Palindromically Rich GT-words
Joint work with Florence Levé, Université de Picardie Jules Verne

Amy Glen
School of Engineering & IT
Murdoch University, Perth, Australia

amy.glen@gmail.com
http://amyglen.wordpress.com

38ACCMCC, Victoria University of Wellington, December 2014

27km run at 4:30am on Monday morning
Complexity of words

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

  The empty word, $\varepsilon$, is the unique word of length 0.

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

- A basic measure of a word’s complexity is the number of distinct blocks (factors) of each length occurring in the word.

- Given a finite or infinite word $w$, the factor complexity function of $w$, denoted by $C_w(n)$, counts the number of distinct factors of $w$ of each length $n \geq 0$.

- Likewise, the palindromic complexity function of $w$, denoted by $P_w(n)$, counts the number of distinct palindromic factors of $w$ of each length $n \geq 0$. 

Complexity of words...

The well-known family of \textit{infinite Sturmian words} are characterised by both their factor complexity and their palindromic complexity.

\textbf{Theorem (Morse, Hedlund,1940)}

An infinite word $w$ is \textbf{Sturmian} if and only if $C_w(n) = n + 1$ for all $n \in \mathbb{N}^+$.

- Morse & Hedlund also showed, in particular, that an infinite word $w$ is eventually periodic $\iff C_w(n) < n + 1$ for some $n \in \mathbb{N}^+$.

- In this sense, Sturmian words are the aperiodic infinite words of minimal complexity.

- Their low complexity accounts for many interesting features, as it induces certain regularities in such words without, however, making them periodic.
Complexity of words . . .

With respect to palindromic complexity, the following is known.

**Theorem (Droubay, Pirillo, 1999)**

An infinite word $w$ is Sturmian if and only if

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

**Remarks:**

- 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.
Trapezoidal Words

So-called trapezoidal words were first introduced by de Luca (1999) when studying the behaviour of the factor complexity of finite Sturmian words (i.e., finite factors of infinite Sturmian words).

Amongst many interesting things, de Luca proved the following result.

Theorem (de Luca 1999)

*If* $w$ *is a finite Sturmian word, then the graph of its complexity* $C_w(n)$ *as a function of* $n$ *(for* $0 \leq n \leq |w|$ *) is that of a regular trapezoid (or possibly an isosceles 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 $r$. 
Example

Graph of the factor complexity of the finite Sturmian word $aabaabab$
This “trapezoidal property” does not characterise Sturmian words. For example, $aaabb$ is trapezoidal ($[1, 2, 3, 3, 2, 1]$), but not Sturmian since it contains 2 palindromes of length 2.

**Note:** If $w$ is a trapezoidal word (i.e., its “complexity graph” is a regular trapezoid on the interval $[0, |w|]$), then necessarily $C_w(1) = 2$. This is because there is 1 factor of length 0, namely the empty word $\varepsilon$.

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

All non-Sturmian trapezoidal words were classified by F. D’Alessandro in 2002.
Definition (Levé, G., 2011)

A finite word $w$ with alphabet $\text{Alph}(w) = \mathcal{A}$ ($|\mathcal{A}| \geq 2$) is said to be a **generalized trapezoidal word** (or **GT-word** for short) if there exist positive integers $m$, $M$ with $m \leq M$ such that the factor complexity function $C_w(n)$ of $w$ increases by 1 for each $n$ in the interval $[1, m]$, is constant for each $n$ in the interval $[m, M]$, and decreases by 1 for each $n$ in the interval $[M, |w|]$.

\[
C_w(n) = |w| - |\mathcal{A}| + 1
\]
Some Examples

Length 10 over $\mathcal{A} = \{a, b, c\}$

<table>
<thead>
<tr>
<th>GT-word</th>
<th>$C(n)$ for $n = 0, 1, 2, \ldots, 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$aaaaaaababc$</td>
<td>$1, 3, 3, 3, 3, 3, 3, 3, 3, 2, 1$</td>
</tr>
<tr>
<td>$abcabcba$</td>
<td>$1, 3, 4, 4, 4, 4, 4, 4, 3, 2, 1$</td>
</tr>
<tr>
<td>$abcabcabcab$</td>
<td>$1, 3, 4, 5, 5, 5, 5, 4, 3, 2, 1$</td>
</tr>
<tr>
<td>$abcabcabcab$</td>
<td>$1, 3, 4, 5, 6, 6, 5, 4, 3, 2, 1$</td>
</tr>
</tbody>
</table>

Length 8 over $\mathcal{A} = \{a, b, c, d\}$

<table>
<thead>
<tr>
<th>GT-word</th>
<th>$C(n)$ for $n = 0, 1, 2, \ldots, 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$aaaaababcd$</td>
<td>$1, 4, 4, 4, 4, 4, 3, 2, 1$</td>
</tr>
<tr>
<td>$aaaabaccd$</td>
<td>$1, 4, 5, 5, 5, 4, 3, 2, 1$</td>
</tr>
<tr>
<td>$aaabcdab$</td>
<td>$1, 4, 5, 6, 5, 4, 3, 2, 1$</td>
</tr>
</tbody>
</table>
Some Basic Properties

The language of all GT-words is closed . . .

