A CHARACTERIZATION OF STURMIAN SEQUENCES BY INDISTINGUISHABLE ASYMPTOTIC PAIRS

SEBASTIÁN BARBIERI, SÉBASTIEN LABBÉ AND ŠTĚPÁN STAROSTA

Abstract. We give a new characterization of biinfinite Sturmian sequences in terms of indistinguishable asymptotic pairs. Two asymptotic sequences on a full $\mathbb{Z}$-shift are indistinguishable if the sets of occurrences of every pattern in each sequence coincide up to a finitely supported permutation. This characterization can be seen as an extension to biinfinite sequences of Pirillo's theorem which characterizes Christoffel words. Furthermore, we provide a full characterization of indistinguishable asymptotic pairs on arbitrary alphabets using substitutions and biinfinite characteristic Sturmian sequences. The proof is based on the well-known notion of derived sequences.

Keywords: Asymptotic pairs, Sturmian sequences, derived sequences, substitutions, Christoffel words.

MSC2010: *Primary:* 37B10, 68R15, *Secondary:* 37C29,

1. Introduction

Let $\alpha \in [0,1]$ and consider the lower and upper sequences c_α and c'_α given respectively by

$$\begin{array}{rccl} c_\alpha : & \mathbb{Z} & \to & \{\mathtt{0},\mathtt{1}\} \\ & n & \mapsto & \lfloor \alpha(n+1) \rfloor - \lfloor \alpha n \rfloor \end{array} \quad \text{and} \quad \begin{array}{rccl} c'_\alpha : & \mathbb{Z} & \to & \{\mathtt{0},\mathtt{1}\} \\ & n & \mapsto & \lceil \alpha(n+1) \rceil - \lceil \alpha n \rceil. \end{array}$$

When α is rational, the sequences c_α and c'_α are periodic and their period corresponds to Christoffel words [9], see Figure 1. More precisely, the shortest periodic pattern and smallest for the lexicographic

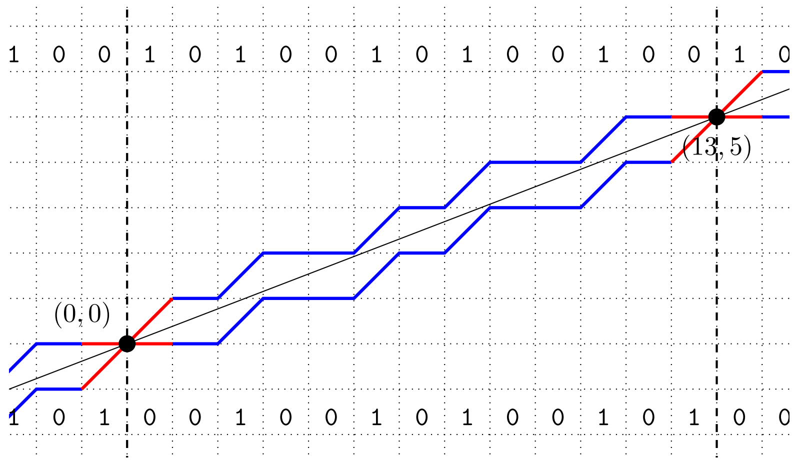

Figure 1. The lower and upper sequences c_α and c'_α when $\alpha = 5/13$ are periodic.

order of c_α is the lower Christoffel word of slope p/q where p and q are nonnegative coprime integers such that $\alpha = p/(p+q)$. For example, when $\alpha = 5/13$, the lower sequence c_α has period `0010010100101` which is the lower Christoffel word of slope $5/8$ and the upper sequence c'_α has period `1010010100100` which is the upper Christoffel word of slope $5/8$. When α is irrational, then c_α and c'_α are not periodic.

The restrictions of c_α and c'_α to $\mathbb{Z}_{\geqslant 1}$ are equal and correspond to the well-known one-sided characteristic Sturmian sequence of slope α [22]. In this work, we consider biinfinite sequences as opposed to one-sided sequences. Over the domain $\mathbb{Z}$, we say that c_α and c'_α are respectively the **lower** and **upper characteristic Sturmian sequences of slope** α whenever α is irrational.

Sturmian sequences have many equivalent definitions, for example, in terms of aperiodic balanced sequences [22], irrational rotations [2, 20], factor complexity [12] or return words [27]. On the other hand, Christoffel words also have many equivalent definitions, including 14 characterizations listed in [7], see also [8, 9]. A recent book [26] gathers exhaustively the combinatorial properties of Christoffel words and uses them to prove two important theorems of Markoff for Diophantine approximations and quadratic forms [21].

In this work, we study a surprising connection between Sturmian sequences and asymptotic pairs satisfying a natural combinatorial property which originates in thermodynamical formalism. This property characterizes asymptotic pairs which induce the trivial linear functional on a space of continuous and shift-invariant cocycles on the asymptotic relation of the full $\mathbb{Z}$-shift. See Section 3 of [4] for further details.

Concretely, given a finite set Σ, we consider the space of sequences $\Sigma^{\mathbb{Z}} = \{x\colon \mathbb{Z} \to \Sigma\}$ endowed with the prodiscrete topology and the shift action $\mathbb{Z} \overset{\sigma}{\curvearrowright} \Sigma^{\mathbb{Z}}$. In this setting, two sequences $x, y \in \Sigma^{\mathbb{Z}}$ are **asymptotic** if x and y differ in finitely many positions of $\mathbb{Z}$. The finite set $F = \{n \in \mathbb{Z} : x_n \neq y_n\}$ is called the **difference set** of (x, y).

Given two asymptotic sequences $x, y \in \Sigma^{\mathbb{Z}}$ with the difference set F, we want to compare the number of occurrences of a fixed pattern. As x and y are asymptotic, occurrences of patterns whose support do not intersect F are the same, so we only need to consider the occurrences of patterns that appear intersecting F. As an example, we can take a fixed symbol $a \in \Sigma$ and define $\Delta_a(x, y)$ as the number of positions $n \in F$ such that $y_n = a$ minus the number of positions $n \in F$ such that $x_n = a$. As F is finite, this value is well defined. More generally, for any given pattern $p\colon S \to \Sigma$ where S is a finite subset of $\mathbb{Z}$, we can consider the difference $\Delta_p(x, y)$ of the number of occurrences of p in y intersecting F minus the number of occurrences of p in x intersecting F.

We say that (x, y) is an **indistinguishable asymptotic pair** if (x, y) is asymptotic and $\Delta_p(x, y) = 0$ for every pattern p. A trivial example of an indistinguishable asymptotic pair is (x, x) for any $x \in \Sigma^{\mathbb{Z}}$. Another simple example is $x, y \in \{\mathtt{0}, \mathtt{1}\}^{\mathbb{Z}}$ where x is equal to $\mathtt{1}$ at the origin, and $\mathtt{0}$ everywhere else, and y is equal to $\mathtt{1}$ at some nonzero $n \in \mathbb{Z}$ and $\mathtt{0}$ everywhere else. Note that in both of these examples x and y lie on the same orbit of $\mathbb{Z} \overset{\sigma}{\curvearrowright} \Sigma^{\mathbb{Z}}$.

In [4] the authors define the following norm on asymptotic sequences of $\Sigma^{\mathbb{Z}}$

$$\|(x, y)\|^*_{\mathsf{NS}} = \sup_{\substack{S \subseteq \mathbb{Z} \\ S \text{ finite}}} \frac{1}{|S|} \sum_{p \in \Sigma^S} |\Delta_p(x, y)|.$$

Every asymptotic pair induces an evaluation map on the space of continuous cocycles on the equivalence relation of asymptotic pairs. The authors show that this norm coincides with the dual norm in the space of linear functionals on the space of continuous cocycles. In other words, the asymptotic pairs which induce the trivial linear functional are precisely the indistinguishable pairs. In this article, we provide a full characterization of which asymptotic pairs induce the trivial linear functional.

Using the notion of indistinguishability, we provide a characterization of the lower and the upper characteristic Sturmian sequences.

Theorem A. *Let $x, y \in \{\mathtt{0}, \mathtt{1}\}^{\mathbb{Z}}$ and assume that x is recurrent. The pair (x, y) is an indistinguishable asymptotic pair with difference set $F = \{-1, 0\}$ such that $x_{-1}x_0 = \mathtt{10}$ and $y_{-1}y_0 = \mathtt{01}$ if and only if* *there exists $\alpha \in [0,1]\backslash\mathbb{Q}$ such that $x = c_\alpha$ and $y = c'_\alpha$ are the lower and upper characteristic Sturmian sequences of slope α.*

Theorem A is proved in Section 3. The auxiliary claims in this section bring a new understanding of Sturmian sequences. In particular, for all $n \in \mathbb{N}$ the words $x_{-n}x_{-n+1}\cdots x_{n-1}$ and $y_{-n}y_{-n+1}\cdots y_{n-1}$ of length $2n$ are optimal representations of the language of c_α as both contain exactly one occurrence of every factor of length n, see Corollary 3.6. Removing the hypothesis that x is recurrent, we obtain a unifying description of the lower and upper characteristic Sturmian sequences and their limits as their slope tends towards a rational value.

Theorem B. *Let $x, y \in \{0,1\}^{\mathbb{Z}}$. The pair (x, y) is an indistinguishable asymptotic pair with difference set $F = \{-1, 0\}$ such that $x_{-1}x_0 = 10$ and $y_{-1}y_0 = 01$ if and only if there exists $(\alpha_n)_{n\in\mathbb{N}}$ with $\alpha_n \in [0,1]\backslash\mathbb{Q}$ such that*

$$x = \lim_{n\to\infty} c_{\alpha_n} \qquad \textit{and} \qquad y = \lim_{n\to\infty} c'_{\alpha_n}.$$

In the case where x is not recurrent, then x and y lie on the same orbit and there exist coprime integers $p, q \in \mathbb{Z}_{\geqslant 0}$ such that (x, y) is the limit of asymptotic pairs formed by the lower and upper characteristic Sturmian sequences of slope α_n as α_n converges toward the rational slope $p/(p+q) \in [0,1]\cap\mathbb{Q}$ either from above or from below, see Theorem 4.5. Limits of the lower and upper characteristic Sturmian sequences as their slope tends to a rational number are expressed in terms of Christoffel words, see Lemma 4.2. The proof of Theorem B follows from Theorem A and Theorem 4.5 and is proved in Section 4.

Theorem A and Theorem B are related to a famous theorem of Pirillo [23] which provides a characterization of Christoffel words of slope p/q where p and q are positive coprime integers. If p and q are nonzero, the lower Christoffel word of slope p/q starts with letter 0 and ends with letter 1, so it can be written as $0m1$ for some finite word $m \in \{0,1\}^*$ and the corresponding upper Christoffel word is $1m0$. Pirillo gave the following elegant characterization of Christoffel words. Recall that two words $w, w' \in \{0,1\}^*$ are **conjugate** if there exists $u, v \in \{0,1\}^*$ such that $w = uv$ and $w' = vu$.

Pirillo's Theorem ([23])**.** *The word $0m1 \in \{0,1\}^*$ is a lower Christoffel word if and only if $0m1$ and $1m0$ are conjugate.*

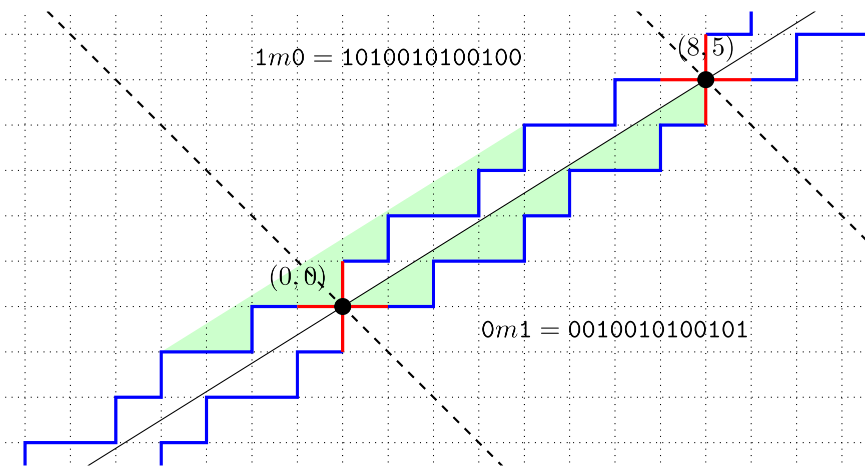

FIGURE 2. Pirillo's theorem characterizes Christoffel words: the lower Christoffel word $0m1 \in \{0,1\}^*$ is conjugate to the upper Christoffel word $1m0$.

Pirillo's theorem is illustrated in Figure 2. We observe that the conjugacy of $\mathtt{0}m\mathtt{1}$ into $\mathtt{1}m\mathtt{0}$ is done via their factorization into a product of two palindromes: $\mathtt{0}m\mathtt{1} = \mathtt{00100} \cdot \mathtt{10100101}$ and $\mathtt{1}m\mathtt{0} = \mathtt{10100101} \cdot \mathtt{00100}$. The factorization of $\mathtt{0}m\mathtt{1}$ as a product of two palindromes and the fact that the central word m is a palindrome [9, Prop. 4.2] is also a characterization of Christoffel words, see [13] and [26, Theorem 12.2.10].

