Combinatorics On Sturmian Words

Amy Glen

Major Review Seminar

February 27, 2004
OUTLINE

• Finite and Infinite Words

• Sturmian Words
  – Mechanical Words and Cutting Sequences
  – Infinite Words and Morphisms
  – The Fibonacci Word & Characteristic Sturmian Words

• Decompositions of Sturmian Words into Palindrome Words

• Powers of Factors of Sturmian Words

• The Tribonacci Sequence

• Future Endeavours – Application to Continued Fractions
FINITE AND INFINITE WORDS

• Combinatorial properties have become significantly important

• Applications in physics, biology, maths & computer science

• Family of Sturmian words – a very active subject of research

• Named after Charles François Sturm (1803–1855)

• Morse and Hedlund (1940): symbolic dynamics

• Largest developments in last twenty years

• Points of view: geometrically, combinatorially, algebraically

• Combinatorial approach is most common
WHAT ARE FINITE AND INFINITE WORDS?

• Alphabet: \( A \) denotes a finite set of symbols called letters.

• Finite/Infinite word: concatenation of letters from \( A \).

Let \( u, w \) denote finite words & \( x, y \) denote infinite words.

If \( x = uwy \rightarrow w \) is a factor of \( x \),

\( u \) is a prefix of \( x \),

\( y \) is a suffix of \( x \).

The length of \( w \) is number of letters it contains, denoted by \( |w| \).

• The empty word of length 0 is denoted by \( \varepsilon \).

• The set \( A^* \) of all finite words over \( A \) is a monoid (\( Id = \varepsilon \)).
**WHAT IS A STURMIAN WORD?**

- An infinite word $s$ with exactly $n + 1$ distinct factors of length $n$, $\forall n \geq 0$ \(\implies s\) is over a two-letter alphabet, say $\mathcal{A} = \{a, b\}$.

- Sturmian words are *not* eventually periodic
  - they are aperiodic words of minimal complexity

- \(\exists\) many characterizations and numerous properties of Sturmian words

- **Applications:** symbolic dynamics, continued fraction expansion, physics (crystallography), computer science (pattern recognition)

- **Equivalent definitions:** mechanical words & cutting sequences
IRRATIONAL MECHANICAL WORDS

• An infinite word $s$ over $\mathcal{A}$ is Sturmian $\iff \exists$ an irrational $\alpha \in (0, 1)$ & $\rho \in \mathbb{R}$ such that $s$ is either:

$$s_{\alpha, \rho} = s(0)s(1)s(2) \cdots \text{ or } s'_{\alpha, \rho} = s'(0)s'(1)s'(2) \cdots,$$

where

$$s(n) = a \text{ if } \lfloor (n + 1)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor = 0, \quad s(n) = b \text{ otherwise;}$$

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

Note: $\lfloor \cdot \rfloor$ denotes floor function & $\lceil \cdot \rceil$ denotes ceiling function

• $\alpha$ is called the slope & $\rho$ is the intercept

• $\rho = 0 \longrightarrow s_{\alpha, 0} = ac_\alpha \text{ & } s'_{\alpha, 0} = bc_\alpha$

where $c_\alpha$ is the characteristic Sturmian word of slope $\alpha$
CUTTING SEQUENCES

• An interpretation of mechanical words

• Line: $y = \beta x + \rho$ with irrational $\beta > 0, \rho \in \mathbb{R}$
  
  – consider this ray in the positive quadrant of $\mathbb{R}^2$
  
  – overlay quadrant with an integer grid & construct infinite word $K_{\beta,\rho}$
  
  – vertical grid-line crossed: label intersection with $a$
  
  – horizontal grid-line crossed: label intersection with $b$

Labels: $x_0, x_1, x_2, \ldots \rightarrow K_{\beta,\rho} = x_0x_1x_2 \cdots$ is a Sturmian word, i.e.

