Rich, Sturmian & trapezoidal words

Amy Glen

School of Chemical & Mathematical Sciences
Murdoch University, Perth, Australia

amy.glen@gmail.com
http://wwwstaff.murdoch.edu.au/~aglen

54th AustMS Annual Meeting @ The University of Queensland
Special Session: Combinatorics

September 28, 2010

Words

In mathematics, words naturally arise when one wants to represent elements from some set in a systematic way.

Words

In mathematics, words naturally arise when one wants to represent elements from some set in a systematic way.

For instance:

- Expansions of real numbers in integer bases (e.g., binary and decimal expansions) or continued fraction expansions allow us to associate with every real number a finite or infinite sequence of digits.

Words

In mathematics, words naturally arise when one wants to represent elements from some set in a systematic way.

For instance:

- Expansions of real numbers in integer bases (e.g., binary and decimal expansions) or continued fraction expansions allow us to associate with every real number a finite or infinite sequence of digits.
- Combinatorial group theory involves the study of words that represent group elements.

Words

In mathematics, words naturally arise when one wants to represent elements from some set in a systematic way.

For instance:

- Expansions of real numbers in integer bases (e.g., binary and decimal expansions) or continued fraction expansions allow us to associate with every real number a finite or infinite sequence of digits.

- Combinatorial group theory involves the study of words that represent group elements.

Formally:

- A *word* is a finite or infinite sequence of symbols (*letters*) taken from a non-empty countable set $\mathcal{A}$ (*alphabet*).

Examples

- 001
- $(001)^\infty = 001001001001001001001001001\cdots$
- $11001111000110111011110011011100101111111101\cdots$
- $10010211012222010211002111110221222220111 2012\cdots$
- $0123456789101112131415\cdots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = \underset{\uparrow}{0.}01001001001001001001001001001\ldots$
- $1100111100011011101111001101110010111111101\cdots$
- $100102110122220102110021111102212222201112012\cdots$
- $0123456789101112131415\cdots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.010010010010010010010010010001\ldots = (2/7)_2$
- $110011110001101110111100110111001011111101\cdots$
- $1001021101222201021100211111022122222201112012\cdots$
- $0123456789101112131415\cdots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.01001001001001001001001001001\ldots = (2/7)_2$
- $1.100111100011011101111001101110010\ldots$
  $\uparrow$
- $100102110122220102110021111102212222201112012\cdots$
- $0123456789101112131415\cdots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.0100100100100100100100100100 1\ldots = (2/7)_2$
- $1.1001111000110111011110011011100 10\ldots = ((1+\sqrt{5})/2)_2$
- $10010211012222010211002111110221222220111 2012\cdots$
- $0123456789101112131415\cdots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.01001001001001001001001001001\ldots = (2/7)_2$
- $1.100111100011011101111001101110010\ldots = ((1+\sqrt{5})/2)_2$
- $10\underset{\uparrow}{.}0102110122220102110021111102212222201112012\ldots$
- $0123456789101112131415\cdots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.010010010010010010010010010010\ldots = (2/7)_2$
- $1.1001111000110111011110011011110010\ldots = ((1+\sqrt{5})/2)_2$
- $10.0102110122220102110021111102212222201112012\ldots = (\pi)_3$
- $0123456789101112131415\cdots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.01001001001001001001001001001001\ldots = (2/7)_2$
- $1.100111100011011101111001101110010\ldots = ((1+\sqrt{5})/2)_2$
- $10.0102110122220102110021111102212222201112012\ldots = (\pi)_3$
- $0.123456789101112131415\ldots$
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^{\infty} = 0.0100100100100100100100100100\ldots = (2/7)_2$
- $1.1001111000110111011110011101110010\ldots = ((1+\sqrt{5})/2)_2$
- $10.0102110122220102110021111102212222201112012\ldots = (\pi)_3$
- $0.123456789101112131415\ldots =$ Champernowne's number (C_{10})
- $1121212121212\cdots$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.0100100100100100100100100100 1\ldots = (2/7)_2$
- $1.10011110001101110111100110111001 0\ldots = ((1+\sqrt{5})/2)_2$
- $10.0102110122220102110021111102212222201112012\ldots = (\pi)_3$
- $0.123456789101112131415\ldots =$ Champernowne's number (C_{10})
- $[1; 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, \ldots] = \sqrt{3}$
- $212114116118\cdots$

Examples

- 001
- $(001)^\infty = 0.010010010010010010010010010\ldots = (2/7)_2$
- $1.1001111000110111011110011011100 10\ldots = ((1+\sqrt{5})/2)_2$
- $10.010211012222010211002111110221222220111 2012\ldots = (\pi)_3$
- $0.123456789101112131415\ldots =$ Champernowne's number (C_{10})
- $[1; 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, \ldots] = \sqrt{3}$
- $[2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots]$

Examples

- 001
- $(001)^\infty = 0.01001001001001001001001001001\ldots = (2/7)_2$
- $1.1001111000110111011110011011110010\ldots = ((1+\sqrt{5})/2)_2$
- $10.0102110122220102110021111102212222201112012\ldots = (\pi)_3$
- $0.123456789101112131415\ldots =$ Champernowne's number (C_{10})
- $[1; 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, \ldots] = \sqrt{3}$
- $[2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots] = e$

Words . . .

- Depending on the problem to be solved, it may be fruitful to study combinatorial and structural properties of the words representing the elements of a particular set or to impose certain combinatorial conditions on such words.

Words . . .

- Depending on the problem to be solved, it may be fruitful to study combinatorial and structural properties of the words representing the elements of a particular set or to impose certain combinatorial conditions on such words.

- Most commonly studied words are those which satisfy one or more strong regularity properties; for instance, words containing many repetitions or palindromes.

Words . . .

- Depending on the problem to be solved, it may be fruitful to study combinatorial and structural properties of the words representing the elements of a particular set or to impose certain combinatorial conditions on such words.

- Most commonly studied words are those which satisfy one or more strong regularity properties; for instance, words containing many repetitions or palindromes.

- The extent to which a word exhibits strong regularity properties is generally inversely proportional to its "complexity".

Words . . .

- Depending on the problem to be solved, it may be fruitful to study combinatorial and structural properties of the words representing the elements of a particular set or to impose certain combinatorial conditions on such words.

- Most commonly studied words are those which satisfy one or more strong regularity properties; for instance, words containing many repetitions or palindromes.

- The extent to which a word exhibits strong regularity properties is generally inversely proportional to its "complexity".

  Basic measure: number of distinct blocks (factors) of each length occurring in the word.

Factor Complexity

- Given a finite or infinite word w, let $\mathcal{F}_n(w)$ denote the set of distinct factors of w of length $n \in \mathbb{N}^+$.

Factor Complexity

- Given a finite or infinite word w, let $\mathcal{F}_n(w)$ denote the set of distinct factors of w of length $n \in \mathbb{N}^+$.
- The function $C_w(n) : \mathbb{N} \to \mathbb{N}$ defined by

$$C_w(n) = Card(\mathcal{F}_n(w))$$

is called the *factor complexity function* of w.

Factor Complexity

- Given a finite or infinite word w, let $\mathcal{F}_n(w)$ denote the set of distinct factors of w of length $n \in \mathbb{N}^+$.
- The function $C_w(n) : \mathbb{N} \to \mathbb{N}$ defined by

$$C_w(n) = Card(\mathcal{F}_n(w))$$

is called the *factor complexity function* of w.

