# Initial nonrepetitive complexity of regular episturmian words and their Diophantine exponents

Jarkko Peltomäki

The Turku Collegium for Science, Medicine and Technology
Department of Mathematics and Statistics
University of Turku

21.6.2021

# Contents

- Intercepts of episturmian words.
- Initial nonrepetitive complexity of episturmian words.
- Diophantine exponents of episturmian words.
- Results on irrationality exponents

# Setup

- We suppose that $\Delta$ is an infinite word, a *directive word*, over a $d$-letter alphabet and write $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ with $a_k \geq 1$ and $x_k \neq x_{k+1}$.

# Setup

- We suppose that $\Delta$ is an infinite word, a *directive word*, over a $d$-letter alphabet and write $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ with $a_k \geq 1$ and $x_k \neq x_{k+1}$.
- The sequence $(a_k)$ is the sequence of *partial quotients*

# Setup

- We suppose that $\Delta$ is an infinite word, a *directive word*, over a $d$-letter alphabet and write $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ with $a_k \geq 1$ and $x_k \neq x_{k+1}$.
- The sequence $(a_k)$ is the sequence of *partial quotients*
- $\Delta$ is *regular* if $x_1 x_2 \cdots$ is of the form $(01 \cdots (d-1))^\omega$.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.
- Set $s_{1-d} = x_2, s_{2-d} = x_3, \ldots, s_{-1} = x_d, s_0 = x_1$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_0^{a_1} x_{k+1}$, $1 \leq k < d$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_{k-(d-1)}^{a_{k-(d-2)}} s_{k-d}$ for $k \geq d$.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.
- Set $s_{1-d} = x_2, s_{2-d} = x_3, \ldots, s_{-1} = x_d, s_0 = x_1$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_0^{a_1} x_{k+1}$, $1 \leq k < d$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_{k-(d-1)}^{a_{k-(d-2)}} s_{k-d}$ for $k \geq d$.
- When $\Delta = (012)^\omega$, then
  - $s_0 = 0$, $s_1 = 01$, $s_2 = 0102$, and $s_k = s_{k-1} s_{k-2} s_{k-3}$ for $k \geq 3$.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.
- Set $s_{1-d} = x_2, s_{2-d} = x_3, \ldots, s_{-1} = x_d, s_0 = x_1$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_0^{a_1} x_{k+1}$, $1 \leq k < d$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_{k-(d-1)}^{a_{k-(d-2)}} s_{k-d}$ for $k \geq d$.
- When $\Delta = (012)^\omega$, then
  - $s_0 = 0$, $s_1 = 01$, $s_2 = 0102$, and $s_k = s_{k-1} s_{k-2} s_{k-3}$ for $k \geq 3$.
- The sequence $(s_k)$ is the sequence of (generalized) *standard words* associated with $\Delta$.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.
- Set $s_{1-d} = x_2, s_{2-d} = x_3, \ldots, s_{-1} = x_d, s_0 = x_1$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_0^{a_1} x_{k+1}$, $1 \leq k < d$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_{k-(d-1)}^{a_{k-(d-2)}} s_{k-d}$ for $k \geq d$.
- When $\Delta = (012)^\omega$, then
  - $s_0 = 0$, $s_1 = 01$, $s_2 = 0102$, and $s_k = s_{k-1} s_{k-2} s_{k-3}$ for $k \geq 3$.
- The sequence $(s_k)$ is the sequence of (generalized) *standard words* associated with $\Delta$.
- The limit $\mathbf{c}_\Delta$ of $(s_k)$ is called the *regular standard episturmian word* with directive word $\Delta$.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.
- Set $s_{1-d} = x_2, s_{2-d} = x_3, \ldots, s_{-1} = x_d, s_0 = x_1$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_0^{a_1} x_{k+1}$, $1 \leq k < d$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_{k-(d-1)}^{a_{k-(d-2)}} s_{k-d}$ for $k \geq d$.
- When $\Delta = (012)^\omega$, then
  - $s_0 = 0$, $s_1 = 01$, $s_2 = 0102$, and $s_k = s_{k-1} s_{k-2} s_{k-3}$ for $k \geq 3$.
- The sequence $(s_k)$ is the sequence of (generalized) *standard words* associated with $\Delta$.
- The limit $\mathbf{c}_\Delta$ of $(s_k)$ is called the *regular standard episturmian word* with directive word $\Delta$.
- An infinite word is a *regular episturmian word* if it has the same language as some regular standard episturmian word.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.
- Set $s_{1-d} = x_2, s_{2-d} = x_3, \ldots, s_{-1} = x_d, s_0 = x_1$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_0^{a_1} x_{k+1}$, $1 \leq k < d$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_{k-(d-1)}^{a_{k-(d-2)}} s_{k-d}$ for $k \geq d$.
- When $\Delta = (012)^\omega$, then
  - $s_0 = 0$, $s_1 = 01$, $s_2 = 0102$, and $s_k = s_{k-1} s_{k-2} s_{k-3}$ for $k \geq 3$.
- The sequence $(s_k)$ is the sequence of (generalized) *standard words* associated with $\Delta$.
- The limit $\mathbf{c}_\Delta$ of $(s_k)$ is called the *regular standard episturmian word* with directive word $\Delta$.
- An infinite word is a *regular episturmian word* if it has the same language as some regular standard episturmian word.
- Combinatorial generalizations of Sturmian words.

