A NEW COMPLEXITY MEASURE FOR WORDS BASED ON PERIODICITY

Antonio Restivo

University of Palermo

Italy

Joint work with Filippo Mignosi

Periods of a word

$w = a_1 a_2 \ldots\ldots a_n$

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

$$a_{i+p} = a_i \,, \qquad \text{for } i = 1,2,\ldots,n-p$$

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

a b a a b a b a a b a has periods 5 and 8

Local periods

$w = a_1 a_2 \ldots\ldots a_n$

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

$$A^* x \cap A^* u \neq \phi \quad \text{and} \quad y A^* \cap u A^* \neq \phi$$

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 a b a a b

1 2 3 4 5 6 7 8 9 10 11 12

a a b a a b a b a a b a a b a b

1 8

p(w,3) = 1 p(w,7) = 8

a b a a b a b a a b a a b

2 3 1 5 2 2 8 1 3 3 1 3

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

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

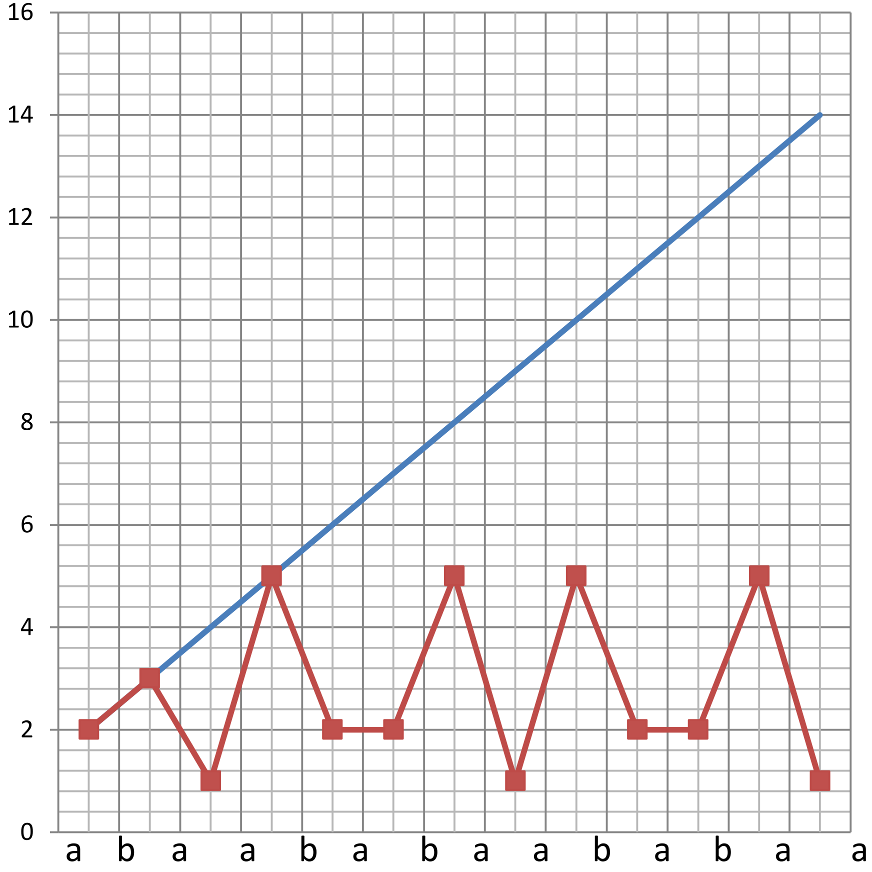
Gap Theorem

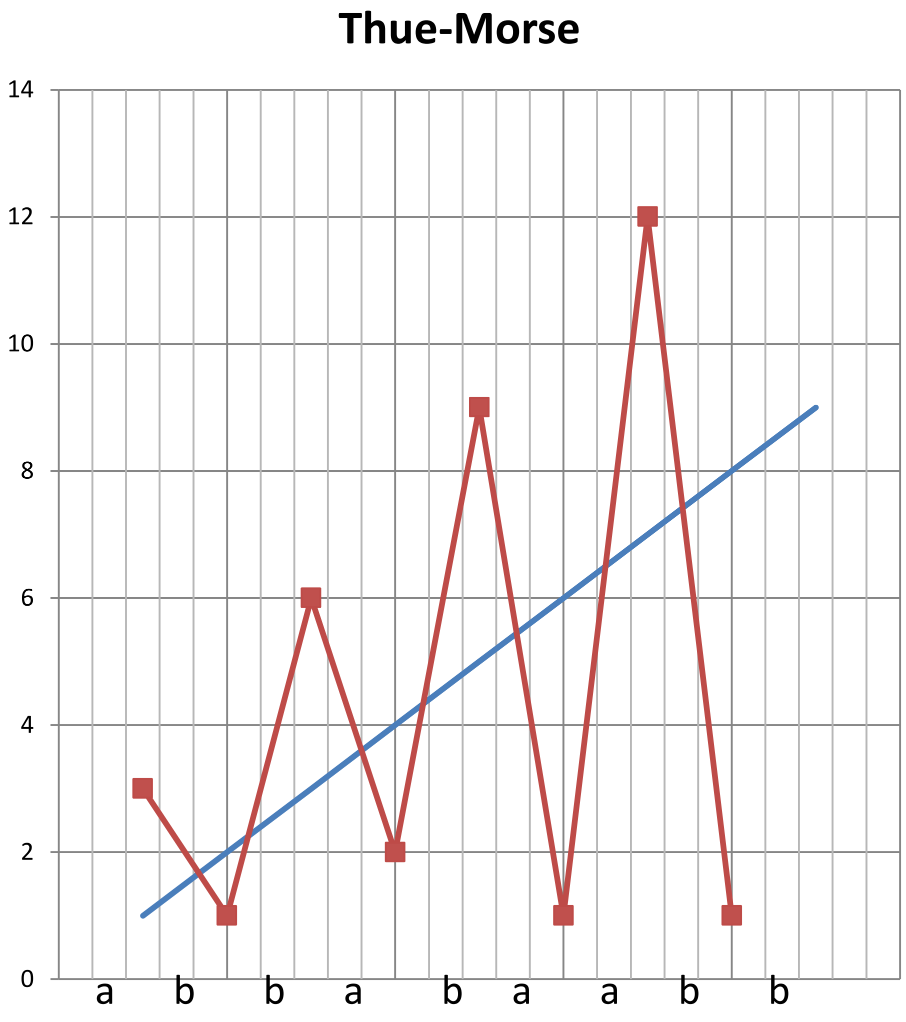
Theorem. Let x be an infinite recurrent word. Then either p_x is bounded, i.e. x is periodic, or $p_x(n) \geq n+1$, for infinitely many integers n.

