A NEW COMPLEXITY MEASURE FOR WORDS BASED ON PERIODICITY

Antonio Restivo
University of Palermo
Italy

Joint work with Filippo Mignosi
Periods of a word

\[ w = a_1 a_2 \ldots a_n \]

A positive integer \( p \leq |w| \) is a period of \( w \) if

\[ a_{i+p} = a_i , \quad \text{for } i = 1,2,\ldots,n-p \]

The smallest period of \( w \) is called the period of \( w \) and is denoted by \( p(w) \)

\[ a \ b \ a \ a \ b \ a \ a \ b \ a \ a \ b \ a \] has periods 5 and 8
Local periods

\[ w = a_1 a_2 \ldots a_n \]

A non-empty word \( u \) is a repetition of \( w \) at the point \( i \) if \( w = xy \), with \( |x| = i \) and the following holds:

\[ A^* x \cap A^* u \neq \emptyset \quad \text{and} \quad y A^* \cap u A^* \neq \emptyset \]

The local period of \( w \) at the point \( i \) is:

\[ p(w,i) = \min \{ |u| : u \text{ is a repetition of } w \text{ at the point } i \} \]
An example of repetition and local period

\[ w = a \ b \ a \ a \ b \ a \ b \ a \ b \ a \ a \ b \ a \ a \ b \]

\[
\begin{array}{cccccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
a & a & b & a & a & b & a & b & a & a & b & a \\
\end{array}
\]

\[ p(w,3) = 1 \quad p(w,7) = 8 \]
A point \( i \) is critical if \( p(w,i) = p(w) \)

**Critical Factorization Theorem (CFT)**

*(Cesari-Vincent, 1978; Duval, 1979)*

If \( |w| \geq 2 \), in any sequence of \( m = \max \{ 1, p(w)-1 \} \) consecutive points there is a critical one, i.e. there exists a positive integer \( i \) such that \( p(w,i) = p(w) \).

A point \( i \) is called **left external** if \( i < p(w,i) \). From CFT, the first critical point is left external.
Local periods in infinite words

**Theorem.** An infinite words is recurrent if and only if at any point there is a repetition

**Periodicity function** of an infinite recurrent word $x$:

$$p_x(n) = \min \{ |u| : u \text{ is a repetition at the point } n \}$$

**Theorem.** An infinite recurrent word $x$ is periodic if and only if the periodicity function $p_x$ is bounded. Moreover

$$p(x) = \sup \{ p_x(n) : n \geq 1 \}$$
Gap Theorem

Theorem. Let $x$ be an infinite recurrent word. Then either $p_x$ is bounded, i.e. $x$ is periodic, or $p_x(n) \geq n+1$, for infinitely many integers $n$.

Analogous to the Coven-Hedlund theorem:

Theorem (Coven-Hedlund). The (factor) complexity function $c_x$ of an infinite word $x$ either is bounded, and in such a case $x$ is periodic, or $c_x(n) \geq n+1$, for all integers $n$. 
Thue-Morse
Fibonacci
Characteristic Sturmian words are extremal for the CFT

Theorem. Let $x$ be an infinite recurrent word. $X$ is a characteristic sturmian word if and only if $p_x(n) \leq n + 1$ for all $n \geq 1$ and $p_x(n) = n + 1$ for infinitely many integers $n$.

Equivalently:

The characteristic sturmian words are exactly the recurrent non periodic words $x$ such that $p_x(n) \leq n + 1$. 
Finite Standard words

Let \( q_0, q_1, q_2, \ldots \) be a sequence of non-negative integers, with \( q_i > 0 \) for \( i > 0 \).

Consider the sequence of words \( \{s_n\}_{n \geq 0} \) defined as follows:

\[
\begin{align*}
    s_0 &= b \\
    s_1 &= a \\
    s_{n+1} &= s_n^{q_{n-1}} s_{n-1}
\end{align*}
\]
Characteristic Sturmian words

The sequence \( \{s_n\}_{n \geq 0} \) converges to a limit \( x \) that is an infinite characteristic Sturmian word.

The sequence \( \{s_n\}_{n \geq 0} \) is called the approximating sequence of \( x \) and \((q_0, q_1, q_2, \ldots)\) is the directive sequence of \( x \).

Each finite word \( s_n \) is called a standard word and it is univocally determined by the (finite) directive sequence \((q_0, q_1, \ldots, q_{n-2})\).
Computation of the periodicity function of a characteristic Sturmian word

If $x$ is (the Fibonacci) a characteristic Sturmian word, then the function $p_x(n)$ can be computed from the (Zeckendorf) Ostrowski representation of the integer $n+1$

(J. Shallit, L. Schaeffer)
Non-characteristic Sturmian words

Remark that the characterization theorem holds true just for characteristic Sturmian words, not for all Sturmian words:

\[ y = a \ a\ b\ a\ b\ a\ a\ b\ a\ a\ b\ a\ b\ a\ a\ b\ a\ a\ b\ \ldots \ldots \]

\[ p_y(2) = 5 \quad p_y(5) = 8 \]
Theorem

The periodicity function characterizes any finite or infinite binary word up to exchange of letters.

Remark: this is not true in alphabets having more than two letters.
Periodicity Complexity

The periodicity function has a strong fluctuation, and this is not convenient for certain purposes.

So, we introduce the periodicity complexity function $h_x(n)$ of an infinite word $x$, defined as follows:

$$h_x(n) = \frac{1}{n} \sum_{j=1}^{n} p_x(j)$$
Theorem

If $x$ is an infinite periodic word, then the periodicity complexity function $h_x(n)$ is bounded.

The converse is not true:

There exist non-periodic recurrent words having bounded periodicity complexity.
A non-periodic word with bounded periodicity complexity

Consider a sequence of finite words recursively defined as follows:

\[ w_0 = \text{ab} \]
\[ w_{n+1} = w_n a^{2|w_n|}w_n \]

\[ w_1 = \text{abaaaaaab} \]
\[ w_2 = \text{abaaaaabaaaaaaaaaaaaaaaaaaaaaaaaaabaaaaaab} \]

\[ w = \lim w_n \]

Theorem \[ \limsup h_w(n) = \sup h_w(n) = 7 \]
The Fibonacci word

\[ f = \text{abaababaabaababaababaabaababaababaababaabab} \ldots \]

Theorem \[ h_f(n) \text{ grows as } \Theta (\log n) \]
The Thue-Morse word

\[ t = \text{abbabaabbaabbbabaababbaabbaabbbabaabbaabbaa} \ldots \]

**Theorem** \( h_t(n) \) grows as \( \Theta(n) \)
An infinite recurrent word with arbitrary high periodicity complexity

Let \( v_n \) be the finite binary word obtained by concatenating in the lexicographic order all the words of length \( n \).

\[
\begin{align*}
v_1 &= ab \\
v_2 &= aaabbabb \\
v_3 &= aaaaaababaabbbbaababbbabbb
\end{align*}
\]

For any function \( f \) from \( \mathbb{N} \) to \( \mathbb{N} \) consider the sequence of words:

\[
\begin{align*}
z_1 &= v_1 \\
z_{n+1} &= z_n \cdot z_n^{[2 \cdot f(|z_n| + 1)]} \cdot v_{n+1}
\end{align*}
\]

Consider the infinite word \( z = \lim z_n \)

**Theorem** For infinitely many \( j \), \( h_z(j) > f(j) \).
Problems

• Does there exist a uniformly recurrent non-periodic word having bounded periodicity complexity?

• Does there exist a uniformly recurrent word with arbitrary high periodicity complexity?

• Evaluate the periodicity complexity of other special words