# Regular Episturmian Words

- Let $\Delta = x_1^{a_1} x_2^{a_2} \cdots$ be regular.
- Set $s_{1-d} = x_2, s_{2-d} = x_3, \ldots, s_{-1} = x_d, s_0 = x_1$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_0^{a_1} x_{k+1}$, $1 \leq k < d$.
- Define $s_k = s_{k-1}^{a_k} s_{k-2}^{a_{k-1}} \cdots s_{k-(d-1)}^{a_{k-(d-2)}} s_{k-d}$ for $k \geq d$.
- When $\Delta = (012)^\omega$, then
  - $s_0 = 0$, $s_1 = 01$, $s_2 = 0102$, and $s_k = s_{k-1} s_{k-2} s_{k-3}$ for $k \geq 3$.
- The sequence $(s_k)$ is the sequence of (generalized) *standard words* associated with $\Delta$.
- The limit $\mathbf{c}_\Delta$ of $(s_k)$ is called the *regular standard episturmian word* with directive word $\Delta$.
- An infinite word is a *regular episturmian word* if it has the same language as some regular standard episturmian word.
- Combinatorial generalizations of Sturmian words.
- Remark: if $\Delta$ is not regular, then we need to use morphisms.

# Generalized Ostrowski Numeration Systems

- Set $q_k = |s_k|$ and view $(q_k)$ as a numeration system, that is, greedily express each positive integer as a sum according to the sequence $(q_k)$.

# Generalized Ostrowski Numeration Systems

- Set $q_k = |s_k|$ and view $(q_k)$ as a numeration system, that is, greedily express each positive integer as a sum according to the sequence $(q_k)$.
- When $\Delta = (012)^\omega$, then $(q_k) = (1, 2, 4, 7, 13, 21, \ldots)$ (the Tribonacci numbers) and $10 = 7 + 3 = 7 + 2 + 1$, so $\mathrm{rep}(10) = 1101$ (we write least significant digit first).

# Generalized Ostrowski Numeration Systems

- Set $q_k = |s_k|$ and view $(q_k)$ as a numeration system, that is, greedily express each positive integer as a sum according to the sequence $(q_k)$.
- When $\Delta = (012)^\omega$, then $(q_k) = (1, 2, 4, 7, 13, 21, \ldots)$ (the Tribonacci numbers) and $10 = 7 + 3 = 7 + 2 + 1$, so $\mathrm{rep}(10) = 1101$ (we write least significant digit first).
- This is the (generalized) *Ostrowski numeration system* associated with $\Delta$.

