha\|$, so all of them have length $\|\alpha\|$ except one that has length $1 - (k - 1)\|\alpha\|$. Thus J has length $1 - (k - 1)\|\alpha\|$. Since R is an isometry, J contains $k - 1$ intervals of length $\|\alpha\|$ (defined by the points of \mathcal{S}), and we must have $(k - 1)\|\alpha\| < 1 - (k - 1)\|\alpha\|$, that is, $\|\alpha\| < 1/(2(k - 1))$. Thus we obtain the following strengthening of Proposition 3.1.

THEOREM 3.6. *Let u and v be two factors of the same length occurring in a Sturmian word of slope α . Then $u \sim_k v$ if and only if they share a common prefix and a common suffix of length $\min\{|u|, k - 1\}$ and $u \sim_1 v$. Moreover, the condition $u \sim_1 v$ may be omitted if $2(k - 1)\|\alpha\| > 1$.*

The slope of the Fibonacci word is approximately 0.38, so Theorem 3.6 says that the condition $u \sim_1 v$ can be omitted when $k \geq 3$. It is rather surprising that such a weak condition is sufficient to establish k -abelian equivalence. This raises the question of whether it is possible to improve

on the Fibonacci word and have an infinite word for which the condition is redundant even when $k = 2$. We study this question in Section 4.

3.2. The k -Lagrange spectrum. Let $\mathcal{A}e_{k,\alpha}(m)$ be the maximum exponent of k -abelian powers of period m occurring in a Sturmian word of slope α . We define the k -abelian critical exponent of slope α to be the quantity

$$\limsup_{m \rightarrow \infty} \frac{\mathcal{A}e_{k,\alpha}(m)}{m},$$

and we denote it by $\mathcal{A}c_k(\alpha)$. It measures the maximal ratio between the exponent and period of a k -abelian power in a Sturmian word of slope α ; it was introduced in the case $k = 1$ in [9] (in the current paper we follow the notation of the dissertation [22] instead of the article [9]). As mentioned in the introduction, the set of finite values of $\mathcal{A}c_1(\alpha)$ is the Lagrange spectrum [9, Theorem 5.10], so the finite values of $\mathcal{A}c_k(\alpha)$ can be viewed as a combinatorial generalization of the Lagrange spectrum. Thus we give the following definition.

DEFINITION 3.7. The k -Lagrange spectrum \mathcal{L}_k is the set

$$\{\mathcal{A}c_k(\alpha) : \alpha \text{ is irrational}\} \cap \mathbb{R}.$$

In order to study \mathcal{L}_k , we begin by showing how to compute $\mathcal{A}e_{k,\alpha}(m)$ especially when m is a denominator of a convergent of α .

Say that a Sturmian word $\mathbf{s}_{x,\alpha}$ of slope α and intercept x begins with a k -abelian power of period m and exponent n . The prefix of $\mathbf{s}_{x,\alpha}$ of length m and the factor of $\mathbf{s}_{x,\alpha}$ of length m starting after this prefix are k -abelian equivalent, so by Theorem 3.4, the points x and $\{x + m\alpha\}$ lie in a common interval of $\mathcal{I}_{k,m}$. The distance between these points is $\|m\alpha\|$. Thus we see that the points $x, \{x + m\alpha\}, \dots, \{x + (n-1)m\alpha\}$ all lie in a common interval of $\mathcal{I}_{k,m}$, which must have length at least $(n-1)\|m\alpha\|$. Conversely, given such points, we see that the word $\mathbf{s}_{x,\alpha}$ begins with a k -abelian power of period m and exponent n . Thus by considering the longest interval in $\mathcal{I}_{k,m}$, we obtain the following result (recall that $\max \mathcal{I}_{k,m}$ means the maximal length of an interval in $\mathcal{I}_{k,m}$).

LEMMA 3.8. We have

$$\mathcal{A}e_{k,\alpha}(m) = \left\lfloor \frac{\max \mathcal{I}_{k,m}}{\|m\alpha\|} \right\rfloor + \gamma,$$

where γ is 1 if $\max \mathcal{I}_{k,m} \neq \|m\alpha\|$ and 0 otherwise.

EXAMPLE 3.9 (Example 3.5 continued). Observe that the interval of the class $\{10010, 10100\}$ has length α , which means by Lemma 3.8 that using the words in the class we can form a 2-abelian power of period 5 and exponent $\lfloor \alpha / \|5\alpha\| \rfloor + 1 = 5$. Indeed, it is straightforward to check that $(10100)^2(10010)^3$ is a factor of the Fibonacci word. Using words from the

class $\{01010\}$ can only lead to 2-abelian powers of exponent $\lfloor \frac{\|3\alpha\|}{\|5\alpha\|} \rfloor + 1 = 2$. The word $(00100)^2$ is not a factor of the Fibonacci word since it contains 000. Indeed, we see using Lemma 3.8 that the exponent for this class is 1.