$$K_{\beta,\rho} = s_{\beta/(1+\beta),\rho/(1+\beta)}.$$

Note: $c_\alpha = K_{\beta,0} \iff \alpha = \beta/(1 + \beta)$. 
\[ y = \frac{\sqrt{5}-1}{2} x \quad \rightarrow \quad \text{Fibonacci word} \]

**Fibonacci word:** \( f = K_{(\sqrt{5}-1)/2,0} = abaabaabaababaababaabababababaab \cdots \)

- special example of a characteristic Sturmian word of slope \( \alpha = \frac{3-\sqrt{5}}{2} \)
- Properties of \( f \) may be extended to Sturmian words
INFINITE WORDS & MORPHISMS

- Let \((u_n)_{n \geq 0}\) be a sequence of words from \(A^*\) s.t. \(u_n\) is a proper prefix of \(u_{n+1}\). This gives obvious meaning to

  \[
  \lim_{n \to \infty} u_n \quad \text{as an infinite word.}
  \]

- A **morphism on** \(A\) is a map \(\psi : A^* \to A^*\) such that

  \[
  \psi(uv) = \psi(u)\psi(v), \quad \forall u, v \in A^*.
  \]

- If \(\psi(c) = cw, \ c \in A, \ w \in A^* \to \psi^n(c)\) is a proper prefix of \(\psi^{n+1}(c)\) and \((\psi^n(c))_{n \geq 0}\) converges to a unique infinite word:

  \[
  x = \lim_{n \to \infty} \psi^n(c).
  \]

We say that \(x\) is **generated** by \(\psi\).
PALINDROME WORDS

• **Palindrome**: a finite word that reads the same backwards as forwards

• **Examples**: $aa$, $abaaba$, $aba$, $ababa$

• Objects of great interest in computer science

• Important tools used in the study of factors of Sturmian words

• **Question**: Where exactly do palindromes occur in a Sturmian word?

• **Example**: $f = ab(aa)bab(aa)b(aa)bab(aa)bab(aa)\cdots$,  
  where $aa$ occurs at positions: 2, 7, 10, 15, 20, …  
  Distances: 5, 3, 5, 5, … $\longrightarrow$ Fibonacci word over the alphabet \{5, 3\}

• I have described precisely where palindromes occur in a Sturmian word
FINITE FIBONACCI WORDS

• Define the **Fibonacci numbers** \((F_n)_{n \geq -1}\) by

\[
F_{-1} = F_0 = 1, \quad F_n = F_{n-1} + F_{n-2}.
\]

• Let \(f_n\) denote the prefix of \(f\) of length \(F_n\) and set \(f_{-1} = b\). Then

\[
f_0 = a, \quad f_n = f_{n-1}f_{n-2} \quad \text{and} \quad f = \lim_{n \to \infty} f_n.
\]

• E.g. \(f_1 = ab, f_2 = aba, f_3 = abaab\)

• **Singular words** of \(f\) : \(w_{2n-1} = af_{2n-1}b^{-1}, \ w_{2n} = bf_{2n}a^{-1}\)

• **Examples:** \(w_0 = b, w_1 = aa, w_2 = bab, w_3 = aabaa \quad \rightarrow \quad \text{palindromes}\)

• Wen & Wen (1994): \(f = aw_0w_1w_2w_3 \cdots = a\ b\ aa\ bab\ aabaa\ \cdots\)
CONTINUED FRACTIONS

• Combinatorial structure of $c_{\alpha}$ has close relationship with the continued fraction (CF) of its slope $\alpha$

• Recall: every irrational $\alpha \in (0, 1)$ has a unique CF expansion

$$\alpha = [0; a_1, a_2, a_3, \ldots] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}},$$

where all the $a_i \in \mathbb{Z}^+$ are called partial quotients.
CHARACTERISTIC STURMIAN WORDS

- Let $\alpha \in (0, 1)$ be irrational with $\alpha = [0; 1 + d_1, d_2, d_3, \ldots]$.