# Intercept of an Episturmian Word

**Theorem (Droubay-Justin-Pirillo (2001))**

*Let* $\mathbf{t}$ *be a regular episturmian word with directive word* $\Delta$*. Then there exists a unique word* $c_1 c_2 \cdots$ *such that for all* $k$*, the word* $c_1 \cdots c_k$ *is the Ostrowski expansion of an integer* $\ell_k$ *and*

$$\begin{aligned}\mathbf{t} &= \lim_{k\to\infty} T^{\ell_k}(\mathbf{c}_\Delta) \\ &= \lim_{k\to\infty} T^{\mathrm{val}(c_1\cdots c_k)}(\mathbf{c}_\Delta).\end{aligned}$$

# Intercept of an Episturmian Word

**Theorem (Droubay-Justin-Pirillo (2001))**

*Let* $\mathbf{t}$ *be a regular episturmian word with directive word* $\Delta$*. Then there exists a unique word* $c_1 c_2 \cdots$ *such that for all* $k$*, the word* $c_1 \cdots c_k$ *is the Ostrowski expansion of an integer* $\ell_k$ *and*

$$\begin{aligned}\mathbf{t} &= \lim_{k\to\infty} T^{\ell_k}(\mathbf{c}_\Delta) \\ &= \lim_{k\to\infty} T^{\mathrm{val}(c_1\cdots c_k)}(\mathbf{c}_\Delta).\end{aligned}$$

- The word $c_1 c_2 \cdots$ is called the *intercept* of $\mathbf{t}$.

# Intercept of an Episturmian Word

**Theorem (Droubay-Justin-Pirillo (2001))**

*Let* $\mathbf{t}$ *be a regular episturmian word with directive word* $\Delta$*. Then there exists a unique word* $c_1 c_2 \cdots$ *such that for all* $k$*, the word* $c_1 \cdots c_k$ *is the Ostrowski expansion of an integer* $\ell_k$ *and*

$$\begin{aligned} \mathbf{t} &= \lim_{k \to \infty} T^{\ell_k}(\mathbf{c}_\Delta) \\ &= \lim_{k \to \infty} T^{\mathrm{val}(c_1 \cdots c_k)}(\mathbf{c}_\Delta). \end{aligned}$$

- The word $c_1 c_2 \cdots$ is called the *intercept* of $\mathbf{t}$.
- In the Sturmian case, this coincides with the usual notion of intercept via Ostrowski expansions of real numbers.

# Intercept of an Episturmian Word

**Theorem (Droubay-Justin-Pirillo (2001))**

*Let* $\mathbf{t}$ *be a regular episturmian word with directive word* $\Delta$*. Then there exists a unique word* $c_1 c_2 \cdots$ *such that for all* $k$*, the word* $c_1 \cdots c_k$ *is the Ostrowski expansion of an integer* $\ell_k$ *and*

$$\begin{aligned} \mathbf{t} &= \lim_{k \to \infty} T^{\ell_k}(\mathbf{c}_\Delta) \\ &= \lim_{k \to \infty} T^{\mathrm{val}(c_1 \cdots c_k)}(\mathbf{c}_\Delta). \end{aligned}$$

- The word $c_1 c_2 \cdots$ is called the *intercept* of $\mathbf{t}$.
- In the Sturmian case, this coincides with the usual notion of intercept via Ostrowski expansions of real numbers.
- Important point: it is in principle possible to reduce the study of a property of episturmian words to studying the property on standard episturmian words.

# Initial Nonrepetitive Complexity

## Definition

Let $\mathbf{x}$ be an infinite word. Its *initial nonrepetitive complexity function* $\mathrm{inrc}(\mathbf{x}, n)$ is defined as

$$\mathrm{inrc}(\mathbf{x}, n) = \max\{m : \mathbf{x}[i, i+n-1] \neq \mathbf{x}[j, j+n-1] \text{ for all } i, j \text{ with } 1 \leq i < j \leq m\}.$$

- I.e.: $\mathrm{inrc}(\mathbf{x}, n)$ is the maximum number of factors of length $n$ seen when $\mathbf{x}$ is read from left to right prior to the first repeated factor of length $n$.