Interestingly, if $m = 7$, then the exponent for each equivalence class is 1. The reason is that $\|7\alpha\|$ is large: we have $\|7\alpha\| \approx 0.33$ whereas $\|5\alpha\| \approx 0.09$. The *k*-abelian equivalence relation for $k > 1$ differs in this respect from abelian equivalence: it follows from [9, Theorem 4.7] that in any Sturmian word there exists an abelian square of period m for each $m \geq 1$.

As the number $\max \mathcal{I}_{k,m}$ is generally difficult to find, let us argue next that when m is chosen suitably then, in order to find $\mathcal{A}_{k,\alpha}(m)$, it is sufficient to study the level $2k - 2$ intervals. As in Subsection 3.1, the points $\mathcal{D}_{k,m} = \{0, \{-\alpha\}, \{-2\alpha\}, \dots, \{-(k-1)\alpha\}\}$ together with the points $\mathcal{S} = R^{-(m-(k-1))}(\mathcal{D}_{k,m}) = \{\{-(m-(k-1))\alpha\}, \dots, \{-m\alpha\}\}$ determine the intervals $\mathcal{I}_{k,m}$ of the *k*-abelian equivalence classes. Suppose now that $\|m\alpha\|$ is sufficiently small. Then the points of $R^m(\mathcal{S}) = R^{k-1}(\mathcal{D}_{k,m})$ are close to the points of \mathcal{S} . In fact, when comparing the intervals $\mathcal{I}_{k,m}$ defined by the points of $\mathcal{D}_{k,m} \cup \mathcal{S}$ to those intervals defined by the points of $\mathcal{D}_{k,m} \cup R^{k-1}(\mathcal{D}_{k,m})$, we see that some intervals are shortened by $\|m\alpha\|$ and some intervals are lengthened by $\|m\alpha\|$, but the order of the points is the same whenever $\|m\alpha\|$ is small enough. The points $\{-m\alpha\}$ and 0 however merge, but this is irrelevant when considering $\max \mathcal{I}_{k,m}$ as we only lose a short interval of length $\|m\alpha\|$. Now

$$\mathcal{D}_{k,m} \cup R^{k-1}(\mathcal{D}_{k,m}) = \{\{-(k-1)\alpha\}, \dots, \{-\alpha\}, 0, \alpha, \dots, \{(k-1)\alpha\}\}.$$

Using the fact that R is an isometry, we can study the set $R^{-(k-1)}(\mathcal{D}_{k,m} \cup R^{k-1}(\mathcal{D}_{k,m}))$ instead. This is the set of endpoints of the level $2k - 2$ intervals. It is quite obvious from the preceding that $\|m\alpha\|$ is small enough whenever $\|m\alpha\| < \min L(2k - 2)$. We have thus argued that whenever $\|m\alpha\| < \min L(2k - 2)$, we have

$$|\max \mathcal{I}_{k,m} - \max L(2k - 2)| \leq \|m\alpha\|.$$

Therefore we have proved the following lemma.

LEMMA 3.10. *Let m be a positive integer and suppose that $\|m\alpha\| < \min L(2k - 2)$. Then*

$$\left| \left\lfloor \frac{\max L(2k - 2)}{\|m\alpha\|} \right\rfloor - \mathcal{A}_{k,\alpha}(m) \right| \leq 1.$$

This lemma shows that the exponents of *k*-abelian powers grow arbitrarily large (as we can make $\|m\alpha\|$ as small as desired). A more general result was obtained in [15, Theorem 5.4].

With the results so far, we are able to show that in order to determine $\mathcal{A}_k(\alpha)$ it is sufficient to consider $\mathcal{A}_{k,\alpha}(m)$ only when m is the denominator

of a convergent. Recall that q_t refers to the denominator of the t th convergent of α .

PROPOSITION 3.11. *For all large enough t ,*

$$\mathcal{A}e_{k,\alpha}(m) \leq \mathcal{A}e_{k,\alpha}(q_t) + 2 \quad \text{for all } 1 \leq m < q_{t+1}.$$

Proof. Let $t \geq 1$, and assume t is large enough that $\|q_t\alpha\| < \min L(2k - 2)$. Suppose that m is an integer such that $1 \leq m < q_{t+1}$. By the best approximation property of the convergents, we have $\|m\alpha\| \geq \|q_t\alpha\|$. Suppose first that $\|m\alpha\| < \min L(2k - 2)$. Then by Lemma 3.10,

$$\mathcal{A}e_{k,\alpha}(m) \leq \frac{\max L(2k - 2)}{\|m\alpha\|} + 1 \leq \frac{\max L(2k - 2)}{\|q_t\alpha\|} + 1,$$

so, by the same lemma, we have $\mathcal{A}e_{k,\alpha}(m) \leq \mathcal{A}e_{k,\alpha}(q_t) + 2$. Suppose next that $\|m\alpha\| \geq \min L(2k - 2)$. Then

$$\frac{\max L(m)}{\|m\alpha\|} \leq \frac{\max L(m)}{\min L(2k - 2)} \leq \frac{1}{\min L(2k - 2)},$$

so $\mathcal{A}e_{k,\alpha}(m)$ is bounded by a constant. Thus $\mathcal{A}e_{k,\alpha}(m) < \mathcal{A}e_{k,\alpha}(q_t)$ for all large enough t . The sequence $(\mathcal{A}e_{k,\alpha}(q_i))_i$ reaches arbitrarily high values due to Lemma 3.10. ■