Example

$x = (\sqrt{2})_2 = 1.0110101000001001111\ldots$

Factor Complexity

- Given a finite or infinite word w, let $\mathcal{F}_n(w)$ denote the set of distinct factors of w of length $n \in \mathbb{N}^+$.
- The function $C_w(n) : \mathbb{N} \to \mathbb{N}$ defined by

$$C_w(n) = Card(\mathcal{F}_n(w))$$

is called the *factor complexity function* of w.

Example

$\mathbf{x} = (\sqrt{2})_2 = 1.0110101000001001111\ldots$

$\mathcal{F}_1(\mathbf{x}) = \{0, 1\}, \ C_{\mathbf{x}}(1) = 2$

Factor Complexity

- Given a finite or infinite word w, let $\mathcal{F}_n(w)$ denote the set of distinct factors of w of length $n \in \mathbb{N}^+$.

- The function $C_w(n) : \mathbb{N} \to \mathbb{N}$ defined by

$$C_w(n) = Card(\mathcal{F}_n(w))$$

is called the *factor complexity function* of w.

Example

$\mathbf{x} = (\sqrt{2})_2 = 1.0110101000001001111\ldots$

$\mathcal{F}_1(\mathbf{x}) = \{0, 1\},\ C_{\mathbf{x}}(1) = 2$

$\mathcal{F}_2(\mathbf{x}) = \{00, 01, 10, 11\},\ C_{\mathbf{x}}(2) = 4$

Factor Complexity

- Given a finite or infinite word w, let $\mathcal{F}_n(w)$ denote the set of distinct factors of w of length $n \in \mathbb{N}^+$.
- The function $C_w(n) : \mathbb{N} \to \mathbb{N}$ defined by

$$C_w(n) = Card(\mathcal{F}_n(w))$$

is called the *factor complexity function* of w.

Example

$\boldsymbol{x} = (\sqrt{2})_2 = 1.0110101000001001111\ldots$

$\mathcal{F}_1(\boldsymbol{x}) = \{0, 1\},\ C_{\boldsymbol{x}}(1) = 2$

$\mathcal{F}_2(\boldsymbol{x}) = \{00, 01, 10, 11\},\ C_{\boldsymbol{x}}(2) = 4$

$\mathcal{F}_3(\boldsymbol{x}) = \{000, 001, 010, 100, 101, 110, 111\},\ C_{\boldsymbol{x}}(3) = 8$

Factor Complexity

- Given a finite or infinite word w, let $\mathcal{F}_n(w)$ denote the set of distinct factors of w of length $n \in \mathbb{N}^+$.
- The function $C_w(n) : \mathbb{N} \to \mathbb{N}$ defined by

$$C_w(n) = Card(\mathcal{F}_n(w))$$

is called the *factor complexity function* of w.

Example

$\mathbf{x} = (\sqrt{2})_2 = 1.0110101000001001111\ldots$

$\mathcal{F}_1(\mathbf{x}) = \{0, 1\},\ C_{\mathbf{x}}(1) = 2$

$\mathcal{F}_2(\mathbf{x}) = \{00, 01, 10, 11\},\ C_{\mathbf{x}}(2) = 4$

$\mathcal{F}_3(\mathbf{x}) = \{000, 001, 010, 100, 101, 110, 111\},\ C_{\mathbf{x}}(3) = 8$

Conjecture: $C_{\mathbf{x}}(n) = 2^n$ for all n as it is believed $\sqrt{2}$ is *normal* in base 2.

Complexity & Periodicity

Theorem (Morse-Hedlund 1940)

An infinite word $\boldsymbol{w}$ is eventually periodic if and only if $C_{\boldsymbol{w}}(n) \leq n$ for some $n \in \mathbb{N}^+$.

Complexity & Periodicity

Theorem (Morse-Hedlund 1940)

An infinite word $\boldsymbol{w}$ is eventually periodic if and only if $C_{\boldsymbol{w}}(n) \leq n$ for some $n \in \mathbb{N}^+$.

That is: $\boldsymbol{w}$ is aperiodic $\Leftrightarrow C_{\boldsymbol{w}}(n) \geq n+1$ for all $n \in \mathbb{N}$.

Complexity & Periodicity

Theorem (Morse-Hedlund 1940)

An infinite word $\mathbf{w}$ is eventually periodic if and only if $C_{\mathbf{w}}(n) \leq n$ for some $n \in \mathbb{N}^{+}$.

That is: $\mathbf{w}$ is aperiodic $\Leftrightarrow C_{\mathbf{w}}(n) \geq n + 1$ for all $n \in \mathbb{N}$.

- Sturmian words are the aperiodic infinite words of minimal complexity, i.e., an infinite word $\mathbf{w}$ is *Sturmian* if and only if $C_{\mathbf{w}}(n) = n + 1$ for each n.

Complexity & Periodicity

Theorem (Morse-Hedlund 1940)

An infinite word $\mathbf{w}$ is eventually periodic if and only if $C_{\mathbf{w}}(n) \leq n$ for some $n \in \mathbb{N}^+$.

That is: $\mathbf{w}$ is aperiodic $\Leftrightarrow C_{\mathbf{w}}(n) \geq n + 1$ for all $n \in \mathbb{N}$.

- Sturmian words are the aperiodic infinite words of minimal complexity, i.e., an infinite word $\mathbf{w}$ is *Sturmian* if and only if $C_{\mathbf{w}}(n) = n + 1$ for each n.
- Low complexity accounts for many interesting features, as it induces certain regularities in such words without, however, making them periodic.

Complexity & Periodicity

Theorem (Morse-Hedlund 1940)

An infinite word $\boldsymbol{w}$ is eventually periodic if and only if $C_{\boldsymbol{w}}(n) \leq n$ for some $n \in \mathbb{N}^{+}$.

That is: $\boldsymbol{w}$ is aperiodic $\Leftrightarrow C_{\boldsymbol{w}}(n) \geq n+1$ for all $n \in \mathbb{N}$.

- Sturmian words are the aperiodic infinite words of minimal complexity, i.e., an infinite word $\boldsymbol{w}$ is *Sturmian* if and only if $C_{\boldsymbol{w}}(n) = n+1$ for each n.
- Low complexity accounts for many interesting features, as it induces certain regularities in such words without, however, making them periodic.
- References in: Combinatorics, Symbolic Dynamics, Number Theory, Discrete Geometry, Theoretical Physics, Theoretical Computer Science.

Complexity & Periodicity

Theorem (Morse-Hedlund 1940)

An infinite word $\boldsymbol{w}$ is eventually periodic if and only if $C_{\boldsymbol{w}}(n) \leq n$ for some $n \in \mathbb{N}^+$.

That is: $\boldsymbol{w}$ is aperiodic $\Leftrightarrow C_{\boldsymbol{w}}(n) \geq n + 1$ for all $n \in \mathbb{N}$.

- Sturmian words are the aperiodic infinite words of minimal complexity, i.e., an infinite word $\boldsymbol{w}$ is *Sturmian* if and only if $C_{\boldsymbol{w}}(n) = n + 1$ for each n.
- Low complexity accounts for many interesting features, as it induces certain regularities in such words without, however, making them periodic.
- References in: Combinatorics, Symbolic Dynamics, Number Theory, Discrete Geometry, Theoretical Physics, Theoretical Computer Science.
- Numerous equivalent definitions & characterisations . . .

