A Characterization of Bispecial Sturmian Words

Gabriele Fici

ICTCS 2012 – Varese, 19 September 2012

Sturmian Words

A word $w = a_1 a_2 \cdots a_n$ is a finite sequence of symbols from a finite set Σ, called the alphabet. Its length n is denoted by $|w|$.

Sturmian Words

A word $w = a_1 a_2 \cdots a_n$ is a finite sequence of symbols from a finite set Σ, called the alphabet. Its length n is denoted by $|w|$.

Let $w = a_1 a_2 \cdots a_n$ be a word. A factor of w is a subword $v = a_i \cdots a_j$, $1 \leq i \leq j \leq n$.

Sturmian Words

A word $w = a_1 a_2 \cdots a_n$ is a finite sequence of symbols from a finite set Σ, called the alphabet. Its length n is denoted by $|w|$.

Let $w = a_1 a_2 \cdots a_n$ be a word. A factor of w is a subword $v = a_i \cdots a_j$, $1 \leq i \leq j \leq n$.

A word w is balanced if for any $u, v \in \mathrm{Fact}(w)$ and for every $a \in \Sigma$, any two factors of the same length contain the same number of a's up to 1:

$$|u| = |v| \Rightarrow ||u|_a - |v|_a| \leq 1$$

Example

aaabaa and *baaaab* are balanced, *aabb* and *aaabab* are not.

Sturmian Words

A word $w = a_1 a_2 \cdots a_n$ is a finite sequence of symbols from a finite set Σ, called the alphabet. Its length n is denoted by $|w|$.

Let $w = a_1 a_2 \cdots a_n$ be a word. A factor of w is a subword $v = a_i \cdots a_j$, $1 \leq i \leq j \leq n$.

A word w is balanced if for any $u, v \in \mathrm{Fact}(w)$ and for every $a \in \Sigma$, any two factors of the same length contain the same number of a's up to 1:

$$|u| = |v| \Rightarrow ||u|_a - |v|_a| \leq 1$$

Example

aaabaa and *baaaab* are balanced, *aabb* and *aaabab* are not.

A binary balanced word is called a Sturmian word. The set of Sturmian words is noted St.

Special Sturmian Words

A Sturmian word w is:

- right special (RS) if $wa, wb \in St$
- left special (LS) if $aw, bw \in St$
- bispecial (BS) if w is both right and left special

Special Sturmian Words

A Sturmian word w is:

- right special (RS) if $wa, wb \in St$
- left special (LS) if $aw, bw \in St$
- bispecial (BS) if w is both right and left special

Moreover, a bispecial Sturmian word w is

- strictly bispecial (SBS) if $awa, awb, bwa, bwb \in St$
- non-strictly bispecial (NBS), otherwise.

Special Sturmian Words

A Sturmian word w is:

- right special (RS) if $wa, wb \in St$
- left special (LS) if $aw, bw \in St$
- bispecial (BS) if w is both right and left special

Moreover, a bispecial Sturmian word w is

- strictly bispecial (SBS) if $awa, awb, bwa, bwb \in St$
- non-strictly bispecial (NBS), otherwise.

Example

aba is strictly bispecial,
ab is non-strictly bispecial,
aab is left special but not right special,
baab is neither left special nor right special.

Special Sturmian Words

Proposition

Let w be a Sturmian word. Then:

- $|\Sigma w \Sigma \cap St| = 4$ *if and only if w is strictly bispecial;*
- $|\Sigma w \Sigma \cap St| = 3$ *if and only if w is non-strictly bispecial;*
- $|\Sigma w \Sigma \cap St| = 2$ *if and only if w is left special or right special but not bispecial;*
- $|\Sigma w \Sigma \cap St| = 1$ *if and only if w is neither left special nor right special;*

Example

aba is strictly bispecial,
ab is non-strictly bispecial,
aab is left special but not right special,
baab is neither left special nor right special.