- Define standard sequence $(s_n)_{n \geq -1}$ of words by:
  
  $s_{-1} = b, \quad s_0 = a, \quad s_n = s_{d_n} s_{n-1} s_{n-2}$.

- $\forall n \geq 0$, $s_n$ is a prefix of $s_{n+1}$ and $\lim_{n \to \infty} s_n$ is a well-defined infinite word
  - in fact, each $s_n$ is a prefix of $c_\alpha$ and $c_\alpha = \lim_{n \to \infty} s_n$
  - Fibonacci word: $\alpha = (3 - \sqrt{5})/2 = [0; 2, 1, 1, 1, \ldots]$}

- Singular words of $c_\alpha$: $w_{2n-1} = a s_{2n-1} b^{-1}$, $w_{2n} = b s_{2n} a^{-1}$

- Melançon (1999):
  - defined palindrome words $v_n$ such that $c_\alpha = av_0 v_1 v_2 v_3 v_4 \cdots$
Consider $c_\alpha$ with $\alpha = (\sqrt{3} - 1)/2 = [0; 2, 1, 2, 1, 2, 1, 2, \ldots]$. 

**Standard sequence:** $s_1 = ab$, $s_2 = s_1s_0 = aba$, $s_3 = s_2^2s_1 = abaabaab, \ldots$

$\longrightarrow c_\alpha = abaabaababaabaabaabaabaabaabaabaab \cdots$

**Singular words:** $w_0 = b$, $w_1 = aa$, $w_2 = bab$, $w_3 = aabaabaa, \ldots$

**Decomposition in terms of $w_2$:**

$$c_\alpha = abaabaa(bab)aabaabaa(bab)aabaabaa(bab)aabaabaa(bab)aabaabaa \cdots$$

$$= abaaabaa(bab)z_1(bab)z_2(bab)z_3(bab)z_4 \cdots ,$$

where $z_1z_2z_3 \cdots$ is the characteristic Sturmian word of slope $\beta = [0; 1, 2] = \sqrt{3} - 1$ over the alphabet $\{aabaa, w_3\}$
CONJUGATES OF $f$ AND $c_\alpha$

- Let $x, x'$ be infinite words such that $x = wx'$, $w$ finite, $|w| = k$. Then $x'$ is called the $k^{th}$ conjugate of $x$; i.e. $x'$ is obtained from $x$ by deleting the first $k$ letters of $x$.

- Levé and Séébold (2003): a factorization of each conjugate of $f$ as an infinite concatenation of generalized singular words

- I have extended this result to conjugates of $c_\alpha$ for

$$\alpha = [0; 2, r, r, r, \ldots] \text{ and } \alpha = [0; 1, 1, r, r, r, \ldots]$$

- Paper accepted for publication (European J. Combinatorics)
POWERS IN STURMIAN WORDS

• Interested in finite words $w$ and integers $p \in \mathbb{Z}^+$ such that
  $$w^p = www \cdots w \ (p \text{ times})$$
  is a factor of $c_\alpha$

• Application: spectral properties of associated quantum mechanical
  quasi-crystal models (Physics)

• Damanik and Lenz (2003): explicitly determined all integer powers
  of words occurring in $c_\alpha$
  – also considered the # of distinct squares of words contained in $s_n$
  – combinatorial method: based on properties of the building blocks $s_n$
EXACT NUMBER OF SQUARES IN $f_n$

- A **square** is a word of the form $uu$; $u$ a finite word

- **Example:** $f_4 = abaababa$ contains 4 squares:
  $$(aba)^2, (ab)^2, (ba)^2, a^2$$

- Fraenkel and Simpson (1999): exact # of distinct squares in $f_n$ is
  $$2(F_{n-2} - 1), \forall n \geq 5$$