Proposition 3.11 can be improved to $\mathcal{A}e_{k,\alpha}(m) \leq \mathcal{A}e_{k,\alpha}(q_t) + 1$ for all $1 \leq m < q_{t+1}$ and t large enough. Proving this would complicate the argument significantly, and we do not need the improved statement in this paper. It is very well possible that $\mathcal{A}e_{k,\alpha}(q_t) > \mathcal{A}e_{k,\alpha}(q_{t+1})$. For example, if $k = 2$ and say $\alpha = [0; 3, 1, 1, 1, 100, \bar{1}]$, then the sequence of denominators of convergents is $1, 3, 4, 7, \dots$, and it is readily computed that $\mathcal{A}e_{k,\alpha}(4) = 6 > 5 = \mathcal{A}e_{k,\alpha}(7)$. On the other hand, if $k = 1$, then $\mathcal{A}e_{k,\alpha}(m) < \mathcal{A}e_{k,\alpha}(q_t)$ for all t and $1 \leq m < q_t$ as can be readily observed from [9, Theorem 4.7].

For t large enough, let m be an integer such that $q_t \leq m < q_{t+1}$. It follows from Proposition 3.11 that

$$\frac{\mathcal{A}e_{k,\alpha}(m)}{m} \leq \frac{\mathcal{A}e_{k,\alpha}(q_t) + 2}{q_t},$$

so we can conclude using Lemma 3.10 that

$$\mathcal{A}c_k(\alpha) = \limsup_{t \rightarrow \infty} \frac{\mathcal{A}e_{k,\alpha}(q_t)}{q_t} = \limsup_{t \rightarrow \infty} \frac{\max L(2k - 2)}{q_t \|q_t\alpha\|}.$$

When $k = 1$, we obtain

$$\mathcal{A}c_1(\alpha) = \limsup_{t \rightarrow \infty} \frac{1}{q_t \|q_t\alpha\|},$$

so

$$\mathcal{A}c_k(\alpha) = \max L(2k - 2) \cdot \mathcal{A}c_1(\alpha).$$

Let us restate the result.

THEOREM 3.12. *We have $\mathcal{A}_{\mathcal{L}_k}(\alpha) = \max L(2k - 2) \cdot \mathcal{A}_{\mathcal{L}_1}(\alpha)$ for all $k \geq 1$.*

Notice that $\mathcal{A}_{\mathcal{L}_1}(\alpha)$ is finite if and only if α has bounded partial quotients; see (4). Therefore $\mathcal{A}_{\mathcal{L}_k}(\alpha)$ is finite if and only if α has bounded partial quotients. As is well-known, the numbers with bounded partial quotients form a set of measure zero.

As mentioned in Section 2, equivalent numbers have the same Lagrange constant. By Theorem 3.12, this is no longer true when $k > 1$ because $\max L(2k - 2)$ depends on α . It is not difficult to convince oneself that the points obtained in Theorem 3.12 from a single class of equivalent numbers form a dense set. This is what we shall prove next. As a corollary we obtain Theorem 3.15, which states that the k -Lagrange spectrum \mathcal{L}_k is itself dense when $k > 1$. In the statement of the following lemma, $\max L_{\beta}(\ell)$ is the maximal length of a level ℓ interval of slope β .

LEMMA 3.13. *Let α be irrational. The set $\{\max L_{\beta}(\ell) : \beta \text{ is equivalent to } \alpha\}$ is contained and dense in $(\frac{1}{\ell+1}, 1)$ for all $\ell > 1$.*

Proof. Clearly $\max L_{\beta}(\ell) > \frac{1}{\ell+1}$ since there are $\ell + 1$ level ℓ intervals. Let $\gamma \in (\frac{1}{\ell+1}, 1)$, and suppose without loss of generality that it is irrational. By cutting the continued fraction expansion of $1 - \gamma$ after finitely many partial quotients, we obtain a fraction that is as close to $1 - \gamma$ as we desire. Thus we can find a rational β such that $\ell\beta$ is arbitrarily close to $1 - \gamma$ (from either side). Now form an irrational β' by continuing the continued fraction expansion of β in such a way that it is equivalent to α . By selecting the partial quotients appropriately, we find that $\ell\beta'$ is arbitrarily close to $1 - \gamma$. Consider now the level ℓ intervals of slope β' . The longest such interval clearly has length $1 - \ell\beta'$ since $\gamma > \frac{1}{\ell+1}$. Since $1 - \ell\beta'$ is as close to γ as we like, the claim follows. ■

As the smallest element of the Lagrange spectrum is $\sqrt{5}$, Theorem 3.12 and Lemma 3.13 imply the following result.

