A **few** words about **skew** and **episkew** words

Amy Glen

Laboratoire de Combinatoire et d'Informatique Mathématique (LaCIM)
Université du Québec à Montréal

amy.glen@gmail.com
http://www.lacim.uqam.ca/∼glen

Recent Progress in Combinatorics on Words – March 2007

Outline

Recall Sturmian words . . .

Definition

An infinite word **s** over $\{a, b\}$ is *Sturmian* if there exist real numbers α, $\rho \in [0, 1]$ such that **s** is equal to one of the following two infinite words:

$$s_{\alpha,\rho},\ s'_{\alpha,\rho} : \mathbb{N} \to \{a, b\}$$

defined by

$$s_{\alpha,\rho}(n) = \begin{cases} a & \text{if } \lfloor (n+1)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor = 0, \\ b & \text{otherwise;} \end{cases}$$

$$(n \geq 0)$$

$$s'_{\alpha,\rho}(n) = \begin{cases} a & \text{if } \lceil (n+1)\alpha + \rho \rceil - \lceil n\alpha + \rho \rceil = 0, \\ b & \text{otherwise.} \end{cases}$$

Sturmian words (cont.)

A *Sturmian word* is:

- *aperiodic* if α is irrational;
- *periodic* if α is rational;
- *standard* (or *characteristic*) if $\rho = \alpha$.

- Nowadays: only the aperiodic ones are considered to be 'Sturmian'.
- Here: *Sturmian* refers to both aperiodic and periodic Sturmian words.

Balanced words

Definition (Balance)

A finite or infinite word w on $\{a, b\}$ is *balanced* if:

$$u,\ v \in F(w),\ |u| = |v| \quad \Rightarrow \quad ||u|_b - |v|_b| \leq 1.$$

Morse & Hedlund (1940)

All balanced infinite words over a 2-letter alphabet are called *Sturmian trajectories*. They belong to three classes:

- aperiodic Sturmian;
- periodic Sturmian;
- ultimately periodic non-recurrent infinite words, called *skew words*: $\varphi(x)^p\varphi(y)\varphi(x)^\omega$, φ a pure standard (Sturmian) morphism, $x, y \in \{a, b\}$, $x \neq y$.
  - Example: $aaabaaaaaaaa\cdots$

Extremal words

- Let $\mathbf{x} = x_0x_1x_2\cdots$ be a (right) infinite word.

Define $\min(\mathbf{x})$ to be the infinite word such that any prefix of $\min(\mathbf{x})$ is the *lexicographically smallest* amongst the factors of $\mathbf{x}$ of the same length.

Similarly define $\max(\mathbf{x})$.

- Let $\min(\mathbf{x}|k)$ denote the lexicographically smallest factor of $\mathbf{x}$ of length k. Then:

$$\min(\mathbf{x}) = \lim_{k\to\infty} \min(\mathbf{x}|k).$$

- Note: $\min(\mathbf{x})$ is an *infinite Lyndon word*.

Proposition (Pirillo, 2005)

An infinite word $\mathbf{s}$ *on* $\{a, b\}$ *(*$a < b$*) is standard Sturmian (aperiodic or periodic)* $\iff$

$$a\mathbf{s} \leq \min(\mathbf{s}) \leq \max(\mathbf{s}) \leq b\mathbf{s}.$$

- That is, standard Sturmian words $\mathbf{s}$ on $\{a, b\}$ are characterized by the inequality:

$$a\mathbf{s} \leq T^k(\mathbf{s}) \leq b\mathbf{s}, \quad \text{for all } k \geq 0,$$

  where T is the *shift map*.
- In particular, an infinite word $\mathbf{s}$ on $\{a, b\}$ ($a < b$) is an *aperiodic standard Sturmian word* $\iff$

$$(\min(\mathbf{s}), \max(\mathbf{s})) = (a\mathbf{s}, b\mathbf{s}).$$

Fine words

Question

What are the infinite words $\mathbf{t}$ on $\{a, b\}$ satisfying

$$(\min(\mathbf{t}), \max(\mathbf{t})) = (a\mathbf{s}, b\mathbf{s}) \quad \text{for some infinite word } \mathbf{s} \text{ ?}$$

Definition (Pirillo, 2005)

An infinite word $\mathbf{t}$ on $\{a, b\}$ ($a < b$) is said to be *fine* if

$$(\min(\mathbf{t}), \max(\mathbf{t})) = (a\mathbf{s}, b\mathbf{s}) \quad \text{for some infinite word } \mathbf{s}.$$

- Fine words on $\{a, b\}$ are exactly the aperiodic Sturmian and skew infinite words.
- Recently generalized to an arbitrary finite alphabet ...