Palindrome Complexity

Given a finite or infinite word w, let $P_w(n)$ denote the *palindromic complexity function* of w, which counts the number of palindromic factors of w of each length $n \geq 0$.

Palindrome Complexity

Given a finite or infinite word w, let $P_w(n)$ denote the *palindromic complexity function* of w, which counts the number of palindromic factors of w of each length $n \geq 0$.

Theorem (Droubay-Pirillo 1999)

An infinite word $\mathbf{w}$ is Sturmian if and only if

$$P_{\mathbf{w}}(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 2 & \text{if } n \text{ is odd} \end{cases}$$

Palindrome Complexity

Given a finite or infinite word w, let $P_w(n)$ denote the *palindromic complexity function* of w, which counts the number of palindromic factors of w of each length $n \geq 0$.

Theorem (Droubay-Pirillo 1999)

An infinite word $\mathbf{w}$ is Sturmian if and only if

$$P_{\mathbf{w}}(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 2 & \text{if } n \text{ is odd} \end{cases}$$

Note:

- Any Sturmian word is over a 2-letter alphabet since it has two distinct factors of length 1.

Palindrome Complexity

Given a finite or infinite word w, let $P_w(n)$ denote the *palindromic complexity function* of w, which counts the number of palindromic factors of w of each length $n \geq 0$.

Theorem (Droubay-Pirillo 1999)

An infinite word $\boldsymbol{w}$ is Sturmian if and only if

$$P_{\boldsymbol{w}}(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 2 & \text{if } n \text{ is odd} \end{cases}$$

Note:

- Any Sturmian word is over a 2-letter alphabet since it has two distinct factors of length 1.
- A Sturmian word over the alphabet $\{a, b\}$ contains either aa or bb, but not both.

Constructing Sturmian words

- Let's consider a nice geometric realisation, starting with a special class of finite words . . .

Constructing Sturmian words

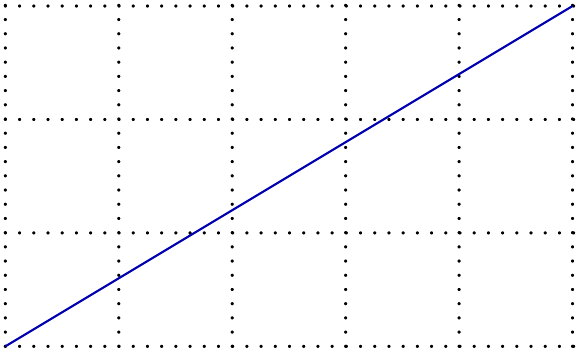
- Let's consider a nice geometric realisation, starting with a special class of finite words ...
- Words over a 2-letter alphabet $\{a, b\}$ that are factors of (infinite) Sturmian words are called *finite Sturmian words* – they are the cyclic shifts of *Christoffel words* obtained via the following construction.

Constructing Sturmian words

- Let's consider a nice geometric realisation, starting with a special class of finite words . . .
- Words over a 2-letter alphabet $\{a, b\}$ that are factors of (infinite) Sturmian words are called *finite Sturmian words* – they are the cyclic shifts of *Christoffel words* obtained via the following construction.
- Using a similar construction we obtain infinite Sturmian words.

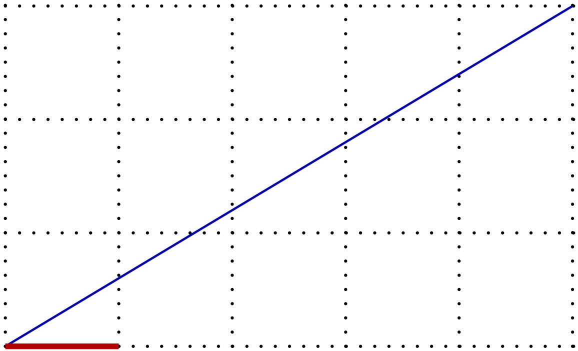
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



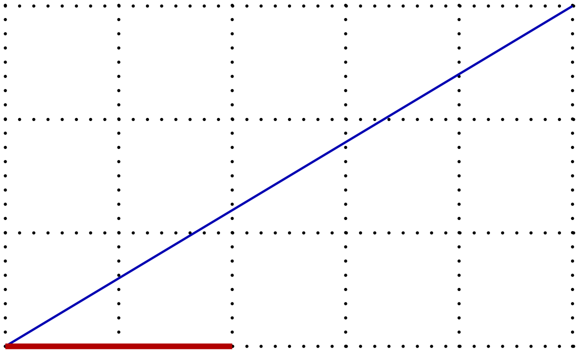
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



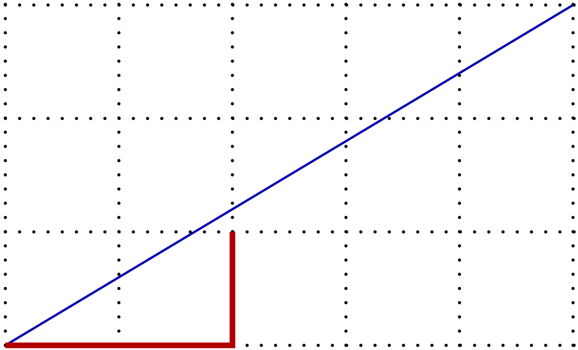
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



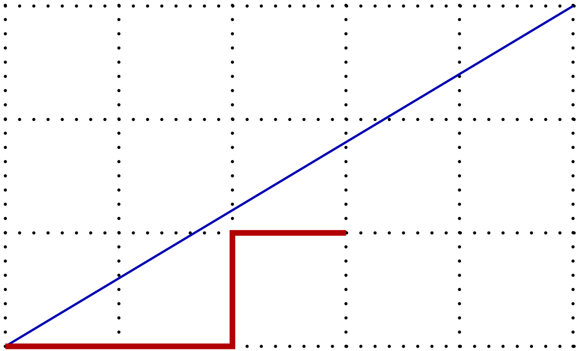
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



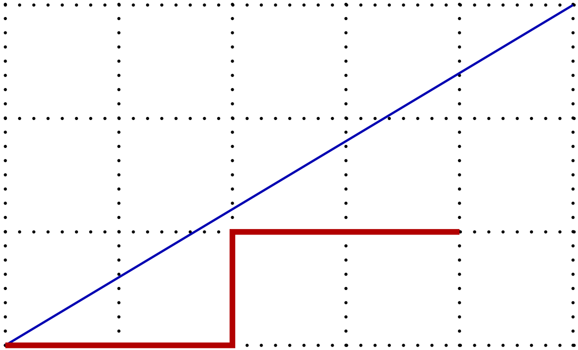
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



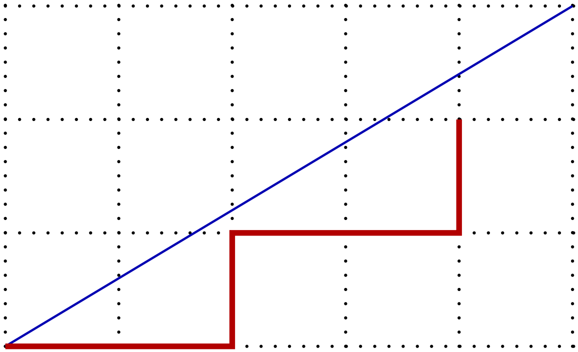
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