THEOREM 3.14. *Let $k > 1$. Then $\mathcal{L}_k \subseteq (\frac{\sqrt{5}}{2k-1}, \infty)$ and $\frac{\sqrt{5}}{2k-1}$ is the least accumulation point of \mathcal{L}_k . In particular, the set \mathcal{L}_k is not closed.*

This proposition should be compared with the fact that \mathcal{L}_1 is closed [6, Ch. 3, Theorem 2]. Notice that it also follows that when $k > 1$, the Fibonacci word no longer has minimal critical k -abelian exponent among all Sturmian words.

Let us now recall some remarkable facts about the Lagrange spectrum. *Hall's ray* is the largest half-line contained in \mathcal{L}_1 . It was proven by Hall [12] that the half-line $[6, \infty)$ is contained in \mathcal{L}_1 . Through a series of improvements by several researchers, it was finally determined by Freiman [11] that Hall's

ray equals $[c_F, \infty)$, where c_F is the *Freiman constant*

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

The detailed history and references can be found in [6, Ch. 4]. Hall's result together with Theorem 3.12 and Lemma 3.13 implies the following theorem.

THEOREM 3.15. *The k -Lagrange spectrum \mathcal{L}_k is dense in $(\frac{\sqrt{5}}{2k-1}, \infty)$ when $k > 1$.*

Proof. By Lemma 3.13 and Hall's result, the intervals $(\frac{\sqrt{5}}{2k-1}, \sqrt{5})$ and $(\frac{c_F}{2k-1}, \infty)$ are dense with points of \mathcal{L}_k . Now $c_F < 6$, so $\frac{c_F}{2k-1} \leq 2 < \sqrt{5}$, meaning that these dense sets overlap. ■

We do not know if \mathcal{L}_k contains a half-line when $k > 1$. If true, this is not a straightforward consequence of Hall's and Freiman's results: the union of the dense subsets obtained from each $\theta \in [c_F, \infty)$ by Lemma 3.13 is not automatically a half-line. This poses an interesting open problem.

QUESTION. *Does \mathcal{L}_k contain a half-line when $k > 1$? If so, what is the largest such half-line? Is it $(\frac{c_F}{2k-1}, \infty)$?*

It is conceivable that a point in \mathcal{L}_1 below c_F could map to $\frac{c_F}{2k-1}$. Moreover, \mathcal{L}_1 could contain an interval below c_F (see below) that could produce an interval in \mathcal{L}_k .

The usual Lagrange spectrum is not dense between $\sqrt{5}$ and c_F . In fact, substantial amount of research has been done on maximal gaps occurring in this interval (see for instance [6, Ch. 5]). It is known for example that the set $[\sqrt{5}, 3] \cap \mathcal{L}_1$ is discrete and that the interior of the interval $[\sqrt{12}, \sqrt{13}]$ does not include any points of \mathcal{L}_1 while its endpoints are in \mathcal{L}_1 . It is unknown if \mathcal{L}_1 contains an interval below c_F . The existence of such an interval could show that \mathcal{L}_k also contains an interval below $\frac{c_F}{2k-1}$, but it is plausible that this could also happen for other reasons. For example, it is possible for uncountably many numbers to have the same Lagrange constant, so an interval could be produced by means of Lemma 3.13. One such example is the number 3; it is the Lagrange constant of uncountably many numbers [29, Sect. IV.6, Theorem 3]. We do not believe that this particular example would provide an interval; we just mention it is a possibility. It is known that the part of \mathcal{L}_1 below $\sqrt{689}/8$ has measure zero [3]. It seems to us that studying intervals in \mathcal{L}_k for $k > 1$ is of comparable difficulty as the study of intervals in \mathcal{L}_1 .

Let us also point out that it is easy to come up with numbers greater than $\sqrt{5}/(2k-1)$ that are not in \mathcal{L}_k . The two smallest elements of \mathcal{L}_1 are $\sqrt{5}$ and $\sqrt{8}$, so any point in \mathcal{L}_k between $\sqrt{5}/(2k-1)$ and $\sqrt{8}/(2k-1)$ is of the form $\max L_\alpha(2k-2) \cdot \sqrt{5}$ for some α equivalent to the golden ratio.

The number $\max L_\alpha(2k-2)$ is always irrational, so rational multiples of $\sqrt{5}$ between $\sqrt{5}/(2k-1)$ and $\sqrt{8}/(2k-1)$ are not in \mathcal{L}_k .