Pirillo's theorem can be restated for biinfinite sequences as follows: *c_α is the lower sequence associated to the rational slope $\alpha = p/(p+q)$ for some coprime nonnegative integers p, q if and only if c_α is a shift of c'_α.* It is natural to ask if there is an analogous statement which holds as we take the limit $\frac{p}{p+q} \to \alpha$ for some irrational $\alpha \in [0,1]\backslash\mathbb{Q}$. In this light, Theorem B can be considered as the extension of Pirillo's theorem to aperiodic biinfinite sequences where the notion of conjugacy of words is replaced by the notion of indistinguishability of an asymptotic pair. This seems to be the correct approach since other alternatives (e.g., having the same language, see Remark 3.8) fail.

The next result provides a full characterization of non-trivial indistinguishable asymptotic pairs for $\mathbb{Z}$ which does not depend upon the form of the difference set or the alphabet. More precisely, we show that every indistinguishable asymptotic pair can be obtained from limits of pairs of lower and upper characteristic Sturmian sequences by means of shifts and substitutions.

Given finite sets Σ, Γ, a map $\varphi\colon \Sigma \to \Gamma^+$ which replaces symbols of Σ by nonempty words on Γ is called a substitution. This map is naturally extended by concatenation to a continuous map $\varphi\colon \Sigma^{\mathbb{Z}} \to \Gamma^{\mathbb{Z}}$.

Theorem C. *Let Σ be a finite alphabet and $x, y \in \Sigma^{\mathbb{Z}}$ a non-trivial asymptotic pair. Then x, y is indistinguishable if and only if either*

- *x is recurrent and there exists $\alpha \in [0,1]\backslash\mathbb{Q}$, a substitution $\varphi\colon \{\mathtt{0},\mathtt{1}\} \to \Sigma^+$ and an integer $m \in \mathbb{Z}$ such that*
$$\{x, y\} = \{\sigma^m\varphi(\sigma(c_\alpha)), \sigma^m\varphi(\sigma(c'_\alpha))\},$$
- *x is not recurrent and there exists a substitution $\varphi\colon \{\mathtt{0},\mathtt{1}\} \to \Sigma^+$ and an integer $m \in \mathbb{Z}$ such that*
$$\{x, y\} = \{\sigma^m\varphi({}^\infty\mathtt{0}.\mathtt{1}\mathtt{0}^\infty), \sigma^m\varphi({}^\infty\mathtt{0}.\mathtt{0}\mathtt{1}\mathtt{0}^\infty)\}.$$

This means that every indistinguishable asymptotic pair in $\mathbb{Z}$ consists either of (1) two sequences in the same orbit, which are shifts of a sequence of the form ${}^\infty v.uv^\infty$ for some $u, v \in \Sigma^+$, or (2) two sequences which, up to translation, can be obtained through a substitution from a pair of lower and upper characteristic Sturmian sequences. In simpler terms, all non-trivial examples of one-dimensional indistinguishable asymptotic pairs arise from irrational circle rotations. The proof of Theorem C is given in Section 5. It is based on the well-known notions of return words and derived sequences [14].

Acknowledgments: The first two authors were supported by the Agence Nationale de la Recherche through the projects CODYS (ANR-18-CE40-0007) and CoCoGro (ANR-16-CE40-0005). S. Barbieri was also supported by the FONDECYT grant 11200037. Š. Starosta acknowledges the support of the OP VVV MEYS funded project CZ.02.1.01/0.0/0.0/16_019/0000765 "Research Center for Informatics". This work originated from a visit of the first two authors to Prague in October 2019 supported by PHC Barrande, a France-Czech Republic bilateral funding and grant no. 7AMB18FR048 of MEYS of Czech Republic.

2. Preliminaries

Let $\mathbb{N}$ denote the set of nonnegative integers. Intervals consisting of integers will be written using the notation $[\![n, m]\!] = [n, m] \cap \mathbb{Z}$, for $n, m \in \mathbb{Z}$.

Let Σ be a finite set to which we refer as an **alphabet**. An element $x \in \Sigma^{\mathbb{Z}} = \{x \colon \mathbb{Z} \to \Sigma\}$ is called a biinfinite **sequence**. We shall often omit the word "biinfinite" and use the word "one-sided" to refer to functions with domain $\mathbb{Z}_{\geqslant 1}$. For $n \in \mathbb{Z}$, we write x_n to denote the value $x(n)$. The set $\Sigma^{\mathbb{Z}}$ of all sequences is endowed with the prodiscrete topology, which is generated by the metric

$$d(x,y) = 2^{-\inf\{|n| \,:\, n \in \mathbb{Z} \text{ and } x_n \neq y_n\}}.$$

The **shift** is the map $\sigma \colon \Sigma^{\mathbb{Z}} \to \Sigma^{\mathbb{Z}}$ where

$$(\sigma(x))_m = x_{m+1} \qquad \text{for every } m \in \mathbb{Z} \text{ and } x \in \Sigma^{\mathbb{Z}}.$$

Let us represent pictorially a sequence $x \in \Sigma^{\mathbb{Z}}$ by marking the position of the zero coordinate with a point as follows:

$$x = \dots x_{-5}x_{-4}x_{-3}x_{-2}x_{-1}.x_0x_1x_2x_3x_4\dots$$

Given words $u, w \in \Sigma^+$ and $y, z \in \Sigma^*$ we shall use the notation

$$x = {}^{\infty}uy.zw^{\infty} \in \Sigma^{\mathbb{Z}}$$

to indicate that the sequence consists of repeated concatenations of u to the left of y, and of repeated concatenations of w to the right of z.

Definition 2.1. *We say that two sequences x, y are **asymptotic**, or that (x, y) is an asymptotic pair, if the set $F = \{n \in \mathbb{Z} \colon x_n \neq y_n\}$ is finite. F is called the **difference set** of (x, y). If $x = y$ we say that the asymptotic pair is **trivial**.*

Equivalently, x, y are asymptotic if for every sequence $\{n_i\}_{i \in \mathbb{N}}$ of elements of $\mathbb{Z}$ such that $|n_i| \to \infty$, the distance $d(\sigma^{n_i}(x), \sigma^{n_i}(y))$ converges to zero.

For finite $S \subseteq \mathbb{Z}$, a function $p \colon S \to \Sigma$ is called a **pattern** and the set S is its **support**. Given a pattern $p \in \Sigma^S$, the **cylinder** centered at p is $[p] = \{x \in \Sigma^{\mathbb{Z}} \colon x|_S = p\}$. We say a pattern p **appears** in $x \in \Sigma^{\mathbb{Z}}$ if there exists $n \in \mathbb{Z}$ such that $\sigma^n(x) \in [p]$. Let us also denote by $\mathsf{occ}_p(x) = \{n \in \mathbb{Z} \colon \sigma^n(x) \in [p]\}$ the set of **occurrences** of p in $x \in \Sigma^{\mathbb{Z}}$.

Given a pattern p and sequences $x, y \in \Sigma^{\mathbb{Z}}$, we want to define a number which counts the difference between the occurrences of p in y compared to the occurrences of p in x. Naively, if $\mathbb{1}_{[p]}$ is the indicator function of $[p]$, we would like to sum over all integers n the difference $\mathbb{1}_{[p]}(\sigma^n(y)) - \mathbb{1}_{[p]}(\sigma^n(x))$. For arbitrary $x, y \in \Sigma^{\mathbb{Z}}$ this sum is not well defined. However, it can be given meaning if x, y are asymptotic. Indeed, if F denotes the difference set of x, y and S denotes the support of p, then $F - S = \{f - s : f \in F, s \in S\}$ is the set of all integers for which there exists an $s \in S$ such that $n + s \in F$. In consequence, we have that if $n \in \mathbb{Z} \setminus (F - S)$ then for every $s \in S$ we have that $n + s \notin F$ and thus $\sigma^n(x)_s = \sigma^n(y)_s$, which implies in turn that $\mathbb{1}_{[p]}(\sigma^n(x)) = \mathbb{1}_{[p]}(\sigma^n(y))$. This motivates the following definition.

Definition 2.2. *Let p be a pattern with finite support $S \subseteq \mathbb{Z}$ and $x, y \in \Sigma^{\mathbb{Z}}$ be asymptotic sequences with difference set F. The **discrepancy** of the pattern p associated to the pair (x, y) is given by*

$$\Delta_p(x,y) = \sum_{n \in F - S} \mathbb{1}_{[p]}(\sigma^n(y)) - \mathbb{1}_{[p]}(\sigma^n(x)).$$

For example, the discrepancy of the pattern $p = abcabc$ in the sequences

$$x = \cdots bcabcbc\underline{abcabc}bc\underline{abc.}\boxed{\underline{abc}}bcabcbc\underline{abcabc}bcabcbc\cdots,$$

$$y = \cdots bcabcbc\underline{abcabc}bcabc.\boxed{bc\underline{a}}\underline{bcabc}bc\underline{abcabc}bcabcbc\cdots$$

is $\Delta_{abcabc}(x, y) = 1 - 1 = 0$, because both x and y contain exactly one occurrence of the pattern $p = abcabc$ intersecting the difference set $F = \{0, 1, 2\}$.

Definition 2.3. *We say that an asymptotic pair (x, y) is an* ***indistinguishable asymptotic pair*** *if the discrepancy of every pattern p of finite support is $\Delta_p(x, y) = 0$.*

A related notion is the one of *local indistinguishability* given in [3, § 5.1] which corresponds symbolically to having two sequences with the same language (as defined in the next section). This notion applies to a more general context as it can be defined for pairs which are not asymptotic. An indistinguishable asymptotic pair is locally indistinguishable but the converse is not true even for asymptotic pairs as explained in Remark 3.8.

Whenever x, y are asymptotic, for every pattern p the sets $\mathsf{occ}_p(x)$ and $\mathsf{occ}_p(y)$ are asymptotic when seen as sequences in $\{\mathtt{0}, \mathtt{1}\}^{\mathbb{Z}}$. In consequence, the condition $\Delta_p(x, y) = 0$ is equivalent to having $\mathsf{occ}_p(x)$ and $\mathsf{occ}_p(y)$ coincide up to a finitely supported permutation of $\mathbb{Z}$. More precisely, we have:

$$\#(\mathsf{occ}_p(x) \cap (F - S)) = \#(\mathsf{occ}_p(y) \cap (F - S)).$$

We are interested in understanding which asymptotic pairs are indistinguishable. In order to avoid simple cases, we will restrict our search to asymptotic pairs which are non-trivial. Notice that non-trivial indistinguishable asymptotic pairs may consist of sequences that lie in the same orbit. For example,

$$x = {}^{\infty}\mathtt{0000}\boxed{\mathtt{00.11}}\mathtt{0000}^{\infty} \qquad \text{and} \qquad y = {}^{\infty}\mathtt{0000}\boxed{\mathtt{11.00}}\mathtt{0000}^{\infty}$$

is an indistinguishable asymptotic pair for $\Sigma = \{\mathtt{0}, \mathtt{1}\}$ where the difference set $F = [\![-2, 1]\!]$ is shown in boxes. Here $y = \sigma^2(x)$.

2.1. **Basic properties of indistinguishable asymptotic pairs.** We defined indistinguishable asymptotic pairs through patterns whose support is an arbitrary subset of $\mathbb{Z}$. Next, we show that we can equivalently characterize indistinguishability using factors, first recalling the definitions.

A pattern w whose support is the set $[\![0, n-1]\!]$ for some $n \in \mathbb{N}$ is a **word**, and we write $w = w_0 \dots w_{n-1}$. The length of w is denoted by $|w| = n$. Let us denote the set of all finite words with symbols in Σ by $\Sigma^* = \bigcup_{n \in \mathbb{N}} \Sigma^{[\![0,n-1]\!]}$.

A word w is a **factor** of x if it appears in x. We write $\mathcal{L}_n(x)$ for the set of all factors of x of length n and the **language of** x is the union $\mathcal{L}(x)$ of the sets $\mathcal{L}_n(x)$ for every $n \in \mathbb{N}$.

Proposition 2.4. *An asymptotic pair $x, y \in \Sigma^{\mathbb{Z}}$ is indistinguishable if and only if for every $w \in \Sigma^*$ we have*

$$\Delta_w(x, y) = 0.$$

Proof. One direction is obvious. Let us suppose that for every $w \in \Sigma^*$ we have $\Delta_w(x, y) = 0$ and let $S \subseteq \mathbb{Z}$ be a support and $p \in \Sigma^S$. For every $m \in \mathbb{Z}$ we can define $p' \in \Sigma^{m+S}$ by $p'(m+s) = p(s)$ for every $s \in S$. For every sequence $x \in \Sigma^{\mathbb{Z}}$ we have that $x \in [p]$ if and only if $\sigma^{-m}(x) \in [p']$. Consequently we have that $\mathsf{occ}_p(x) = \mathsf{occ}_{p'}(x) + m$ and thus,

$$\Delta_{p'}(x, y) = \Delta_p(x, y) \text{ for every asymptotic pair } x, y \in \Sigma^{\mathbb{Z}}.$$

By the former argument, without loss of generality, we may assume that $S \subseteq [\![0, n-1]\!]$ for some large enough n.

