Combinatorics on Words

Gabriele Fici

CWI, Amsterdam — April 2024

Part 4: Complexity

Some Notions of Symbolic Dynamics

The set Σ^ω of infinite words over the alphabet Σ is an ultrametric[1] space for the distance between two words x and y defined by $2^{-\delta}$, where δ is the length of the longest common prefix of x and y.

Given an infinite word $x = x_0x_1\cdots$, where $x_i \in \Sigma$, we let $S(x) = x_1x_2\cdots$ denote the shift map.

The orbit of x is the infinite set $\mathcal{O}(x) = \{S^n(x) \mid n \geq 0\}$.

The shift orbit closure $\overline{\mathcal{O}(x)}$ of x is the topological closure of the orbit of x and coincides with the set of infinite words y such that $\mathit{Fact}(y) \subseteq \mathit{Fact}(x)$.

[1] Recall that d is ultrametric if instead of the triangular inequality it verifies the stronger property $d(x, y) \leq \max\{d(x, z), d(z, y)\}$.

Some Notions of Symbolic Dynamics

For example, the complement $\bar{t} = 100101100110100 1 \cdots$ of the Thue–Morse word is in the shift orbit closure of the Thue–Morse word (since the set of factors of the Thue–Morse word is closed under taking the complement), but not in its orbit.

Every binary infinite word x is in the shift orbit closure of the Champernowne word $C_2 = 0110111001011101111000 \cdots$, since $\mathit{Fact}(x) \subset \mathit{Fact}(C_2)$.

Note that $\overline{\mathcal{O}(x)}$ is finite if and only if x is (eventually) periodic. If instead x is aperiodic, then $\overline{\mathcal{O}(x)}$ is uncountable.

Some Notions of Symbolic Dynamics

The pair $(\overline{\mathcal{O}(x)}, S)$ is called the (one-sided) subshift generated by x.

A subshift is minimal if it does not contain properly another subshift.

Theorem 1

Let x be an infinite word. The following are equivalent:

1. *the subshift $(\overline{\mathcal{O}(x)}, S)$ is minimal;*
2. *x is uniformly recurrent;*
3. *for any $y \in \overline{\mathcal{O}(x)}$, $\overline{\mathcal{O}(y)} = \overline{\mathcal{O}(x)}$;*
4. *for any $y \in \overline{\mathcal{O}(x)}$, $\mathit{Fact}(y) = \mathit{Fact}(x)$.*

Some Notions of Symbolic Dynamics

The shift orbit closure of an aperiodic word can, in general, contain a periodic word. For example, the shift orbit closure of the Sierpiński word s, fixed point of $0 \mapsto 010$, $1 \mapsto 111$, contains the word 1^ω, since $\mathit{Fact}(1^\omega) \subseteq \mathit{Fact}(s)$.

However, by Theorem 1, if x is uniformly recurrent this cannot happen (remember that the Sierpiński word is not uniformly recurrent).

Remark 2

A theorem of Furstenberg says that if x is an infinite word, then there exists a uniformly recurrent word x' such that $\mathit{Fact}(x') \subseteq \mathit{Fact}(x)$.

Recurrence Function

Recall that an infinite word x is uniformly recurrent if and only if for every finite factor u of x there exits an integer m (that depends on u) such that u occurs in every factor of x of length $m|u|$.

So for uniformly recurrent we can define the recurrence function

$R_x(n) = \inf\{m \mid \text{every factor of length } n \text{ occurs in every factor of length } m\}$

For example, the first few values of the recurrence function of the Thue–Morse word $t = 01101001100101101\cdots$ are $R_t(0) = 0$, $R_t(1) = 3$ (every factor of length 3 contains both 0 and 1), $R_t(2) = 9$, $R_t(3) = 11$, etc.

Recurrence Function

A word x is linearly recurrent (with constant m) if and only if its recurrent function is a linear function (with constant m), that is, $R_x(n) \leq mn$, for every $n \geq 0$.

