

Open and closed complexity of infinite words

Mickaël Postic (joint work with Olga Parshina)

Université Lyon 1 - Institut Camille Jordan

Monday, March 1st 2021

Table of contents

- 1 Complexity functions
- 2 Open and closed complexity functions
- 3 Words with bounded open complexity are ultimately periodic
- 4 Words with bounded closed complexity are ultimately periodic

Table of content

- 1 Complexity functions
- 2 Open and closed complexity functions
- 3 Words with bounded open complexity are ultimately periodic
- 4 Words with bounded closed complexity are ultimately periodic

Classical complexity function $p_w(n)$

Definition

For an infinite word w , we define its **complexity function** p_w which associates to every non-negative integer n the number of distinct factors of length n in $\text{Fact}(w)$; so we get

$$p_w(n) = \text{Card}(\text{Fact}_n(w)).$$

Classical complexity function $p_w(n)$

Definition

For an infinite word w , we define its **complexity function** p_w which associates to every non-negative integer n the number of distinct factors of length n in $\text{Fact}(w)$; so we get

$$p_w(n) = \text{Card}(\text{Fact}_n(w)).$$

Example

For the regular paperfolding word

$$\mathbf{p} = 0010011000110110001001110011011 \dots$$

we have $p_{\mathbf{p}}(n) = 4n$ for $n \geq 7$ (Allouche).

Classical complexity function $p_w(n)$

Definition

For an infinite word w , we define its **complexity function** p_w which associates to every non-negative integer n the number of distinct factors of length n in $\text{Fact}(w)$; so we get

$$p_w(n) = \text{Card}(\text{Fact}_n(w)).$$

Example

For the regular paperfolding word

$$\mathbf{p} = 0010011000110110001001110011011 \dots$$

we have $p_{\mathbf{p}}(n) = 4n$ for $n \geq 7$ (Allouche). For the Fibonacci word

$$\mathbf{f} = 0100101001001010010100100101001001 \dots$$

the complexity function is $p_{\mathbf{f}}(n) = n + 1$ for every integer n .

Morse and Hedlund's theorem

Theorem (Morse-Hedlund)

An infinite word w is aperiodic if and only if its complexity function is unbounded. If w is aperiodic, we have, for any integer n , $p_w(n) \geq n + 1$.

Morse and Hedlund's theorem

Theorem (Morse-Hedlund)

An infinite word w is aperiodic if and only if its complexity function is unbounded. If w is aperiodic, we have, for any integer n , $p_w(n) \geq n + 1$.

Remark

- ▶ The proof is quite direct since the complexity function is non-decreasing.

Morse and Hedlund's theorem

Theorem (Morse-Hedlund)

An infinite word w is aperiodic if and only if its complexity function is unbounded. If w is aperiodic, we have, for any integer n , $p_w(n) \geq n + 1$.

Remark

- ▶ The proof is quite direct since the complexity function is non-decreasing.
- ▶ The Fibonacci word has the minimal complexity for an aperiodic infinite word. Words verifying $p_w(n) \geq n + 1$ for every integer n are called *Sturmian words* and have many other interesting properties.

Some generalizations

Other complexity functions have been studied : *abelian complexity*,

Some generalizations

Other complexity functions have been studied : *abelian complexity*,
palindromic complexity,

Some generalizations

Other complexity functions have been studied : *abelian complexity*, *palindromic complexity*, *cyclic complexity*, ...

Some generalizations

Other complexity functions have been studied : *abelian complexity*, *palindromic complexity*, *cyclic complexity*, ...

These complexity function are usually *not* non-decreasing.

Some generalizations

Other complexity functions have been studied : *abelian complexity*, *palindromic complexity*, *cyclic complexity*, ...

These complexity function are usually *not* non-decreasing.

So Morse and Hedlund's theorem is not always valid.

Some generalizations

Other complexity functions have been studied : *abelian complexity*, *palindromic complexity*, *cyclic complexity*, ...

These complexity function are usually *not* non-decreasing.

So Morse and Hedlund's theorem is not always valid.

Example

For the Fibonacci word, the abelian complexity is constant, and for the Thue-Morse word, $a_t(2) = 3 > a_t(3) = 2 > a_t(4) = 1$.

Table of content

- 1 Complexity functions
- 2 Open and closed complexity functions
- 3 Words with bounded open complexity are ultimately periodic
- 4 Words with bounded closed complexity are ultimately periodic

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Example

The words a and aa are borders of $aabaaa$

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Example

The words a and aa are borders of $aabaaa$

Definition

We say $w \in \mathbb{A}^+$ is **closed** if either $w \in \mathbb{A}$ or w admits a border u which occurs precisely twice in w .