Notice that $[p]$ is the disjoint union of all $[w]$ where w is a word of length n such that $w|_S = p$. It follows that for any $z \in \Sigma^{\mathbb{Z}}$ we have $\mathbb{1}_{[p]}(z) = 1$ if and only if there is a unique such w such that $[w] \subseteq [p]$ and $\mathbb{1}_{[w]}(z) = 1$. Letting F be the difference set of x, y we obtain,

$$\begin{aligned}\Delta_p(x,y) &= \sum_{n\in F-S} \mathbb{1}_{[p]}(\sigma^n(y)) - \mathbb{1}_{[p]}(\sigma^n(x))\\ &= \sum_{n\in F-\llbracket 0,n-1\rrbracket} \mathbb{1}_{[p]}(\sigma^n(y)) - \mathbb{1}_{[p]}(\sigma^n(x))\\ &= \sum_{n\in F-\llbracket 0,n-1\rrbracket} \sum_{\substack{w\in\Sigma^{\llbracket 0,n-1\rrbracket}\\ [w]\subseteq[p]}} \mathbb{1}_{[w]}(\sigma^n(y)) - \mathbb{1}_{[w]}(\sigma^n(x)).\end{aligned}$$

Exchanging the order of the sums yields

$$\Delta_p(x,y) = \sum_{\substack{w\in\Sigma^{\llbracket 0,n-1\rrbracket}\\ [w]\subseteq[p]}} \Delta_w(x,y) = 0.$$

And thus x, y is an indistinguishable asymptotic pair. □

Given a sequence $x \in \Sigma^{\mathbb{Z}}$, define its **reversal** x^R by setting $x^R(n) = x(-n)$ for every $n \in \mathbb{N}$. The next proposition states that indistinguishable asymptotic pairs are stable under actions of the affine group of $\mathbb{Z}$.

Proposition 2.5. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be an indistinguishable asymptotic pair.*

(1) For every $n \in \mathbb{Z}$, $(\sigma^n(x), \sigma^n(y))$ is an indistinguishable asymptotic pair.
(2) (x^R, y^R) is an indistinguishable asymptotic pair.

Proof. Given a pattern p with support S, denote by $-p$ the pattern with support $-S$ such that $-p(-s) = p(s)$. It is clear that if $x \in \Sigma^{\mathbb{Z}}$ and $n \in \mathbb{N}$, then for every pattern p,

$$\mathsf{occ}_p(x) = \mathsf{occ}_p(\sigma^n(x)) - n \quad \text{ and } \quad \mathsf{occ}_p(x) = -\mathsf{occ}_{-p}(x^R).$$

Note that if x, y is an asymptotic pair with difference set F, then the difference set for $\sigma^n(x), \sigma^n(y)$ is $F - n$ and the difference set for x^R, y^R is $-F$.

From the relations on the occurrence sets, we obtain that for every pattern p we have that

$$\Delta_p(x,y) = \Delta_p(\sigma^n(x), \sigma^n(y)) = \Delta_{-p}(x^R, y^R).$$

In particular, x, y is indistinguishable if and only if $\sigma^n(x), \sigma^n(y)$ is indistinguishable if and only if x^R, y^R is indistinguishable. □

Let us recall that a sequence $(x_m)_{m\in\mathbb{N}}$ of sequences in $\Sigma^{\mathbb{Z}}$ converges to $\bar{x} \in \Sigma^{\mathbb{Z}}$ if for every $n \in \mathbb{Z}$ we have that $(x_m)_n = \bar{x}_n$ for all large enough $m \in \mathbb{N}$. If $(x_m, y_m)_{m\in\mathbb{N}}$ is a sequence of asymptotic pairs, it is natural to ask that both $(x_m)_{m\in\mathbb{N}}$ and $(y_m)_{m\in\mathbb{N}}$ converge to say that $(x_m, y_m)_{m\in\mathbb{N}}$ converges. However, if we only asked for that there would be no guarantee that the limit is also an asymptotic pair. We shall consider a slightly stronger notion of convergence for asymptotic pairs which ensures that the limit is also an asymptotic pair.

Definition 2.6. *We say that a sequence $(x_n, y_n)_{n\in\mathbb{N}}$ of asymptotic pairs* ***converges*** *to a pair (x, y) if $(x_n)_{n\in\mathbb{N}}$ converges to x, $(y_n)_{n\in\mathbb{N}}$ converges to y, and there exists a finite set $F \subseteq \mathbb{Z}$ so that $x_n|_{\mathbb{Z}\setminus F} = y_n|_{\mathbb{Z}\setminus F}$ for all large enough $n \in \mathbb{N}$.*

This notion of convergence is also used in the theory of topological orbit equivalence of Cantor minimal systems. An interested reader can refer to [24] for further information. The advantage of this notion in our context is that it preserves indistinguishability.

Proposition 2.7. *Let $(x_n, y_n)_{n\in\mathbb{N}}$ be a sequence of asymptotic pairs in $\Sigma^{\mathbb{Z}}$ which converges to (x, y). If for every $n \in \mathbb{N}$ we have that (x_n, y_n) is indistinguishable, then (x, y) is indistinguishable.*

Proof. Let $p \in \Sigma^S$ be a pattern. As $(x_n, y_n)_{n\in\mathbb{N}}$ converges to (x, y), there exists a finite set $F \subseteq \mathbb{Z}$ and $N_1 \in \mathbb{N}$ so that $x_n|_{\mathbb{Z}\setminus F} = y_n|_{\mathbb{Z}\setminus F}$ for every $n \geqslant N_1$. In particular we have that the difference sets of (x, y) and (x_n, y_n) for $n \geqslant N_1$ are contained in F. It suffices thus to show that

$$\#\{\mathsf{occ}_p(x) \cap (F - S)\} = \#\{\mathsf{occ}_p(y) \cap (F - S)\}.$$

As $(x_n)_{n\in\mathbb{N}}$ converges to x and $(y_n)_{n\in\mathbb{N}}$ converges to y, there exists $N_2 \in \mathbb{N}$ such that

$$x_n|_{(F-S)+S} = x|_{(F-S)+S} \text{ and } y_n|_{(F-S)+S} = y|_{(F-S)+S} \text{ for all } n \geqslant N_2.$$

This implies that $\mathsf{occ}_p(x) \cap (F - S) = \mathsf{occ}_p(x_n) \cap (F - S)$ and $\mathsf{occ}_p(y) \cap (F - S) = \mathsf{occ}_p(y_n) \cap (F - S)$ for every $n \geqslant N_2$. Taking $N = \max\{N_1, N_2\}$, as (x_n, y_n) is indistinguishable with the difference set contained in F, it follows that for $n \geqslant N$ we have $\#\{\mathsf{occ}_p(x_n) \cap (F - S)\} = \#\{\mathsf{occ}_p(y_n) \cap (F - S)\}$. Therefore we obtain $\#\{\mathsf{occ}_p(x) \cap F - S\} = \#\{\mathsf{occ}_p(y) \cap F - S\}$. As this argument holds for every pattern p, we conclude that (x, y) is indistinguishable. $\square$

2.2. **Recurrence of indistinguishable asymptotic pairs.** In this section, we study the recurrence of indistinguishable asymptotic pairs. To that end, we shall first show that if x, y is an indistinguishable asymptotic pair, then every pattern with support $I \subseteq \mathbb{Z}$ which appears in x must necessarily appear at some position $n \in \mathbb{Z}$ so that $n + I$ intersects the difference set of x, y. More precisely, let us say that the **occurrences of a word** $w \in \mathcal{L}(x)$ **intersect** F **in** x if $\mathsf{occ}_w(x) \cap (F - [\![0, |w| - 1]\!]) \neq \varnothing$. Equivalently, there exist $i \in F$ and $j \in [\![0, |w| - 1]\!]$ such that $\sigma^{i-j}(x) \in [w]$.

Lemma 2.8. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be a non-trivial indistinguishable asymptotic pair with the difference set F. The occurrences of every $w \in \mathcal{L}(x)$ intersect F in x.*

Proof. Without loss of generality, let us suppose that F is contained in $[\![0, k - 1]\!]$, $x_0 \neq y_0$, and $x_{k-1} \neq y_{k-1}$. Let us write $I = [\![0, |w| - 1]\!]$. If $(F - I) \cap \mathsf{occ}_w(x) = \varnothing$, then there is $u \in \mathsf{occ}_w(x)$ such that either (1) u is the smallest value satisfying $u \geqslant k$ or (2) u is the largest value satisfying $u \leqslant -|w|$. We shall deal with case the (1), the second case is analogous.

As $[\![0, k - 1]\!]$ contains the difference set of x, y we have that $x_u \dots, x_{u+|w|-1} = y_u \dots, y_{u+|w|-1}$. Let $w' = y_{k-1} \dots y_{u+|w|-1}$. As $x_{k-1} \neq y_{k-1}$, we obtain $\sigma^{k-1}(y) \in [w']$ but $\sigma^{k-1}(x) \notin [w']$. As $\Delta_{w'}(x, y) = 0$, there must be some $j' \in F$ and $i' \in [\![0, u + |w| - k]\!]$ such that $j' - i' \neq k - 1$ and $\sigma^{j'-i'}(x) \in [w']$ and, since w is a suffix of w', $\sigma^{j'-i'+u-(k-1)}(x) \in [w]$.

Let $u' = j' - i' + u - (k - 1)$. On one hand, we have $j' - i' < k - 1$ and thus $u' < u$. On the other hand, $u' \geqslant 0 - (u + |w| - k) + u - (k - 1) = -|w| + 1$. Since u is the smallest value of $\mathsf{occ}_w(x)$ satisfying $u \geqslant k$, we obtain $u' \in F - I$ which is a contradiction. $\square$

A sequence $x \in \Sigma^{\mathbb{Z}}$ is called **recurrent** if every $w \in \mathcal{L}(x)$ occurs at least twice in x. It is quite easy to see that x is recurrent if and only if $\mathsf{occ}_w(x)$ is in fact an infinite set for every $w \in \mathcal{L}(x)$.

We say that x is **uniformly recurrent** if every $w \in \mathcal{L}(x)$ appears with bounded gaps, that is, for every $w \in \mathcal{L}(x)$ there exists an integer $g \geqslant 1$ such that for every $n \in \mathbb{Z}$ there is $0 \leqslant m \leqslant g$ such that $\sigma^{n+m}(x) \in [w]$.

It is clear that if (x, y) is an indistinguishable asymptotic pair, then x is (uniformly) recurrent if and only if y is (uniformly) recurrent.

Proposition 2.9. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be an indistinguishable asymptotic pair. If x is not recurrent, then x and y lie in the same orbit.*

Proof. If x is not recurrent, there is a word $w \in \mathcal{L}(x)$ such that $\#(\mathsf{occ}_w(x)) = 1$. Without loss of generality let us assume that w occurs at the origin, that is, $\mathsf{occ}_w(x) = \{0\}$. As $\Delta_w(x, y) = 0$, it follows that $\#(\mathsf{occ}_w(y)) = 1$ as well. Let m be the only integer such that $\sigma^m(y) \in [w]$.

Let $n \in \mathbb{N}$ be larger than the length of w. Let $q_n = x|_{[\![-n,n]\!]}$. As $x \in [q_n]$ and $\Delta_{q_n}(x,y) = 0$, there exists $k \in \mathbb{Z}$ so that $\sigma^k(y) \in [q_n]$. Furthermore, as $q_n|_{[\![0,|w|-1]\!]} = w$, it follows that $\sigma^k(y) \in [w]$ and thus $k = m$. Therefore we obtain that $\sigma^m(y) \in [q_n]$ for every large enough n. As $\bigcap_{n\in\mathbb{N}}[q_n] = \{x\}$ we deduce that $\sigma^m(y) = x$. □

Next we are going to show that recurrent indistinguishable asymptotic pairs are in fact uniformly recurrent. To that end, we recall the notions of return word [14, 18, 27] and complete return word [15].

Definition 2.10. *A word $u \in \Sigma^+$ is a* ***complete return word*** *to $w \in \Sigma^+$ in $x \in \Sigma^{\mathbb{Z}}$ if u appears in x, $u = ws = pw$ for some nonempty words $p, s \in \Sigma^+$, and there are only two occurrences of w in u, one as a prefix and one as a suffix. The word p is called a* ***return word*** *to w in x.*

Note that the two occurrences of w in a complete return word $u = ws = pw$ to w may overlap. Denote the set of all complete return words to w in x by $\texttt{CRW}_w(x)$ and the set of all return words to w in x by $\texttt{RW}_w(x)$. The following fact is elementary.

Lemma 2.11. *Let $x \in \Sigma^{\mathbb{Z}}$. The following are equivalent.*

(1) x is uniformly recurrent.

(2) x is recurrent and for every $w \in \mathcal{L}(x)$ we have that $\texttt{CRW}_w(x)$ is finite.

(3) x is recurrent and for every $w \in \mathcal{L}(x)$ we have that $\texttt{RW}_w(x)$ is finite.

Lemma 2.12. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be a non-trivial indistinguishable asymptotic pair. If x is recurrent, then x is uniformly recurrent.*

Proof. Without loss of generality, suppose the difference set of (x,y) is contained in $F = [\![0, k-1]\!]$. Suppose x is not uniformly recurrent. By Lemma 2.11 there is a word $w \in \mathcal{L}(x)$ such that $\texttt{CRW}_x(w)$ is infinite. As $\texttt{CRW}_x(w)$ is infinite, there exist distinct $v_1, v_2, v_3 \in \texttt{CRW}_x(w)$ such that $\min\{|v_1|, |v_2|\} > k + 2|w|$ and $|v_3| > k + 2\max\{|v_1|, |v_2|\}$. By Lemma 2.8 the words v_1, v_2 and v_3 must occur in x at positions such that their support intersects $F = [\![0, k-1]\!]$.

As $\min\{|v_1|, |v_2|\} > k + 2|w|$, exactly one of the two occurrences of w in v_1 must be completely contained in the support $L_1 = [\![-|v_1|+1, -1]\!]$ or the support $R_1 = [\![k, k+|v_1|-1]\!]$. Similarly, exactly one occurrence of w in v_2 appears in $L_2 = [\![-|v_2|+1, -1]\!]$ or $R_2 = [\![k, k+|v_2|-1]\!]$. As v_1, v_2 are distinct complete return words, if an occurrence of w coming from v_1 appears in L_1, then another coming from v_2 appears in R_2. Analogously, if there is an occurrence of w coming from v_1 in R_1, then an occurrence of w coming from v_2 appears in L_2.