The recurrence quotient of an infinite word x is defined as

$$\rho_x = \limsup_{n \to \infty} \frac{R_x(n)}{n}$$

It can be proved that for every aperiodic word x, $\rho_x \geq 3$.

Conjecture 3 (Rauzy, 1982)

For every aperiodic word x, $\rho_x \geq \varphi + 2 \approx 3.618$

The value $\varphi + 2$ is exactly the recurrence quotient of the Fibonacci word, so if the conjecture holds, then it is optimal.

Recurrence Function

A variation of the recurrence function is the prefix recurrence function

$R'_x(n) = \inf\{m \mid \text{every factor of length } n \text{ occurs in the prefix of length } m\}$

Of course, for every n, one has $R'_x(n) \leq R_x(n)$.

Now, one can define the prefix recurrence quotient of x as

$$\rho'_x = \limsup_{n \to \infty} \frac{R'_x(n)}{n}$$

Recurrence Function

With respect to ρ'_x, we can state the following conjecture, analogous Conjecture 3.

Conjecture 4

For every aperiodic word x, $\rho'_x \geq \varphi + 1 = \varphi^2 \approx 2.618$

The value $\varphi + 1$ is exactly the prefix recurrence quotient of the Fibonacci word.

But actually Cassaigne disproved this conjecture, by proving that for every aperiodic word x,

$$\rho'_x \geq \frac{29 - 2\sqrt{10}}{9} \approx 2.52$$

and this value is actually attained by the fixed point of $0 \mapsto 01001010,\ 1 \mapsto 010$.

Recurrence Function

Cassaigne also defined the factor recurrence function

$R''_x(n) = \inf\{|v| \mid v \text{ is a factor and every factor of length } n \text{ occurs in } v\}$

For the quantity

$$\rho''_x = \limsup_{n\to\infty} \frac{R''_x(n)}{n}$$

there is no analogous conjecture; in fact for every aperiodic word x one has $\rho''_x \geq 2$ and the value 2 actually characterizes aperiodic words with minimal factor complexity (i.e., Sturmian words).

Return Words

Definition 5

Let x be a recurrent word, and u a nonempty factor of x. We say that a word w is a return to u if wu is a factor of x and contains exactly two occurrences of u (one as a prefix and one as a suffix), i.e., if wu is a complete return to u.

In other words, given a factor u of a recurrent word x, we know that u must eventually reoccur in x, and we consider the portions of x between two consecutive occurrences of u in x.

For example, in the Fibonacci word

$$f = 0100\mathbf{101}00100\mathbf{101}00\mathbf{101}00100\mathbf{101}00100\mathbf{101}00\cdots$$

the returns to $u = 101$ are $w = 10100$ and $w' = 10100100$.

The factor w can be of any positive length. For example, if a^{n+1} is a factor of x, for a letter $a \in \Sigma$, then $w = a$ is a return word to $u = a^n$.

Return Words

We let $\mathcal{R}_x(u)$ denote the set of returns to u in x. We will omit the subscript when the word x is clear from the context.

Notice that if x is uniformly recurrent, then for every factor u of x, $\mathcal{R}_x(u)$ is finite, since u occurs in x with bounded gaps.

Proposition 6

Let $w, w' \in \mathcal{R}_x(u)$ be distinct. Then w and w' do not overlap in x.

Proof.

If w and w' occur overlapping in x, one of the two contains an internal occurrence of u, against the definition of return word. □

Derived Words

We know that in a uniformly recurrent word, the set of returns to any factor is finite.

So, for any nonempty prefix u of a uniformly recurrent word x, we can factor x using returns to u. If we call these returns $0, 1, \ldots, k-1$, we get a new word over the alphabet Σ_k, called the derived word of x w.r.t. the prefix u, noted $D_u(x)$.

Since x is uniformly recurrent, also $D_u(x)$ is.