Special Sturmian Words

Proposition

Let w be a Sturmian word. Then:

- $|\Sigma w\Sigma \cap St| = 4$ *if and only if w is strictly bispecial;*
- $|\Sigma w\Sigma \cap St| = 3$ *if and only if w is non-strictly bispecial;*
- $|\Sigma w\Sigma \cap St| = 2$ *if and only if w is left special or right special but not bispecial;*
- $|\Sigma w\Sigma \cap St| = 1$ *if and only if w is neither left special nor right special;*

Proposition

A bispecial Sturmian word is strictly bispecial if and only if it is a palindrome.

What are the best grid approximations of a segment with integer coordinates in the Euclidean plane?

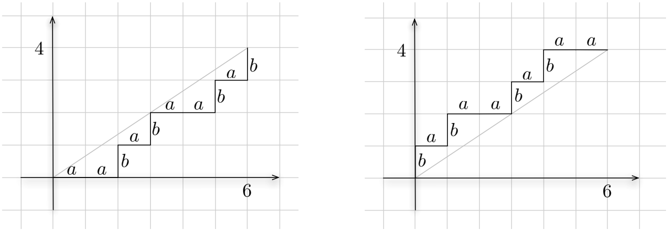
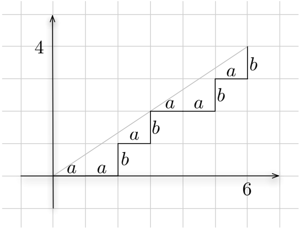


Figure: The lower Christoffel word $w_{6,4} = aababaabab$ (left) and the upper Christoffel word $w'_{6,4} = babaababaa$ (right).

Christoffel Words

What are the best grid approximations of a segment with integer coordinates in the Euclidean plane?



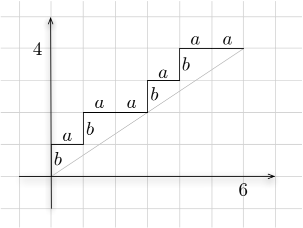
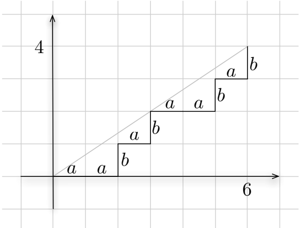


Figure: The lower Christoffel word $w_{6,4} = aababaabab$ (left) and the upper Christoffel word $w'_{6,4} = babaababaa$ (right).

$$w_{p,q}[i] = \begin{cases} a & \text{if } iq \bmod (p+q) > (i-1)q \bmod (p+q), \\ b & \text{if } iq \bmod (p+q) < (i-1)q \bmod (p+q). \end{cases}$$

Christoffel Words

What are the best grid approximations of a segment with integer coordinates in the Euclidean plane?



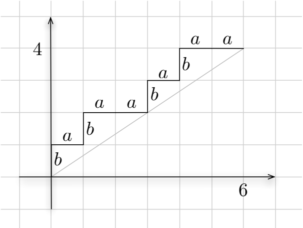


Figure: The lower Christoffel word $w_{6,4} = aababaabab$ (left) and the upper Christoffel word $w'_{6,4} = babaababaa$ (right).

$$\{i4 \bmod(10) \mid i = 0, 1, \ldots, 10\} = \{0, 4, 8, 2, 6, 0, 4, 8, 2, 6, 0\}$$

Christoffel Words

What are the best grid approximations of a segment with integer coordinates in the Euclidean plane?

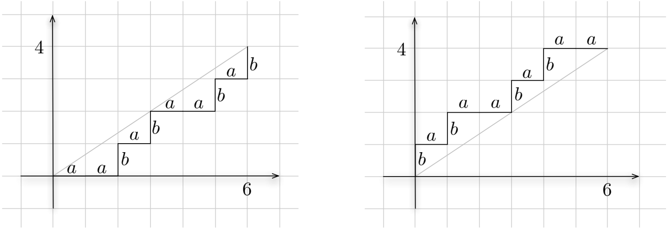


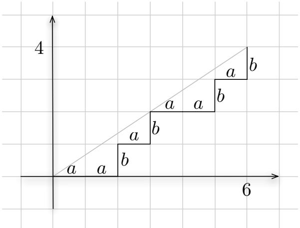
Figure: The lower Christoffel word $w_{6,4} = aababaabab$ (left) and the upper Christoffel word $w'_{6,4} = babaababaa$ (right).