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Example

The words a and aa are borders of $aabaaa$

Definition

We say $w \in \mathbb{A}^+$ is **closed** if either $w \in \mathbb{A}$ or w admits a border u which occurs precisely twice in w . Otherwise w is said to be **open**.

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Example

The words a and aa are borders of $aabaaa$

Definition

We say $w \in \mathbb{A}^+$ is **closed** if either $w \in \mathbb{A}$ or w admits a border u which occurs precisely twice in w . Otherwise w is said to be **open**.

The longest border of a closed word is called its **frontier**.

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Example

The words a and aa are borders of $aabaaa$

Definition

We say $w \in \mathbb{A}^+$ is **closed** if either $w \in \mathbb{A}$ or w admits a border u which occurs precisely twice in w . Otherwise w is said to be **open**.

The longest border of a closed word is called its **frontier**.

- ▶ $w = abaaaab$ is closed (with frontier ab),

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Example

The words a and aa are borders of $aabaaa$

Definition

We say $w \in \mathbb{A}^+$ is **closed** if either $w \in \mathbb{A}$ or w admits a border u which occurs precisely twice in w . Otherwise w is said to be **open**.

The longest border of a closed word is called its **frontier**.

- ▶ $w = abaaaab$ is closed (with frontier ab),
- ▶ $aabab$ is open,

Open and closed words

Given $u, w \in \mathbb{A}^+$, we say u is a **border** of w if u is both a proper prefix and suffix of w .

Example

The words a and aa are borders of $aabaaa$

Definition

We say $w \in \mathbb{A}^+$ is **closed** if either $w \in \mathbb{A}$ or w admits a border u which occurs precisely twice in w . Otherwise w is said to be **open**.

The longest border of a closed word is called its **frontier**.

- ▶ $w = abaaaab$ is closed (with frontier ab),
- ▶ $aabab$ is open,
- ▶ $aabaaa$ is open.

Remark

- ▶ Closed words are frequently called **complete first returns**, they play a central part in symbolic dynamics.
- ▶ In a recurrent word, every factor admits complete first returns to itself.
- ▶ In a uniformly recurrent word, every factor admits finitely many complete first returns to itself and for a prefix of the word, those give a new coding of the word.

Open and closed complexity functions

Definition

For a word $x \in \mathbb{A}^{\mathbb{N}}$ we define the **closed complexity function** Cl_x which associates to each $n \in \mathbb{N}$, the number of closed factors of x of length n .

Open and closed complexity functions

Definition

For a word $x \in \mathbb{A}^{\mathbb{N}}$ we define the **closed complexity function** Cl_x which associates to each $n \in \mathbb{N}$, the number of closed factors of x of length n . Similarly, we define the **open complexity function** Op_x which associates to each $n \in \mathbb{N}$, the number of open factors of x of length n .

Open and closed complexity functions

Definition

For a word $x \in \mathbb{A}^{\mathbb{N}}$ we define the **closed complexity function** Cl_x which associates to each $n \in \mathbb{N}$, the number of closed factors of x of length n . Similarly, we define the **open complexity function** Op_x which associates to each $n \in \mathbb{N}$, the number of open factors of x of length n .

Example

$Cl_{(abba)^\omega}(2) = 2$ since aa and bb are the only closed factors of length 2

Open and closed complexity functions

Definition

For a word $x \in \mathbb{A}^{\mathbb{N}}$ we define the **closed complexity function** Cl_x which associates to each $n \in \mathbb{N}$, the number of closed factors of x of length n . Similarly, we define the **open complexity function** Op_x which associates to each $n \in \mathbb{N}$, the number of open factors of x of length n .

Example

$\text{Cl}_{(abba)^\omega}(2) = 2$ since aa and bb are the only closed factors of length 2 while $\text{Cl}_{(abba)^\omega}(3) = 0$.

Open and closed complexity functions

Definition

For a word $x \in \mathbb{A}^{\mathbb{N}}$ we define the **closed complexity function** Cl_x which associates to each $n \in \mathbb{N}$, the number of closed factors of x of length n . Similarly, we define the **open complexity function** Op_x which associates to each $n \in \mathbb{N}$, the number of open factors of x of length n .

Example

$\text{Cl}_{(abba)^\omega}(2) = 2$ since aa and bb are the only closed factors of length 2 while $\text{Cl}_{(abba)^\omega}(3) = 0$.