For example, the prefix $u = 011$ of the Thue–Morse word t has 4 returns: $\mathcal{R}_t(011) = \{011010, 011001, 01101001, 0110\}$.
So, we get

$$t = 011010 \cdot 011001 \cdot 01101001 \cdot 0110 \cdot 011010 \cdot 011001 \cdot 0110 \cdot 011 \cdots$$

so that

$$D_{011}(t) = 0123013 \cdots$$

Derived Words

A word x is primitive morphic if x is obtained by applying a coding to a fixed point of a primitive morphism.

So, primitive morphic words are a subclass of morphic words.

Theorem 7 (Durand, 1998)

Let x be a uniformly recurrent word. Then, x is primitive morphic if and only if the set $\{D_u(x) \mid u \in \mathit{Pref}(x), u \neq \varepsilon\}$ of its derived words is finite.

Exercise 8

What is the cardinality of the set of derived words of the Thue–Morse word?

Factor Complexity

In order to study the complexity of a word one can define a quantitative measure of the degree of variety of patterns that can appear in the word.

The most natural patterns are factors, and the most natural measure is counting the number of distinct factors of each length.

Definition 9

The factor complexity of a finite or infinite word w over the alphabet Σ is the function defined by $f_w(n) = |\mathit{Fact}(w) \cap \Sigma^n|$, for every $n \geq 0$.

Notice that $f_w(1)$ is the number of distinct letters occurring in w.

Factor Complexity

Example 10 (Brlek)

The factor complexity $f_t(n)$ of the Thue–Morse word t is the function defined as follows: $f_t(1) = 2$, $f_t(2) = 4$ and for $n \geq 3$, let $n = 2^r + q + 1$, $r \geq 0$, $0 \leq q < 2^r$; then:

$$f_t(n) = \begin{cases} 6 \cdot 2^{r-1} + 4q & \text{if } \ 0 \leq q \leq 2^{r-1}; \\ 2^{r+2} + 2q & \text{if } \ 2^{r-1} < q < 2^r. \end{cases}$$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$f_t(n)$	2	4	6	10	12	16	20	22	24	28	32	36	40	42	44	46

Table: The first few values of the complexity function of the Thue–Morse word $t = 01101001100101101\cdots$.

Factor Complexity

It is possible to construct infinite words with maximal factor complexity $f_w(n) = |\Sigma|^n$. For example, over Σ_k, the infinite word obtained by concatenating the k-ary representation of n for every $n \geq 0$ is called the Champernowne word C_k and has factor complexity k^n.

Any word with maximal factor complexity is recurrent but cannot be uniformly recurrent (it contains arbitrarily large powers of a single letter, so no other letter can occur with bounded gaps).

However, if a word does not contain Σ^* as set of factors, then it cannot have arbitrary factor complexity:

Theorem 11

Let x be an infinite word over Σ such that Fact$(x) \neq \Sigma^*$. Then there exists a real number α, with $1 < \alpha < |\Sigma|$, such that $f_x(n) = O(\alpha^n)$.

For example, a binary infinite word cannot have factor complexity $2^n/n$.

Factor Complexity

The basic result on the factor complexity is the following theorem of Hedlund and Morse:

Theorem 12 (Hedlund, Morse, 1938)

An infinite word x is aperiodic if and only if $f_x(n) \geq n + 1$ for every $n \geq 0$.

The previous theorem establishes a threshold on the factor complexity function that allows one to distinguish between periodic and aperiodic words.

Factor Complexity

The factor complexity of the Chacon word is $2n - 1$ for every $n \geq 2$.

The factor complexity of the Rudin–Shapiro word r verifies $f_r(n) = 8n - 8$ for every $n \geq 7$.

The factor complexity of the von Neumann word $v = 0010011001001110010011001\cdots$ is

$$f_v(n) = \sum_{i=0}^{n} \min(2^i, n - i + 1)$$

which is exactly the maximal number of distinct factors of a binary word of length n, although no explicit bijection is known. Notice that one has $f_v(n) \in \Theta(n^2)$.

A conjecture of Dekking is that the factor complexity of the Oldenburger–Kolakoski word k verifies $f_k(n) = \Theta(n^q)$, for $q = \frac{\log 3}{\log \frac{3}{2}} \approx 2.7095$.

