Characterizations of finite and infinite episturmian words via lexicographic orderings

Amy Glen*, Jacques Justin, Giuseppe Pirillo

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

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

The Fields Institute – Monday 12 February, 2007
Outline

1. Introduction

2. Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words

3. Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words

4. Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words

A. Glen*, J. Justin, G. Pirillo
Characterizations of episturmian words
Episturmian words

- An interesting natural generalization of the well-known \textit{Sturmian words}.
- Share many properties with Sturmian words.
- Include the well-known \textit{Arnoux-Rauzy sequences}.
We characterize by *lexicographic order* all:

- *finite* Sturmian and episturmian words;
- episturmian words in a *wide sense* (recurrent, episkew);
- *balanced* infinite words over a 2-letter alphabet.
Outline

1. Introduction
2. Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words
3. Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words
4. Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words

A. Glen*, J. Justin, G. Pirillo
Characterizations of episturmian words
Let $\mathcal{A}$ be a finite alphabet and let $u = x_1x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- **Length:** $|u| = m$
- $|u|_{a}$: number of occurrences of the letter $a$ in $u$
- **Reversal:** $\tilde{u} = x_mx_{m-1} \cdots x_1$
- $u$ is a palindrome if $u = \tilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- $\varepsilon$: the empty word ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all non-empty finite words over $\mathcal{A}$
Let $A$ be a *finite alphabet* and let $u = x_1x_2 \cdots x_m$, each $x_i \in A$.

- **Length**: $|u| = m$
- $|u|_a$: number of occurrences of the letter $a$ in $u$
- **Reversal**: $\tilde{u} = x_mx_{m-1} \cdots x_1$
- $u$ is a *palindrome* if $u = \tilde{u}$
- $A^*$: set of all finite words over $A$
- $\varepsilon$: the *empty word* ($|\varepsilon| = 0$)
- $A^+ = A^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $A$
Let $\mathcal{A}$ be a finite alphabet and let $u = x_1x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- **Length**: $|u| = m$
- $|u|_a$: number of occurrences of the letter $a$ in $u$
- **Reversal**: $\tilde{u} = x_mx_{m-1}\cdots x_1$
- $u$ is a palindrome if $u = \tilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- $\varepsilon$: the empty word ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all non-empty finite words over $\mathcal{A}$
Let $\mathcal{A}$ be a finite alphabet and let $u = x_1 x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- **Length**: $|u| = m$
- $|u|_a$: number of occurrences of the letter $a$ in $u$
- **Reversal**: $\tilde{u} = x_m x_{m-1} \cdots x_1$
- $u$ is a palindrome if $u = \tilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- $\varepsilon$: the empty word ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all non-empty finite words over $\mathcal{A}$
Let $\mathcal{A}$ be a finite alphabet and let $u = x_1x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- **Length**: $|u| = m$
- $|u|_a$: number of occurrences of the letter $a$ in $u$
- **Reversal**: $\tilde{u} = x_mx_{m-1} \cdots x_1$
- $u$ is a *palindrome* if $u = \tilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- $\varepsilon$: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$
Let $\mathcal{A}$ be a *finite alphabet* and let $u = x_1x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- **Length**: $|u| = m$
- $|u|_a$: number of occurrences of the letter $a$ in $u$
- **Reversal**: $\tilde{u} = x_mx_{m-1} \cdots x_1$
- $u$ is a *palindrome* if $u = \tilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- $\varepsilon$: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$
Let $A$ be a finite alphabet and let $u = x_1x_2 \cdots x_m$, each $x_i \in A$.

- **Length**: $|u| = m$
- $|u|_a$: number of occurrences of the letter $a$ in $u$
- **Reversal**: $\tilde{u} = x_mx_{m-1} \cdots x_1$
- $u$ is a palindrome if $u = \tilde{u}$
- $A^*$: set of all finite words over $A$
- $\varepsilon$: the empty word ($|\varepsilon| = 0$)
- $A^+ = A^* \setminus \{\varepsilon\}$: set of all non-empty finite words over $A$
Let $\mathcal{A}$ be a finite alphabet and let $u = x_1 x_2 \cdots x_m$, each $x_i \in \mathcal{A}$.