Theorem (Levé, G., 2011)

If \( w \) is a GT-word, then each factor of \( w \) (containing at least two different letters) is also a GT-word.

That is, the set of all GT-words is closed by factors.

Moreover, the set of all GT-words is closed under reversal.

Theorem (Levé, G., 2011)

A finite word \( w \) is a GT-word if and only if its reversal is a GT-word.
Back to the Binary Case

In the case when $|\mathcal{A}| = 2$, we have proved the following.

**Theorem (de Luca, G., Zamboni, 2008)**

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

**Theorem (de Luca, G., Zamboni, 2008)**

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

That is, binary trapezoidal words (and hence finite Sturmian words) are “rich” in palindromes in the sense that they contain the maximum number of distinct palindromic factors, according to the following result.

**Theorem (Droubay, Justin, Pirillo, 2001)**

A finite word $w$ contains at most $|w| + 1$ distinct palindromes (including $\varepsilon$).
Rich Words

Definition (G., Justin, Widmer, Zamboni, 2009)
A finite word $w$ is said to be rich if $w$ contains exactly $|w| + 1$ distinct palindromes (including $\varepsilon$).

Examples:
- $abac$ is rich, whereas $abca$ is not rich.
- The word $rich$ is rich . . . and $poor$ is rich too!
- Any binary trapezoidal word is rich, but not conversely.
  E.g., $aabbaa$ is rich, but not trapezoidal ($C(1) = 2$, $C(2) = 4$).

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

Example:
$abaabaaabaaaaaabab\cdots$ $abaabaaabaaaaaabab\cdots$ $abaaaabaaaaaabab\cdots$ $abaabaaabaaaaaabab\cdots$
Here are some other characteristic properties of rich words that were previously established by Droubay, Justin, and Pirillo (2001) and G., Justin, Widmer, and Zamboni (2009).

**Characteristic Properties of Rich Words**

For any finite or infinite word \( w \), the following conditions are equivalent:

i) \( w \) is rich;

ii) every prefix of \( w \) has a unioccurrent palindromic suffix (and equivalently, when \( w \) is finite, every suffix of \( w \) has a unioccurrent palindromic prefix);

iii) for each factor \( u \) of \( w \), every prefix (resp. suffix) of \( u \) has a unioccurrent palindromic suffix (resp. prefix);

iv) for each palindromic factor \( p \) of \( w \), every complete return to \( p \) in \( w \) is a palindrome.
Palindromically Rich GT-words when $|\mathcal{A}| \geq 3$

- Unlike in the binary case ($|\mathcal{A}| = 2$), **not** all GT-words are palindromically rich.

**Example**

The GT-word $ababadbc$ is not rich since it contains a non-palindromic complete return to $b$, namely $badb$.

- However, it is easy to show, for instance, that any ternary word that is formed by appending a new letter to a binary GT-word is a GT-word, and all such ternary words are rich.

- More generally, we can describe all the rich GT-words with respect to their longest palindromic prefixes and suffixes.

- To do this, we need the notion of the heart of a word, which has proved useful for establishing several characterisations of GT-words.
Heart of a GT-word

Definition

If $w$ is a finite word with $|w| > |\text{Alph}(w)|$ (i.e., for any finite word $w$ that contains at least two occurrences of some letter), the heart of $w$ is defined to be the unique (non-empty) factor of $w$ that remains if we delete the longest prefix and the longest suffix of $w$ that contain letters only occurring once in $w$.

On the other hand, if $|w| = |\text{Alph}(w)|$, the heart of $w$ is $w$ itself.

Example

Consider the GT-word $w = ebbacbadf$. By deleting from $w$ the longest prefix and the longest suffix that contain letters only occurring once in $w$, we determine that the heart of $w$ is $v = bbacba$.

Note

All binary trapezoidal words are equal to their own hearts, except those of the form $a^n b$ or $ab^n$ where $a$, $b$ are distinct letters and $n$ is a positive integer. Such (binary) trapezoidal words have hearts of the form $x^n$ for some letter $x \in \{a, b\}$.
Some Examples

For a given word $w$, let $p$ denote its longest palindromic prefix and let $q$ denote its longest palindromic suffix.