Similarly, $\text{Op}_{(abba)^\omega}(2) = 2$ since ab and ba are open ;

Open and closed complexity functions

Definition

For a word $x \in \mathbb{A}^{\mathbb{N}}$ we define the **closed complexity function** Cl_x which associates to each $n \in \mathbb{N}$, the number of closed factors of x of length n . Similarly, we define the **open complexity function** Op_x which associates to each $n \in \mathbb{N}$, the number of open factors of x of length n .

Example

$\text{Cl}_{(abba)^\omega}(2) = 2$ since aa and bb are the only closed factors of length 2 while $\text{Cl}_{(abba)^\omega}(3) = 0$.

Similarly, $\text{Op}_{(abba)^\omega}(2) = 2$ since ab and ba are open ; $\text{Op}_{(abba)^\omega}(3) = 4$.

Open and closed complexity functions

Definition

For a word $x \in \mathbb{A}^{\mathbb{N}}$ we define the **closed complexity function** Cl_x which associates to each $n \in \mathbb{N}$, the number of closed factors of x of length n . Similarly, we define the **open complexity function** Op_x which associates to each $n \in \mathbb{N}$, the number of open factors of x of length n .

Example

$Cl_{(abba)^\omega}(2) = 2$ since aa and bb are the only closed factors of length 2 while $Cl_{(abba)^\omega}(3) = 0$.

Similarly, $Op_{(abba)^\omega}(2) = 2$ since ab and ba are open ; $Op_{(abba)^\omega}(3) = 4$.

Remark

For every infinite word w and integer n we have :

$$Op_w(n) + Cl_w(n) = p_w(n).$$

Theorem (O. Parshina, P. 2020)

Let $x \in \mathbb{A}^{\mathbb{N}}$ be a right-infinite word over a finite alphabet \mathbb{A} . The following are equivalent :

1. x is aperiodic ;
2. $\liminf_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty$.
3. $\limsup_{n \rightarrow +\infty} \text{Cl}_x(n) = +\infty$;

Remarks

Using Morse and Hedlund, it is easy to get and

$$\liminf_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty \implies x \text{ is aperiodic}$$

and

$$\limsup_{n \rightarrow +\infty} \text{Cl}_x(n) = +\infty \implies x \text{ is aperiodic.}$$

Remarks

Using Morse and Hedlund, it is easy to get and

$$\liminf_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty \implies x \text{ is aperiodic}$$

and

$$\limsup_{n \rightarrow +\infty} \text{Cl}_x(n) = +\infty \implies x \text{ is aperiodic.}$$

With $\liminf \text{Op}_x(n) = +\infty \Leftrightarrow x$ is aperiodic, we get that

$$\liminf_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty \Leftrightarrow \limsup_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty.$$

Remarks

Using Morse and Hedlund, it is easy to get and

$$\liminf_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty \implies x \text{ is aperiodic}$$

and

$$\limsup_{n \rightarrow +\infty} \text{Cl}_x(n) = +\infty \implies x \text{ is aperiodic.}$$

With $\liminf \text{Op}_x(n) = +\infty \Leftrightarrow x$ is aperiodic, we get that

$$\liminf_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty \Leftrightarrow \limsup_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty.$$

We can't have \liminf in $\limsup \text{Cl}_x(n) = +\infty \Leftrightarrow x$ is aperiodic : for the paperfolding words, it is known that

$$\liminf_{n \rightarrow +\infty} \text{Cl}_x(n) = 0.$$

Table of content

- 1 Complexity functions
- 2 Open and closed complexity functions
- 3 Words with bounded open complexity are ultimately periodic
- 4 Words with bounded closed complexity are ultimately periodic

First part of the theorem : Open complexity function

Theorem

Let x be an infinite word over a finite alphabet \mathbb{A} . Let $k \in \mathbb{N}$ be such that $\liminf \text{Op}_x(n) = k$. Then x is ultimately periodic.

Rauzy graphs

For $x \in \mathbb{A}^{\mathbb{N}}$ and $n \in \mathbb{N}$, the **Rauzy graph** of order n of $x \in \mathbb{A}^{\mathbb{N}}$ is the directed graph whose set of vertices (resp. edges) consists of all factors of x of length n (resp. $n + 1$). There is a directed edge from u to v labeled w if u is a prefix of w and v a suffix of w .