Generalized fine words

Definition

An *acceptable pair* is a pair $(a, <)$ where a is a letter and $<$ is a order on $\mathcal{A}$ such that $a = \min(\mathcal{A})$.

Definition (Glen, 2006)

An infinite word $\mathbf{t}$ on $\mathcal{A}$ is said to be *fine* if there exists an infinite word $\mathbf{s}$ such that $\min(\mathbf{t}) = a\mathbf{s}$ for any acceptable pair $(a, <)$.

Proposition (Glen, 2006)

An infinite word $\mathbf{t}$ *is fine* $\iff$ $\mathbf{t}$ *is either a strict episturmian word, or a "strict episkew word" (i.e., a certain kind of non-recurrent infinite word, all of whose factors are finite episturmian).*

Episturmian words

- Introduced by X. Droubay, J. Justin, and G. Pirillo (2001).
- Interesting natural generalization Sturmian words.
- Share many properties with Sturmian words.
- Include the well-known *Arnoux-Rauzy sequences*.

Episturmian words (cont.)

Definition

A (right-, left-, bi-) infinite word **t** on $\mathcal{A}$ is *episturmian* if:

- $F(\mathbf{t})$ (its set of factors) is *closed under reversal*, and
- **t** has at most one *right special factor* of each length.

t is *standard* if all of its left special factors are prefixes of it.

- Gives Sturmian words (both aperiodic and periodic) when $|\mathcal{A}| = 2$.
- Episturmian words are *uniformly recurrent*.

Episturmian morphisms

- For each $a \in \mathcal{A}$, define the morphisms ψ_a, $\overline{\psi}_a$ on $\mathcal{A}$ by

$$\psi_a : \begin{cases} a & \mapsto & a \\ x & \mapsto & ax \end{cases}, \quad \overline{\psi}_a : \begin{cases} a & \mapsto & a \\ x & \mapsto & xa \end{cases} \quad \text{for all } x \in \mathcal{A} \setminus \{a\}.$$

- *Monoid of episturmian morphisms*: generated by all the ψ_a, $\overline{\psi}_a$, and permutations on $\mathcal{A}$.
- *Monoid of epistandard morphisms*: generated by all the ψ_a and permutations on $\mathcal{A}$.
- Submonoids:
  - Pure episturmian morphisms: generated by the ψ_a, $\overline{\psi}_a$;
  - Pure epistandard morphisms: generated by the ψ_a only.
- Pure epistandard morphisms are precisely the *pure standard (Sturmian) morphisms* when $|\mathcal{A}| = 2$.

Characterization by morphisms

Proposition (Justin & Pirillo, 2002)

An infinite word $\mathbf{t}$ on $\mathcal{A}$ is standard episturmian $\iff$ there exists an infinite sequence $(\mathbf{t}^{(i)})_{i\geq 0}$ of recurrent infinite words and a *directive word* $\Delta = x_1x_2x_3\cdots$ ($x_i \in \mathcal{A}$) such that

$$\mathbf{t}^{(0)} = \mathbf{t} \quad \text{and} \quad \mathbf{t}^{(i-1)} = \psi_{x_i}(\mathbf{t}^{(i)}) \quad \text{for all } i > 0.$$

Moreover, each $\mathbf{t}^{(i)}$ is standard episturmian with directive word $T^i(\Delta) = x_{i+1}x_{i+2}x_{i+3}\cdots$.

Example: Fibonacci word, $\Delta = (ab)^\omega$

- $\mathbf{f} = abaababaabaaba\cdots$
- $\mathbf{f} = \varphi^\omega(a)$ where $\varphi : a \mapsto ab, b \mapsto a$. Note that

$$\varphi = \psi_a E = E\psi_b$$

  where E is the *exchange morphism*: $a \leftrightarrow b$
- For all $n \geq 0$,

$$\mathbf{f} = (\psi_a E)^{2n}(\mathbf{f}) = (\psi_a E \psi_a E)^n(\mathbf{f}) = (\psi_a\psi_b)^n(\mathbf{f})$$

- $\mathbf{f} = \psi_a(\mathbf{f}^{(1)})$ where

$$\mathbf{f}^{(1)} = E(\mathbf{f}), \text{ the Fibonacci word directed by } T(\Delta) = (ba)^\omega$$

- $\mathbf{f}^{(1)} = \psi_b(\mathbf{f}^{(2)})$ where $\mathbf{f}^{(2)} = \mathbf{f}$, directed by $T^2(\Delta) = (ab)^\omega$, etc. ...