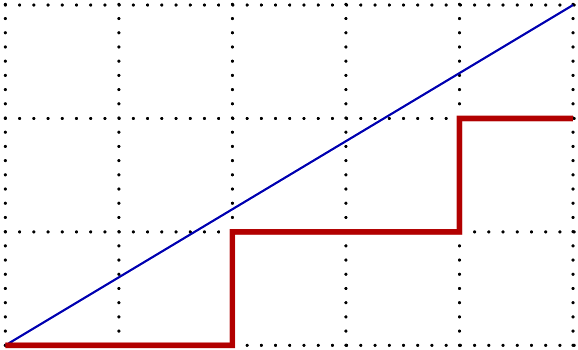
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



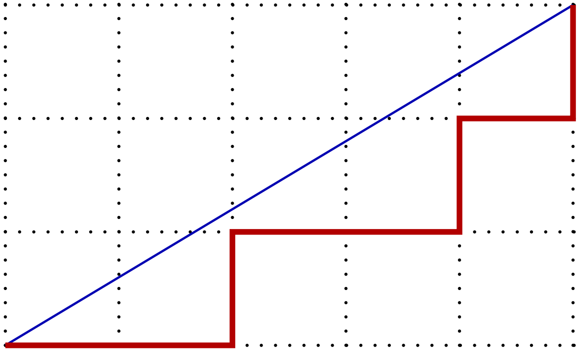
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



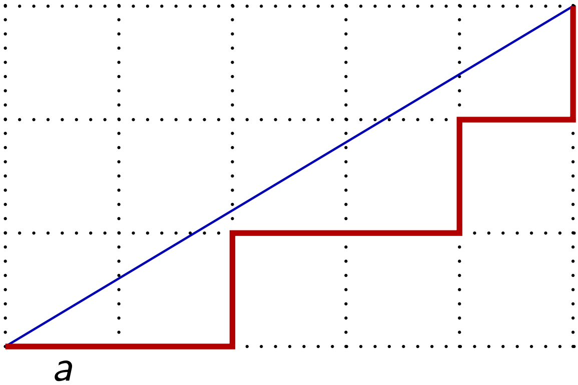
Christoffel words: Construction by example

Lower Christoffel word of slope $\frac{3}{5}$



Christoffel words: Construction by example

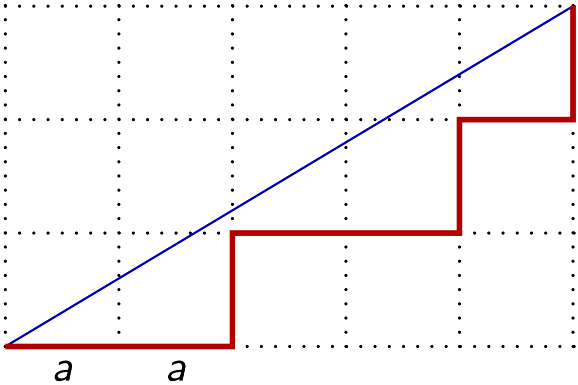
Lower Christoffel word of slope $\frac{3}{5}$



$$L(3,5) = a$$

Christoffel words: Construction by example

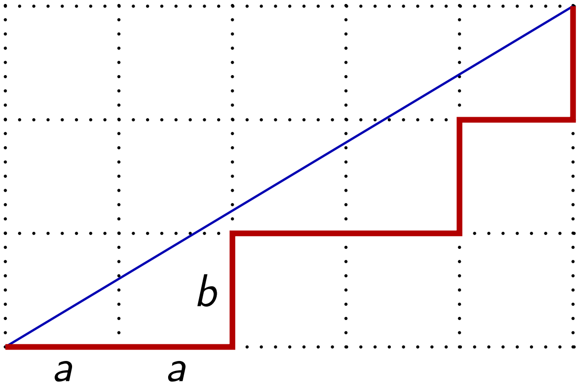
Lower Christoffel word of slope $\frac{3}{5}$



$L(3,5) = aa$

Christoffel words: Construction by example

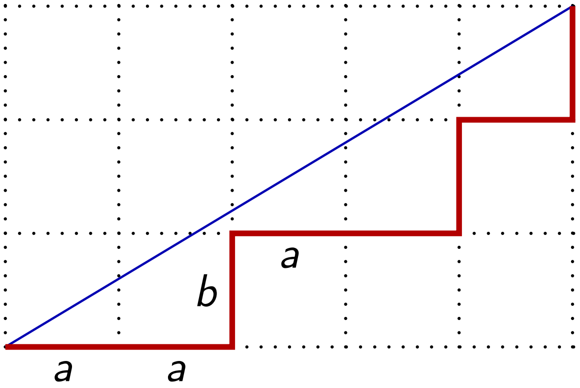
Lower Christoffel word of slope $\frac{3}{5}$



$L(3,5) = aab$

Christoffel words: Construction by example

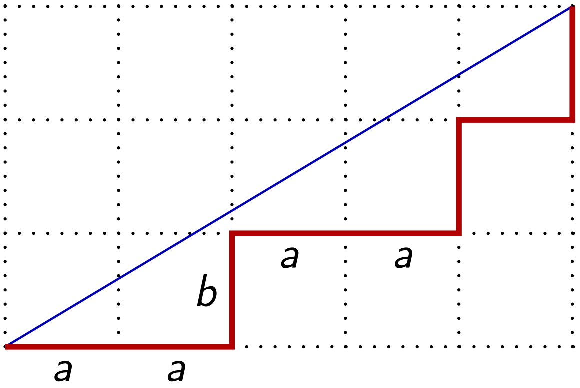
Lower Christoffel word of slope $\frac{3}{5}$



$L(3,5) = aaba$

Christoffel words: Construction by example

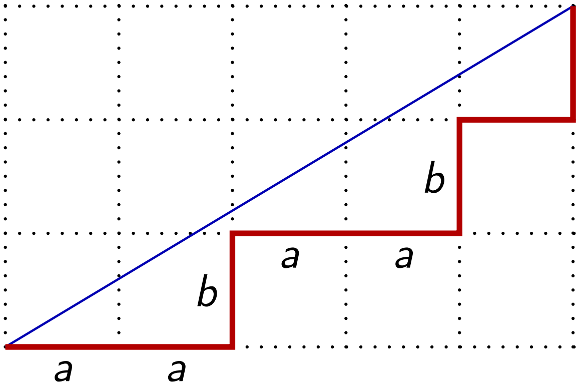
Lower Christoffel word of slope $\frac{3}{5}$



$L(3,5) = aabaa$

Christoffel words: Construction by example

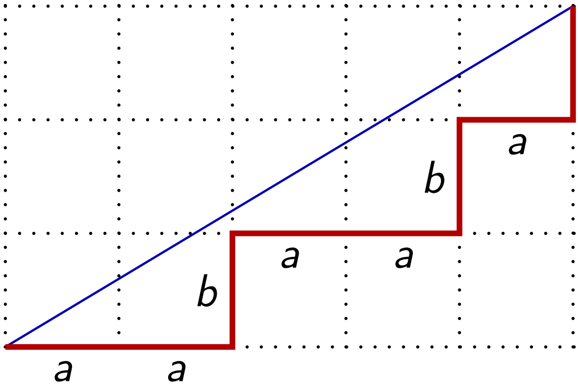
Lower Christoffel word of slope $\frac{3}{5}$