# Initial Nonrepetitive Complexity

**Definition**

Let $\mathbf{x}$ be an infinite word. Its *initial nonrepetitive complexity function* $\mathrm{inrc}(\mathbf{x}, n)$ is defined as

$$\mathrm{inrc}(\mathbf{x}, n) = \max\{m : \mathbf{x}[i, i+n-1] \neq \mathbf{x}[j, j+n-1] \text{ for all } i, j \text{ with } 1 \leq i < j \leq m\}.$$

- I.e.: $\mathrm{inrc}(\mathbf{x}, n)$ is the maximum number of factors of length $n$ seen when $\mathbf{x}$ is read from left to right prior to the first repeated factor of length $n$.
- Introduced by Moothatu (2012), studied further by Nicholson and Rampersad (2016) and Medková et al. (2020).

# Initial Nonrepetitive Complexity

- Wojcik (2020) finds a formula for $\mathrm{inrc}(\mathbf{t}, n)$ for an arbitrary Sturmian word $\mathbf{t}$ based on its intercept.
- I generalize this to all **regular** episturmian words, but this formula is too complicated to display here.

# Motivation

- Why find complicated formulas for the initial nonrepetitive complexity?
- Answer: they can be used to determine Diophantine exponents.

# Diophantine Exponents

**Definition**

Let $\mathbf{x}$ be an infinite word. We let its *Diophantine exponent*, denoted by $\mathrm{dio}(\mathbf{x})$, to be the supremum of all real numbers $\rho$ for which there exist arbitrarily long prefixes of $\mathbf{x}$ of the form $UV^e$, where $U$ and $V$ are finite words and $e$ is a real number, such that

$$\frac{|UV^e|}{|UV|} \geq \rho.$$

# Diophantine Exponents

**Proposition (Bugeaud-Kim (2019))**

*If* $\mathbf{x}$ *is an infinite word, then*

$$\operatorname{dio}(\mathbf{x}) = 1 + \limsup_{n\to\infty} \frac{n}{\operatorname{inrc}(\mathbf{x}, n)}.$$

# Diophantine Exponents

**Proposition (Bugeaud-Kim (2019))**

*If* $\mathbf{x}$ *is an infinite word, then*

$$\mathrm{dio}(\mathbf{x}) = 1 + \limsup_{n \to \infty} \frac{n}{\mathrm{inrc}(\mathbf{x}, n)}.$$

- Thus we may in principle compute the Diophantine exponent of a given regular episturmian word.
- In practice, this is doable only when the intercept and $\Delta$ are “nice”.

# Results

**Theorem (P. (2021))**

*Let* $\mathbf{t}$ *be a regular episturmian word. Then* $\mathrm{dio}(\mathbf{t}) < \infty$ *if and only if the sequence* $(a_k)$ *of partial quotients in bounded.*

- Proved for Sturmian words by Adamczewski and Bugeaud (2011).

# Results

**Theorem (P. (2021))**

*Let* $\mathbf{t}$ *be a regular episturmian word of period* $d$. *If* $d = 2$ *or* $\limsup_k a_k \geq 3$, *then* $\mathrm{dio}(\mathbf{t}) > 2$.

- For Sturmian words follows from Adamczewski (2010) and Berthé et al. (2006).

# Results

## Proposition (P. (2021))

*Let* $\mathbf{t}$ *be the episturmian word with directive word* $(001122)^\omega$ *having intercept* $1^\omega$*. Then*

$$\mathrm{dio}(\mathbf{t}) = 1 + \frac{1}{2}(\beta - 1) \approx 1.9156$$

*where* $\beta \approx 2.8312$ *is the real root of the polynomial* $x^3 - 2x^2 - 2x - 1$*.*