- **Length**: $|u| = m$
- $|u|_a$: number of occurrences of the letter $a$ in $u$
- **Reversal**: $\tilde{u} = x_m x_{m-1} \cdots x_1$
- $u$ is a *palindrome* if $u = \tilde{u}$
- $\mathcal{A}^*$: set of all finite words over $\mathcal{A}$
- $\varepsilon$: the *empty word* ($|\varepsilon| = 0$)
- $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$: set of all *non-empty* finite words over $\mathcal{A}$
Words (cont.)

Let $x = x_0x_1x_2 \cdots$ be an *infinite word* over $\mathcal{A}$.

- **Factor of $x$**: a finite string of consecutive letters in $x$
- **Prefix of $x$**: factor occurring at the beginning of $x$
- **$F(x)$**: set of all factors of $x$
- **Ult($x$)**: set of letters occurring infinitely often in $x$
- **Alph($x$)** := $F(x) \cap \mathcal{A}$, the *alphabet* of $x$
- $w \in F(x)$ is *recurrent* in $x$ if $w$ occurs infinitely often in $x$
- $x$ is *recurrent* if all of its factors are recurrent in it
Let $x = x_0x_1x_2 \cdots$ be an *infinite word* over $A$.

- **Factor of $x$**: a finite string of consecutive letters in $x$
- **Prefix of $x$**: factor occurring at the beginning of $x$
- **$F(x)$**: set of all factors of $x$
- **Ult($x$)**: set of letters occurring infinitely often in $x$
- **Alph($x$) := $F(x) \cap A$**, the *alphabet* of $x$
- $w \in F(x)$ is *recurrent* in $x$ if $w$ occurs infinitely often in $x$
- $x$ is *recurrent* if all of its factors are recurrent in it
Introduction
Preliminaries
Previous Work
Characterizations via Lexicographic Orderings

Terminology & Notation
Sturmian & Episturmian Words
Episkew Words

Words (cont.)

Let $x = x_0x_1x_2 \cdots$ be an *infinite word* over $A$.

- **Factor of $x$**: a finite string of consecutive letters in $x$
- **Prefix of $x$**: factor occurring at the beginning of $x$
- $F(x)$: set of all factors of $x$
- $Ult(x)$: set of letters occurring infinitely often in $x$
- $Alph(x) := F(x) \cap A$, the *alphabet* of $x$
- $w \in F(x)$ is *recurrent* in $x$ if $w$ occurs infinitely often in $x$
- $x$ is *recurrent* if all of its factors are recurrent in it
Words (cont.)

Let $x = x_0 x_1 x_2 \cdots$ be an infinite word over $A$.

- **Factor of** $x$: a finite string of consecutive letters in $x$
- **Prefix of** $x$: factor occurring at the beginning of $x$
- $F(x)$: set of all factors of $x$
- $\text{Ult}(x)$: set of letters occurring infinitely often in $x$
- $\text{Alph}(x) := F(x) \cap A$, the alphabet of $x$
- $w \in F(x)$ is recurrent in $x$ if $w$ occurs infinitely often in $x$
- $x$ is recurrent if all of its factors are recurrent in it
Let $x = x_0 x_1 x_2 \cdots$ be an **infinite word** over $A$.

- **Factor of $x$**: a finite string of consecutive letters in $x$
- **Prefix of $x$**: factor occurring at the beginning of $x$
- $F(x)$: **set of all factors** of $x$
- $\text{Ult}(x)$: set of letters occurring infinitely often in $x$
- $\text{Alph}(x) := F(x) \cap A$, the **alphabet** of $x$
- $w \in F(x)$ is **recurrent** in $x$ if $w$ occurs infinitely often in $x$
- $x$ is **recurrent** if all of its factors are recurrent in it
Let $x = x_0x_1x_2 \cdots$ be an infinite word over $A$.

- **Factor of $x$**: a finite string of consecutive letters in $x$
- **Prefix of $x$**: factor occurring at the beginning of $x$
- $F(x)$: set of all factors of $x$
- Ult$(x)$: set of letters occurring infinitely often in $x$
- Alph$(x) := F(x) \cap A$, the alphabet of $x$
- $w \in F(x)$ is recurrent in $x$ if $w$ occurs infinitely often in $x$
- $x$ is recurrent if all of its factors are recurrent in it
Let $x = x_0 x_1 x_2 \cdots$ be an infinite word over $\mathcal{A}$.