3.3. The spectrum \mathcal{L}_∞ . As mentioned in the introduction, when the critical exponent is considered for the equality relation, it is typical to just measure the supremum of fractional exponents, not the ratio of the exponent and the period. In this subsection, we briefly remark what happens if we look at the ratio instead.

Analogous to what we have done already, we set

$$\mathcal{A}_\infty(\alpha) = \limsup_{m \rightarrow \infty} \frac{\mathcal{A}_{\infty, \alpha}(m)}{m},$$

where $\mathcal{A}_{\infty, \alpha}(m)$ is the maximum integer exponent of a power of period m occurring in a Sturmian word of slope α . We further set

$$\mathcal{L}_\infty = \{\mathcal{A}_\infty(\alpha) : \alpha \text{ is irrational}\} \cap \mathbb{R}.$$

We show next that the set \mathcal{L}_∞ contains every nonnegative real number.

PROPOSITION 3.16. *We have $\mathcal{L}_\infty = \mathbb{R}_{\geq 0}$.*

Proof. Consider powers occurring in a Sturmian word of slope α having continued fraction expansion $[0; a_1, a_2, \dots]$ and a sequence of convergents $(p_t/q_t)_t$. It is well-known that if m is not a denominator of a convergent of α , then any power of period m has exponent at most 2 (see, e.g., [7, Lemma 3.6] or [22, Theorem 4.6.5]). Moreover, if $m = q_t$ with $t > 1$, then the highest integer exponent of a power of period m is $a_{t+1} + 2$ [7, Lemma 3.4], [22, Theorem 4.6.5]. Given that we have chosen the partial quotients a_1, a_2, \dots, a_t and thus determined the convergent q_t , we have complete freedom to choose a_{t+1} to make the ratio $(a_{t+1} + 2)/q_t$ behave as we like.

If the sequence $(a_t)_t$ of partial quotients is bounded, then we clearly have $\mathcal{A}_\infty(\alpha) = 0$ because the sequence $(q_t)_t$ is increasing. Hence $0 \in \mathcal{L}_\infty$. Let then λ be a fixed positive real number, and let k_1 be the least integer such that $k_1 > 1$ and that there exist nonnegative integers r_1 and s_1 such that $0 \leq s_1 < q_{k_1}$ and $\lambda - (r_1 + s_1/q_{k_1}) < 1/2$. Set $a_{1,1} = a_1$, $a_{1,2} = a_2, \dots, a_{1,k_1} = a_{k_1}, a_{1,k_1+1} = \max\{1, q_{k_1}(r_1 + s_1/q_{k_1}) - 2\}$, and let $a_{1,t} = 1$ for $t > k_1 + 1$ to obtain a new number α_1 with continued fraction expansion $[0; a_{1,1}, a_{1,2}, \dots]$. Analogously, select then k_2 to be the least positive integer such that $k_2 > k_1$ and that there exist nonnegative integers r_2 and s_2 such that $\lambda - (r_2 + s_2/q_{1,k_2}) < 1/4$ where q_{1,k_2} is the denominator of the k_2 th convergent of α_1 . Set $a_{2,1} = a_{1,1}, \dots, a_{2,k_2} = a_{1,k_2}, a_{2,k_2+1} = \max\{1, q_{1,k_2}(r_2 + s_2/q_{1,k_2}) - 2\}$, and let $a_{2,t} = 1$ for $t > k_2 + 1$ to again obtain a number α_2 with continued fraction expansion $[0; a_{2,1}, a_{2,2}, \dots]$. Repeating this procedure yields sequences $(k_t), (r_t), (s_t)$ and a number β with continued fraction expansion $[0; b_1, b_2, \dots]$ and a subsequence $(p'_t/q'_t)_t$

of its convergents such that

$$\lambda - \frac{b_{k_t+1} + 2}{q'_{k_t}} < \frac{1}{2^t}$$

for all $t \geq 1$ (the numbers a_{t,k_t+1} will grow arbitrarily large since $\lambda > 0$). We conclude that

$$\limsup_{t \rightarrow \infty} \frac{\mathcal{A}_{\infty, \beta}(q'_{k_t})}{q'_{k_t}} = \lambda,$$

so $\mathcal{A}_{\infty}(\beta) \geq \lambda$. As we have constructed the sequence $(b_t)_t$ in such a way that $b_t = 1$ whenever $k_i < t < k_{i+1}$ for some i , it follows that for such i and t large enough

$$\frac{b_t + 2}{q'_{t-1}} \leq \frac{b_{k_i} + 2}{q'_{t-1}} < \frac{b_{k_i} + 2}{q'_{k_i-1}} \leq \lambda.$$

Therefore $\mathcal{A}_{\infty}(\beta) = \lambda$ and $\lambda \in \mathcal{L}_{\infty}$. ■

4. Additional questions. At the end of Subsection 3.1, we asked if there exist infinite words for which the condition of Theorem 3.6 on abelian equivalence is redundant. The next proposition tells us that such binary words exist but that they are rather uninteresting.