$L(3,5) = aabaab$

Christoffel words: Construction by example

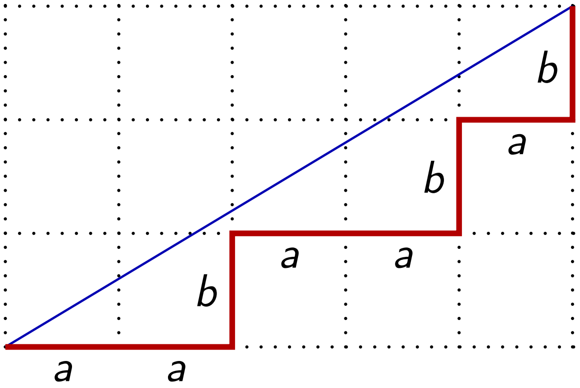
Lower Christoffel word of slope $\frac{3}{5}$



L(3,5) = aabaaba

Christoffel words: Construction by example

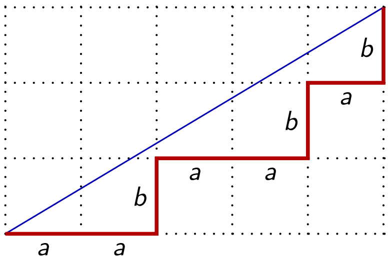
Lower Christoffel word of slope $\frac{3}{5}$

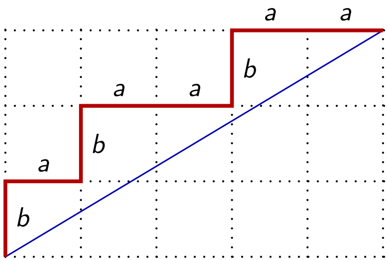


L(3,5) = aabaabab

Christoffel words: Construction by example

Lower & Upper Christoffel words of slope $\frac{3}{5}$





L(3,5) = aabaabab

U(3,5) = babaabaa

From Christoffel words to Sturmian words

Sturmian words: Obtained **similarly** by replacing the line segment by a half-line:

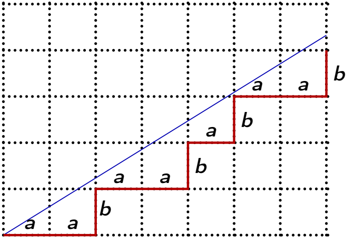
$$y = \alpha x + \rho \text{ with irrational } \alpha \in (0,1),\ \rho \in \mathbb{R}.$$

From Christoffel words to Sturmian words

Sturmian words: Obtained **similarly** by replacing the line segment by a half-line:

$$y = \alpha x + \rho \text{ with irrational } \alpha \in (0,1),\ \rho \in \mathbb{R}.$$

Example: $y = \frac{\sqrt{5}-1}{2}x \longrightarrow$ *Fibonacci word*

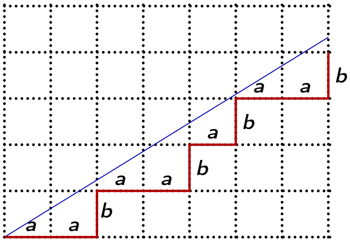


From Christoffel words to Sturmian words

Sturmian words: Obtained **similarly** by replacing the line segment by a half-line:

$$y = \alpha x + \rho \text{ with irrational } \alpha \in (0,1),\ \rho \in \mathbb{R}.$$

Example: $y = \frac{\sqrt{5}-1}{2}x \longrightarrow$ *Fibonacci word*



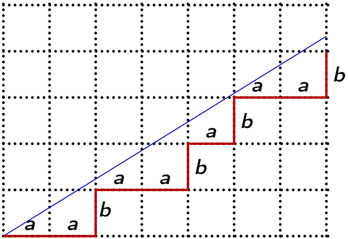
- $\boldsymbol{f} = abaababaabaababaaba\cdots$ (note: disregard 1st a in construction)

From Christoffel words to Sturmian words

Sturmian words: Obtained **similarly** by replacing the line segment by a half-line:

$$y = \alpha x + \rho \text{ with irrational } \alpha \in (0,1),\ \rho \in \mathbb{R}.$$

Example: $y = \frac{\sqrt{5}-1}{2}x \longrightarrow$ *Fibonacci word*



- $\boldsymbol{f} = abaababaabaababaaba\cdots$ (note: disregard 1st a in construction)
- *Standard Sturmian word* of slope $\frac{\sqrt{5}-1}{2}$, golden ratio conjugate

Factor complexity of finite Sturmian words

In 1999, A. de Luca studied the behaviour of the factor complexity of finite words.

Factor complexity of finite Sturmian words

In 1999, A. de Luca studied the behaviour of the factor complexity of finite words. He showed:

Theorem (de Luca 1999)

If w is a finite Sturmian word of length $|w|$ (i.e., a cyclic shift of a Christoffel word), then the graph of $C_w(n)$ as a function of n (for $0 \leq n \leq |w|$) is that of a regular trapezoid (possibly degenerated to a triangle).

Factor complexity of finite Sturmian words

In 1999, A. de Luca studied the behaviour of the factor complexity of finite words. He showed:

Theorem (de Luca 1999)

If w is a finite Sturmian word of length $|w|$ (i.e., a cyclic shift of a Christoffel word), then the graph of $C_w(n)$ as a function of n (for $0 \leq n \leq |w|$) is that of a regular trapezoid (possibly degenerated to a triangle).

That is:

- $C_w(n)$ increases by 1 with each n on some interval of length r.

Factor complexity of finite Sturmian words

In 1999, A. de Luca studied the behaviour of the factor complexity of finite words. He showed:

Theorem (de Luca 1999)

If w is a finite Sturmian word of length $|w|$ (i.e., a cyclic shift of a Christoffel word), then the graph of $C_w(n)$ as a function of n (for $0 \leq n \leq |w|$) is that of a regular trapezoid (possibly degenerated to a triangle).

That is:

- $C_w(n)$ increases by 1 with each n on some interval of length r.
- Then $C_w(n)$ is constant on some interval of length s.

Factor complexity of finite Sturmian words

In 1999, A. de Luca studied the behaviour of the factor complexity of finite words. He showed:

Theorem (de Luca 1999)

If w is a finite Sturmian word of length $|w|$ (i.e., a cyclic shift of a Christoffel word), then the graph of $C_w(n)$ as a function of n (for $0 \leq n \leq |w|$) is that of a regular trapezoid (possibly degenerated to a triangle).

That is:

- $C_w(n)$ increases by 1 with each n on some interval of length r.
- Then $C_w(n)$ is constant on some interval of length s.
- Finally $C_w(n)$ decreases by 1 with each n on an interval of length r.

Factor complexity of finite Sturmian words

In 1999, A. de Luca studied the behaviour of the factor complexity of finite words. He showed:

Theorem (de Luca 1999)

If w is a finite Sturmian word of length $|w|$ (i.e., a cyclic shift of a Christoffel word), then the graph of $C_w(n)$ as a function of n (for $0 \leq n \leq |w|$) is that of a regular trapezoid (possibly degenerated to a triangle).