Rauzy graphs

For $x \in \mathbb{A}^{\mathbb{N}}$ and $n \in \mathbb{N}$, the **Rauzy graph** of order n of $x \in \mathbb{A}^{\mathbb{N}}$ is the directed graph whose set of vertices (resp. edges) consists of all factors of x of length n (resp. $n + 1$). There is a directed edge from u to v labeled w if u is a prefix of w and v a suffix of w .

The **distance** between two vertices in a Rauzy graph is the length of the shortest directed path between them.

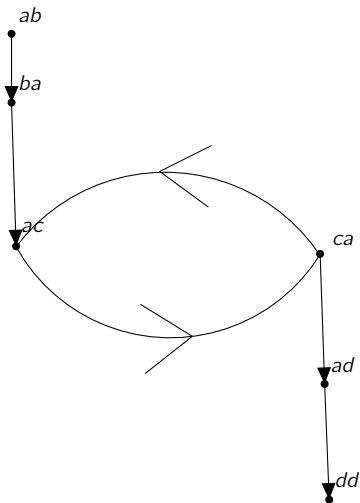
Rauzy graphs

For $x \in \mathbb{A}^{\mathbb{N}}$ and $n \in \mathbb{N}$, the **Rauzy graph** of order n of $x \in \mathbb{A}^{\mathbb{N}}$ is the directed graph whose set of vertices (resp. edges) consists of all factors of x of length n (resp. $n + 1$). There is a directed edge from u to v labeled w if u is a prefix of w and v a suffix of w .

The **distance** between two vertices in a Rauzy graph is the length of the shortest directed path between them. This path does not necessarily correspond to a factor in the word.

Rauzy graph

Rauzy graph of $abacacaddddd \dots$ of order 2 :



Open complexity function

Proposition

*Let $x \in \mathbb{A}^{\mathbb{N}}$ and $N \in \mathbb{N}$. Let w_1 and w_2 be two factors of x , such that there is a path of length i from w_1 to w_2 in the Rauzy graph of order N of x . Suppose w_1 and w_2 are closed with frontiers u_1 and u_2 respectively. Then $||u_1| - |u_2|| < i$.
In particular, if $i = 1$ the frontiers are of the same length : $|u_1| = |u_2|$.*

Open complexity function

Proposition

Let $x \in \mathbb{A}^{\mathbb{N}}$ and $N \in \mathbb{N}$. Let w_1 and w_2 be two factors of x , such that there is a path of length i from w_1 to w_2 in the Rauzy graph of order N of x . Suppose w_1 and w_2 are closed with frontiers u_1 and u_2 respectively. Then $||u_1| - |u_2|| < i$.

In particular, if $i = 1$ the frontiers are of the same length : $|u_1| = |u_2|$.

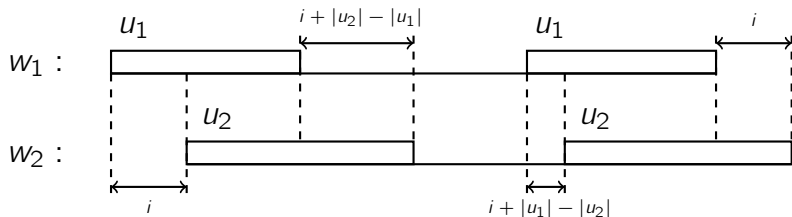


Figure – Factors w_1 and w_2 .

Some other lemmas

Lemma

Suppose that x is aperiodic with $\liminf \text{Op}_x = k$. There exists an integer N such that $u^N \notin \text{Fact}(x)$ for any choice of u with $|u| < 2k$.

Some other lemmas

Lemma

Suppose that x is aperiodic with $\liminf \text{Op}_x = k$. There exists an integer N such that $u^N \notin \text{Fact}(x)$ for any choice of u with $|u| < 2k$.

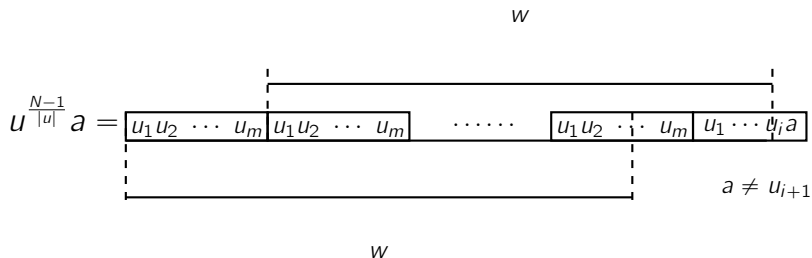


Figure – The frontier should be longer than $w = u^{\frac{N-1}{|u|} - 1}$.