Factor Complexity

Definition 13

A word x such that $f_x(n) = n + 1$ for every $n \geq 0$ is called a Sturmian word.

Sturmian words are therefore aperiodic words with minimal factor complexity.

The Fibonacci word f is an example of a Sturmian word.

The Thue–Morse word, instead, is not Sturmian, since it has 4 distinct factors of length 2.

Note that a Sturmian word is a binary word, since it must verify $f(1) = 2$.

Although the Fibonacci word is a fixed point of a morphism, not all Sturmian words are. Note also that every infinite suffix of a Sturmian word is a Sturmian word, by definition.

Factor Complexity

Definition 14

A word x is quasi-Sturmian if it has factor complexity $f_x(n) = n + c$ for some $c \geq 1$ and for every $n \geq n_0$.

The word

$$\delta(f) = 0100010100010001010001010001000101000100010100010001010001\cdots$$

where δ is the period-doubling morphism, is a quasi-Sturmian word, since it has factor complexity $n + 2$ for every $n \geq 3$.

The word

$$\mu(f) = 0011000011001100001100001100110000110011000011000011000\cdots$$

image of the Fibonacci word under the doubling letter morphism $\mu : 0 \mapsto 00,\ 1 \mapsto 11$, has factor complexity $n + 3$ for every $n \geq 4$.

Cassaigne characterized quasi-Sturmian words as those words that are of the form $w\mu(x)$ where w is a finite word, x is a Sturmian word, and μ is an acyclic binary morphism (i.e., such that that $\mu(01) \neq \mu(10)$).

Factor Complexity

Definition 15

A word x such that $f_x(n) = 2n$ for every $n \geq 0$ is called a Rote word.

All Stewart words have factor complexity $2n$, i.e., they are Rote words.

An example of a Rote word that is not uniformly recurrent is the fixed point of $0 \mapsto 001, 1 \mapsto 111$

$$0010011110010011111111111110010011110010011 1 \cdots$$

Factor Complexity

With respect to the recurrence functions, Cassaigne proved the following inequality

Theorem 16

Let x be an infinite word. For every $n \geq 0$,

$$f_x(n) + n - 1 \leq R''_x(n) \leq R'_x(n) \leq R_x(n).$$

Factor Complexity of Linearly Recurrent Words

For linearly recurrent words, the factor complexity is at most linear.

Recall that $R_x(n) \leq mn$ for every $n \geq 0$ if and only if the distance between two consecutive occurrences of a factor of length n of x is at most $(m-1)n+1$.

Theorem 17 (Durand, Host, Skau, 1999)

Let x be an aperiodic linearly recurrent infinite word, i.e., there exists an integer $m > 0$ such that $R_x(n) \leq mn$ for every $n \geq 0$. Then:

1. *x is m-free;*
2. *for every $n \geq 0$, $f_x(n) \leq (m-1)n+1$;*
3. *for every nonempty $u \in \mathit{Fact}(x)$, if $w \in \mathcal{R}(u)$, then $\frac{|u|}{m} < |w| \leq m|u|$;*
4. *for every nonempty $u \in \mathit{Fact}(x)$, $|\mathcal{R}(u)| \leq m^3$;*

Factor Complexity of Linearly Recurrent Words

For 1, suppose x contains v^m for some nonempty v. Since $|v^m| = m|v|$, v^m must contain all factors of x of length $|v|$. But v^m contains at most $|v|$ distinct factors of length $|v|$, and an aperiodic word must contain at least $n+1$ distinct factors of length n for every n by Theorem 12, contradiction.

For 2, let v be any factor of length mn of x. Hence, v contains all factors of length n of x. Since $|v| = mn$, v contains at most $mn - n + 1$ distinct factors of length n (in general, a word of length ℓ contains at most $\ell - t + 1$ factors of length t for every $t \leq \ell$). Therefore, the distinct factors of x of length n cannot be more than $mn - n + 1 = (m-1)n + 1$.