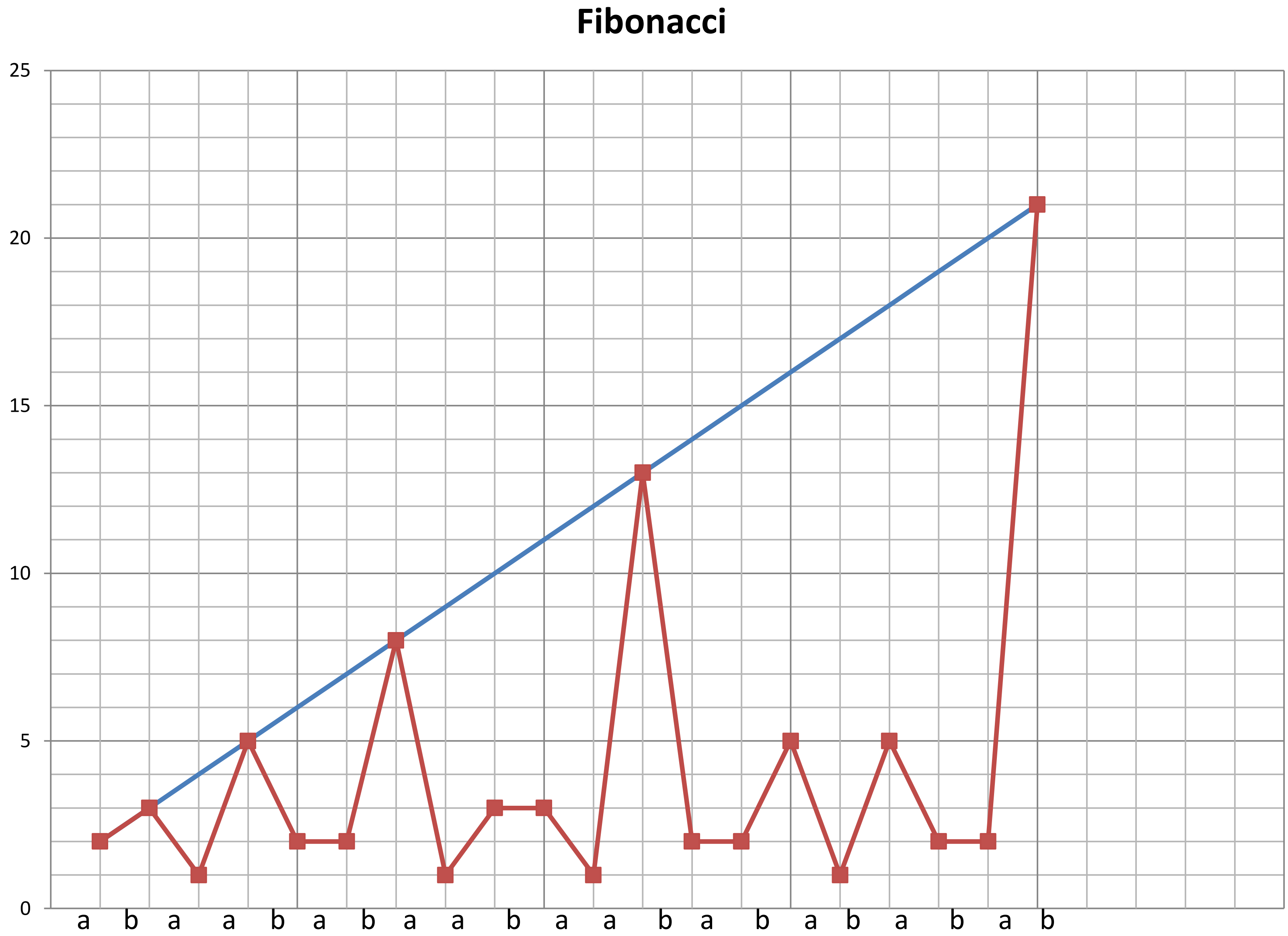
Analogous to the Coven-Hedlund theorem:

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

Periodic




Thue-Morse
14
12
10
8
6
4
2
0
a
b
b
a
b
a
a
b
b



Fibonacci
25
20
15
10
5
0
a b a a b a b a a b a a b a b a b a b a b

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 $q_0, q_1, q_2, \ldots\ldots$ 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:

$$s_0 = b$$

$$s_1 = a$$

$$s_{n+1} = s_n^{q_{n-1}} s_{n-1}$$

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 . . . . .

$p_y(2) = 5$ $p_y(5) = 8$

Theorem

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

Remark: this is not true in alphabeths having more than two letters.

b		b		c		a		c		b		c		a		b		b
	1		8		8		8		8		8		8		1			
b		b		c		a		c		b		a		c		b		b

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:

$w_0 = ab$

$w_{n+1} = w_n \, a^{2|w_n|} w_n$

$w_1 = abaaaaab$

$w_2 = abaaaaabaaaaaaaaaaaaaaaaabaaaaab$

$\mathbf{w} = \lim \; w_n$

Theorem $\limsup \; h_{\mathbf{w}}(n) = \sup h_{\mathbf{w}}(n) = 7$

The Fibonacci word

f = abaababaabaababaababaabaabab........

Theorem $h_{\mathbf{f}}(n)$ grows as Θ (log n)

The Thue-Morse word

$\mathbf{t}$ = abbabaabbaababbabaababbaabba…….

Theorem $h_{\mathbf{t}}(n)$ 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.

v_1 = ab

v_2 = aaabbabb

v_3 = aaaaababaabbbaababbbabbb

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

$z_1 = v_1$

$z_{n+1} = z_n \, b \, z_n^{[2 f(|z_n|+1)]} \, v_{n+1}$

Consider the infinite word $\mathbf{z} = \lim z_n$

Theorem For infinitely many j, $h_{\mathbf{z}}(j) > f(j)$.

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