- Recently extended to the case of the factors $s_n$ of $c_\alpha$

- $u^p$ ($p \geq 2$) is a factor of $f \implies |u| = F_n$ for some $n$

- I have extended the above results to the Tribonacci sequence
  - paper in preparation
TRIBONACCI SEQUENCE $\xi$

- Infinite word over $A = \{a, b, c\}$: $\xi = abacabaabacababacabaabacab \cdots$
  - a generalization of the Fibonacci word

- Define the Tribonacci numbers $(T_n)_{n \geq -1}$ by
  
  $$T_{-1} = T_0 = 1, \quad T_1 = 2, \quad T_n = T_{n-1} + T_{n-2} + T_{n-3}.$$ 

- Let $A_n$ denote the prefix of $\xi$ of length $T_n$ and set $A_{-1} = c$. Then
  
  $$A_0 = a, \quad A_1 = ab, \quad A_n = A_{n-1}A_{n-2}A_{n-3} \quad \text{and} \quad \xi = \lim_{n \to \infty} A_n.$$ 

- E.g. $A_2 = (ab)ac$, $A_3 = (abac)(ab)a$, $A_4 = (abacaba)(abac)(ab)$

- Yet to determine exactly where palindromes occur in $\xi$
EXACT NUMBER OF SQUARES IN $A_n$

- Define $P_0 = \varepsilon$, $P_n = A_{n-1}A_{n-2}\cdots A_1A_0$, $n \geq 1$

- E.g. $P_1 = a$, $P_2 = aba$, $P_3 = abacaba$ $\longrightarrow$ palindrome prefixes of $\xi$

- Exact # of distinct squares in $A_n$ ($n \geq 6$):

  $$\sum_{m=0}^{n-3} (|P_m| + 1) + |P_{n-5}| + |P_{n-6}| + 1$$

- Example: $A_5 = abacabaabacabacabaabacabacabaabac$ contains 7 squares:

  $$aa, abab, baba, abaaab, (abacab)^2, (bacaba)^2, (abacaba)^2$$

  $A_6$ has $(0 + 1) + (1 + 1) + (3 + 1) + (7 + 1) + 1 + 0 + 1 = 17$ squares
SQUARES AND HIGHER POWERS IN $\xi$

• Let $u$ be a factor of $\xi$ with $T_n \leq |u| < T_{n+1}$.

| $|u|$   | $p$   | $\# u \text{ s.t. } u^p \text{ in } \xi$ |
|--------|-------|----------------------------------------|
| $T_n$  | 2     | $T_n$                                  |
| $T_n + T_{n-1}$ | 2     | $|P_{n-1}| + 1$                       |
| $T_n$  | $\geq 3$ | $|P_{n-3}| + 1$                     |
| $T_n + T_{n-1}$ | $\geq 3$ | 0                                      |

• I am now extending this to episturmian words

• Episturmian words:
  
  – a generalization of Sturmian words to an arbitrary finite alphabet
FUTURE ENDEAVOURS

• The study of episturmian words is a new area of research
  – I hope to extend my results to the family of episturmian words

• Application to continued fractions and transcendence:
  \( \gamma \in \mathbb{R} \) is \textit{transcendental} if it is not a zero of a polynomial with integer coefficients
  – Allouche \textit{et al.} (2001): Let \( a, b \in \mathbb{Z}^+ (a \neq b) \) & let \( x = (x_n)_{n \geq 0} \) be an infinite sequence over \( \{a, b\} \). Then \( \gamma := [0; x_0, x_1, x_2, \ldots] \) is transcendental if \( x \) is a Sturmian word over \( \{a, b\} \).
  – Question: What if the partial quotients form an infinite episturmian word over a \( k \)-letter alphabet?
SOME INTERESTING READING

- Allouche *et al.* (2001): Transcendence of Sturmian or morphic continued fractions


- Justin and Pirillo (2002):
  - Episturmian words & episturmian morphisms

- Melançon (1999): Lyndon words & singular factors of Sturmian words

- Wen and Wen (1994): Some properties of the Fibonacci word