Strictness

Definition

A standard episturmian word **t**, or any equivalent (episturmian) word, is *strict* if

$$\mathrm{Ult}(\Delta) = \mathrm{Alph}(\Delta).$$

That is, every letter in Alph(**t**) appears infinitely often in Δ.

Examples

1. $\Delta = c(ab)^{\omega}$ directs the *non-strict* standard episturmian word :

$$\psi_c(\mathbf{f}) = cacbcacacbcacbcacacbcacacbca\cdots$$

2. $\Delta = (abc)^{\omega}$ directs the *Tribonacci word* :

$$\mathbf{r} = abacabaabacababacabaabacabacabaabaca\cdots$$

*Also known as the *Rauzy word* (1982).

Arnoux-Rauzy sequences

- The strict episturmian words are exactly the well-known *Arnoux-Rauzy sequences*.
- The family of episturmian words over $\{a, b, c\}$ consists of:
  - Arnoux-Rauzy sequences over $\{a, b, c\}$;
  - Sturmian words over $\{a, b\}$, $\{b, c\}$, $\{a, c\}$ & certain morphic images of them;
  - periodic infinite words of the form $\mu(x)^\omega$ where μ is an episturmian morphism on $\{a, b, c\}$ and $x \in \{a, b, c\}$.

Recall: An infinite word $\mathbf{t}$ on $\mathcal{A}$ is *fine* if there exists an infinite word $\mathbf{s}$ such that $\min(\mathbf{t}) = a\mathbf{s}$ for any acceptable pair $(a, <)$.

Notation

v_p: prefix of length p of a given finite or infinite word v

Proposition (Glen, 2006)

An infinite word $\mathbf{t}$ *with* $Alph(\mathbf{t}) = \mathcal{A}$ *is fine* $\iff$ *one of the following holds:*

- *(i)* $\mathbf{t}$ *is a strict episturmian word;*
- *(ii) there exists a letter* $x \in \mathcal{A}$ *and a strict standard episturmian word* $\mathbf{s}$ *on* $\mathcal{A} \setminus \{x\}$ *such that* $\mathbf{t} = v\varphi(\mathbf{s})$*, where* φ *is a pure epistandard morphism on* $\mathcal{A}$ *and* v *is a non-empty suffix of* $\varphi(\widetilde{\mathbf{s}_p}x)$ *for some* $p \in \mathbb{N}$*.*

- An infinite word $\mathbf{t}$ of the form (ii) is called a *strict episkew word*.
- All factors of such infinite words are *finite episturmian*.

Example

Suppose **f** is the Fibonacci word on $\{a, b\}$. Then, the following infinite words are fine on $\{a, b, c\}$.

- $\mathbf{f} = abaababaabaaba\cdots$
- $c\mathbf{f} = cabaababaabaaba\cdots$
- $\widetilde{\mathbf{f}_4}c\mathbf{f} = aabacabaababaabaaba\cdots$
- $\psi_a(\mathbf{f}) = aabaaabaabaaabaaaba\cdots$
- $\psi_c(c\mathbf{f}) = ccacbcacacbcacbcacacbcacacbca\cdots$
- $\psi_c(\widetilde{\mathbf{f}_4}c\mathbf{f}) = cacacbcaccacbcacacbcacbcacacbcaca\cdots$

Note

$\psi_c(\mathbf{f})$ is **not** fine since it is a *non-strict* standard episturmian word with directive word $\Delta = c(ab)^\omega$.

Equivalent definitions

Theorem

An infinite word $\mathbf{t}$ *with* $Alph(\mathbf{t}) = \mathcal{A}$ *is episkew if equivalently:*

(i) $\mathbf{t}$ *is non-recurrent and all of its factors are (finite) episturmian;*

(ii) there exists an infinite sequence $(\mathbf{t}^{(i)})_{i\geq 0}$ *of non-recurrent infinite words and a directive word* $x_1x_2x_3\cdots$ *(*$x_i \in \mathcal{A}$*) such that* $\mathbf{t}^{(0)} = \mathbf{t}$, $\ldots$, $\mathbf{t}'^{(i-1)} = \psi_{x_i}(\mathbf{t}^{(i)})$*, where* $\mathbf{t}'^{(i-1)} = \mathbf{t}^{(i-1)}$ *if* $\mathbf{t}^{(i-1)}$ *begins with* x_i *and* $\mathbf{t}'^{(i-1)} = x_i\mathbf{t}^{(i-1)}$ *otherwise;*

(iii) there exists a letter $x \in \mathcal{A}$ *and a standard episturmian word* $\mathbf{s}$ *on* $\mathcal{A}\setminus\{x\}$ *such that* $\mathbf{t} = v\varphi(\mathbf{s})$*, where* φ *is a pure epistandard morphism on* $\mathcal{A}$ *and* v *is a non-empty suffix of* $\varphi(\widetilde{\mathbf{s}_p}x)$ *for some* $p \in \mathbb{N}$.