In consequence with the reasoning above, the word w appears completely contained both in the interval $[\![-\max\{|v_1|, |v_2|\}+1, -1]\!]$ and in the interval $[\![k, k+\max\{|v_1|, |v_2|\}-1]\!]$. As v_3 is also a complete return word which appears intersecting F and $|v_3| > k + 2\max\{|v_1|, |v_2|\}$, this means that there are no copies of w completely contained in either $[\![-\max\{|v_1|, |v_2|\}+1, -1]\!]$ or $[\![k, k+\max\{|v_1|, |v_2|\}-1]\!]$, contradicting the above statement. □

Gathering Proposition 2.9 and Lemma 2.12 we obtain the following beautiful dichotomy.

Corollary 2.13. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be a non-trivial asymptotic indistinguishable pair. Then exactly one of the following statements holds*

(1) $x = \sigma^n(y)$ for some nonzero $n \in \mathbb{Z}$,

(2) x and y are uniformly recurrent.

Proof. If x is not recurrent, then by Proposition 2.9 we obtain that $x = \sigma^n(y)$ for some $n \in \mathbb{Z}\backslash\{0\}$. If x is recurrent, then y is also recurrent. Applying Lemma 2.12 we obtain that x and y are uniformly recurrent.

Let us assume that both conditions happen at the same time. As $x = \sigma^n(y)$ for some nonzero $n \in \mathbb{Z}$ and x, y are asymptotic, we obtain that x is eventually periodic. Furthermore, as x is uniformly recurrent, we obtain that x is a periodic sequence. Hence the only possibility to have a finite difference set is having $x = y$, which contradicts the non-triviality assumption. $\square$

3. Lower and upper characteristic Sturmian sequences on $\mathbb{Z}$

The purpose of this section is to prove Theorem A, that is, that recurrent indistinguishable asymptotic pairs $x, y \in \{0,1\}^{\mathbb{Z}}$ whose difference set is of size 2 consist of lower and upper characteristic Sturmian sequences and vice versa. The proof of the first implication is based on the description of Sturmian sequences as lower and upper mechanical words, a terminology introduced by Morse and Hedlund [22]. Notice that here the word "mechanical word" is used to refer to a biinfinite sequence in our context. The proof of the reciprocal is based on the description of Sturmian sequences by their factor complexity [12]. Schematically, the proofs in this section are done as follows:

$$\begin{tikzcd}[column sep=small]
 & x,y \text{ is an indistinguishable asymptotic pair} \arrow[dr] & \\
x,y \text{ are mechanical words} \arrow[ur] & & x,y \text{ have complexity } n+1 \arrow[ll]
\end{tikzcd}$$

The fact that recurrent sequences with complexity $n+1$ are mechanical words of irrational slope implies that all the aforementioned are equivalent.

3.1. **Mechanical words.** Given two real numbers α and ρ with $0 \leqslant \alpha < 1$, we define two sequences

$$s_{\alpha,\rho} : \mathbb{Z} \to \{0,1\}, \quad s'_{\alpha,\rho} : \mathbb{Z} \to \{0,1\}$$

by

$$s_{\alpha,\rho}(n) = \lfloor \alpha(n+1) + \rho \rfloor - \lfloor \alpha n + \rho \rfloor,$$
$$s'_{\alpha,\rho}(n) = \lceil \alpha(n+1) + \rho \rceil - \lceil \alpha n + \rho \rceil.$$

The sequence $s_{\alpha,\rho}$ is the **lower mechanical word** and $s'_{\alpha,\rho}$ is the **upper mechanical word** with **slope** α and **intercept** ρ, see Chapter 2 of [20]. It is clear that if $\rho - \rho'$ is an integer, then $s_{\alpha,\rho} = s_{\alpha,\rho'}$ and $s'_{\alpha,\rho} = s'_{\alpha,\rho'}$. Thus we may always assume $0 \leqslant \rho < 1$.

The mechanical words $s_{\alpha,\rho}, s'_{\alpha,\rho}$ are in fact codings of trajectories of irrational circle rotations, namely, consider the isometry $R_\alpha \colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$, where $R_\alpha(\rho) = (\rho + \alpha) \bmod 1$ for every $\rho \in \mathbb{R}/\mathbb{Z}$. Consider the partition $\mathcal{P} = \{I_0, I_1\}$ of $\mathbb{R}/\mathbb{Z}$ given by $I_0 = [0, 1-\alpha)$ and $I_1 = [1-\alpha, 1)$. For $\rho \in \mathbb{R}/\mathbb{Z}$, define

$$\nu(\rho) = i \quad \text{ if } \rho \in I_i \quad \text{ for } i \in \{0,1\}.$$

We obtain

$$s_{\alpha,\rho}(n) = \nu(R^n_\alpha(\rho)) \quad \text{ for every } n \in \mathbb{Z},$$

i.e., $s_{\alpha,\rho}$ is the coding of the trajectory of ρ with respect to the partition $\mathcal{P}$, see Section 2.2.2 of [20]. Similarly, $s_{\alpha,\rho}$ is the coding of the trajectory of ρ with respect to the partition $\mathcal{P}' = \{I'_0, I'_1\}$ of $\mathbb{R}/\mathbb{Z}$ given by $I'_0 = (0, 1-\alpha]$ and $I'_1 = (1-\alpha, 1]$.

Since $1 + \lfloor \alpha n + \rho \rfloor = \lceil \alpha n + \rho \rceil$ whenever $\alpha n + \rho$ is not an integer, one has $s_{\alpha,\rho} = s'_{\alpha,\rho}$ except when $\alpha n + \rho$ is an integer for some $n \in \mathbb{Z}$. As $\alpha \notin \mathbb{Q}$, this can happen for at most one $n \in \mathbb{Z}$, in this case,

$$s_{\alpha,\rho}(n-1) = 1, \qquad s'_{\alpha,\rho}(n-1) = 0,$$
$$s_{\alpha,\rho}(n) = 0, \qquad s'_{\alpha,\rho}(n) = 1,$$

and elsewhere

$$s_{\alpha,\rho}(k) = s'_{\alpha,\rho}(k) \qquad \text{whenever} \qquad k \notin \{n-1, n\}.$$

In what follows, we say that the sequence $c_\alpha = s_{\alpha,0}$ is the **lower characteristic Sturmian sequence of slope** α and the sequence $c'_\alpha = s'_{\alpha,0}$ is the **upper characteristic Sturmian sequence of slope** α. Notice that $c_\alpha(n) = c'_\alpha(n)$ if and only if $n \in \mathbb{Z}\backslash\{-1, 0\}$.

Remark 3.1. For one-sided sequences, the characteristic Sturmian sequence of slope α is usually the one having two distinct extensions to the left, see [6], [5] or [1, §9]. Here, we consider biinfinite Sturmian sequences as in [2, §6.2], and we believe it is more natural to define $s_{\alpha,0}$ and $s'_{\alpha,0}$ with intercept $\rho = 0$ as the lower and upper "characteristic" ones.

3.2. **Pairs of characteristic Sturmian sequences are indistinguishable.**

Proposition 3.2. *The lower and upper characteristic Sturmian sequences* (c_α, c'_α) *form a non-trivial indistinguishable asymptotic pair for every irrational* $\alpha \in [0, 1]\backslash\mathbb{Q}$.

Proof. From the above discussion, it follows that c_α and c'_α are asymptotic with the difference set $\{-1, 0\}$. The pair is non-trivial since the difference set is nonempty. Note that in this case,

$$c_\alpha(-1) = \mathtt{1}, \qquad\qquad c'_\alpha(-1) = \mathtt{0},$$
$$c_\alpha(0) = \mathtt{0}, \qquad\qquad c'_\alpha(0) = \mathtt{1}.$$

Let $m \in \mathbb{N}$ and $w \in \{\mathtt{0}, \mathtt{1}\}^m$. We shall show that

$$\sum_{i=0}^{m} \mathbb{1}_{[w]}\sigma^{-i}(c_\alpha) = \sum_{i=0}^{m} \mathbb{1}_{[w]}\sigma^{-i}(c'_\alpha). \tag{1}$$

Note that this sum above has $m + 1$ indexes. As c_α, c'_α are Sturmian of angle α, we have that $\mathcal{L}_m(c_\alpha) = \mathcal{L}_m(c'_\alpha)$ is of size $m + 1$. Together with showing that for each word $w \in \mathcal{L}_m(c_\alpha)$ there exist $i, i' \in [\![0, m]\!]$ such that $\sigma^{-i}(c_\alpha) \in [w]$ and $\sigma^{-i'}(c'_\alpha) \in [w]$, it implies that such i and i' are unique, which implies (1).

Let us consider the refinement $\mathcal{P}^m = \bigvee_{j\in[\![0,m-1]\!]} R_\alpha^{-j}(\mathcal{P})$. That is, the partition obtained by intersecting the semiclosed intervals of each shifted partition $R_\alpha^{-j}(\mathcal{P})$ between themselves. By definition of the coding ν, for each $w \in \mathcal{L}_m(c_\alpha)$ there is $I \in \mathcal{P}^m$ such that for every $x \in I$ we have

$$\nu(x)\nu(R_\alpha(x)) \cdots \nu(R_\alpha^{m-1}(x)) = w.$$

In consequence, there are $m+1$ semiclosed intervals in $\mathcal{P}^m$ representing each word in $\mathcal{L}_m(c_\alpha)$. Note that the set of (closed) endpoints of the semiclosed intervals in $\mathcal{P}^m$ is given by the collection:

$$\{0, -\alpha \bmod 1, -2\alpha \bmod 1, \ldots, -m\alpha \bmod 1\}.$$

As for $i \in [\![0, m]\!]$ we have $\sigma^{-i}(c_\alpha) = s_{\alpha,-i\alpha}$, we obtain that each of these shifts $\sigma^{-i}(c_\alpha)$ begins in one of the above endpoints. This proves that there exists $i \in [\![0, m]\!]$ such that $\sigma^{-i}(c_\alpha) \in [w]$.

The situation for the upper characteristic word c'_α is analogous with the following distinction: all intervals are left-open right-closed, that is, the initial partition is $I'_{\mathtt{0}} = (0, 1 - \alpha]$ and $I'_{\mathtt{1}} = (1 - \alpha, 1]$. The analogous partition $(\mathcal{P}')^m$ has the same set of endpoints as $\mathcal{P}^m$ and thus the same conclusion follows.

Equality (1) implies $\Delta_w(x, y) = 0$ for all $w \in \mathcal{L}_m(c_\alpha)$ and all m. By Proposition 2.4, the lower and upper characteristic words form an indistinguishable pair. $\square$

Remark 3.3. If $\{\alpha n + \rho : n \in \mathbb{Z}\} \cap \mathbb{Z} = \varnothing$, then $s_{\alpha,\rho} = s'_{\alpha,\rho}$ and then $s_{\alpha,\rho}$ and $s'_{\alpha,\rho}$ form a trivial indistinguishable asymptotic pair. Otherwise, if there exists $n \in \mathbb{Z}$ such that $\alpha n + \rho \in \mathbb{Z}$, then

$c_\alpha = \sigma^n(s_{\alpha,\rho})$ and $c'_\alpha = \sigma^n(s'_{\alpha,\rho})$. By Propositions 2.5 and 3.2 it follows that $s_{\alpha,\rho}$ and $s'_{\alpha,\rho}$ form a non-trivial indistinguishable asymptotic pair.

3.3. **Recurrent indistinguishable asymptotic pairs are Sturmian.** The goal of this subsection is to prove the reciprocal about recurrent indistinguishable asymptotic pairs with difference set $F = \{-1, 0\}$. In order to do that, we shall first show that their factor complexity, which counts the number of words of each length in their language, coincides with that of a Sturmian sequence.

The **factor complexity** of a sequence $x \in \{0,1\}^{\mathbb{Z}}$ is the mapping $n \mapsto \#\mathcal{L}_n(x)$. Let us recall that a sequence $x \in \{0,1\}^{\mathbb{Z}}$ is **Sturmian** if its factor complexity $\#\mathcal{L}_n(x) = n+1$ for every $n \in \mathbb{N}$ and it is not eventually periodic [2, Def. 6.2.4, Prop. 6.2.5]. Moreover, a sequence is Sturmian if and only if it is a lower or upper mechanical word for some irrational slope α [22, 12], see also [20, Theorem 2.1.13].

The study of factor complexity is closely related to special factors, a notion which is used in the next proof to provide a lower bound. A word $w \in \mathcal{L}_n(x)$ is called **right special** (**left special** resp.) in x if there exists at least two distinct letters $a, b \in \Sigma$ such that $wa, wb \in \mathcal{L}_{n+1}(x)$ (such that $aw, bw \in \mathcal{L}_{n+1}(x)$ resp.), see [11].

A consequence of Lemma 2.8 is that the factor complexity of indistinguishable asymptotic pairs can be bounded above by the size of the smallest interval that contains their difference set.