- **Factor of $x$**: a finite string of consecutive letters in $x$
- **Prefix of $x$**: factor occurring at the beginning of $x$
- **$F(x)$**: set of all factors of $x$
- **Ult($x$)**: set of letters occurring infinitely often in $x$
- **Alph($x$) := $F(x) \cap \mathcal{A}$**, the alphabet of $x$
- $w \in F(x)$ is recurrent in $x$ if $w$ occurs infinitely often in $x$
- $x$ is recurrent if all of its factors are recurrent in it
Let $\mathbf{x} = x_0x_1x_2 \cdots$ be an *infinite word* over $\mathcal{A}$.

- **Factor of $\mathbf{x}$**: a finite string of consecutive letters in $\mathbf{x}$
- **Prefix of $\mathbf{x}$**: factor occurring at the beginning of $\mathbf{x}$
- **$F(\mathbf{x})$**: set of all factors of $\mathbf{x}$
- **Ult($\mathbf{x}$)**: set of letters occurring infinitely often in $\mathbf{x}$
- **Alph($\mathbf{x}$) := $F(\mathbf{x}) \cap \mathcal{A}$**, the *alphabet* of $\mathbf{x}$
- $w \in F(\mathbf{x})$ is **recurrent** in $\mathbf{x}$ if $w$ occurs infinitely often in $\mathbf{x}$
- $\mathbf{x}$ is **recurrent** if all of its factors are recurrent in it
Suppose $A$ is totally ordered by the relation $<$. Then we can totally order $A^+$ by the *lexicographic order* $<$. That is:

**Definition**

Given two words $u, v \in A^+$, we have $u < v \iff$ either $u$ is a proper prefix of $v$ or $u = xau'$ and $v = xbv'$, for some $x, u', v' \in A^*$ and letters $a, b$ with $a < b$.

- This is the usual alphabetic ordering in a dictionary.
- We say that $u$ is *lexicographically less* than $v$.
- This notion naturally extends to infinite words.
Suppose $\mathcal{A}$ is totally ordered by the relation $<$. Then we can totally order $\mathcal{A}^+$ by the **lexicographic order** $<$. That is:

**Definition**

Given two words $u, v \in \mathcal{A}^+$, we have $u < v \iff$ either $u$ is a proper prefix of $v$ or $u = xau'$ and $v = xbv'$, for some $x, u', v' \in \mathcal{A}^*$ and letters $a, b$ with $a < b$.

- This is the usual alphabetic ordering in a dictionary.
- We say that $u$ is **lexicographically less** than $v$.
- This notion naturally extends to infinite words.
Suppose $\mathcal{A}$ is totally ordered by the relation $<$. Then we can totally order $\mathcal{A}^+$ by the *lexicographic order* $<$. That is:

**Definition**

Given two words $u, v \in \mathcal{A}^+$, we have $u < v \iff$ either $u$ is a proper prefix of $v$ or $u = xau'$ and $v = xbv'$, for some $x, u', v' \in \mathcal{A}^*$ and letters $a, b$ with $a < b$.

- This is the usual alphabetic ordering in a dictionary.
- We say that $u$ is *lexicographically less* than $v$.
- This notion naturally extends to infinite words.
1. Introduction

2. Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words

3. Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words

4. Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words

A. Glen*, J. Justin, G. Pirillo
Characterizations of episturmian words
Sturmian words

**Definition**

An infinite word $s$ over $\{a, b\}$ is **Sturmian** if there exist real numbers $\alpha, \rho \in [0, 1]$ 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

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

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

- *aperiodic* if $\alpha$ is irrational;
- *periodic* if $\alpha$ is rational;
- *standard* if $\rho = \alpha$.

Here, **Sturmian** refers to both aperiodic and periodic Sturmian words.
A Sturmian word is:
- *aperiodic* if $\alpha$ is irrational;
- *periodic* if $\alpha$ is rational;
- *standard* if $\rho = \alpha$.

Here, *Sturmian* refers to both aperiodic and periodic Sturmian words.
A finite or infinite word \( w \) on \( \{a, b\} \) is \textit{balanced} if:

\[
u, v \in F(w), \ |u| = |v| \implies ||u||_b - |v|_b| \leq 1.
\]

Morse & Hedlund (1940):

All balanced infinite words over a 2-letter alphabet are called \textit{Sturmian trajectories}.