- **Rich GT-word**: $\underline{abacaba}de [1, 5, 6, 6, 5, 4, 3, 2, 1]$
  Heart: $v = p = q = abacaba$ (a palindrome).

- **Rich GT-word**: $\underline{ababada}c [1, 4, 5, 5, 5, 4, 3, 2, 1]$
  Heart: $v = ababada$ where $p = ababa$, $q = ada$ ($p$ and $q$ overlap)

- **Rich GT-word**: $aaabab [1, 2, 3, 4, 3, 2, 1]$
  Heart: $v = pq$ where $p = aaa$ and $q = bab$ ($v$ is a product of $p$ and $q$)

- **Rich GT-word**: $acacbcb [1, 3, 4, 5, 4, 3, 2, 1]$
  Heart: $v = pcq$ where $p = aca$ and $q = bcb$

Here $\text{Alph}(p) \cap \text{Alph}(q) \neq \emptyset$ and $p$, $q$ are separated by a letter in $\text{Alph}(p) \cup \text{Alph}(q)$.

- **Rich GT-word**: $aaadcbcb [1, 4, 5, 6, 5, 4, 3, 2, 1]$
  Heart: $v = puq$ where $p = aaa$, $u = dc$, and $q = bcb$

Here $\text{Alph}(p) \cap \text{Alph}(q) = \emptyset$ and $p$, $q$ are separated by the word $dc$ where the letter $d$ is not in $\text{Alph}(p) \cup \text{Alph}(q)$ and $c \in \text{Alph}(q)$. 
Some Examples . . .

- **Non-Rich GT-word**: \textit{ababadbc} [1, 4, 5, 5, 5, 4, 3, 2, 1]
  Heart: \( v = pdq \) where \( p = ababa, q = b \)

Here \( \text{Alph}(q) \subset \text{Alph}(p) \) and \( p, q \) are separated by a letter not in \( \text{Alph}(p) \). Note that \( v \) ends with a non-palindromic complete return to \( b \).

- **Non-Rich GT-word**: \textit{dcdbacdc} [1, 4, 5, 6, 5, 4, 3, 2, 1]
  Heart: \( v = puq \) where \( p = dcd, u = ba, q = cdc \)

Here \( \text{Alph}(p) = \text{Alph}(q) \) and \( p, q \) are separated by a product of two distinct letters, neither of which they contain themselves. Note that \( v \) contains non-palindromic complete returns to each of the letters \( c \) and \( d \).

- **Non-Rich GT-word**: \textit{aaaaaadebcad} [1, 5, 6, 7, 7, 7, 6, 5, 4, 3, 2, 1]
  Heart: \( v = puq \) where \( p = aaaaaa, u = debca, q = d \)

Here \( \text{Alph}(p) \cap \text{Alph}(q) = \emptyset \) and \( p, q \) are separated by a product of mutually distinct letters, with last letter in \( \text{Alph}(p) \) and first letter in \( \text{Alph}(q) \). Note that \( v \) begins with a non-palindromic complete return to \( a \) and also ends with a non-palindromic complete return to \( d \).
A Characterisation of Rich GT-words

Theorem (Levé, G., 2014)

Let \( w \) be a GT-word and suppose that its heart \( v \) has longest palindromic prefix \( p \) and longest palindromic suffix \( q \). Then \( w \) is rich if and only if one of the following conditions is satisfied.

(i) \( p \) and \( q \) are unseparated in \( v \) (i.e., either \( v = pq \), \( v = p = q \), or \( p \) and \( q \) overlap in \( v \)).

(ii) \( v = pxq \) where \( \text{Alph}(p) \cap \text{Alph}(q) \neq \emptyset \) and \( x \in \text{Alph}(p) \cup \text{Alph}(q) \).

(iii) \( v = puq \) where \( \text{Alph}(p) \cap \text{Alph}(q) = \emptyset \) and \( u = u_1 Zu_2 \) where \( u_1, u_2, Z \) are words (at least one of which is non-empty) such that \( \text{Alph}(u_1) \subseteq \text{Alph}(p) \), \( \text{Alph}(u_2) \subseteq \text{Alph}(q) \), and \( Z \) contains no letters in common with \( p \) and \( q \).

In particular, by part (i), all trapezoidal palindromes are rich.

In the case of a binary alphabet, every trapezoidal word has unseparated \( p \) and \( q \) (proved independently).
Thank You!