A characterization of Sturmian sequences by indistinguishable asymptotic pairs

Sebastián Barbieri*, Sébastien Labbé† and Štěpán Starosta‡

* Universidad de Santiago de Chile
† CNRS, LaBRI, Université de Bordeaux
‡ Czech Technical University in Prague

One World Combinatorics on Words Seminar
December 14th 2020
Outline

- Sturmian words
  *Mechanical words, Christoffel words, Pirillo’s theorem*

- Terminology
  *Symbolic dynamics, Asymptotic pairs, Pattern discrepancy, Indistinguishable asymptotic pairs*

- Results
  *Theorem A, Theorem B, Theorem C*
Mechanical words (Morse, Hedlund, 1940)

Let $\alpha \in [0, 1]$ and $c_\alpha, c'_\alpha : \mathbb{Z} \rightarrow \{0, 1\}$ be the configurations

$$c'_\alpha(n) = \lceil \alpha(n + 1) \rceil - \lceil \alpha n \rceil \quad \text{(upper mechanical word)}$$

$$c_\alpha(n) = \lfloor \alpha(n + 1) \rfloor - \lfloor \alpha n \rfloor \quad \text{(lower mechanical word)}$$
Sturmian words (Morse, Hedlund, 1940)

If \( \alpha \in [0, 1] \setminus \mathbb{Q} \), then the mechanical words are not periodic:

\[
\begin{align*}
c'_\alpha(n) &= \lceil \alpha(n + 1) \rceil - \lceil \alpha n \rceil \quad \text{(upper characteristic Sturmian word)} \\
c_\alpha(n) &= \lfloor \alpha(n + 1) \rfloor - \lfloor \alpha n \rfloor \quad \text{(lower characteristic Sturmian word)}
\end{align*}
\]
Christoffel words

If $\alpha \in [0, 1] \cap \mathbb{Q}$, then the mechanical words are periodic:

$c'_\alpha(n) = \infty w'^\infty$ where $w'$ is the upper Christoffel word of slope $p/q$,

$c_\alpha(n) = \infty w^\infty$ where $w$ is the lower Christoffel word of slope $p/q$,

where $\alpha = p/(p + q)$ with $p, q \in \mathbb{Z}_{\geq 0}$ coprime integers.

Moreover $w \leq_{\text{lex}} p \leq_{\text{lex}} w'$ for all primitive period $p$ of $c_\alpha$ and $c'_\alpha$. 
Books

- Chapter 2 of Lothaire’s book (2002), by Berstel and Séébold
- Chapter 6 of Pytheas Fogg’s book (2002), by Arnoux
- Christophe Reutenauer’s book (2019)
Pirillo’s theorem (2001)

Let $w = 0m1$ and $w' = 1m0$ for some $m \in \{0, 1\}^*$.

Theorem

The word $w$ is a lower Christoffel word iff $w$ and $w'$ are conjugate.

A $d$-dimensional extension of Christoffel words

Pirillo’s theorem (restated for $\infty w^\infty$)

Let $w = 0m1$ and $w' = 1m0$ for some $m \in \{0, 1\}^*$.

**Theorem**

$\infty w^\infty = c_\alpha$ is a lower mechanical word of slope $\alpha = \frac{p}{p + q}$ iff $\infty w^\infty$ is a shift of $\infty w'^\infty$.

**Question**:

Let $\alpha \in [0, 1] \setminus \mathbb{Q}$.

Does $\lim_{p \to \infty \atop p + q} \frac{p}{p + q} \to \alpha$ (Pirillo’s theorem) exist?
Symbolic dynamics

We consider
- a finite set $\Sigma$: the alphabet,
- the space of configurations $\Sigma^\mathbb{Z} = \{ x : \mathbb{Z} \to \Sigma \}$,
- $\Sigma^\mathbb{Z}$ endowed with the prodiscrete topology,

$$x = \cdots 0010111011100.01100100000110110001 \cdots$$
$$y = \cdots 10010111011100.01100100000110110001 \cdots$$
- to a pattern $p : S \to \Sigma$ with finite support $S \subset \mathbb{Z}$, a cylinder

$$[p] = \left\{ x \in \Sigma^\mathbb{Z} : x|_S = p \right\}.$$  
- the shift action $\mathbb{Z} \acts \Sigma^\mathbb{Z}$.