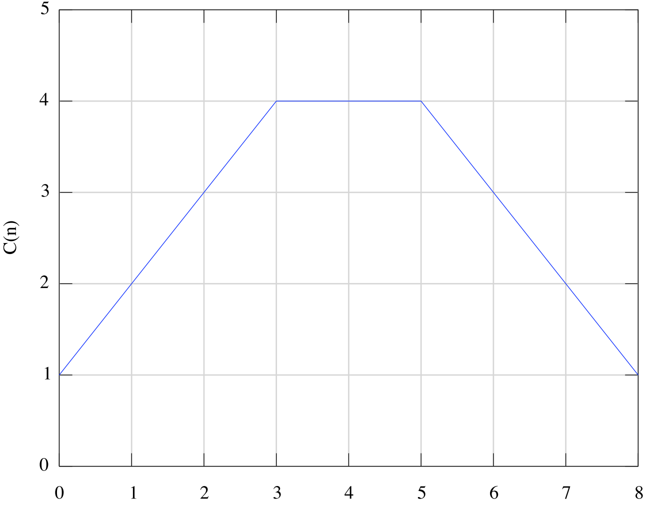
That is:

- $C_w(n)$ increases by 1 with each n on some interval of length r.
- Then $C_w(n)$ is constant on some interval of length s.
- Finally $C_w(n)$ decreases by 1 with each n on an interval of length r.

So if we set $D_w(n) = C_w(n+1) - C_w(n)$ for each n with $0 \leq n \leq |w| - 1$, then the word $D_w(0)D_w(1)\cdots D_w(|w|-1)$ takes the form $1^r 0^s (-1)^r$.

Example

Graph of the factor complexity of the Christoffel word $L(3,5) = aabaabab$



Trapezoidal words

- This "trapezoidal property" does not characterise Sturmian words.

Trapezoidal words

- This "trapezoidal property" does not characterise Sturmian words.

  For example, *aabb* is trapezoidal, but not Sturmian.

Trapezoidal words

- This "trapezoidal property" does not characterise Sturmian words.

  For example, *aabb* is trapezoidal, but not Sturmian.

- Note: If w is a *trapezoidal word* (i.e., its 'complexity' graph has the same behaviour as that of Sturmian words), then necessarily $C_w(1) = 2$.

Trapezoidal words

- This "trapezoidal property" does not characterise Sturmian words.

  For example, *aabb* is trapezoidal, but not Sturmian.

- Note: If w is a *trapezoidal word* (i.e., its 'complexity' graph has the same behaviour as that of Sturmian words), then necessarily $C_w(1) = 2$.

  This is because there is 1 factor of length 0, namely the *empty word* ε.

Trapezoidal words

- This "trapezoidal property" does not characterise Sturmian words.

  For example, *aabb* is trapezoidal, but not Sturmian.

- Note: If w is a *trapezoidal word* (i.e., its 'complexity' graph has the same behaviour as that of Sturmian words), then necessarily $C_w(1) = 2$.

  This is because there is 1 factor of length 0, namely the *empty word* ε.

- So any trapezoidal word is on a binary alphabet and the family of trapezoidal words properly contains all finite Sturmian words.

Trapezoidal words

- This "trapezoidal property" does not characterise Sturmian words.

  For example, *aabb* is trapezoidal, but not Sturmian.

- Note: If w is a *trapezoidal word* (i.e., its 'complexity' graph has the same behaviour as that of Sturmian words), then necessarily $C_w(1) = 2$.

  This is because there is 1 factor of length 0, namely the *empty word* ε.

- So any trapezoidal word is on a binary alphabet and the family of trapezoidal words properly contains all finite Sturmian words.

- F. D'Alessandro (2002): classified all non-Sturmian trapezoidal words.

Characterisation of Sturmian palindromes

We have shown:

Theorem (de Luca-G.-Zamboni)

Let w be a palindrome. Then w is Sturmian if and only if w is trapezoidal.

Characterisation of Sturmian palindromes

We have shown:

Theorem (de Luca-G.-Zamboni)

Let w be a palindrome. Then w is Sturmian if and only if w is trapezoidal.

Theorem (de Luca-G.-Zamboni)

Let w be a trapezoidal word. Then w contains $|w| + 1$ distinct palindromes (including ε).

Characterisation of Sturmian palindromes

We have shown:

Theorem (de Luca-G.-Zamboni)

Let w be a palindrome. Then w is Sturmian if and only if w is trapezoidal.

Theorem (de Luca-G.-Zamboni)

Let w be a trapezoidal word. Then w contains $|w| + 1$ distinct palindromes (including ε).

That is, trapezoidal words (and hence finite Sturmian words) are "rich" in palindromes in the sense that they contain the maximum number of distinct palindromic factors since:

Characterisation of Sturmian palindromes

We have shown:

Theorem (de Luca-G.-Zamboni)

Let w be a palindrome. Then w is Sturmian if and only if w is trapezoidal.

Theorem (de Luca-G.-Zamboni)

Let w be a trapezoidal word. Then w contains $|w| + 1$ distinct palindromes (including ε).

That is, trapezoidal words (and hence finite Sturmian words) are "rich" in palindromes in the sense that they contain the maximum number of distinct palindromic factors since:

Theorem (Droubay-Justin-Pirillo 2001)

A finite word w contains at most $|w| + 1$ distinct palindromes (including ε).

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- *abac* is rich, whereas *abca* is **not** rich.

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- *abac* is rich, whereas *abca* is **not** rich.
- The word `rich` is rich . . . and `poor` is rich too!

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- *abac* is rich, whereas *abca* is **not** rich.
- The word `rich` is rich ... and `poor` is rich too!
- Any trapezoidal word is rich, but not conversely.

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w|+1$ distinct palindromes.

Examples

- *abac* is rich, whereas *abca* is **not** rich.
- The word `rich` is rich . . . and `poor` is rich too!
- Any trapezoidal word is rich, but not conversely.

  E.g., *aabbaa* is rich, but not trapezoidal

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- *abac* is rich, whereas *abca* is **not** rich.
- The word `rich` is rich ... and `poor` is rich too!
- Any trapezoidal word is rich, but not conversely.

  E.g., *aabbaa* is rich, but not trapezoidal ($C(1) = 2$, $C(2) = 4$)

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- $abac$ is rich, whereas $abca$ is **not** rich.
- The word `rich` is rich ... and `poor` is rich too!
- Any trapezoidal word is rich, but not conversely.

  E.g., $aabbaa$ is rich, but not trapezoidal ($C(1) = 2$, $C(2) = 4$)

Definition (G.-Justin 2007)

An infinite word is *rich* iff all of its factors are rich.

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- *abac* is rich, whereas *abca* is **not** rich.
- The word `rich` is rich . . . and `poor` is rich too!
- Any trapezoidal word is rich, but not conversely.

  E.g., *aabbaa* is rich, but not trapezoidal ($C(1) = 2$, $C(2) = 4$)

Definition (G.-Justin 2007)

An infinite word is *rich* iff all of its factors are rich.

Examples

- $a^\omega = aaaaaa\cdots$ and $ab^\omega = abbb\cdots$ are rich.

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- *abac* is rich, whereas *abca* is **not** rich.
- The word `rich` is rich ... and `poor` is rich too!
- Any trapezoidal word is rich, but not conversely.

  E.g., *aabbaa* is rich, but not trapezoidal ($C(1) = 2$, $C(2) = 4$)

Definition (G.-Justin 2007)

An infinite word is *rich* iff all of its factors are rich.

Examples

- $a^\omega = aaaaaa\cdots$ and $ab^\omega = abbb\cdots$ are rich.
- $(ab)^\omega = abababab\cdots$ and $(aba)^\omega = abaabaaba\cdots$ are rich.