Moreover, $\mathbf{t}$ *is said to be strict episkew if* $\mathbf{s}$ *is strict on* $\mathcal{A}\setminus\{x\}$.

Infinite words with episturmian factors

Recall: Balanced infinite words are precisely the infinite words whose factors are finite Sturmian.

Fact 1

{Infinite words whose factors are Sturmian} $=$

{(recurrent) Sturmian} $\cup$ {(non-recurrent) skew}

Fact 2

{Infinite words whose factors are episturmian} $=$

{(recurrent) episturmian} $\cup$ {(non-recurrent) episkew}

Terminology

Notation

- Let w be a finite or infinite word on $\mathcal{A}$.
- $\min(w|k)$ denotes the lexicographically smallest factor of w of length k for the given order (where $|w| \geq k$ for w finite).

Definition

- For a finite word $w \in \mathcal{A}^+$ and a given order, $\min(w)$ will denote $\min(w|k)$ where k is maximal such that all $\min(w|j)$, $j = 1, 2, \ldots, k$, are prefixes of $\min(w|k)$.
- In the case $\mathcal{A} = \{a, b\}$, $\max(w)$ is defined similarly.

Example

Suppose $w = baababacababac$.

For the orders $b < a < c$ and $b < c < a$ on $\{a, b, c\}$:

$$
\begin{aligned}
\min(w|1) &= b \\
\min(w|2) &= ba \\
\min(w|3) &= bab \\
\min(w|4) &= baba \\
\min(w|5) &= babac = \min(w)
\end{aligned}
$$

Note: $\min(w)$ is a suffix of w, which is true in general.

Characterizations

Theorem

A finite word w on $\mathcal{A}$ is episturmian $\iff$ there exists a finite word u such that, for any acceptable pair $(a, <)$, we have

$$au_{|m|-1} \leq m \qquad (1)$$

where $m = \min(w)$ for the considered order.

Recall $w = baabacababac$.

- For the different orders on $\{a, b, c\}$:
  - $a < b < c$ or $a < c < b$: $\min(w) = aabacababac$;
  - $b < a < c$ or $b < c < a$: $\min(w) = babac$;
  - $c < a < b$ or $c < b < a$: $\min(w) = cababac$.
- $u = abacaaaaaa$ satisfies (1) $\Rightarrow$ w is finite episturmian.

Characterizations (cont.)

Corollary

A finite word w on $\{a, b\}$, $a < b$, is not Sturmian (i.e., not balanced) $\iff$ there exists a finite word u such that

$$aua \text{ is a prefix of } \min(w) \quad \& \quad bub \text{ is a prefix of } \max(w).$$

Example (1)

For $w = aababababaabaab$:

- $\min(w) = aabaab$, $\max(w) = bababaabaab$.
- $\min(w) = auab$ and $\max(w) = bubaabaab$ where $u = aba$.
- Thus w is **not** Sturmian.

Characterizations (cont.)

Corollary

A finite word w on $\{a,b\}$, $a<b$, is not Sturmian (i.e., not balanced) $\iff$ there exists a finite word u such that

$$aua \text{ is a prefix of } \min(w) \quad \& \quad bub \text{ is a prefix of } \max(w).$$

Example (2)

For $w = ababaabaabab$:

- $\min(w) = aabaabab$, $\max(w) = babaabaabab$.
- $abaaba$ is the longest common prefix of $a^{-1}\min(w)$ and $b^{-1}\max(w)$.
- $abaaba$ is followed by b in $\min(w)$ and a in $\max(w)$.
- Thus w is Sturmian.

Characterizations

Definition

An infinite word is said to be *episturmian in the wide sense* if all of its factors are (finite) episturmian.

Corollary (1)

An infinite word $\mathbf{t}$ *on* $\mathcal{A}$ *is episturmian in the wide sense (episturmian or episkew)* $\iff$ *there exists an infinite word* $\mathbf{u}$ *such that*

$$a\mathbf{u} \leq \min(\mathbf{t}) \quad \textit{for any acceptable pair } (a, <).$$

Corollary (2)

An infinite word $\mathbf{t}$ *on* $\{a, b\}$*,* $a < b$*, is balanced (Sturmian or skew)* $\iff$ *there exists an infinite word* $\mathbf{u}$ *such that*

$$a\mathbf{u} \leq \min(\mathbf{t}) \leq \max(\mathbf{t}) \leq b\mathbf{u}.$$