They belong to three classes:

- aperiodic Sturmian;
- periodic Sturmian;
- ultimately periodic non-recurrent infinite words, called \textit{skew words}.
Sturmian words (cont.)

**Definition (Balance)**

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

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

**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*.
Episturmian words

Definition

An infinite word $t$ is *episturmian* if:

- $F(t)$ 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.

Episturmian words are recurrent.
Definition

An infinite word $t$ is \textit{episturmian} if:

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

$t$ is \textit{standard} if all of its left special factors are prefixes of it.

- Episturmian words are recurrent.
Let $t$ be a standard episturmian word over $A$ and let

$$u_1 = \varepsilon, \ u_2, \ u_3, \ u_4, \ldots$$

be the infinite sequence of its palindromic prefixes.

$\exists$ an infinite word $\Delta(t) = x_1x_2x_3\ldots (x_i \in A)$ such that

$$u_{n+1} = (u_nx_n)^{(+)}, \quad n \in \mathbb{N}^+$$

where $w^{(+)}$ is the shortest palindrome having $w$ as a prefix.

$\Delta(t)$ is called the \textit{directive word} of $t = \lim_{n \to \infty} u_n$. 
Let $t$ be a standard episturmian word over $A$ and let

$$u_1 = \varepsilon, \ u_2, \ u_3, \ u_4, \ldots$$

be the infinite sequence of its palindromic prefixes.

$\exists$ an infinite word $\Delta(t) = x_1 x_2 x_3 \cdots \ (x_i \in A)$ such that

$$u_{n+1} = (u_n x_n)^{(+)}, \ n \in \mathbb{N}^+$$

where $w^{(+)}$ is the shortest palindrome having $w$ as a prefix.

$\Delta(t)$ is called the directional word of $t = \lim_{n \to \infty} u_n$. 
Let \( t \) be a standard episturmian word over \( A \) and let

\[
    u_1 = \varepsilon, \ u_2, \ u_3, \ u_4, \ldots
\]

be the infinite sequence of its palindromic prefixes.

∃ an infinite word \( \Delta(t) = x_1 x_2 x_3 \cdots (x_i \in A) \) such that

\[
    u_{n+1} = (u_n x_n)^{(+)}, \quad n \in \mathbb{N}^+
\]

where \( w^{(+) \ldots} \) is the shortest palindrome having \( w \) as a prefix.

\( \Delta(t) \) is called the \textit{directive word} of \( t = \lim_{n \to \infty} u_n \).
A standard episturmian word $t$ over $A$, or any equivalent (episturmian) word, is said to be $B$-strict (or $k$-strict if $|B| = k$) if

$$\text{Alph}(\Delta(t)) = \text{Ult}(\Delta(t)) = B \subseteq A$$

- The $k$-strict episturmian words have complexity $(k - 1)n + 1$ for each $n \in \mathbb{N}$.
- Such words are exactly the $k$-letter Arnoux-Rauzy sequences.
- Example: $k$-bonacci word, $k \geq 2$. 

A. Glen*, J. Justin, G. Pirillo
Characterizations of episturmian words
Strict episturmian words

**Definition**

A standard episturmian word $t$ over $\mathcal{A}$, or any equivalent (episturmian) word, is said to be $\mathcal{B}$-strict (or $k$-strict if $|\mathcal{B}| = k$) if

$$\text{Alph}(\Delta(t)) = \text{Ult}(\Delta(t)) = \mathcal{B} \subseteq \mathcal{A}$$

- The $k$-strict episturmian words have **complexity** $(k - 1)n + 1$ for each $n \in \mathbb{N}$.
- Such words are exactly the $k$-letter **Arnoux-Rauzy sequences**.
- **Example:** $k$-bonacci word, $k \geq 2$. 