Rich words

Definition (G.-Justin 2007)

A finite word w is *rich* iff w contains exactly $|w| + 1$ distinct palindromes.

Examples

- $abac$ is rich, whereas $abca$ is **not** rich.
- The word `rich` is rich ... and `poor` is rich too!
- Any trapezoidal word is rich, but not conversely.

  E.g., $aabbaa$ is rich, but not trapezoidal $(C(1) = 2,\ C(2) = 4)$

Definition (G.-Justin 2007)

An infinite word is *rich* iff all of its factors are rich.

Examples

- $a^\omega = aaaaaa\cdots$ and $ab^\omega = abbb\cdots$ are rich.
- $(ab)^\omega = abababab\cdots$ and $(aba)^\omega = abaabaaba\cdots$ are rich.
- abc is rich, but $(abc)^\omega = abcabcabc\cdots$ is **not** rich.

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $abaabaaabaaaabaaaaab \cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $\underline{a}baabaaabaaaabaaaaab\cdots$

Rich words ...

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $a\underline{b}aabaaabaaaabaaaaab\cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $\underline{aba}abaaabaaaabaaaaab\cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $ab\underline{aa}baaabaaaabaaaaab\cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $a\underline{baab}aaabaaaabaaaaab\cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $\underline{abaaba}aabaaaabaaaaab\cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $ab\underline{aabaa}abaaaabaaaaab\cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $aba ab\underline{aaa}baaaabaaaaab \cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $aba a\underline{baaab}aaaabaaaaab\cdots$

Rich words . . .

Roughly speaking, a finite or infinite word is rich if and only if a new palindrome is introduced at each new position.

Example: $abaabaaabaaaabaaaaab\cdots$

A characterisation of rich words

- Let u be a factor of a finite or infinite word w.

A characterisation of rich words

- Let u be a factor of a finite or infinite word w.
- A *complete return* to u in w is a factor of w having exactly two occurrences of u, one as a prefix and one as a suffix.

A characterisation of rich words

- Let u be a factor of a finite or infinite word w.
- A *complete return* to u in w is a factor of w having exactly two occurrences of u, one as a prefix and one as a suffix.
- Example: $aabcbaa$ is a complete return to aa in $aabcbaaba$ (rich).

A characterisation of rich words

- Let u be a factor of a finite or infinite word w.
- A *complete return* to u in w is a factor of w having exactly two occurrences of u, one as a prefix and one as a suffix.
- Example: $aabcbaa$ is a complete return to aa in $aabcbaaba$ (rich).

Characteristic Property (G.-Justin 2007)

A finite or infinite word w is rich if and only if for each palindromic factor p of w, every *complete return* to p in w is a palindrome.

A characterisation of rich words

- Let u be a factor of a finite or infinite word w.
- A *complete return* to u in w is a factor of w having exactly two occurrences of u, one as a prefix and one as a suffix.
- Example: $aabcbaa$ is a complete return to aa in $aabcbaaba$ (rich).

Characteristic Property (G.-Justin 2007)

A finite or infinite word w is rich if and only if for each palindromic factor p of w, every *complete return* to p in w is a palindrome.

In short, a word is rich if and only if all complete returns to palindromes are palindromes.

More General Examples

Rich words have appeared in many different contexts; they include:

- Sturmian and episturmian words
  Droubay-Justin-Pirillo (2001)
  Anne-Zamboni-Zorca (2005)
  Bucci-De Luca-G.-Zamboni (2008)
- Complementation-symmetric Rote sequences
  Allouche-Baake-Cassaigne-Damanik (2003) + Bucci-De Luca-G.-Zamboni (2008)
- Symbolic codings of trajectories of symmetric interval exchange transformations – Ferenczi-Zamboni (2008)
- A certain class of words associated with β-expansions where β is a simple Parry number
  Ambrož-Frougny-Masáková-Pelantová (2006) + Bucci-De Luca-G.-Zamboni (2008)
- Infinite words with "abundant palindromic prefixes"
  Introduced by Fischler in 2006 in relation to Diophantine approximation

A Connection Between Palindromic & Factor Complexity

Allouche-Baake-Cassaigne-Damanik, 2003: for any aperiodic infinite word $\boldsymbol{w}$,

$$P_{\boldsymbol{w}}(n) \leq \frac{16}{n}\, C_{\boldsymbol{w}}\left(n + \left\lfloor \frac{n}{4} \right\rfloor\right) \quad \text{for all } n \in \mathbb{N}.$$

A Connection Between Palindromic & Factor Complexity

Allouche-Baake-Cassaigne-Damanik, 2003: for any aperiodic infinite word $\boldsymbol{w}$,

$$P_{\boldsymbol{w}}(n) \leq \frac{16}{n} C_{\boldsymbol{w}}\left(n + \left\lfloor \frac{n}{4} \right\rfloor\right) \quad \text{for all } n \in \mathbb{N}.$$

Baláži-Masáková-Pelantová, 2007: for any uniformly recurrent infinite word $\boldsymbol{w}$ with $\mathcal{F}(\boldsymbol{w})$ closed under reversal,

$$P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1) \leq C_{\boldsymbol{w}}(n+1) - C_{\boldsymbol{w}}(n) + 2 \quad \text{for all } n \in \mathbb{N}. \qquad (*)$$

A Connection Between Palindromic & Factor Complexity

Allouche-Baake-Cassaigne-Damanik, 2003: for any aperiodic infinite word $\boldsymbol{w}$,

$$P_{\boldsymbol{w}}(n) \leq \frac{16}{n} C_{\boldsymbol{w}}\left(n + \left\lfloor \frac{n}{4} \right\rfloor\right) \quad \text{for all } n \in \mathbb{N}.$$

Baláži-Masáková-Pelantová, 2007: for any uniformly recurrent infinite word $\boldsymbol{w}$ with $\mathcal{F}(\boldsymbol{w})$ closed under reversal,

$$P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1) \leq C_{\boldsymbol{w}}(n+1) - C_{\boldsymbol{w}}(n) + 2 \quad \text{for all } n \in \mathbb{N}. \qquad (*)$$

Bucci-De Luca-G.-Zamboni, 2008: infinite words $\boldsymbol{w}$ for which $P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1)$ reaches the upper bound in $(*)$ for every n are **rich** . . .

A Connection Between Palindromic & Factor Complexity

Theorem (Bucci-De Luca-G.-Zamboni 2008)

For any infinite word $\boldsymbol{w}$ with set of factors $\mathcal{F}(\boldsymbol{w})$ closed under reversal, the following conditions are equivalent:

(I) $\boldsymbol{w}$ is rich;

(II) $P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1) = C_{\boldsymbol{w}}(n+1) - C_{\boldsymbol{w}}(n) + 2$ for all $n \in \mathbb{N}$.

A Connection Between Palindromic & Factor Complexity

Theorem (Bucci-De Luca-G.-Zamboni 2008)

For any infinite word $\boldsymbol{w}$ with set of factors $\mathcal{F}(\boldsymbol{w})$ closed under reversal, the following conditions are equivalent:

(I) $\boldsymbol{w}$ is rich;

(II) $P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1) = C_{\boldsymbol{w}}(n+1) - C_{\boldsymbol{w}}(n) + 2$ for all $n \in \mathbb{N}$.

Complementation-symmetric Rote sequences:

- Infinite words over $\{a, b\}$ with factors closed under both complementation and reversal, and such that $C(n) = 2n$ for all n.