$$\sigma^{-1}(x) = \cdots 10001011101110.0011100100000110110000 \cdots$$
$$x = \cdots 00010111011100.01100100000110110001 \cdots$$
$$\sigma(x) = \cdots 00101110111000.110010000011011000010 \cdots$$
$$\sigma^2(x) = \cdots 01011101110001.100100000110110000100 \cdots$$
Asymptotic pairs

Let $x, y \in \Sigma^\mathbb{Z}$ be two configurations $x, y \in \Sigma^\mathbb{Z}$, e.g.,

\begin{align*}
x &= \cdots 001011101100.0000101100000011\cdots \\
y &= \cdots 0010111011100.01110010000011\cdots
\end{align*}

unequal at positions $F = \{-4\} \cup \{1, 2, 3, 4\} \cup \{7, 8\}$.

**Definition**

$x, y \in \Sigma^\mathbb{Z}$ are **asymptotic** if $x$ and $y$ differ in finitely many sites of $\mathbb{Z}$.

The set $F = \{n \in \mathbb{Z} : x_n \neq y_n\}$ is called the **difference set** of $(x, y)$. 
Pattern discrepancy

- Two asymptotic configurations \( x, y \in \Sigma^\mathbb{Z} \) with difference set \( F \).
- A pattern \( p : S \rightarrow \Sigma \) with finite support \( S \subseteq \mathbb{Z} \).

**Goal**: compare the # of occurrences of \( p \) in \( x \) and \( y \) : \( |y|_p - |x|_p \).

**Example**: pattern \( p = \cdot.1001 \) with support \( S = \{0, 1, 2, 3\} \)

\[
x = \cdots 1010010100101.010010100101 \cdots
\]

\[
y = \cdots 1010010100100.1010010100101 \cdots
\]

with difference set \( F = \{-1, 0\} \).

**Definition**

The **\( p \)-discrepancy** associated to \((x, y)\) is given by

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

**Note**: \( n \in \mathbb{Z} \setminus (F - S) \) if and only if \((n + S) \cap F = \emptyset\).
Indistinguishable asymptotic pairs

Let \( x, y \in \Sigma^\mathbb{Z} \) be asymptotic configurations.

**Definition**

\( x, y \) are **indistinguishable** if \( \Delta_p(x, y) = 0 \) for every finite pattern \( p \).

Example 1: The **trivial** asymptotic pair \((x, x)\) is indistinguishable.

Example 2:

\[
\begin{align*}
x &= \cdots 00000000000000.100000000000000000\cdots \\
y &= \cdots 00000000000000.000000100000000000\cdots
\end{align*}
\]

In both of these examples, \( x \) and \( y \) lie on the **same orbit** of \( \mathbb{Z} \acts \Sigma^\mathbb{Z} \).

**Question**: Can we find other examples?
Indistinguishable asymptotic pairs

Let \( x, y \in \Sigma \mathbb{Z} \) be asymptotic configurations.

**Definition**

\( x, y \) are **indistinguishable** if \( \Delta_p(x, y) = 0 \) for every finite pattern \( p \).

Non-Example 3, because \( \Delta_1(x, y) = -7 \):

\[
\begin{align*}
x &= \cdots 00000000000000.\overline{1111111111111111} \cdots \\
y &= \cdots 00000000000000.\overline{000000011111111111} \cdots 
\end{align*}
\]

Example 4, with \( \Delta_{abcabc}(x, y) = 1 - 1 = 0 \):

\[
\begin{align*}
x &= \cdots bcabc\underline{bcabc}\abc\bcabc\cdots \\
y &= \cdots bcabc\underline{bcabc}\bcabc\bcabc\cdots 
\end{align*}
\]
Let $x, y \in \{0, 1\}^\mathbb{Z}$ and assume that $x$ is \textit{recurrent}. The pair $(x, y)$ is an \textit{indistinguishable asymptotic pair} with difference set $F = \{-1, 0\}$ such that $x_{-1}x_0 = 10$ and $y_{-1}y_0 = 01$

\textit{if and only if}

there exists $\alpha \in [0, 1] \setminus \mathbb{Q}$ such that $x = c_\alpha$ and $y = c'_\alpha$ are the lower and upper \textit{characteristic Sturmian words} of slope $\alpha$. 