A. Glen*, J. Justin, G. Pirillo

Characterizations of episturmian words
**Strict episturmian words**

**Definition**

A standard episturmian word $t$ over $A$, or any equivalent (episturmian) word, is said to be $B$-strict (or $k$-strict if $|B| = k$) if

$$\text{Alph}(\Delta(t)) = \text{Ult}(\Delta(t)) = B \subseteq A$$

- The $k$-strict episturmian words have complexity $(k - 1)n + 1$ for each $n \in \mathbb{N}$.
- Such words are exactly the $k$-letter Arnoux-Rauzy sequences.
- Example: $k$-bonacci word, $k \geq 2$. 
Strict episturmian words

Definition

A standard episturmian word \( t \) over \( A \), or any equivalent (episturmian) word, is said to be \( B\)-strict (or \( k\)-strict if \(|B| = k\)) if

\[
\text{Alph}(\Delta(t)) = \text{Ult}(\Delta(t)) = B \subseteq A
\]

- The \( k \)-strict episturmian words have complexity \((k - 1)n + 1\) for each \( n \in \mathbb{N} \).
- Such words are exactly the \( k \)-letter Arnoux-Rauzy sequences.
- Example: \( k\)-bonacci word, \( k \geq 2 \).
Outline

1. Introduction

2. Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words

3. Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words

4. Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words

A. Glen*, J. Justin, G. Pirillo

Characterizations of episturmian words
A finite word $w$ is said to be \textit{finite Sturmian} or \textit{finite episturmian} if $w$ is a factor of some infinite Sturmian or episturmian word.

It suffices to consider strict standard episturmian words. Finite episturmian words are exactly the \textit{finite Arnoux-Rauzy words} (Mignosi and Zamboni, 2002).

An infinite word $t$ on a finite alphabet is said to be \textit{episkew} if $t$ is non-recurrent and all of its factors are (finite) episturmian.

There are a number of equivalent definitions of episkew words.
**Terminology**

**Definition**

A finite word $w$ is said to be *finite Sturmian* or *finite episturmian* if $w$ is a factor of some infinite Sturmian or episturmian word.

- It suffices to consider strict standard episturmian words.
- Finite episturmian words are exactly the *finite Arnoux-Rauzy words* (Mignosi and Zamboni, 2002).

**Definition (Episkew)**

An infinite word $t$ on a finite alphabet is said to be *episkew* if $t$ is non-recurrent and all of its factors are (finite) episturmian.

- There are a number of equivalent definitions of episkew words.
A finite word $w$ is said to be \textit{finite Sturmian} or \textit{finite episturmian} if $w$ is a factor of some infinite Sturmian or episturmian word.

- It suffices to consider strict standard episturmian words.
- Finite episturmian words are exactly the \textit{finite Arnoux-Rauzy words} (Mignosi and Zamboni, 2002).

An infinite word $t$ on a finite alphabet is said to be \textit{episkew} if $t$ is non-recurrent and all of its factors are (finite) episturmian.

- There are a number of equivalent definitions of episkew words.
A finite word $w$ is said to be *finite Sturmian* or *finite episturmian* if $w$ is a factor of some infinite Sturmian or episturmian word.

- It suffices to consider strict standard episturmian words.
- Finite episturmian words are exactly the *finite Arnoux-Rauzy words* (Mignosi and Zamboni, 2002).

An infinite word $t$ on a finite alphabet is said to be *episkew* if $t$ is non-recurrent and all of its factors are (finite) episturmian.

- There are a number of equivalent definitions of episkew words.
Outline

1. Introduction

2. Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words

3. Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words

4. Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words
Let $t$ be an infinite word.

**Definition**

Define $\text{min}(t)$ to be the infinite word such that any prefix of $\text{min}(t)$ is the *lexicographically* smallest amongst the factors of $t$ of the same length. Similarly define $\text{max}(t)$.

Our main results extend the following recent work . . .
Let $t$ be an infinite word.

**Definition**

Define $\text{min}(t)$ to be the infinite word such that any prefix of $\text{min}(t)$ is the *lexicographically* smallest amongst the factors of $t$ of the same length. Similarly define $\text{max}(t)$.

Our main results extend the following recent work . . .
Outline

1. Introduction

2. Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words

3. Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words

4. Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words
Proposition (Pirillo, 2005)

Let \( s \) be an infinite word over a finite alphabet \( A \). The following properties are equivalent:

(i) \( s \) is standard episturmian,

(ii) for any \( a \in A \) and order \( < \) such that \( a = \min(A) \), we have \( as \leq \min(s) \).

Similarly, for an infinite word \( s \) on \( \{a, b\} \) (\( a < b \)), the inequality:

\[
as \leq \min(s) \leq \max(s) \leq bs
\]

characterizes standard Sturmian words (aperiodic and periodic).
Proposition (Pirillo, 2005)

Let $s$ be an infinite word over a finite alphabet $A$. The following properties are equivalent:

(i) $s$ is standard episturmian,

(ii) for any $a \in A$ and order $<$ such that $a = \min(A)$, we have $as \leq \min(s)$.