PROPOSITION 4.1. *Let \mathbf{w} be an infinite binary word such that for each of its factors u and v of equal length we have $u \sim_1 v$ if they share a common prefix and a common suffix of length 1. Then \mathbf{w} is ultimately periodic.*

Proof. Suppose for a contradiction that \mathbf{w} is aperiodic, so either 00 or 11 occurs in \mathbf{w} . By symmetry, we assume that 00 is a factor of \mathbf{w} . If 0011 occurs also, then 001 and 011 occur. This is impossible as then by our assumption we should have $001 \sim_1 011$; this is clearly absurd. Thus 0010^n1 occurs in \mathbf{w} for some $n \geq 1$. The factors 000 and 010 are also incompatible, so 000 cannot occur in \mathbf{w} . Hence 101 and 1001 are the only possible factors of the form 10^n1 with $n \geq 1$. Since $(100)^\omega$ is not a suffix of \mathbf{w} , either 101 occurs or 10011 must occur. We already ruled out the latter case, so 101 occurs, meaning that 111 is not a factor of \mathbf{w} . If 11 is not a factor, then \mathbf{w} has a suffix that is a concatenation of the words 10 and 100. Suppose then that 11 is a factor. The only way this is possible is that we have an occurrence of 1011. This means that we do not see the incompatible factor 1001. Hence 00 occurs only as a prefix of \mathbf{w} . We have concluded that \mathbf{w} has a suffix that is a product of the words 01 and 011. Thus by mapping \mathbf{w} with the coding $0 \mapsto 1, 1 \mapsto 0$, we obtain a word satisfying the assumptions and which has a suffix that is a product of 10 and 100. Thus without loss of generality, we may assume that \mathbf{w} has such a suffix.

If $100(10)^n100$ occurs in \mathbf{w} for two distinct values of n , then for some $m \geq 0$ both $00(10)^m100$ and $0(10)^{m+1}10$ are factors of \mathbf{w} . By our assump-

tion, we must have $00(10)^m 100 \sim_1 0(10)^{m+1} 10$, but this is false. Therefore $100(10)^n 100$ can occur only for a single value n , and \mathbf{w} must have either of the words $(10)^\omega$ or $(100(10)^n)^\omega$ as a suffix. This is a contradiction. ■

However, if we allow more than two letters, then aperiodicity is possible, as is shown by the next proposition. Let A and B be alphabets. Recall that a substitution $f: A^* \rightarrow B^*$ is a mapping such that $f(uv) = f(u)f(v)$. The image of the infinite word $a_0 a_1 \cdots$ under f is the infinite word $f(a_0)f(a_1)\cdots$. If $w = uv$, then by wv^{-1} we mean the word u . In the next proof, we need to know some properties of Sturmian words; these can be found in [17, Ch. 2]. Firstly, Sturmian words are *balanced*. This means that for any two factors u and v of equal length occurring in some Sturmian word, we have $||u|_0 - |v|_0| \leq 1$. Secondly, in a Sturmian word, there exists exactly one right special factor of length n for all $n \geq 0$. A factor u of an infinite word \mathbf{w} is *right special* if ua and ub occur in \mathbf{w} for distinct letters a and b .

Let σ be the substitution defined by $\sigma(0) = 02$, $\sigma(1) = 1$. It is easy to see that the word $\sigma(\mathbf{s})$ is aperiodic for any Sturmian word \mathbf{s} .

PROPOSITION 4.2. *Let $k \geq 2$ and \mathbf{s} be a Sturmian word containing 00 . Let u and v be two factors of the same length occurring in $\sigma(\mathbf{s})$. Then $u \sim_k v$ if and only if they share a common prefix and a common suffix of length $\min\{|u|, k - 1\}$.*

Proof. Suppose that u and v share a common prefix and a common suffix of length $\min\{|u|, k - 1\}$. We proceed as in the proof of [15, Proposition 2.8] (this is the proof of Proposition 3.1). In this proof it is assumed that $u \sim_1 v$ and a counting argument is used to show that $u \sim_{\ell+1} v$ if $u \sim_\ell v$ for $1 \leq \ell < k$. By a careful analysis, it can be seen that this counting argument only uses the fact that there exists at most one right special factor of length n for each n . Let w and w' be two right special factors of equal length occurring in $\sigma(\mathbf{s})$. It is clear that both w and w' must end with 2. By the form of the substitution σ , there exist words a and b and unique factors x and y of \mathbf{s} such that $a, b \in \{\varepsilon, 0\}$, $|x| \geq |y|$, $aw = \sigma(x)$, and $bw' = \sigma(y)$. Since w and w' are right special, so are x and y . It follows that y is a suffix of x , so w and w' are suffixes of $\sigma(x)$. Since $|w| = |w'|$, they are equal. Thus we argued that $u \sim_k v$ if and only if they share a common prefix and a common suffix of length $\min\{|u|, k - 1\}$ and $u \sim_1 v$. Thus it suffices to show that $u \sim_1 v$.