Proposition 3.4. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be a non-trivial indistinguishable asymptotic pair whose difference set F is contained in an interval I. We have that for every $n \geqslant 1$*

$$n + 1 \leqslant \#\mathcal{L}_n(x) \leqslant n + \#(I) - 1.$$

Proof. By Lemma 2.8, the occurrences of every $w \in \mathcal{L}_n(x)$ intersect F in x. In other words, for each $w \in \mathcal{L}_n(x)$ there exists $u \in F - [\![0, n-1]\!]$ such that $\sigma^u(x) \in [w]$. Without loss of generality, by shifting x and y we may assume that $F \subseteq [\![0, k-1]\!]$ (hence $I = [\![0, k-1]\!]$), so there exists a surjective function from $[\![-n+1, k-1]\!]$ to $\mathcal{L}_n(x)$. In particular, $\#\mathcal{L}_n(x) \leqslant n + k - 1 = n + \#(I) - 1$.

In order to obtain the lower bound, notice that as $x \neq y$, we have $F \neq \varnothing$ and thus we can define $m = \max\{i \in \mathbb{Z} \mid x_i \neq y_i\}$. For all $n \geqslant 0$, we have that the word

$$w = x_{m+1} \dots x_{m+n} = y_{m+1} \dots y_{m+n}$$

can be left-extended to a word of length $n+1$ in $\mathcal{L}_{n+1}(x)$ in two different ways, namely

$$w' = x_m w, \quad w'' = y_m w.$$

Thus, w is a left special factor in x. Since every factor in $\mathcal{L}_n(x)$ can be extended to the left by one symbol to get a word in $\mathcal{L}_{n+1}(x)$ and for every n there exists a left special factor of length n in x, it implies that $\#\mathcal{L}_{n+1}(x) - \#\mathcal{L}_n(x) \geqslant 1$ for every $n \geqslant 0$. Since $x \neq y$, then $\#\mathcal{L}_1(x) = \#\Sigma \geqslant 2$ and we conclude $\#\mathcal{L}_n(x) \geqslant n + 1$. □

As a consequence, when the difference set of x and y is of size 2, the factor complexity must be $n + 1$.

Corollary 3.5. *If $x, y \in \Sigma^{\mathbb{Z}}$ is a non-trivial indistinguishable asymptotic pair with difference set $F = \{-1, 0\}$, then $\#\mathcal{L}_n(x) = n + 1$.*

Proof. By Proposition 3.4 we deduce that $n + 1 \leqslant \#\mathcal{L}_n(x) = \#\mathcal{L}_n(y) \leqslant n + \#F - 1 = n + 1$ for every $n \in \mathbb{N}$ and thus $\#\mathcal{L}_n(x) = \#\mathcal{L}_n(y) = n + 1$ for every $n \in \mathbb{N}$. □

It is known that for each Sturmian sequence and each nonnegative integer n, some factor of length $2n$ of the sequence contains the $n+1$ factors of length n of the sequence, see for instance [10, Corollary 5.2]. It turns out that the central factors of x and y of length $2n$ provide two such words.

Corollary 3.6. *If $x, y \in \Sigma^{\mathbb{Z}}$ is a non-trivial indistinguishable asymptotic pair with difference set $F = \{-1, 0\}$, then each of the words $x_{-n}x_{-n+1}\cdots x_{n-1}$ and $y_{-n}y_{-n+1}\cdots y_{n-1}$ contain exactly one occurrence of each word in $\mathcal{L}_n(x)$.*

Proof. From Corollary 3.5, $\#\mathcal{L}_n(x) = \#\mathcal{L}_n(y) = n + 1$ for every $n \in \mathbb{N}$. From Lemma 2.8, both $x_{-n}x_{-n+1}\cdots x_{n-1}$ and $y_{-n}y_{-n+1}\cdots y_{n-1}$ contain an occurrence of every factor of $\mathcal{L}_n(x)$. All of the occurrences must be distinct or otherwise $\#\mathcal{L}_n(x) < n + 1$. □

For example, the following two words of length 26 contains the same 14 factors of length 13:

$$\texttt{1010010100101.0010010100101}$$

$$\texttt{1010010100100.1010010100101}$$

It is well-known that one-sided sequences of complexity $n + 1$ are not eventually periodic, see [2, Th. 6.1.8] and [20, Th. 2.1.13]. This is no longer true for biinfinite sequences of complexity $n+1$, e.g., consider ${}^{\infty}\texttt{0.1}^{\infty}$ or ${}^{\infty}\texttt{0.10}^{\infty}$. A way to exclude eventually periodic sequences of complexity $n + 1$ in the biinfinite setting is to consider recurrent sequences. For completeness, we provide a proof of the following result which can be considered as folklore even if not mentioned in [2, §6.2].

Proposition 3.7. *$x \in \{\texttt{0}, \texttt{1}\}^{\mathbb{Z}}$ is Sturmian if and only if x is recurrent and $\#\mathcal{L}_n(x) = n + 1$.*

Proof. Sturmian sequences are recurrent, see [2, Exercise 6.2.10].

Assume now that x is recurrent and $\#\mathcal{L}_n(x) = n+1$. For contradiction, assume that x is eventually periodic. Let $x = vp^{\infty}$ where p is the shortest such word (and v is a one-sided left infinite word). The choice of p implies that the set of factors of length $|p|$ of the word p^{∞} has exactly $|p|$ elements. As $\mathcal{L}_{|p|}(x) = |p|+1$, there is a word $u \in \mathcal{L}(x)$ which does not occur in p^{∞}, and there is a last occurrence of u in x. The last occurrence of u in x is followed by an arbitrarily long factor s with no occurrence of u, and, as x is recurrent, the factor us has infinitely many occurrences in v. Therefore, u has unbounded gaps between its occurrences. In particular, there are at least three distinct complete return words $\{r_1, r_2, r_3\}$ to u in $\mathcal{L}(x)$. More precisely, for each $a \in \{1, 2, 3\}$, r_a contains exactly two occurrences of u, one as a prefix and one as a suffix. Let p_{ab} denote the longest common prefix of r_a and r_b. Up to some permutation of the complete return words, we have

$$|p_{12}| = |p_{13}| < |p_{23}|.$$

The word p_{12} is a right special factor in $\mathcal{L}(x)$, that is, there exist two distinct letters $a, b \in \{\texttt{0}, \texttt{1}\}$ such that $p_{12}a, p_{12}b \in \mathcal{L}(x)$. The suffix s of p_{23} of length $|p_{12}|$ is also a right special factor in $\mathcal{L}(x)$. Moreover, $p_{12} \neq s$ since u is a prefix of p_{12} and s contains no occurrence of u. This implies that $\#\mathcal{L}_{N+1}(x) - \#\mathcal{L}_N(x) \geqslant 2$ for some integer $N \in \mathbb{N}$, a contradiction.

The case with w being eventually periodic to the left, i.e., $w = {}^{\infty}pv$, is analogous. We conclude that x is not eventually periodic, and thus it is Sturmian by definition. □

These complexity bounds are the main tools to provide the characterization of pairs consisting of lower and upper characteristic Sturmian sequences by recurrent indistinguishable asymptotic pairs.

Proof of Theorem A. By Proposition 3.2, the lower characteristic word c_α and the upper characteristic word c'_α form an indistinguishable asymptotic pair for every irrational α with $F = \{-1, 0\}$ as their difference set where $x_{-1}x_0 = \texttt{10}$ and $y_{-1}y_0 = \texttt{01}$ for $x = c_\alpha$ and $y = c'_\alpha$. If $\alpha \in [0, 1]\backslash\mathbb{Q}$, then x and y are recurrent.

Conversely, from Corollary 3.5 we have $\#\mathcal{L}_n(x) = n + 1$. From Proposition 3.7, recurrent sequences on $\mathbb{Z}$ of complexity $n + 1$ are Sturmian sequences. We conclude that x and y are Sturmian sequences with the same language associated to some irrational slope $\alpha \in [0,1]\backslash\mathbb{Q}$. Since $x|_{\mathbb{Z}\backslash F} = y|_{\mathbb{Z}\backslash F}$ with $F = \{-1,0\}$ such that $x_{-1}x_0 = \texttt{10}$ and $y_{-1}y_0 = \texttt{01}$ we conclude that $x = c_\alpha$ and $y = c'_\alpha$ are respectively the lower and upper characteristic Sturmian sequences with slope α. $\square$

Remark 3.8. One might wonder if it is possible to prove Theorem A with weaker assumptions, for instance, by asking just that $\mathcal{L}(x) = \mathcal{L}(y)$ instead of indistinguishability. This particular condition does not suffice, even if we further ask that the sequences are uniformly recurrent. Indeed, let $z \in \{\texttt{0},\texttt{1}\}^{\mathbb{Z}\backslash\{0\}}$ be defined by $z(n) = k \bmod 2$ whenever $k \geqslant 1$ and $n = 2^{k-1} \bmod 2^k$. Notice that z is well defined for every nonzero integer and looks as follows

$$z = \dots \texttt{101110101011101.?101110101011101}\dots$$

Let $x, y \in \{\texttt{0},\texttt{1}\}^{\mathbb{Z}}$ be the asymptotic pair defined by

$$x(n) = \begin{cases} \texttt{0} & \text{if } n = 0 \\ z(n) & \text{otherwise} \end{cases} \quad \text{and} \quad y(n) = \begin{cases} \texttt{1} & \text{if } n = 0 \\ z(n) & \text{otherwise} \end{cases}.$$

The sequences x and y are limits of **Toeplitz sequences** which were defined in [17]. They are uniformly recurrent (see for instance [19, Section 4.6]), have the same language and are not Sturmian. Furthermore, if one wishes to construct an example with difference set $\{-1,0\}$, one can consider the Thue-Morse substitution $\varphi\colon \{\texttt{0},\texttt{1}\} \to \{\texttt{0},\texttt{1}\}^*$ given by $\varphi(\texttt{0}) = \texttt{01}$ and $\varphi(\texttt{1}) = \texttt{10}$ and consider $x' = \sigma(\varphi(x))$ and $y' = \sigma(\varphi(y))$. Then x', y' are uniformly recurrent, form an asymptotic pair with the same language, and have difference set $\{-1,0\}$. A direct inspection of their language shows that they are not Sturmian.

4. Limits of Sturmian sequences toward rational slopes

In this section, we describe limits of Sturmian sequences toward a rational slope from above or from below and we show that they also constitute indistinguishable asymptotic pairs in $\mathbb{Z}$. Such words were already considered for instance in [25] and in [16] (see condition B_4 and Figure 2). We prove Theorem 4.5 about non-recurrent sequences which, together with Theorem A, implies Theorem B.

4.1. **Christoffel words.**

Christoffel words have many equivalent definitions, see [8, 7] and the books [9, 26]. Let $p, q \in \mathbb{Z}$ be coprime integers such that $p/q \in \mathbb{Q}_{\geqslant 0} \cup \{\infty\}$ where the limit cases are written as $0 = 0/1$ and $\infty = 1/0$. The **lower Christoffel word of slope** p/q is the factor of length $p+q$ of the lower mechanical word of slope $\alpha = p/(p+q)$ and intercept $\rho = 0$ starting at index 0:

$$c_\alpha(0)c_\alpha(1)\cdots c_\alpha(p+q-1).$$

Similarly, the **upper Christoffel word of slope** p/q is the factor of length $p+q$ of the upper mechanical word of slope $\alpha = p/(p+q)$ and intercept $\rho = 0$ starting at index 0:

$$c'_\alpha(0)c'_\alpha(1)\cdots c'_\alpha(p+q-1).$$

If $p = 0$ and $q = 1$, then the lower and upper Christoffel word of slope $p/q = 0$ is $\texttt{0}$. If $p = 1$ and $q = 0$, then the lower and upper Christoffel word of slope $p/q = \infty$ is $\texttt{1}$.

4.2. **Limits of Sturmian sequences.**

The lower and upper characteristic Sturmian sequences are related to each other as one is the shifted reversal of the other. More precisely we have the following elementary result based on the symmetry of floor and ceiling functions.

Lemma 4.1. *The lower characteristic Sturmian sequence is the shifted reversal of the upper characteristic Sturmian sequence in the sense that $c_\alpha(n) = c'_\alpha(-n-1)$ for every $n \in \mathbb{Z}$. Moreover*

$$c_\alpha(n) = c_\alpha(-n-1) \qquad \textit{and} \qquad c'_\alpha(n) = c'_\alpha(-n-1)$$

for every $n \in \mathbb{Z}\backslash\{-1, 0\}$.

Proof. For all $x \in \mathbb{R}$, we have $\lfloor x\rfloor = -\lceil -x\rceil$. Thus $c_\alpha(n) = \lfloor\alpha(n+1)\rfloor - \lfloor\alpha n\rfloor = \lceil\alpha(-n)\rceil - \lceil\alpha(-n-1)\rceil = c'_\alpha(-n-1)$. Also, if $n \in \mathbb{Z}\backslash\{-1,0\}$, then $c_\alpha(n) = c'_\alpha(-n-1) = c_\alpha(-n-1)$. The same holds for c'_α. □

Limits of lower or upper characteristic Sturmian sequences toward rational slopes are eventually periodic sequences of complexity $n+1$ which can be expressed in terms of Christoffel words.