Similarly, for an infinite word $s$ on $\{a, b\}$ ($a < b$), the inequality:

$$as \leq \min(s) \leq \max(s) \leq bs$$

characterizes standard Sturmian words (aperiodic and periodic).
Proposition (Justin & Pirillo, 2002)

Let $s$ be an infinite word over a finite alphabet $A$. The following properties are equivalent:

(i) $s$ is a standard Arnoux-Rauzy sequence,

(ii) for any $a \in A$ and order $<$ such that $a = \min(A)$, we have $as = \min(s)$.

That is: $s$ is a strict standard episturmian word $\iff$ (ii) holds.

2-letters: $s$ is an aperiodic standard Sturmian word $\iff$ $(\min(s), \max(s)) = (as, bs)$ for $a < b$. 
Proposition (Justin & Pirillo, 2002)

Let $s$ be an infinite word over a finite alphabet $A$. The following properties are equivalent:

(i) $s$ is a standard Arnoux-Rauzy sequence,

(ii) for any $a \in A$ and order $<$ such that $a = \min(A)$, we have $as = \min(s)$.

That is: $s$ is a strict standard episturmian word $\iff$ (ii) holds.

2-letters: $s$ is an aperiodic standard Sturmian word $\iff (\min(s), \max(s)) = (as, bs)$ for $a < b$. 
Proposition (Justin & Pirillo, 2002)

Let $s$ be an infinite word over a finite alphabet $A$. The following properties are equivalent:

(i) $s$ is a standard Arnoux-Rauzy sequence,
(ii) for any $a \in A$ and order $<$ such that $a = \min(A)$, we have $as = \min(s)$.

That is: $s$ is a strict standard episturmian word $\iff$ (ii) holds.

2-letters: $s$ is an aperiodic standard Sturmian word $\iff (\min(s), \max(s)) = (as, bs)$ for $a < b$. 
Introduction

Preliminaries
- Terminology & Notation
- Sturmian & Episturmian Words
- Episkew Words

Previous Work
- Extremal Words
- Extremal Properties
- Fine Words

Characterizations via Lexicographic Orderings
- Finite Episturmian Words
- Infinite Episturmian Words

A. Glen*, J. Justin, G. Pirillo

Characterizations of episturmian words
Fine words

Definition (Pirillo, 2005)
An infinite word $t$ over a 2-letter alphabet $\{a, b\}$ ($a < b$) is fine if $(\min(t), \max(t)) = (as, bs)$ for some infinite word $s$.

- Fine words on $\{a, b\}$ are exactly the aperiodic Sturmian and skew infinite words.
- Recently generalized to an arbitrary finite alphabet. . .
Definition (Pirillo, 2005)

An infinite word $t$ over a 2-letter alphabet $\{a, b\}$ ($a < b$) is fine if $(\min(t), \max(t)) = (as, bs)$ for some infinite word $s$.

- Fine words on $\{a, b\}$ are exactly the aperiodic Sturmian and skew infinite words.

- Recently generalized to an arbitrary finite alphabet . . .
Fine words

Definition (Pirillo, 2005)

An infinite word $t$ over a 2-letter alphabet $\{a, b\}$ ($a < b$) is fine if $(\min(t), \max(t)) = (as, bs)$ for some infinite word $s$.

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

**Generalized fine words**

**Definition**

An *acceptable pair* is a pair $(a, <)$ where $a$ is a letter and $<$ is a lexicographic order on $A^+$ such that $a = \min(A)$.

**Definition (Glen, 2006)**

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

**Proposition (Glen, 2006)**

An infinite word $t$ is fine if and only if $t$ is either a strict episturmian word, or a strict episkew word.
Generalized fine words

Definition

An acceptable pair is a pair \((a, <)\) where \(a\) is a letter and \(<\) is a lexicographic order on \(A^+\) such that \(a = \min(A)\).

Definition (Glen, 2006)

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

Proposition (Glen, 2006)

An infinite word \(t\) is fine if and only if \(t\) is either a strict episturmian word, or a strict episkew word.
Generalized fine words

Definition
An *acceptable pair* is a pair \((a, <)\) where \(a\) is a letter and \(<\) is a lexicographic order on \(A^+\) such that \(a = \min(A)\).