As above, there exist words a and b and unique factors x and y of \mathbf{s} such that $a \in \{\varepsilon, 0\}$, $b \in \{\varepsilon, 2\}$, $aub = \sigma(x)$, and $avb = \sigma(y)$. Let us show next that x and y are abelian equivalent. The claim follows from this. Since $k \geq 2$, the words x and y end in a common letter c . Now $x \sim_1 y$ if and only if $xc^{-1} \sim_1 yc^{-1}$, so by replacing x with xc^{-1} and y with yc^{-1} if necessary, we may assume that x and y end with 0 (1 is always preceded by 0 since \mathbf{s} is balanced). For each binary word w , we have $|\sigma(w)| = |w| + |w|_0$. Since

$|u| = |v|$ (if x and y were replaced as above, we must replace u and v respectively by $\sigma(xc^{-1})$ and $\sigma(yd^{-1})$), we have

$$(5) \quad |x| + |x|_0 = |y| + |y|_0.$$

Suppose without loss of generality that $|x| \geq |y|$, and write $x = zt$ with $|z| = |y|$. By plugging this into (5), we deduce that $|t| + |t|_0 = |y|_0 - |z|_0$. Since \mathbf{s} is balanced, we see that $|t| + |t|_0 \leq 1$. Thus $t = \varepsilon$ or $t = 1$. The latter case is impossible as x ends with 0, so $t = \varepsilon$. Thus $|x| = |y|$ and so $|x|_0 = |y|_0$ by (5). This means that $x \sim_1 y$. ■

Sturmian and episturmian words have the property of Proposition 3.1 and it was shown in [5] that so does the Cantor word. The authors of [5] asked what sort of words have this property. As we remarked above in the proof of Proposition 4.2, any infinite word having at most one right special factor of each length also has this property. Proposition 4.2 provides more examples of such words.

5. Further open problems. It would be nice if our combinatorial generalization of the Lagrange spectrum had some number-theoretic interpretation, perhaps in connection with rational approximations of irrational numbers. We are unaware of such a connection.

QUESTION. *Is there an arithmetical characterization of the k -Lagrange spectrum \mathcal{L}_k ?*

An obvious open problem is to determine the k -abelian critical exponent of non-Sturmian infinite words. For example: what is the k -abelian critical exponent of the Tribonacci word, the fixed point of the substitution $0 \mapsto 012$, $1 \mapsto 02$, $2 \mapsto 0$? What about the Thue–Morse word? The case $k = 1$ is clear for the Thue–Morse word as the whole infinite word is an abelian power of infinite exponent and period 2.

Instead of looking at particular words or classes of words, it would be interesting to determine the set of critical exponents of all infinite words. In [16], Krieger and Shallit show that every real number greater than 1 is a critical exponent of some infinite word. The result of Freiman [11] shows that every real number greater than c_F is the abelian critical exponent of some infinite word. Our result in Theorem 3.15 shows that a dense subset of $(\frac{c_F}{2k-1}, \infty)$ is attainable as k -abelian critical exponents when $k > 1$. We are thus led to ask the following question ⁽¹⁾.

QUESTION. *Is every nonnegative real number the k -abelian critical exponent of some infinite word?*

⁽¹⁾ Note added in proof: The question has been solved in the positive; see [23].

In [9], the abelian periods of factors of Sturmian words were studied (see that paper for definitions). It was proven for example that the abelian period of a factor of the Fibonacci word is always a Fibonacci number. The same sort of questions could be asked in the *k*-abelian setting for Sturmian words more generally. We have not attempted to study this.

Acknowledgements. We thank the referee for a careful reading of the paper, which improved the presentation.