Lemma 4.2. *Let $p, q \in \mathbb{Z}_{\geqslant 0}$ be coprime integers. Limits of lower or upper characteristic Sturmian sequences as their slope tends towards $p/(p+q)$ are of one of the following forms depending on the value of p and q. If $p \neq 0$ and $q \neq 0$, then*

$$\begin{aligned}
\lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha &= {}^\infty(1m0)(1m1).(0m1)(0m1)^\infty,\\
\lim_{\alpha\to\frac{p}{p+q}^+} c'_\alpha &= {}^\infty(1m0)(1m0).(1m1)(0m1)^\infty,\\
\lim_{\alpha\to\frac{p}{p+q}^-} c_\alpha &= {}^\infty(0m1)(0m1).(0m0)(1m0)^\infty,\\
\lim_{\alpha\to\frac{p}{p+q}^-} c'_\alpha &= {}^\infty(0m1)(0m0).(1m0)(1m0)^\infty,
\end{aligned}$$

where $0m1$ and $1m0$ are respectively the lower and upper Christoffel word of slope p/q with $m \in \{0,1\}^*$. When $p = 0$ and $q = 1$ and the limit is done from above, then

$$\begin{aligned}
\lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha &= \lim_{\alpha\to 0^+} c_\alpha = {}^\infty 01.00^\infty,\\
\lim_{\alpha\to\frac{p}{p+q}^+} c'_\alpha &= \lim_{\alpha\to 0^+} c'_\alpha = {}^\infty 00.10^\infty.
\end{aligned}$$

When $p = 1$ and $q = 0$ and the limit is done from below, then

$$\begin{aligned}
\lim_{\alpha\to\frac{p}{p+q}^-} c_\alpha &= \lim_{\alpha\to 1^-} c_\alpha = {}^\infty 11.01^\infty,\\
\lim_{\alpha\to\frac{p}{p+q}^-} c'_\alpha &= \lim_{\alpha\to 1^-} c'_\alpha = {}^\infty 10.11^\infty.
\end{aligned}$$

Proof. Recall that $c_\alpha(n) = \lfloor\alpha(n+1)\rfloor - \lfloor\alpha n\rfloor$. Let $p/q \in \mathbb{Q}_{>0}$ where $p, q \in \mathbb{Z}_{>0}$ are coprime integers. We compute the values of $\lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha(n)$ at $n = -1$, $n = 0$ and $n = p+q-1$:

$$\begin{aligned}
\lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha(-1) &= \lim_{\alpha\to\frac{p}{p+q}^+} \lfloor\alpha(-1+1)\rfloor - \lfloor\alpha(-1)\rfloor = 0 - (-1) = 1,\\
\lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha(0) &= \lim_{\alpha\to\frac{p}{p+q}^+} \lfloor\alpha(0+1)\rfloor - \lfloor\alpha(0)\rfloor = 0 - 0 = 0,\\
\lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha(p+q-1) &= \lim_{\alpha\to\frac{p}{p+q}^+} \lfloor\alpha(p+q-1+1)\rfloor - \lfloor\alpha(p+q-1)\rfloor = p - (p-1) = 1.
\end{aligned}$$

For $0 \leqslant n \leqslant p+q-1$, we have $\lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha(1)\cdots c_\alpha(p+q-1) = 0m1 \in \{0,1\}^*$ is the lower Christoffel word of slope p/q. We now prove that $p+q$ is a period of $n \mapsto \lim_{\alpha\to\frac{p}{p+q}^+} c_\alpha(n)$ on the domain $\mathbb{Z}_{\geqslant 0}$.

Let $n \geqslant 0$, we have

$$\begin{aligned}\lim_{\alpha \to \frac{p}{p+q}^+} c_\alpha(n+p+q) &= \lim_{\alpha \to \frac{p}{p+q}^+} \lfloor \alpha(n+p+q+1) \rfloor - \lfloor \alpha(n+p+q) \rfloor \\ &= \lim_{\alpha \to \frac{p}{p+q}^+} \lfloor \alpha(n+1) \rfloor + p - \lfloor \alpha n \rfloor - p \\ &= \lim_{\alpha \to \frac{p}{p+q}^+} \lfloor \alpha(n+1) \rfloor - \lfloor \alpha n \rfloor = \lim_{\alpha \to \frac{p}{p+q}^+} c_\alpha(n).\end{aligned}$$

From Lemma 4.1, we have $c_\alpha(n) = c_\alpha(-n-1)$ for every $n \in \mathbb{Z} \backslash \{-1, 0\}$ and this shows the first equality since m is a palindrome:

$$\lim_{\alpha \to \frac{p}{p+q}^+} c_\alpha = {}^\infty(1m0)(1m1).(0m1)(0m1)^\infty.$$

The other equalities are proved similarly. $\square$

Remark 4.3. In general, the pair $\left({}^\infty(1m0)(1m1).(0m1)(0m1)^\infty, {}^\infty(1m0)(1m0).(1m1)(0m1)^\infty\right)$ is not indistinguishable. For instance, when $m = 0011$ the asymptotic pair

$$\begin{aligned} x &= {}^\infty(100110)(100111).(000111)(000111)^\infty, \\ y &= {}^\infty(100110)(100110).(100111)(000111)^\infty \end{aligned}$$

is not indistinguishable because the pattern 00111 appears in x intersecting the difference set $F = \{-1, 0\}$, but it does not appear in y intersecting the difference set F.

4.3. **Non-recurrent indistinguishable asymptotic pairs.** We first prove that limits of pairs consisting of lower and upper characteristic Sturmian sequences whose slope tends towards a rational number are indistinguishable asymptotic pairs. Then, we show in Theorem 4.5 that non-recurrent indistinguishable asymptotic pairs whose difference set is of size 2 are limits of Sturmian sequences. This result together with Theorem A implies Theorem B.

Proposition 4.4. *Let $p, q \in \mathbb{Z}_{\geqslant 0}$ be coprime integers. The limits of pairs consisting of lower and upper characteristic Sturmian sequences whose slope tends towards $p/(p+q)$ from above or from below*

$$\left(\lim_{\alpha \to \frac{p}{p+q}^+} c_\alpha, \lim_{\alpha \to \frac{p}{p+q}^+} c'_\alpha \right) \qquad \text{and} \qquad \left(\lim_{\alpha \to \frac{p}{p+q}^-} c_\alpha, \lim_{\alpha \to \frac{p}{p+q}^-} c'_\alpha \right).$$

form two indistinguishable asymptotic pairs in $\mathbb{Z}$.

Proof. From Lemma 4.2, we observe that $\lim_{\alpha \to \frac{p}{p+q}^+} c_\alpha$ and $\lim_{\alpha \to \frac{p}{p+q}^+} c'_\alpha$ form an asymptotic pair whose difference set is $\{-1, 0\}$. From Proposition 2.7, the property of being an indistinguishable pair is preserved by the limit. Therefore, it is an indistinguishable asymptotic pair. The same holds for the second pair. $\square$

Theorem 4.5. *Let $x, y \in \{\mathtt{0}, \mathtt{1}\}^\mathbb{Z}$ and assume that x is not recurrent. The pair (x, y) is an indistinguishable asymptotic pair with difference set $F = \{-1, 0\}$ such that $x_{-1}x_0 = \mathtt{10}$ and $y_{-1}y_0 = \mathtt{01}$ if and only if there exist coprime nonnegative integers p, q such that (x, y) is the limit of pairs consisting of lower and upper characteristic Sturmian sequences whose slope tends to $p/(p+q) \in [0, 1] \cap \mathbb{Q}$ either*

- *from above, that is, $(x, y) = \lim_{\alpha \to \frac{p}{p+q}^+}(c_\alpha, c'_\alpha)$ and $x = \sigma^{p+q}(y)$ is a shift of y, or,*
- *from below, that is, $(x, y) = \lim_{\alpha \to \frac{p}{p+q}^-}(c_\alpha, c'_\alpha)$ and $y = \sigma^{p+q}(x)$ is a shift of x.*

Proof. Let $p, q \in \mathbb{Z}_{\geqslant 0}$ be coprime integers. From Proposition 4.4, the limits of pairs consisting of lower and upper characteristic Sturmian sequences whose slope tends to $p/(p+q)$ from above or from below form an indistinguishable asymptotic pair (x, y) with $F = \{-1, 0\}$ as a difference set where $x_{-1}x_0 = \mathtt{10}$ and $y_{-1}y_0 = \mathtt{01}$.

Since x is not recurrent, from Proposition 2.9, we have that x is a shift of y, i.e. $x = \sigma^k(y)$ for some $k \in \mathbb{Z}$. We know that $k \neq 0$ since $x \neq y$. If $k = 1$, then $x = {}^\infty 01.00^\infty$ and $y = {}^\infty 00.10^\infty$. From Lemma 4.2, we conclude that

$$x = \lim_{\alpha \to 0^+} c_\alpha \qquad \text{and} \qquad y = \lim_{\alpha \to 0^+} c'_\alpha$$

are limits of pairs consisting of lower and upper characteristic Sturmian sequences whose slope tends to 0 from above. Similarly, if $k = -1$, then $x = {}^\infty 11.01^\infty$ and $y = {}^\infty 10.11^\infty$. From Lemma 4.2, we conclude that $x = \lim_{\alpha \to 1^-} c_\alpha$ and $y = \lim_{\alpha \to 1^-} c'_\alpha$ are limits of pairs consisting of lower and upper characteristic Sturmian sequences whose slope tends to 1 from below.

Assume now that $k \geqslant 2$. Thus $y_{k-1}y_k = \mathtt{10}$. But $y_{k-1}y_k = x_{k-1}x_k$ so that $x_{k-1}x_k = \mathtt{10}$. Thus $y_{nk-1}y_{nk} = x_{nk-1}x_{nk} = \mathtt{10}$ for all $n > 0$. Moreover $x_{-k-1}x_{-k} = y_{-1}y_0 = \mathtt{01}$ and $x_{nk-1}x_{nk} = y_{nk-1}y_{nk} = \mathtt{01}$ for all $n < 0$. Let $m = x_1 \cdots x_{k-1}$. We have $m = x_{nk+1} \cdots x_{nk-2} = y_{nk+1} \cdots y_{nk-2}$ for every $n \in \mathbb{Z}$. Thus we have

$$\begin{aligned} x &= {}^\infty(1m0)(1m1)\boldsymbol{.}(0m1)(0m1)^\infty, \\ y &= {}^\infty(1m0)(1m0)\boldsymbol{.}(1m1)(0m1)^\infty. \end{aligned}$$

We observe that the factor $1m0$ appears in y intersecting the difference set F. By the hypothesis, it must appear in x intersecting the difference set F. Thus $1m0$ is a factor of $1m1.0m1$, but certainly not as a prefix. Therefore $1m0$ is a factor of $m1.0m1$. We conclude that $1m0$ is a factor of $0m1.0m1 = (0m1)^2$. This implies that $1m0$ and $0m1$ are conjugate. From Pirillo's Theorem, we conclude that $0m1$ is a lower Christoffel word of slope p/q for some coprime integers $p, q \in \mathbb{Z}_{\geqslant 0}$ satisfying $p + q = k$. From Lemma 4.2, we conclude that

$$x = \lim_{\alpha \to \frac{p}{p+q}^+} c_\alpha \qquad \text{and} \qquad y = \lim_{\alpha \to \frac{p}{p+q}^+} c'_\alpha$$

are limits of pairs consisting of lower and upper characteristic Sturmian sequences whose slope tends to $p/(p+q)$ from above.

The proof for $k \leqslant -2$ follows the same line as when $k \geqslant 2$ or can even be deduced from it by considering the reversal of x and y. We obtain that $x = \lim_{\alpha \to \frac{p}{p+q}^-} c_\alpha$ and $y = \lim_{\alpha \to \frac{p}{p+q}^-} c'_\alpha$ are the limits of a sequence of lower and upper characteristic Sturmian sequences respectively whose slope converges to $p/(p+q)$ from below. □

We may now deduce Theorem B.

Proof of Theorem B. We have two cases to consider depending on whether x is recurrent or not. If x is recurrent, then from Theorem A, we have that the pair (x, y) is an indistinguishable asymptotic pair with difference set $\{-1, 0\}$ such that $x_{-1}x_0 = \mathtt{10}$ and $y_{-1}y_0 = \mathtt{01}$ if and only if there exists $\alpha \in [0,1]\backslash\mathbb{Q}$ such that $x = c_\alpha$ and $y = c'_\alpha$ are the lower and upper characteristic Sturmian sequences respectively. In this case, we consider the constant sequence $(\alpha_n)_{n \in \mathbb{N}}$ where $\alpha_n = \alpha$ for every $n \in \mathbb{N}$. We have $x = c_\alpha = \lim_{n \to \infty} c_{\alpha_n}$ and $y = c'_\alpha = \lim_{n \to \infty} c'_{\alpha_n}$.

If x is not recurrent, then from Theorem 4.5 the pair (x, y) is an indistinguishable asymptotic pair with difference set $\{-1, 0\}$ such that $x_{-1}x_0 = \mathtt{10}$ and $y_{-1}y_0 = \mathtt{01}$ if and only if there exist coprime nonnegative integers p and q such that (x, y) is the limit of a sequence of pairs of lower and upper characteristic Sturmian sequences whose slope tends toward the rational slope $p/(p+q) \in [0,1] \cap \mathbb{Q}$ from above or from below. Let $(\alpha_n)_{n \in \mathbb{N}}$ be the sequence defined as $\alpha_n = \frac{p}{p+q} + \frac{1}{\sqrt{2}n}$ if the limit is from above or $\alpha_n = \frac{p}{p+q} - \frac{1}{\sqrt{2}n}$ if the limit is from below. Then $\alpha_n \in [0,1)\backslash\mathbb{Q}$ for all $n \in \mathbb{N}$ and $(x, y) = \lim_{n \to \infty}(c_{\alpha_n}, c'_{\alpha_n})$. □

5. Indistinguishable asymptotic pairs on an arbitrary alphabet

The purpose of this section is to prove Theorem C and hence provide a full characterization of indistinguishable asymptotic pairs in the case where the alphabet and difference set are arbitrary. Theorem C will follow from Propositions 5.3, 5.8 and 5.9.