Definition (Glen, 2006)
An infinite word \(t\) on \(A\) is said to be *fine* if there exists an infinite word \(s\) such that \(\min(t) = as\) for any acceptable pair \((a, <)\).

Proposition (Glen, 2006)
An infinite word \(t\) is fine if and only if \(t\) is either a *strict episturmian word*, or a *strict episkew word*.
Let $w$ be a finite or infinite word on $A$.

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

For the orders \( b < a < c \) and \( b < c < a \) on the 3-letter alphabet \( \{a, b, c\} \):

\[
\begin{align*}
\min(w|1) & = b \\
\min(w|2) & = ba \\
\min(w|3) & = bab \\
\min(w|4) & = baba \\
\min(w|5) & = babac = \min(w)
\end{align*}
\]

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

Notation
\[ \nu_p : \text{prefix of length } p \text{ of a given finite or infinite word } \nu. \]

Theorem
A finite word \( w \) on \( A \) is episturmian if and only if there exists a finite word \( u \) such that, for any acceptable pair \( (a, <) \), we have

\[ au_{|m|-1} \leq m \]  \hspace{1cm} (1)

where \( m = \min(w) \) for the considered order.
A finite word $w$ on $A$ is episturmian if and only if there exists a finite word $u$ such that, for any acceptable pair $(a, <)$, we have

$$au_{|m|-1} \leq m$$

(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.
A new characterization of finite Sturmian words (i.e., finite balanced words):

**Corollary**

*A finite word* $w$ *on* $A = \{a, b\}$, $a < b$, *is not Sturmian (i.e., not balanced)* if and only if there exists a finite word $u$ such that $aua$ is a prefix of $\min(w)$ and $bub$ is a prefix of $\max(w)$. 
Introduction
Preliminaries
Previous Work
Characterizations via Lexicographic Orderings

Finite Episturmian Words
Infinite Episturmian Words

Examples

Example (1)
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.

Example (2)
For $w = aabababaabaab$:

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

Example (1)

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.

Example (2)

For \( w = aabababaabaab \):

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

1 Introduction

2 Preliminaries
   - Terminology & Notation
   - Sturmian & Episturmian Words
   - Episkew Words

3 Previous Work
   - Extremal Words
   - Extremal Properties
   - Fine Words

4 Characterizations via Lexicographic Orderings
   - Finite Episturmian Words
   - Infinite Episturmian Words
A characterization of **episturmian words in a wide sense** (recurrent, episkew):

**Corollary**

An infinite word $t$ on $A$ is episturmian in the wide sense if and only if there exists an infinite word $u$ such that $au \leq \min(t)$ for any acceptable pair $(a, <)$.

A characterization of **balanced infinite words on a 2-letter alphabet** (i.e., Sturmian and skew words):

**Corollary**

An infinite word $t$ on $\{a, b\}$, $a < b$, is balanced (i.e., Sturmian or skew) if and only if there exists an infinite word $u$ such that

$$au \leq \min(t) \leq \max(t) \leq bu.$$
A characterization of *episturmian words in a wide sense* (recurrent, episkew):

**Corollary**

An infinite word \( t \) on \( A \) is episturmian in the wide sense if and only if there exists an infinite word \( u \) such that \( au \leq \min(t) \) for any acceptable pair \( (a, <) \).

A characterization of *balanced infinite words on a 2-letter alphabet* (i.e., Sturmian and skew words):

**Corollary**

An infinite word \( t \) on \( \{a, b\} \), \( a < b \), is balanced (i.e., Sturmian or skew) if and only if there exists an infinite word \( u \) such that

\[
a u \leq \min(t) \leq \max(t) \leq b u.
\]
A characterization of episturmian words in a *wide sense* (recurrent, episkew):

**Corollary**

An infinite word $t$ on $A$ is episturmian in the wide sense if and only if there exists an infinite word $u$ such that $au \leq \min(t)$ for any acceptable pair $(a, <)$.

A characterization of balanced infinite words on a 2-letter alphabet (i.e., Sturmian and skew words):

**Corollary**

An infinite word $t$ on $\{a, b\}$, $a < b$, is balanced (i.e., Sturmian or skew) if and only if there exists an infinite word $u$ such that

$$au \leq \min(t) \leq \max(t) \leq bu.$$