Factor Complexity of Linearly Recurrent Words

Let us now prove 3. Since two consecutive occurrences of u in x are at distance at most $(m-1)|u|+1$, every word $w \in \mathcal{R}(u)$ has a length $|w| \leq m|u|$. For the other inequality, suppose $|w| \leq |u|/m$. Since wu has u as a border, wu has period $|wu| - |u| = |w| \leq |u|/m$ (i.e., the two occurrences of u in wu overlap). We deduce that wu has w^m as a prefix, but this contradicts 1.

For 4, let v be any factor of length $m^2|u|$ of x. Every word $w \in \mathcal{R}(u)$ has a length $|w| \leq m|u|$ (by 3), hence it must have an occurrence in v. Always by 3, every word $w \in \mathcal{R}(u)$ has a length $|w| \geq |u|/m$, and since by Proposition 6 returns to the same word do not overlap, v can contain at most $|\mathcal{R}(u)| \leq |v|/(|u|/m) = m|v|/|u| = m^3$ returns to u.

Factor Complexity of Pure Morphic Words

If a word is a fixed point of a primitive morphism, then it is linearly recurrent, and in this case its factor complexity is sublinear.

In general, though, words that are fixed points of a morphism, not necessarily primitive, cannot have arbitrary complexity:

Theorem 18 (Pansiot, 1984)

Let x be a fixed point of a morphism. Then one of the following holds true:

1. $f_x(n) = \Theta(1)$ *(e.g.,* $0 \mapsto 01$, $1 \mapsto 01$*);*
2. $f_x(n) = \Theta(n)$ *(e.g.,* $0 \mapsto 01$, $1 \mapsto 0$, *Fibonacci);*
3. $f_x(n) = \Theta(n \log\log n)$ *(e.g.,* $0 \mapsto 0101$, $1 \mapsto 11$*);*
4. $f_x(n) = \Theta(n \log n)$ *(e.g.,* $0 \mapsto 012$, $1 \mapsto 0$, $2 \mapsto 23$*);*
5. $f_x(n) = \Theta(n^2)$ *(e.g.,* $0 \mapsto 012$, $1 \mapsto 12$, $2 \mapsto 2$*; or* $0 \mapsto 001$, $1 \mapsto 1$, *von Neumann).*

Moreover, if the morphism is uniform (the images of the letters all have the same length) then only Cases 1 and 2 are possible.

Factor Complexity of Pure Morphic Words

However, if a word is morphic but not pure morphic (that is, it is obtained from a fixed point of a morphism after applying a coding), then it can have other factor complexities.

For example, Pansiot proved that there exists a binary morphic word whose factor complexity is $\Theta(n\sqrt{n})$.

Devyatov proved that there exist morphic words with factor complexity $\Theta(n^{1+\frac{1}{\ell}})$, for every positive integer ℓ.

On the other hand, also words that are not fixed points of a morphism can have linear factor complexity:

Proposition 19

All the paperfolding words have the same factor complexity $f(n)$; moreover, $f(n) = 4n$ for every $n \geq 7$.

Factor Complexity of Toeplitz Words

Recall that the regular paperfolding word is a simple Toeplitz word, that is, it is generated by a single partial word.

Theorem 20 (Cassaigne, Karumäki, 1997)

Let x be an aperiodic simple Toeplitz word generated by a partial word P of length p containing q occurrences of ?.

Let $d = \gcd(p, q)$ and $p' = p/d$, $q' = q/d$.

Then $f_x(n) = \Theta(n^r)$, where $r = \dfrac{\log p'}{\log p' - \log q'}$.

In particular, if q divides p, then the factor complexity is linear. We know that in this case the Toeplitz word is q-automatic (all automatic words have linear factor complexity).

Topological Entropy

Let x be an infinite word over Σ_k. For every n, any factor of x of length n can be extended on the right by one letter into a factor of length $n+1$ in at most k ways. Therefore, $f_x(n) \leq k f_x(n-1)$.