A Connection Between Palindromic & Factor Complexity

Theorem (Bucci-De Luca-G.-Zamboni 2008)

For any infinite word $\boldsymbol{w}$ with set of factors $\mathcal{F}(\boldsymbol{w})$ closed under reversal, the following conditions are equivalent:

(I) $\boldsymbol{w}$ is rich;

(II) $P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1) = C_{\boldsymbol{w}}(n+1) - C_{\boldsymbol{w}}(n) + 2$ for all $n \in \mathbb{N}$.

Complementation-symmetric Rote sequences:

- Infinite words over $\{a, b\}$ with factors closed under both complementation and reversal, and such that $C(n) = 2n$ for all n.
- Allouche-Baake-Cassaigne-Damanik (2003): $P(n) = 2$ for all n.

A Connection Between Palindromic & Factor Complexity

Theorem (Bucci-De Luca-G.-Zamboni 2008)

For any infinite word $\boldsymbol{w}$ with set of factors $\mathcal{F}(\boldsymbol{w})$ closed under reversal, the following conditions are equivalent:

(I) $\boldsymbol{w}$ is rich;

(II) $P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1) = C_{\boldsymbol{w}}(n+1) - C_{\boldsymbol{w}}(n) + 2$ for all $n \in \mathbb{N}$.

Complementation-symmetric Rote sequences:

- Infinite words over $\{a, b\}$ with factors closed under both complementation and reversal, and such that $C(n) = 2n$ for all n.
- Allouche-Baake-Cassaigne-Damanik (2003): $P(n) = 2$ for all n.
- Hence $P(n) + P(n+1) = 4 = C(n+1) - C(n) + 2$ for all $n \Rightarrow$ RICH.

A Connection Between Palindromic & Factor Complexity

Theorem (Bucci-De Luca-G.-Zamboni 2008)

For any infinite word $\mathbf{w}$ with set of factors $\mathcal{F}(\mathbf{w})$ closed under reversal, the following conditions are equivalent:

(I) $\mathbf{w}$ is rich;

(II) $P_{\mathbf{w}}(n) + P_{\mathbf{w}}(n+1) = C_{\mathbf{w}}(n+1) - C_{\mathbf{w}}(n) + 2$ for all $n \in \mathbb{N}$.

Sturmian words:

- Morse-Hedlund (1940): $C(n) = n + 1$ for all n

A Connection Between Palindromic & Factor Complexity

Theorem (Bucci-De Luca-G.-Zamboni 2008)

For any infinite word $\mathbf{w}$ with set of factors $\mathcal{F}(\mathbf{w})$ closed under reversal, the following conditions are equivalent:

(I) $\mathbf{w}$ is rich;

(II) $P_{\mathbf{w}}(n) + P_{\mathbf{w}}(n+1) = C_{\mathbf{w}}(n+1) - C_{\mathbf{w}}(n) + 2$ for all $n \in \mathbb{N}$.

Sturmian words:

- Morse-Hedlund (1940): $C(n) = n + 1$ for all n
- Droubay-Pirillo (1999): $P(n) = 1$ for n even, $P(n) = 2$ for n odd

A Connection Between Palindromic & Factor Complexity

Theorem (Bucci-De Luca-G.-Zamboni 2008)

For any infinite word $\boldsymbol{w}$ with set of factors $\mathcal{F}(\boldsymbol{w})$ closed under reversal, the following conditions are equivalent:

(I) $\boldsymbol{w}$ is rich;

(II) $P_{\boldsymbol{w}}(n) + P_{\boldsymbol{w}}(n+1) = C_{\boldsymbol{w}}(n+1) - C_{\boldsymbol{w}}(n) + 2$ for all $n \in \mathbb{N}$.

Sturmian words:

- Morse-Hedlund (1940): $C(n) = n + 1$ for all n
- Droubay-Pirillo (1999): $P(n) = 1$ for n even, $P(n) = 2$ for n odd
- Hence $P(n) + P(n+1) = 3 = C(n+1) - C(n) + 2$ for all $n \Rightarrow$ RICH.

Finite Case

Using completely different methods . . .

Theorem (de Luca-G.-Zamboni)

For any finite word w, the following two conditions are equivalent:

i) w is a rich palindrome;

ii) $P_w(n) + P_w(n+1) = C_w(n+1) - C_w(n) + 2$ for all n, $0 \leq n \leq |w|$.

More Stuff on Rich Words

G.-Justin-Widmer-Zamboni, Palindromic richness, 2008

- *almost rich words*: a new palindrome is introduced at all, but a finite number of positions

More Stuff on Rich Words

G.-Justin-Widmer-Zamboni, Palindromic richness, 2008

- *almost rich words*: a new palindrome is introduced at all, but a finite number of positions

  Example: $(pq)^{\omega} = pqpqpq\cdots$ where p, q are palindromes

More Stuff on Rich Words

G.-Justin-Widmer-Zamboni, Palindromic richness, 2008

- *almost rich words*: a new palindrome is introduced at all, but a finite number of positions

  Example: $(pq)^\omega = pqpqpq\cdots$ where p, q are palindromes

- *weakly rich words*: all complete returns to letters are palindromes

More Stuff on Rich Words

G.-Justin-Widmer-Zamboni, Palindromic richness, 2008

- *almost rich words*: a new palindrome is introduced at all, but a finite number of positions

  Example: $(pq)^\omega = pqpqpq\cdots$ where p, q are palindromes

- *weakly rich words*: all complete returns to letters are palindromes

  Example: $(aacbccbcacbc)^\omega = aacbccbcacbc\,aacbccbcacbc\cdots$

More Stuff on Rich Words

G.-Justin-Widmer-Zamboni, Palindromic richness, 2008

- *almost rich words*: a new palindrome is introduced at all, but a finite number of positions

  Example: $(pq)^{\omega} = pqpqpq \cdots$ where p, q are palindromes
- *weakly rich words*: all complete returns to letters are palindromes

  Example: $(aacbccbcacbc)^{\omega} = aacbccbcacbc\,aacbccbcacbc \cdots$
- action of morphisms on (almost) rich words

More Stuff on Rich Words

G.-Justin-Widmer-Zamboni, Palindromic richness, 2008

- *almost rich words*: a new palindrome is introduced at all, but a finite number of positions

  Example: $(pq)^\omega = pqpqpq\cdots$ where p, q are palindromes

- *weakly rich words*: all complete returns to letters are palindromes

  Example: $(aacbccbcacbc)^\omega = aacbccbcacbc\,aacbccbcacbc\cdots$

- action of morphisms on (almost) rich words

- morphisms that preserve (almost) richness

More Stuff on Rich Words

G.-Justin-Widmer-Zamboni, Palindromic richness, 2008

- *almost rich words*: a new palindrome is introduced at all, but a finite number of positions

  Example: $(pq)^{\omega} = pqpqpq\cdots$ where p, q are palindromes

- *weakly rich words*: all complete returns to letters are palindromes

  Example: $(aacbccbcacbc)^{\omega} = aacbccbcacbc\,aacbccbcacbc\cdots$

- action of morphisms on (almost) rich words
- morphisms that preserve (almost) richness

Open Problems

- Characterisation of morphisms that preserve (almost) richness
- Enumeration of rich words

Dammit, I'm mad!

U R 2 R U?

* Both phrases are (rich) palindromes! *