\textbf{Theorem A}
Idea of Proof of Theorem A

(Recall that $x$ is recurrent)

$x, y$ indistinguishable pair

$x, y$ mechanical words

$x, y$ have complexity $n + 1$

Coven-Hedlund, 73'

$\begin{array}{c}
-1\alpha \\
-2\alpha \\
1001 \\
1010 \\
0101 \\
0100 \\
-3\alpha \\
-4\alpha \\
-0\alpha \\
\end{array}$

$x = \cdots 101001010010 \boxed{1.0} 010010100101 \cdots$

$y = \cdots 101001010010 \boxed{0.1} 010010100101 \cdots$

Proposition

Let $x, y \in \Sigma^\mathbb{Z}$ be a non-trivial indistinguishable asymptotic pair whose difference set $F$ is contained in an interval $I$. For every $n \geq 1$

\[ n + 1 \leq \#\mathcal{L}_n(x) \leq n + \#I - 1. \]
Example using Christoffel words

Let $0m1$ be a lower Christoffel word of slope $p/q$ with $p + q = n$. The 2 words of length $2n$:

$$1m1.0m1$$

$$1m0.1m1$$

both contain $n + 1$ factors of size $n$ (one occurrence of each):

```python
sage: u = Word('10100101001010010010100101')
sage: v = Word('10100101001010010010100101')
sage: v.factor_set(13) == u.factor_set(13)
True
sage: len(u.factor_set(13))
14
```

But their binomial coefficients are not equal:

```python
sage: u.number_of_subword_occurrences(Word('01'))
82
sage: v.number_of_subword_occurrences(Word('01'))
83
```
Let \( x, y \in \{0, 1\}^\mathbb{Z} \).

The pair \((x, y)\) is an \textit{indistinguishable asymptotic pair} with difference set \( F = \{-1, 0\} \) such that \( x_{-1}x_0 = 10 \) and \( y_{-1}y_0 = 01 \)

\[
\text{if and only if} \quad \text{there exists a monotone sequence } (\alpha_n)_{n \in \mathbb{N}} \text{ with } \alpha_n \in [0, 1] \setminus \mathbb{Q} \text{ s.t.}
\]
\[
x = \lim_{n \to \infty} c_{\alpha_n} \quad \text{and} \quad y = \lim_{n \to \infty} c'_{\alpha_n}.
\]

are the limits of \textit{characteristic Sturmian words} of slope \( \alpha_n \).

If \( \alpha = \lim_{n \to \infty} \alpha_n \in [0, 1] \setminus \mathbb{Q} \), then
\[
x = \lim_{n \to \infty} c_{\alpha_n} = c_{\alpha} \quad \text{and} \quad y = \lim_{n \to \infty} c'_{\alpha_n} = c'_{\alpha}
\]

and it corresponds to Theorem A.
Theorem B: limit towards a rational slope

Assume \( \lim_{n \to \infty} \alpha_n = \frac{p}{p+q} \in [0, 1] \cap \mathbb{Q} \), with \( p, q \in \mathbb{Z}_{\geq 0} \) coprime.

If \( p \neq 0 \) and \( q \neq 0 \) and the limit is from above, then
\[
\lim_{\alpha \to \frac{p}{p+q}^+} c_\alpha = \infty \left(1m0 \right) \left(1m1 \right) \left(0m1 \right) \left(0m1 \right) \infty
\]
\[
\lim_{\alpha \to \frac{p}{p+q}^+} c'_\alpha = \infty \left(1m0 \right) \left(1m0 \right) \left(1m1 \right) \left(0m1 \right) \infty
\]

or the limit is from below, then
\[
\lim_{\alpha \to \frac{p}{p+q}^-} c_\alpha = \infty \left(0m1 \right) \left(0m1 \right) \left(0m0 \right) \left(1m0 \right) \infty
\]
\[
\lim_{\alpha \to \frac{p}{p+q}^-} c'_\alpha = \infty \left(0m1 \right) \left(0m0 \right) \left(1m0 \right) \left(1m0 \right) \infty
\]

Limit cases: when \( p = 0 \) and \( q = 1 \) or \( p = 1 \) and \( q = 0 \), then
\[
\lim_{\alpha \to 0^+} c_\alpha = \infty 01.00 \infty
\]
\[
\lim_{\alpha \to 0^+} c'_\alpha = \infty 00.10 \infty
\]
\[
\lim_{\alpha \to 1^-} c_\alpha = \infty 11.01 \infty
\]
\[
\lim_{\alpha \to 1^-} c'_\alpha = \infty 10.11 \infty
\]
Idea of Proof of Theorem B ($\implies$)