5.1. **Substitutions preserve indistinguishability.** We shall now show that indistinguishable asymptotic pairs are preserved under substitutions.

Definition 5.1. *Let Σ, Γ be alphabets. A* ***substitution*** *is a map $\varphi\colon \Sigma \to \Gamma^+$.*

The extension of φ to a morphism from $\Sigma^+ \to \Gamma^+$ by concatenation is denoted (by abuse of notation) again φ. Moreover, every substitution induces a continuous map denoted (again, by abuse of notation) $\varphi\colon \Sigma^{\mathbb{Z}} \to \Gamma^{\mathbb{Z}}$ in the following way:

$$\varphi(x) := \dots \varphi(x_{-5})\varphi(x_{-4})\varphi(x_{-3})\varphi(x_{-2})\varphi(x_{-1}).\varphi(x_0)\varphi(x_1)\varphi(x_2)\varphi(x_3)\varphi(x_4)\dots$$

Lemma 5.2. *Let $\varphi\colon \Sigma \to \Gamma^+$ be a substitution and $x, y \in \Sigma^{\mathbb{Z}}$. If (x, y) is an indistinguishable asymptotic pair such that its difference set F is contained in $[\![0, k-1]\!]$, then $(\varphi(x), \varphi(y))$ is an indistinguishable asymptotic pair.*

Proof. From $F \subseteq [\![0, k-1]\!]$ it follows immediately that for all $m < 0$, $\varphi(x)_m = \varphi(y)_m$. Let $a \in \Sigma$. As $\Delta_a(x, y) = 0$, we deduce that a appears the same number of times N_a in both x and y in $[\![0, k-1]\!]$. We deduce that

$$K := |\varphi(x_0)\dots\varphi(x_{k-1})| = \sum_{i=0}^{k-1} |\varphi(x_i)| = \sum_{a\in\Sigma} N_a|\varphi(a)| = \sum_{i=0}^{k-1} |\varphi(y_i)| = |\varphi(y_0)\dots\varphi(y_{k-1})|,$$

and thus $\varphi(x)_m = \varphi(y)_m$ for every $m \geqslant K$. This shows that $\varphi(x)$ and $\varphi(y)$ are asymptotic and their difference set is contained in $D = [\![0, K-1]\!]$.

As $\mathcal{L}(x) = \mathcal{L}(y)$, we conclude that $\mathcal{L}(\varphi(x)) = \mathcal{L}(\varphi(y))$. Fix $w \in \mathcal{L}_n(\varphi(x))$. It suffices to show that

$$\#(\mathsf{occ}_w(\varphi(x)) \cap (D - [\![0, n-1]\!])) = \#(\mathsf{occ}_w(\varphi(y)) \cap (D - [\![0, n-1]\!])). \tag{2}$$

Every $i \in \mathsf{occ}_w(\varphi(x)) \cap (D - [\![0, n-1]\!])$ can be uniquely associated to a word $u_i \in \Sigma^+$ and a nonnegative integer k_i such that $\sigma^{i-k_i}(\varphi(x)) \in [\varphi(u_i)]$, $\varphi(u_i)_{k_i}\dots\varphi(u_i)_{k_i+n-1} = w$ and so that u_i is the shortest such word. As $\Delta_{u_i}(x, y) = 0$, we have that u_i occurs the same number of times in x and y in the support $F - [\![0, |u_i|-1]\!]$. Therefore there is a bijection between $\mathsf{occ}_{u_i}(x) \cap (F - [\![0, |u_i|-1]\!])$ and $\mathsf{occ}_{u_i}(y) \cap (F - [\![0, |u_i|-1]\!])$ which induces a bijection between the set

$$A_{w,u,k}(x) = \{i \in (\mathsf{occ}_w(\varphi(x)) \cap (D - [\![0, n-1]\!])) : i \text{ is associated to the pair } (u, k) \in \Sigma^+ \times \mathbb{N}\},$$

and the set

$$A_{w,u,k}(y) = \{i \in (\mathsf{occ}_w(\varphi(y)) \cap (D - [\![0, n-1]\!])) : i \text{ is associated to the pair } (u, k) \in \Sigma^+ \times \mathbb{N}\}.$$

As $\mathsf{occ}_w(\varphi(x)) \cap (D - [\![0, n-1]\!])$ can be written as the union of the $A_{w,u,k}(x)$ over all pairs (u, k), and the same holds exchanging x by y, we obtain that Equation (2) holds. Thus $\Delta_w(\varphi(x), \varphi(y)) = 0$. By Proposition 2.4, this implies that $\varphi(x), \varphi(y)$ form an indistinguishable asymptotic pair. □

We may now prove part of Theorem C based on Proposition 3.2 and Lemma 5.2.

Proposition 5.3. *Let α be irrational and c_α, c'_α be the lower and upper characteristic words of slope α respectively. For any substitution $\varphi\colon \{\mathtt{0}, \mathtt{1}\} \to \Sigma^+$ the sequences $\varphi(\sigma^1(c_\alpha))$ and $\varphi(\sigma^1(c'_\alpha))$ form an indistinguishable asymptotic pair.*

Proof. By Proposition 3.2, we have that c_α, c'_α form a non-trivial indistinguishable asymptotic pair. By Proposition 2.5, $\sigma^{-1}(c_\alpha)$ and $\sigma^{-1}(c'{}_\alpha)$ are also a non-trivial indistinguishable asymptotic pair with difference set $F = \{0, 1\}$. By Lemma 5.2, we have that $\varphi(\sigma^{-1}(c_\alpha)), \varphi(\sigma^{-1}(c'_\alpha))$ is an indistinguishable asymptotic pair.

Note that if we let $m = |\varphi(\mathtt{0})| + |\varphi(\mathtt{1})|$ then $\sigma^m(\varphi(\sigma^{-1}(c_\alpha))) = \varphi(\sigma^1(c_\alpha))$ and $\sigma^m(\varphi(\sigma^{-1}(c'_\alpha))) = \varphi(\sigma^1(c'_\alpha))$. Then again by Proposition 2.5, we obtain that $\varphi(\sigma^1(c_\alpha))$ and $\varphi(\sigma^1(c'_\alpha))$ form an indistinguishable asymptotic pair. □

Note that in Proposition 5.3 we do not ensure that after applying the substitution the words remain non-trivial. For instance, we may consider a substitution sending all symbols to a fixed symbol and thus trivialize the pair.

5.2. **Derived sequences preserve indistinguishability.** We shall find a sequence of inverse substitutions which will allow us to "desubstitute" the asymptotic pair until we arrive to a Sturmian sequence. The main tool is the notion of derived sequence introduced by Durand [14].

Definition 5.4. *Let $x \in \Sigma^{\mathbb{Z}}$ and $w \in \mathcal{L}(x)$ which appears with bounded gaps in x. Let $\{i_k\}_{k \in \mathbb{Z}}$ be the enumeration of $\mathtt{occ}_w(x)$ which is strictly increasing and such that i_0 is the smallest value of $\mathtt{occ}_w(x)$ such that $i_0 > -|w|$*

The ***derived sequence*** *$D_w(x) \in (\mathtt{RW}_w(x))^{\mathbb{Z}}$ is the sequence given by*

$$(D_w(x))_k = x_{i_k} \dots x_{i_{k+1}-1}.$$

The derived sequence of a uniformly recurrent sequence is also uniformly recurrent. Note that the alphabet of the derived sequence consists of symbols in $\mathtt{RW}_w(x)$ which formally are words in Σ^*. It is possible to recover the original sequence x (up to a shift) by applying the morphism $\varphi : \mathtt{RW}_w(x) \to \Sigma^*$ such that $\varphi(u) = u_0 \dots u_{|u|-1}$.

Lemma 5.5. *Let $x, y \in \Sigma^*$ and assume that $a \in \Sigma$ appears with bounded gaps in x. If (x, y) is an indistinguishable asymptotic pair whose difference set F is contained in $[\![0, k-1]\!]$, then $(D_a(x), D_a(y))$ is an indistinguishable asymptotic pair.*

Moreover, the difference set of $(D_a(x), D_a(y))$ is contained in $[\![0, N_a]\!]$ where

$$N_a = \#(\{i \in [\![0, k-1]\!] : x_i = a\}).$$

Proof. Rewrite the sets $\mathtt{occ}_a(x)$ and $\mathtt{occ}_a(y)$ in increasing order as sequences $\{i_t\}_{t \in \mathbb{Z}}$ and $\{j_t\}_{t \in \mathbb{Z}}$ as in Definition 5.4 respectively. As $F \subseteq [\![0, k-1]\!]$, and x, y are asymptotic, we have that $i_t = j_t$ for every $t < 0$ and so $(D_a(x))_t = (D_a(y))_t$ for every $t < 0$.

As $\Delta_a(x, y) = 0$, a occurs the same number of times N_a in the interval $[\![0, k-1]\!]$ in x and y, Therefore using again that x, y are asymptotic, we have that $i_t = j_t$ for every $t \geqslant N_a + 1$ and thus $(D_a(x))_t = (D_a(y))_t$ for every $t \geqslant N_a + 1$. This shows that $D_a(x)$ and $D_a(y)$ are asymptotic and that their difference set is contained in $[\![0, N_a]\!]$.

Let $\varphi : \mathtt{RW}_a(x) \to \Sigma^*$ be the morphism such that $\varphi(u) = u_0 \dots u_{|u|-1}$. Given a word $w = w_1 \dots w_m \in (\mathtt{RW}_a(x))^*$, let $|\varphi(w)| = \sum_{i=1}^{m} |\varphi(w_i)|$. It follows that

$$\begin{aligned}
\Delta_w(D_a(x), D_a(y)) &= \sum_{\ell=-(|w|-1)}^{N_a} \mathbb{1}_{[w]}(\sigma^\ell(D_a(y))) - \mathbb{1}_{[w]}(\sigma^\ell(D_a(x))) \\
&= \sum_{\ell=-(|\varphi(w)|-1)}^{k-1} \mathbb{1}_{[\varphi(w)]}(\sigma^\ell(y)) - \mathbb{1}_{[\varphi(w)]}(\sigma^\ell(x)) \\
&= \Delta_{\varphi(w)}(x, y) = 0.
\end{aligned}$$

It follows that $D_a(x)$,$D_a(y)$ form an indistinguishable asymptotic pair. $\square$

Remark 5.6. An analogous statement holds if instead of considering $a \in \Sigma$ we take an arbitrary $w \in \mathcal{L}(x)$ and we consider the pair $D_w(x), D_w(y)$. We shall not need this general statement.

5.3. **Proof of Theorem C.** We will first show that, as long as the difference set of an indistinguishable asymptotic pair is contained in an interval of length at least 3, we can use derived sequences to construct a new indistinguishable pair whose difference set is contained in a strictly smaller interval. This will later provide a way to reduce the general case to the case where the difference set is $\{-1, 0\}$.

Lemma 5.7. *Suppose $x, y \in \Sigma^{\mathbb{Z}}$ is a non-trivial indistinguishable asymptotic pair whose difference set is contained in an interval $F = [\![0, k-1]\!]$. If x is recurrent, there is $a \in \Sigma$ such that $D_a(x)$ and $D_a(y)$ form a non-trivial indistinguishable asymptotic pair with a difference set contained in the interval $[\![0, \lfloor \frac{k}{2} \rfloor]\!]$.*

Proof. As x, y are non-trivial, we have that $\#(\Sigma) \geqslant 2$. Let $a \in \Sigma$ be the symbol such that $\mathtt{occ}_a(x) \cap [\![0, k-1]\!]$ is the smallest. By the pigeonhole principle, $N_a := \#(\mathtt{occ}_a(x) \cap [\![0, k-1]\!]) \leqslant \lfloor \frac{k}{2} \rfloor$.

By Lemma 2.12, both x and y are uniformly recurrent and so the sequences $D_a(x)$ and $D_a(y)$ are well defined. By Lemma 5.5 $D_a(x)$ and $D_a(y)$ form an indistinguishable asymptotic pair with difference set contained in $[\![0, N_a]\!] \subseteq [\![0, \lfloor \frac{k}{2} \rfloor]\!]$, which is clearly non-trivial as x, y are non-trivial. $\square$

Proposition 5.8. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be a non-trivial indistinguishable asymptotic pair. If x is recurrent, then there exists $\alpha \in [0, 1)\backslash\mathbb{Q}$, a substitution $\varphi\colon \{\mathtt{0}, \mathtt{1}\} \to \Sigma^+$ and an integer $m \in \mathbb{Z}$ such that*

$$\{x, y\} = \{\sigma^m(\varphi(\sigma(c_\alpha))), \sigma^m(\varphi(\sigma(c'_\alpha)))\}$$

where c_α and c'_α are the lower and upper characteristic Sturmian sequences of slope α respectively.

Proof. Let $[\![\ell, \ell + k - 1]\!]$ be the smallest interval containing the difference set F of (x, y). Since (x, y) is non-trivial and indistinguishable, we have that $k \geqslant 2$. We shall proceed by induction on k. If $k = 2$, by Proposition 3.4 we have that the alphabet has size at most $\#\mathcal{L}_1(x) \leqslant 1 + k - 1 = 2$. Therefore, up to a relabeling of the alphabet and a shift, we have an asymptotic pair of sequences which satisfies the assumptions of Theorem A and therefore $\{\sigma^{\ell+1}(x), \sigma^{\ell+1}(y)\} = \{\varphi(c_\alpha), \varphi(c'_\alpha)\}$ are the lower and upper characteristic Sturmian sequences for some $\alpha \in [0, 1)\backslash\mathbb{Q}$ up to some function $\varphi\colon \{\mathtt{0}, \mathtt{1}\} \to \Sigma$. We conclude that $\{x, y\} = \{\sigma^{-\ell-2}(\varphi(\sigma(c_\alpha))), \sigma^{-\ell-2}(\varphi(\sigma(c'_\alpha)))\}$.