## Proposition (P. (2021))

*Let* $\mathbf{t}$ *be the episturmian word with directive word* $(0123)^\omega$ *having intercept* $(001)^\omega$*,* $(010)^\omega$*, or* $(100)^\omega$*. Then*

$$\mathrm{dio}(\mathbf{t}) = 1 + \frac{1}{27}(-7\zeta^3 + 15\zeta^2 + 13\zeta - 4) \approx 1.9873$$

*where* $\zeta \approx 1.9276$ *is the positive real root of the polynomial* $x^4 - x^3 - x^2 - x - 1$*.*

# Irrationality Exponents

- If $\mathbf{x}$ is an infinite word over the alphabet $\{0, 1, \ldots, b-1\}$, $b \geq 2$, let $\xi_{\mathbf{x}}$ be the real number with $\mathbf{x}$ as a fractional part.

# Irrationality Exponents

- If **x** is an infinite word over the alphabet $\{0, 1, \ldots, b-1\}$, $b \geq 2$, let $\xi_{\mathbf{x}}$ be the real number with **x** as a fractional part.

**Problem**

*What can we infer about the arithmetic properties of $\xi_{\mathbf{x}}$ given combinatorial properties of* **x***?*

# Irrationality Exponents

- If $\mathbf{x}$ is an infinite word over the alphabet $\{0, 1, \ldots, b-1\}$, $b \geq 2$, let $\xi_{\mathbf{x}}$ be the real number with $\mathbf{x}$ as a fractional part.

**Problem**

*What can we infer about the arithmetic properties of $\xi_{\mathbf{x}}$ given combinatorial properties of $\mathbf{x}$?*

- Here we consider the irrationality exponent of $\xi_{\mathbf{x}}$.

# Irrationality Exponents

## Definition

The *irrationality exponent* $\mu(\xi)$ of a real number $\xi$ is the supremum of the real numbers $\rho$ such that the inequality

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{q^\rho}$$

has infinitely many rational solutions $p/q$. If $\mu(\xi) = \infty$, then we say that $\xi$ is a *Liouville number*.

# Irrationality Exponents

## Definition

The *irrationality exponent* $\mu(\xi)$ of a real number $\xi$ is the supremum of the real numbers $\rho$ such that the inequality

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{q^\rho}$$

has infinitely many rational solutions $p/q$. If $\mu(\xi) = \infty$, then we say that $\xi$ is a *Liouville number*.

- $\mu(\xi) = 2$ for almost all $\xi$
- $\mu(\xi) < 2$ if and only if $\xi$ is rational
- $\mu(\xi) = 2$ if $\xi$ is an algebraic irrational (Roth's Theorem)

# Link to Diophantine Exponents

**Proposition (Adamczewski (2010))**

$\mu(\xi_{\mathbf{x}}) \geq \mathrm{dio}(\mathbf{x})$

**Proof Sketch.**

The definition of $\mathrm{dio}$ provides arbitrarily long prefixes of $\mathbf{x}$ of the form $UV^e$ with $|UV^e|/|UV|$ arbitrarily close to $\mathrm{dio}(\mathbf{x})$. Select the rational $p/q$ to have fractional part $UV^\omega$ and work out the details. □

# Corollaries

**Theorem**

*Let* $\mathbf{t}$ *be a regular episturmian word with directive word* $\Delta$*. Then* $\xi_{\mathbf{t}}$ *is a Liouville number if and only if the sequence* $(a_k)$ *of partial quotients is bounded.*

- Proved for Sturmian words by Komatsu (1996).

**Theorem**

*Let* $\mathbf{t}$ *be a regular episturmian word of period* $d$*. If* $d = 2$ *or* $\limsup_k a_k \geq 3$*, then* $\mu(\xi_{\mathbf{t}}) > 2$*.*

- $\xi_{\mathbf{t}}$ is transcendental (follows from Bugeaud-Adamczewski (2007)).
- $\xi_{\mathbf{t}}$ is an atypical number (belongs to a set of measure 0).

# Thank You

Thank you for your attention!

J. Peltomäki.
Initial nonrepetitive complexity of regular episturmian words and their Diophantine exponents
Preprint (2021), arXiv:2103.08351