Let $x, y \in \{0, 1\}^\mathbb{Z}$ and assume $x$ is not recurrent. If the pair $(x, y)$ is an indistinguishable asymptotic pair with difference set $F = \{-1, 0\}$ such that $x_{-1}x_0 = 10$ and $y_{-1}y_0 = 01$, then

$$x = \sigma^k(y)$$

for some $k \in \mathbb{Z}$. If $k \geq 2$, then there exists $m \in \{0, 1\}^{k-2}$ s.t.

$$x = \infty (1m0)(1m1).(0m1)(0m1)\infty$$
$$y = \infty (1m0)(1m0).(1m1)(0m1)\infty$$

- $1m0$ appears in $y$ intersecting the difference set $F$
- 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 $m1.0m1$ and $0m1.0m1 = (0m1)^2$
- This implies that $1m0$ and $0m1$ are conjugate
- Pirillo's Theorem $\implies$ $0m1$ is a lower Christoffel word of slope $p/q$ for some coprime integers $p, q \in \mathbb{Z}_{\geq 0}$ satisfying $p + q = k$. 
Theorem C

Let $\Sigma$ 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 $\alpha \in [0, 1] \setminus \mathbb{Q}$, a substitution $\varphi : \{0, 1\} \rightarrow \Sigma^+$ and an integer $m \in \mathbb{Z}$ such that
  \[
  \{x, y\} = \{\sigma^m \varphi(\sigma^1(c_\alpha)), \sigma^m \varphi(\sigma^1(c'_\alpha))\},
  \]

- $x$ is not recurrent and there exists a substitution $\varphi : \{0, 1\} \rightarrow \Sigma^+$ and an integer $m \in \mathbb{Z}$ such that
  \[
  \{x, y\} = \{\sigma^m \varphi(\infty 0.10^\infty), \sigma^m \varphi(\infty 0.010^\infty)\}.\]
Idea of Proof of Theorem C \( \iff \)

Let \( x, y \in \Sigma^\mathbb{Z} \) be an asymptotic pair such that their difference set \( F \) is contained in \([0, k - 1]\).

**Lemma**

Let \( \varphi: \Sigma \to \Gamma^+ \) be a substitution on \( \Sigma^\mathbb{Z} \).

If \((x, y)\) is indistinguishable, then \((\varphi(x), \varphi(y))\) is indistinguishable.

\[
\begin{align*}
x &= \cdots 010010.0110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010110010
Idea of Proof of Theorem C ( \( \iff \) )

**Lemma**

Assume \( a \in \Sigma \) appears with bounded gaps in \( x \). Let \( D_a(x) \) be the **derived sequence** of \( x \) wrt return words to \( a \in \Sigma \). If \( (x, y) \) **indistinguishable**, then \( (D_a(x), D_a(y)) \) **indistinguishable**.

**Return words** to letter \( c \) in \( x \) and \( y \) are \( cab \) and \( cb \):

\[
\begin{align*}
  x &= \cdots bcabcbcabcabcabcabc. \boxed{abc} bcabcbcabcabcabcabc \cdots \\
  y &= \cdots bcabcbcabcabcabcabcabc. \boxed{bca} bcabcbcabcabcabcabcabc \cdots
\end{align*}
\]

Replacing \( cab \mapsto 0, \ cb \mapsto 1 \), we obtain the **derived sequences**:

\[
\begin{align*}
  D_c(x) &= \cdots 010010. \boxed{01} 0100101 \cdots \\
  D_c(y) &= \cdots 010010. \boxed{10} 0100101 \cdots
\end{align*}
\]

with a **smaller** difference set.
They defined the following **norm** on asymptotic configurations of $\Sigma^\mathbb{Z}$:

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

- Every asymptotic pair induces an evaluation map on the space of **continuous cocycles** on the equiv. relation of asymptotic pairs.
- They 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 null operator are precisely the **indistinguishable pairs**.
- Thus, our results provide a full characterization of which asymptotic pairs induce the **null operator**.
We are currently working to extend Theorem A from $\mathbb{Z}$ to $\mathbb{Z}^d$.

Extending Theorem B and Theorem C to $\mathbb{Z}^d$ seems more difficult.