References

- [1] M. Aigner, *Markov's Theorem and 100 Years of the Uniqueness Conjecture: A Mathematical Journey from Irrational Numbers to Perfect Matchings*, Springer, 2013.
- [2] J. Berstel, *Sturmian and episturmian words*, in: Algebraic Informatics, S. Bozapolidis and G. Rahonis (eds.), Lecture Notes in Computer Sci. 4728, Springer, 2007, 23–47.
- [3] R. T. Bumby, *Hausdorff dimensions of Cantor sets*, J. Reine Angew. Math. 331 (1982), 192–206.
- [4] J. Cassaigne, J. Karhumäki, and A. Saarela, *On growth and fluctuation of *k*-abelian complexity*, Eur. J. Combin. 65 (2017), 92–105.
- [5] J. Chen, X. T. Lü, and W. Wu, *On the *k*-abelian complexity of the Cantor sequence*, J. Combin. Theory Ser. A 155 (2018), 287–303.
- [6] T. W. Cusick and M. E. Flahive, *The Markoff and Lagrange Spectra*, Math. Surveys Monogr. 30, Amer. Math. Soc., Providence, RI, 1989.
- [7] D. Damanik and D. Lenz, *The index of Sturmian sequences*, Eur. J. Combin. 23 (2002), 23–29.
- [8] P. Erdős, *Some unsolved problems*, Michigan Math. J. 4 (1957), 291–300.
- [9] G. Fici, A. Langiu, T. Lecroq, A. Lefebvre, F. Mignosi, J. Peltomäki, and É. Prieur-Gaston, *Abelian powers and repetitions in Sturmian words*, Theoret. Computer Sci. 635 (2016), 16–34.
- [10] G. Fici, F. Mignosi, and J. Shallit, *Abelian-square-rich words*, Theoret. Computer Sci. 684 (2017), 29–42.
- [11] G. A. Freiman, *Diophantine Approximations and Geometry of Numbers (Markov's Problem)*, Kalinin. Gos. Univ., Kalinin, 1975 (in Russian).
- [12] M. Hall, Jr., *On the sum and product of continued fractions*, Ann. of Math. 48 (1947), 966–993.
- [13] J. Justin and G. Pirillo, *Fractional powers in Sturmian words*, Theoret. Computer Sci. 255 (2001), 363–376.
- [14] J. Karhumäki, *Generalized Parikh mappings and homomorphisms*, Information and Control 47 (1980), 155–165.
- [15] 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 (2013), 2189–2206.
- [16] D. Krieger and J. Shallit, *Every real number greater than 1 is a critical exponent*, Theoret. Computer Sci. 381 (2007), 177–182.
- [17] M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia Math. Appl. 90, Cambridge Univ. Press, 2002.
- [18] A. Markoff, *Sur les formes quadratiques binaires indéfinies*, Math. Ann. 15 (1879), 381–406.

- [19] A. Markoff, *Sur les formes quadratiques binaires indéfinies (second mémoire)*, Math. Ann. 17 (1880), 379–399.
- [20] F. Mignosi and G. Pirillo, *Repetitions in the Fibonacci infinite word*, RAIRO Inform. Théor. Appl. 26 (1992), 199–204.
- [21] H. M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. 22 (1921), 84–100.
- [22] J. Peltomäki, *Privileged words and Sturmian words*, Ph.D. dissertation, Turku Centre Computer Sci., Univ. of Turku, 2016.
- [23] J. Peltomäki and M. A. Whiteland, *Every nonnegative real number is an abelian critical exponent*, in: Combinatorics on Words, R. Mercas and D. Reidenbach (eds.), Lecture Notes in Computer Sci. 11682, Springer, 2019, 275–285.
- [24] S. Puzynina and L. Q. Zamboni, *Abelian returns in Sturmian words*, J. Combin. Theory Ser. A 120 (2013), 390–408.
- [25] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lecture Notes in Math. 1794, Springer, 2002.
- [26] G. Richomme, K. Saari, and L. Q. Zamboni, *Abelian complexity of minimal subshifts*, J. London Math. Soc. 83 (2011), 79–95.
- [27] M. Rigo and P. Salimov, *Another generalization of abelian equivalence: Binomial complexity of infinite words*, Theoret. Computer Sci. 601 (2015), 47–57.
- [28] M. Rigo, P. Salimov, and É. Vandomme, *Some properties of abelian return words*, J. Integer Sequences 16 (2013), art. 13.2.5, 21 pp.
- [29] A. M. Rockett and P. Szűsz, *Continued Fractions*, World Sci., 1992.
- [30] A. Thue, *Über unendliche Zeichenreihen*, Christiana Vidensk.-Selsk. Skr. I. Math.-naturv. Kl. 7 (1906), 1–22.

Jarkko Peltomäki
 The Turku Collegium
 for Science and Medicine, TCSM
 Turku Centre for Computer Science, TUCS
 Department of Mathematics and Statistics
 University of Turku
 20014 Turku, Finland
 E-mail: r@turambar.org

Markus A. Whiteland
 Department of Mathematics and Statistics
 University of Turku
 20014 Turku, Finland
 E-mail: mawhit@mpi-sws.org

Abstract (will appear on the journal's web site only)

We study the set of k -abelian critical exponents of all Sturmian words. It has been proven that in the case $k = 1$ this set coincides with the Lagrange spectrum. Thus the sets obtained when $k > 1$ can be viewed as generalized Lagrange spectra. We characterize these generalized spectra in terms of the usual Lagrange spectrum and prove that when $k > 1$ the spectrum is a dense nonclosed set. This is in contrast with the case $k = 1$, where the spectrum is a closed set containing a discrete part and a half-line. We describe explicitly the least accumulation points of the generalized spectra. Our geometric approach allows the study of k -abelian powers in Sturmian words by means of continued fractions.