Now suppose $k \geqslant 3$ and the result holds for all $2 \leqslant j < k$. By Proposition 2.5, $x^* = \sigma^\ell(x)$ and $y^* = \sigma^\ell(y)$ are an indistinguishable asymptotic pair whose difference set is contained in $[\![0, k-1]\!]$. By Lemma 5.7 there is $a \in \Sigma$ such that $x' := D_a(x^*)$ and $y' := D_a(y^*)$ are a non-trivial indistinguishable pair in $(\mathtt{RW}_a(x))^{\mathbb{Z}}$ and their difference set is contained in the interval $[\![0, \lfloor \frac{k}{2} \rfloor]\!]$. As $k \geqslant 3$, we have that $\lfloor \frac{k}{2} \rfloor < k - 1$ and thus the result holds for x', y'. It follows that there is a substitution $\varphi'\colon \{\mathtt{0}, \mathtt{1}\} \to (\mathtt{RW}_a(x))^+$ and $m' \in \mathbb{Z}$ so that

$$\{x', y'\} = \{\sigma^{m'}(\varphi'(\sigma(c_\alpha))), \sigma^{m'}(\varphi'(\sigma(c'_\alpha)))\}.$$

Let $\phi\colon \mathtt{RW}_a(x) \to \Sigma^+$ and $s \in \mathbb{Z}$ be respectively the substitution and integer such that $\sigma^s(\phi(x')) = x^*$ and $\sigma^s(\phi(y')) = y^*$. Let $\varphi := \phi \circ \varphi'$. Note that the difference set of x', y' is contained in $[\![0, \lfloor \frac{k}{2} \rfloor]\!]$ and the difference set of $\sigma(\varphi'(c_\alpha)), \sigma(\varphi'(c'_\alpha))$ is contained in $[\![-(|\varphi'(\mathtt{0})| + |\varphi'(\mathtt{1})|), -1]\!]$. Let us first show that there is $N \in \mathbb{N}$ so that $\phi(\sigma^{m'}(\varphi'(\sigma(c_\alpha)))) = \sigma^{-N}(\varphi(\sigma(c_\alpha)))$ and $\phi(\sigma^{m'}(\varphi'(\sigma(c'_\alpha)))) = \sigma^{-N}(\varphi(\sigma(c'_\alpha)))$.

Let $K \in \mathbb{N}$ be the smallest positive integer such that $[\![-K, -1]\!]$ contains the difference set of $\varphi'(\sigma(c_\alpha)), \varphi'(\sigma(c'_\alpha))$. As the difference set of x', y' is contained in $[\![0, \lfloor \frac{k}{2} \rfloor]\!]$, we obtain that $m' \leqslant -K+1$.

Consider the words

$$w_1 = \varphi'(\sigma(c_\alpha))_{m'} \dots \varphi'(\sigma(c_\alpha))_{-1}$$
$$w_2 = \varphi'(\sigma(c'_\alpha))_{m'} \dots \varphi'(\sigma(c'_\alpha))_{-1}.$$

By construction, every symbol in $\mathtt{RW}_a(x)$ occurs the same number of times in w_1 and w_2. Letting $N = |\phi(w_1)| = |\phi(w_2)|$ we obtain that $\phi(\sigma^{m'}(\varphi'(\sigma(c_\alpha)))) = \sigma^{-N}(\varphi(\sigma(c_\alpha)))$ and $\phi(\sigma^{m'}(\varphi'(\sigma(c'_\alpha)))) = \sigma^{-N}(\varphi(\sigma(c'_\alpha)))$.

Finally, we conclude that

$$\{x, y\} = \{\sigma^{\ell+s}(\phi(x')), \sigma^{\ell+s}(\phi(y'))\} = \{\sigma^{\ell+s-N}(\varphi(\sigma(c_\alpha))), \sigma^{\ell+s-N}(\varphi(\sigma(c'_\alpha)))\}$$

which is what we wanted to prove. $\square$

We deal with the case when x is non-recurrent in the following proposition.

Proposition 5.9. *Let $x, y \in \Sigma^{\mathbb{Z}}$ be a non-trivial indistinguishable asymptotic pair. If x is not recurrent, then there exists a substitution $\varphi\colon \{\mathtt{0}, \mathtt{1}\} \to \Sigma^+$ and an integer $m \in \mathbb{Z}$ such that*

$$\{x, y\} = \{\sigma^m \varphi({}^\infty\mathtt{0}.\mathtt{1}\mathtt{0}^\infty), \sigma^m \varphi({}^\infty\mathtt{0}.\mathtt{0}\mathtt{1}\mathtt{0}^\infty)\}.$$

Proof. By Proposition 2.9, x and y lie on the same orbit, i.e., there exists $s \in \mathbb{Z}\backslash\{0\}$ with $x = \sigma^s(y)$. Possibly exchanging x and y, we may assume $s > 0$. Let $r = \min\{i \in \mathbb{Z} : x_i \neq y_i\}$, then the difference set of $\sigma^r(x), \sigma^r(y)$ is contained in the interval $[\![0, k-1]\!]$ for some $k > s$. Let us denote $x' = \sigma^r(x)$ and $y' = \sigma^r(y)$.

A word u that cannot be written as a repeated concatenation of another word is called **primitive**. Let u be a primitive word such that $u^n = x'_k \dots x'_{k+s-1}$ for some positive integer n. Since $x'|_{[\![k,\infty]\!]} = y'|_{[\![k,\infty]\!]} = \sigma^s(x')|_{[\![k,\infty]\!]}$, we obtain that x' and y' are eventually periodic to the right, more precisely

$$x'|_{[\![k-s,\infty]\!]} = u^\infty \quad \text{and} \quad y'|_{[\![k,\infty]\!]} = u^\infty$$

Similarly, let w be a primitive word such that $w^{n'} = x'_{-s} \dots x'_{-1}$ for some positive integer n'. Since we have $y'|_{[\![-\infty,-1]\!]} = x'|_{[\![-\infty,-1]\!]} = \sigma^{-s}(y')|_{[\![-\infty,-1]\!]}$ we obtain that x' and y' are eventually periodic to the left, more precisely

$$x'|_{[\![-\infty,-1]\!]} = {}^\infty w \quad \text{and} \quad y'|_{[\![-\infty,s-1]\!]} = {}^\infty w$$

As we took $k > s$, letting $v = x'_0 \dots x'_{k-s-1}$ we can write

$$x' = {}^\infty w.vuu^\infty, \quad \text{and} \quad y' = {}^\infty w.wvu^\infty.$$

We claim that w and u are conjugate. Indeed, as x' and y' are indistinguishable, we have that

$$\#(\mathsf{occ}_w(x') \cap [\![-|w|+1, k-1]\!]) = \#(\mathsf{occ}_w(y') \cap [\![-|w|+1, k-1]\!]).$$

First, note that there cannot be any occurrence of w in y' on the indexes $[\![-|w|+1, -1]\!]$, otherwise we would have that w occurs as a factor of ww neither as a prefix nor suffix, which would contradict the primitivity of w. Also note that every occurrence of w in y' on an index $j \in [\![1, k-1]\!]$ can be mapped uniquely to an occurrence of w in x' on index $j - |w|$. Therefore, the only remaining occurrence of w in y' at index 0 must necessarily be mapped to an occurrence of w in x' as a factor of uu. Therefore w is a factor of uu. Since u and w have the same length, it implies that they are conjugate, that is, $u = pz$ and $w = zp$ for some words z, p. Letting $t = vp$ we can rewrite x' and y' in the following way

$$x' = {}^\infty w.tw^\infty, \quad \text{and} \quad y' = {}^\infty w.wtw^\infty.$$

Setting $\varphi : \mathtt{0} \mapsto w, \mathtt{1} \mapsto t$ we get $x' = \varphi({}^{\infty}\mathtt{0.10}^{\infty})$ and $y' = \varphi({}^{\infty}\mathtt{0.010}^{\infty})$. Letting $m = -r$ we obtain

$$x = \sigma^m(\varphi({}^{\infty}\mathtt{0.10}^{\infty})) \quad \text{and} \quad y = \sigma^m({}^{\infty}\mathtt{0.010}^{\infty}).$$

Which is what we wanted, modulo exchanging x and y. $\square$

Proof of Theorem C. Let $\varphi\colon \{\mathtt{0},\mathtt{1}\} \to \Sigma^+$ be a substitution. From Proposition 5.3, if α is irrational, then the sequences $\varphi(\sigma(c_\alpha))$ and $\varphi(\sigma(c'_\alpha))$ form an indistinguishable asymptotic pair and thus by Proposition 2.5, the asymptotic pair $\sigma^m\varphi(\sigma(c_\alpha)), \sigma^m\varphi(\sigma(c'_\alpha))$ is indistinguishable. Similarly, as ${}^{\infty}\mathtt{0.10}^{\infty}, {}^{\infty}\mathtt{0.010}^{\infty}$ is indistinguishable, we have that for every integer $m \in \mathbb{Z}$, the pair $\sigma^m\varphi({}^{\infty}\mathtt{0.10}^{\infty}), \sigma^m\varphi({}^{\infty}\mathtt{0.010}^{\infty})$ is indistinguishable.

Conversely, if x is recurrent, the result is proved in Proposition 5.8. If x is non-recurrent, the result is proved in Proposition 5.9. $\square$

S. Barbieri, DMCC, Universidad de Santiago de Chile, Las Sophoras 173. Estación Central. Santiago. Chile.

E-mail address: `sebastian.barbieri@usach.cl`

S. Labbé, LaBRI, Université de Bordeaux, 351, cours de la Libération, F-33405, Talence, France.

E-mail address: `sebastien.labbe@labri.fr`

Š. Starosta, Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 160 00 Praha 6, Czech Republic

E-mail address: `stepan.starosta@fit.cvut.cz`

References

[1] J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[2] P. Arnoux. Sturmian sequences. In *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Math.*, pages 143–198. Springer, Berlin, 2002.

[3] M. Baake and U. Grimm. *Aperiodic order. Volume 1. A mathematical invitation.*, volume 149. Cambridge: Cambridge University Press, 2013.

[4] S. Barbieri, R. Gómez, B. Marcus, T. Meyerovitch, and S. Taati. Gibbsian representations of continuous specifications: the theorems of Kozlov and Sullivan revisited. *arXiv:2001.03880*, 2020.

[5] J. Berstel. Recent results in Sturmian words. In *Developments in language theory, II (Magdeburg, 1995)*, pages 13–24. World Sci. Publ., River Edge, NJ, 1996.

[6] J. Berstel. Recent results on extensions of Sturmian words. volume 12, pages 371–385. 2002. International Conference on Geometric and Combinatorial Methods in Group Theory and Semigroup Theory (Lincoln, NE, 2000).

[7] J. Berstel. Sturmian and episturmian words (a survey of some recent results). In *Algebraic informatics*, volume 4728 of *Lecture Notes in Comput. Sci.*, pages 23–47. Springer, Berlin, 2007.

[8] J. Berstel and A. de Luca. Sturmian words, Lyndon words and trees. *Theoret. Comput. Sci.*, 178(1-2):171–203, 1997.

[9] J. Berstel, A. Lauve, C. Reutenauer, and F. V. Saliola. *Combinatorics on words*, volume 27 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 2009. Christoffel words and repetitions in words.

[10] J.-P. Borel and C. Reutenauer. On Christoffel classes. *Theor. Inform. Appl.*, 40(1):15–27, 2006.

[11] J. Cassaigne and F. Nicolas. Factor complexity. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 163–247. Cambridge Univ. Press, Cambridge, 2010.

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

[13] A. de Luca and F. Mignosi. Some combinatorial properties of Sturmian words. *Theoret. Comput. Sci.*, 136(2):361–385, 1994.

[14] F. Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3):89–101, 1998.

[15] A. Glen, J. Justin, S. Widmer, and L. Q. Zamboni. Palindromic richness. *European J. Combin.*, 30(2):510–531, 2009.

[16] A. Glen, A. Lauve, and F. V. Saliola. A note on the Markoff condition and central words. *Inform. Process. Lett.*, 105(6):241–244, 2008.

[17] K. Jacobs and M. Keane. 0-1-sequences of toeplitz type. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 13:123–131, 1969.

[18] J. Justin and L. Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000.

[19] P. Kůrka. *Topological and Symbolic Dynamics.* Societé Mathématique de France, 2003.

[20] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 2002.

[21] A. Markoff. Sur les formes quadratiques binaires indéfinies. *Math. Ann.*, 17(3):379–399, 1880.

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

[23] G. Pirillo. A curious characteristic property of standard Sturmian words. In *Algebraic combinatorics and computer science*, pages 541–546. Springer Italia, Milan, 2001.

[24] I. F. Putnam. *Cantor minimal systems*, volume 70 of *University Lecture Series.* American Mathematical Society, 2018.

[25] C. Reutenauer. On Markoff's property and Sturmian words. *Math. Ann.*, 336(1):1–12, 2006.

[26] C. Reutenauer. *From Christoffel words to Markoff numbers.* Oxford University Press, Oxford, 2019.

[27] L. Vuillon. A characterization of Sturmian words by return words. *European J. Combin.*, 22(2):263–275, 2001.