Remark

$w'_{p,q}$ is the reversal of $w_{p,q}$.

Christoffel Words

What are the best grid approximations of a segment with integer coordinates in the Euclidean plane?



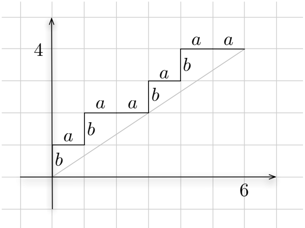


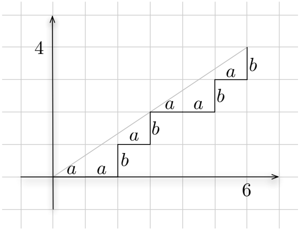
Figure: The lower Christoffel word $w_{6,4} = aababaabab$ (left) and the upper Christoffel word $w'_{6,4} = babaababaa$ (right).

Remark

If $\frac{p}{q} = r\frac{p'}{q'}$*, then* $w_{p,q} = (w_{p',q'})^r$.

Christoffel Words

What are the best grid approximations of a segment with integer coordinates in the Euclidean plane?



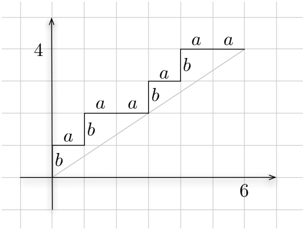


Figure: The lower Christoffel word $w_{6,4} = aababaabab$ (left) and the upper Christoffel word $w'_{6,4} = babaababaa$ (right).

Remark

A Christoffel word is primitive if and only if $\gcd(p, q) = 1$.

Main result

The maximal internal factor of a word $w = a_1 a_2 \cdots a_n$, $n \geq 2$, is the factor $a_2 a_3 \cdots a_{n-1}$.

$BS =$ bispecial Sturmian words = $SBS \cup NBS$

$SBS =$ strictly bispecial Sturmian words
$NBS =$ non-strictly bispecial Sturmian words

Main result

The maximal internal factor of a word $w = a_1 a_2 \cdots a_n$, $n \geq 2$, is the factor $a_2 a_3 \cdots a_{n-1}$.

$BS =$ bispecial Sturmian words = $SBS \cup NBS$

$SBS =$ strictly bispecial Sturmian words
$NBS =$ non-strictly bispecial Sturmian words

Theorem (Berstel, de Luca, 1997)

SBS is the set of maximal internal factors of primitive Christoffel words.

Main result

The maximal internal factor of a word $w = a_1 a_2 \cdots a_n$, $n \geq 2$, is the factor $a_2 a_3 \cdots a_{n-1}$.

$BS =$ bispecial Sturmian words = $SBS \cup NBS$

$SBS =$ strictly bispecial Sturmian words
$NBS =$ non-strictly bispecial Sturmian words

Theorem (Berstel, de Luca, 1997)

SBS is the set of maximal internal factors of primitive Christoffel words.

Theorem (F., 2012)

BS is the set of maximal internal factors of Christoffel words.