More generally:

Proposition 21

Let x be an infinite word. Then, for all integers m and n, one has $f_x(m+n) \leq f_x(m) f_x(n)$.

Proof.

Every factor of length $m+n$ has a prefix of length m that can be chosen in $f_x(m)$ possible ways and a suffix of length n that can be chosen in $f_x(n)$ possible ways. □

In other words, the real-valued function $\log f_x$ is subadditive.
Theorem 11 is in fact a direct consequence of the previous proposition.

Topological Entropy

Lemma 22 (Fekete, 1923)

Let (a_n) be a sequence of real numbers such that $a_{m+n} \leq a_m + a_n$ (subadditive). Then $\lim_{n\to\infty} \frac{a_n}{n}$ exists and is equal to $\inf \frac{a_n}{n}$.

Since by Proposition 21 the function $\log f_x$ is subadditive, we therefore have that

$$\lim_{n\to\infty} \frac{\log f_x(n)}{n}$$

always exists and is a constant h such that $0 \leq h \leq \log k$, where k is the size of the alphabet of x; the constant h is called the topological entropy of the infinite word x. Moreover,

$$h = \liminf \frac{\log f_x(n)}{n}.$$

Notice that $h > 0$ if and only if $f_x(n)$ is exponential. Words with polynomial factor complexity have null topological entropy.

Topological Entropy

So, a natural question is whether it is possible to construct words with arbitrary topological entropy.

Theorem 23 (Grillenberger, 1972)

For every real number $h \in (0, \log k)$ there exists a uniformly recurrent word over Σ_k, $k \geq 2$, whose topological entropy is equal to h.

The proof is constructive and the word obtained is actually a Toeplitz word.

Special Factors

The factor complexity is strictly related to the special factors.

Definition 24

A factor v of a finite or infinite word w is right (resp., left) special if there exist two different letters a and b in Σ such that va and vb (resp., av and bv) are both factors of w.

A factor is called bispecial if it is right and left special.

Special Factors

Theorem 25

An infinite word is aperiodic if and only if it has at least one right special factor for each length.

Proof.

Suppose x is aperiodic and take $n \in \mathbb{N}$. Since the set of factors of length n of x is finite, there is one factor w of x of length n that occurs at positions i and j, with $i < j$. There exists m such that the letters x_{i+m} and x_{j+m} are different, otherwise $j - i$ would be a period of x; and $m \geq n$ since the same word w of length n occurs at positions i and j. Let $w' = x_i x_{i+1} \cdots x_{i+m-1}$. Then w' is right special, and so is its suffix of length n. So, there is at least one right special factor of each length.

If x is not aperiodic, then it has the form $x = uv^\omega$ for some u and v. Now, for sufficiently large n, the number of distinct factors of length n in x is constant (at most $|u| + |v|$). Hence, there are no right special factors of length n, since a right special factor of length n makes the number of factors of length $n + 1$ strictly greater than those of length n. □

Special Factors

Sturmian words have exactly one right special factor for each length. Indeed, in general, the number of factors of length n of an infinite binary infinite word is equal to the number of factors of length $n-1$ plus the number of right special factors of length $n-1$ (since a right special factor can be extended in exactly two ways, while any other factor in one way).

Since the factor complexity of a Sturmian word is $n+1$ for every n, there must be exactly one right special factor for each length.

And conversely, if a binary word has exactly one right special factor for each length, then its factor complexity must be $n+1$ for every n, hence it is a Sturmian word.

The same also holds, of course, if one takes left special factors instead of right special ones (assuming recurrence).

Special Factors

Proof of Morse–Hedlund Theorem.

If x is aperiodic, x has at least one right special factor for each length by Theorem 25. Hence, $f_x(n+1) > f_x(n)$ for every n, since every right special factor can be extended by at least two different letters. Since $f_x(1) \geq 2$ (if $f_x(1) = 1$, x cannot be aperiodic), we have $f_x(n) > n$ for every n.

If x is not aperiodic, then it has the form $x = uv^\omega$ for some words u and v. For sufficiently large n, the number of distinct factors of length n in x is constant (at most $|u| + |v|$). Hence, we cannot have $f_x(n) > n$ for every n. □

Special Factors

Extensions of the class of Sturmian word to larger alphabet exist based on the notion of special factors.

Definition 26

An Arnoux–Rauzy word is a word over Σ_k, $k \geq 2$, having exactly one left and one right special factor of each length, and these can be extended with every letter of the alphabet (i.e., they have degree k).

Equivalently, an Arnoux–Rauzy word is a recurrent infinite word having factor complexity $(k-1)n+1$ and exactly one left special and one right special factor of each length n.

An example of an Arnoux–Rauzy word is the Tribonacci word tr (or more generally, all m-bonacci words).

Special Factors

A remarkable property that is shared by Sturmian and Arnoux–Rauzy words is that the set of factors is closed under reversal.

This nice property inspired Droubay, Justin, and Pirillo to introduce the following generalization of Sturmian words:

Definition 27

An infinite word is an episturmian word if it is closed under reversal and has at most one left special factor of each length, but not necessarily of degree k.

Arnoux–Rauzy words are sometimes called strict episturmian words.

Note that, by definition, an episturmian word may be periodic.

Special Factors

Bispecial factors and their returns can be used to compute the critical exponent.

Theorem 28

Let x be a recurrent aperiodic word. Let (b_n) be a sequence of all bispecial factors of x ordered by their length. For every $n \in \mathbb{N}$, let r_n be a shortest return word to b_n in x. Then the critical exponent $ce(x)$ of x is equal to

$$ce(x) = 1 + \sup_{n \in \mathbb{N}} \left\{ \frac{|b_n|}{|r_n|} \right\}$$

As a consequence, for aperiodic linearly recurrent words the critical exponent is always finite.

Palindromic Complexity

Given an infinite word x, one can count the number of factors of x of length n that are palindromes, for every n. This function, $PAL_x(n)$, is called the palindromic complexity of the word x.

Trivially, one has $PAL_x(n) \leq f_x(n)$, but the following bound, due to J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, is not trivial:

Theorem 29

Let x be an aperiodic word. Then

$$PAL_x(n) \leq \frac{16}{n} f_x\left(n + \left\lfloor \frac{n}{4} \right\rfloor\right).$$

Palindromic Complexity

Damanik and Zare proved that the palindromic complexity of a fixed point of a primitive morphism is bounded. The same holds for fixed point of uniform morphisms (not necessarily primitive).

For example, the palindromic complexity of the Fibonacci word f is 2 for odd n and 1 for even n. (Actually, this is the palindromic complexity of any Sturmian word).

The palindromic complexity of the Tribonacci word tr is 3 for odd n and 1 for even n.

Palindromic Complexity

Another consequence is that if the factor complexity is at most linear, then the palindromic complexity is bounded (but the converse is not true in general).

Baláži, Masáková, and Pelantová proved that for uniformly recurrent infinite words such that the set of factors is closed under reversal (otherwise $PAL_x(n)$ eventually vanishes), one has for all $n \in \mathbb{N}$,

$$PAL_x(n) + PAL_x(n+1) \leq f_x(n+1) - f_x(n) + 2.$$

Palindromic Complexity

Infinite words closed under reversal for which $PAL_x(n) + PAL_x(n+1)$ always reach the upper bound given in the previous relation are precisely recurrent infinite rich words.

This result can be viewed as a characterization of recurrent rich infinite words, since any rich infinite word is recurrent if and only if its set of factors is closed under reversal.

As a consequence, any infinite word with sublinear factor complexity has bounded palindromic complexity, since the first difference $f_x(n+1) - f_x(n)$ is bounded for any such infinite word.

On the other hand, there also exist recurrent infinite rich words with superlinear factor complexity. An example is the von Neumann word $v = 0010011001001110010011001 00 \cdots$, fixed point of the morphism $0 \mapsto 001, 1 \mapsto 1$, whose factor complexity grows like n^2.