Example

pair (p, q)	lower Christoffel word $w_{p,q}$	upper Christoffel word $w'_{p,q}$
$(11, 1)$	aaaaaaaaaaab	baaaaaaaaaaa
$(10, 2)$	aaaaabaaaaab	baaaaabaaaaa
$(9, 3)$	aaabaaabaaab	baaabaaabaaa
$(8, 4)$	aabaabaabaab	baabaabaabaa
$(7, 5)$	aababaababab	babababaabaa
$(6, 6)$	abababababab	babababababa
$(5, 7)$	abababbababb	bbababbababa
$(4, 8)$	abbabbabbabb	bbabbabbabba
$(3, 9)$	abbbabbbabbb	bbbabbbabbba
$(2, 10)$	abbbbbabbbbb	bbbbbabbbbba
$(1, 11)$	abbbbbbbbbbb	bbbbbbbbbbba

Table: The Christoffel words of length 12. Their maximal internal factors are the bispecial Sturmian words of length 10. Strictly bispecial Sturmian words (the palindromes) are underlined.

Enumerative formula

Let ϕ be the Euler totient function.
The number of strictly bispecial Sturmian words is known to be

$$SBS(n) = \phi(n+2)$$

Enumerative formula

Let ϕ be the Euler totient function.
The number of strictly bispecial Sturmian words is known to be

$$SBS(n) = \phi(n+2)$$

By the main theorem we have:

$$NBS(n) = 2\,(n+1-\phi(n+2))$$

and so

$$BS(n) = 2\,(n+1) - \phi(n+2)$$

Minimal Forbidden Words

Let $L \subseteq \Sigma^*$ be a factorial ($uv \in L \Rightarrow u, v \in L$) language.

Minimal Forbidden Words

Let $L \subseteq \Sigma^*$ be a factorial ($uv \in L \Rightarrow u, v \in L$) language.

Definition

A word $w \in \Sigma^*$ is a minimal forbidden word for $L \subseteq \Sigma^*$ if $w \notin L$ but every proper factor of w belongs to L.

Example

Ex. $L = \text{Fact}(aba, aab)$. Its set of m.f.w. is $\{bb, aaa, bab, baa, aaba\}$.

Minimal Forbidden Words

Let $L \subseteq \Sigma^*$ be a factorial ($uv \in L \Rightarrow u, v \in L$) language.

Definition

A word $w \in \Sigma^*$ is a minimal forbidden word for $L \subseteq \Sigma^*$ if $w \notin L$ but every proper factor of w belongs to L.

Example

Ex. $L = \text{Fact}(aba, aab)$. Its set of m.f.w. is $\{bb, aaa, bab, baa, aaba\}$.

MFSt = set of m.f.w. for the language *St* of Sturmian words.

Minimal Forbidden Words

Let $L \subseteq \Sigma^*$ be a factorial ($uv \in L \Rightarrow u, v \in L$) language.

Definition

A word $w \in \Sigma^*$ is a minimal forbidden word for $L \subseteq \Sigma^*$ if $w \notin L$ but every proper factor of w belongs to L.

Example

Ex. $L = \text{Fact}(aba, aab)$. Its set of m.f.w. is $\{bb, aaa, bab, baa, aaba\}$.

MFSt = set of m.f.w. for the language *St* of Sturmian words.

Theorem (F., 2012)

$MFSt = \{ywx \mid xwy \textit{ is a non-primitive Christoffel word, } x, y \in \Sigma\}$.

Minimal Forbidden Words

Corollary

For every $n > 1$, one has $MFSt(n) = 2(n - 1 - \phi(n))$.

Minimal Forbidden Words

Corollary

For every $n > 1$, one has $MFSt(n) = 2(n - 1 - \phi(n))$.

It is known that $St(n) = O(n^3)$, as a consequence of the estimation

$$\sum_{i=1}^{n} \phi(i) = \frac{3n^2}{\pi^2} + O(n \log n) \qquad (1)$$

Minimal Forbidden Words

Corollary

For every $n > 1$, one has $MFSt(n) = 2(n - 1 - \phi(n))$.

It is known that $St(n) = O(n^3)$, as a consequence of the estimation

$$\sum_{i=1}^{n} \phi(i) = \frac{3n^2}{\pi^2} + O(n \log n) \qquad (1)$$

From (**??**) and from the formula of the Corollary above, we have that

$$\sum_{i=1}^{n} MFSt(n) = O(n^2)$$