Lemma

Suppose that x is aperiodic with $\liminf \text{Op}_x = k$. For every integer m , there exists an integer N such that the frontier of any closed factor of length N is at least m .

Proof - open complexity

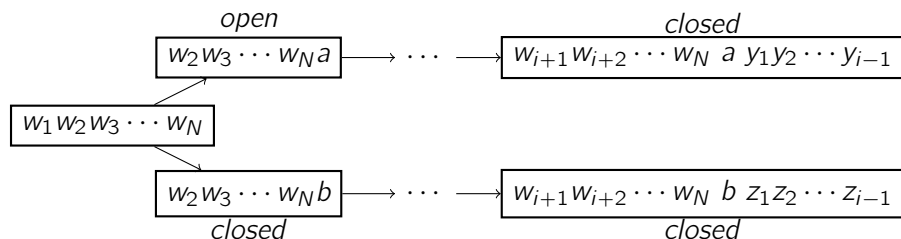


Figure – Open and closed factors in the Rauzy graph of order N .

Table of content

- 1 Complexity functions
- 2 Open and closed complexity functions
- 3 Words with bounded open complexity are ultimately periodic
- 4 Words with bounded closed complexity are ultimately periodic

Theorem

Let $x \in \mathbb{A}^{\mathbb{N}}$ be such that $\limsup_{n \in \mathbb{N}} Cl_x(n) < k$. Then x is ultimately periodic.

The branching lemma

Lemma

Let x be aperiodic with closed complexity bounded by k . At least one of the two following assertions holds :

1. $\forall u \in \text{RecFact}(x), \exists (p \leq k, a_1, \dots, a_p, a, b) \in \mathbb{N} \times \mathbb{A}^{p+2}, a \neq b$, such that

$$aa_1 \cdots a_p u \in \text{RecFact}(x) \quad \text{and} \quad ba_1 \cdots a_p u \in \text{RecFact}(x);$$

2. $\forall u \in \text{RecFact}(x), \exists (p \leq k, a_1, \dots, a_p, a, b) \in \mathbb{N} \times \mathbb{A}^{p+2}, a \neq b$, such that

$$ua_1 \cdots a_p a \in \text{RecFact}(x) \quad \text{and} \quad ua_1 \cdots a_p b \in \text{RecFact}(x).$$

Sketch of proof

Let u be such that u is always preceded by $a_1 a_2 \cdots a_k$.

Sketch of proof

Let u be such that u is always preceded by $a_1 a_2 \cdots a_k$. Let v be such that v is always followed by $b_1 b_2 \cdots b_k$.

Sketch of proof

Let u be such that u is always preceded by $a_1 a_2 \cdots a_k$. Let v be such that v is always followed by $b_1 b_2 \cdots b_k$.

Let w be such that uwv is a recurrent factor of x . We consider a complete first return to uwv .

Sketch of proof

Let u be such that u is always preceded by $a_1 a_2 \cdots a_k$. Let v be such that v is always followed by $b_1 b_2 \cdots b_k$.

Let w be such that uwv is a recurrent factor of x . We consider a complete first return to uwv .

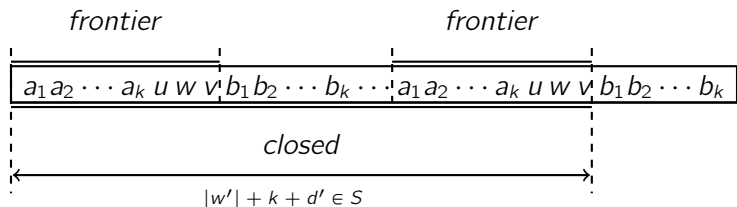


Figure – k closed factors of x .

Corollary

Let $x \in \mathbb{A}^{\mathbb{N}}$ with bounded closed complexity. If x is uniformly recurrent, then it is periodic.

Corollary

Let $x \in \mathbb{A}^{\mathbb{N}}$ with bounded closed complexity. If x is uniformly recurrent, then it is periodic.

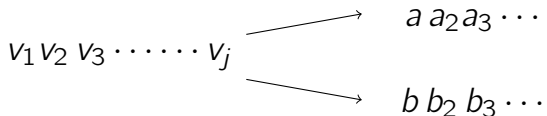


Figure – It is possible to avoid seeing u .

Lemma

Let $x \in \mathbb{A}^{\mathbb{N}}$ with $\limsup Cl_x(n) < k$. Let u be a primitive word in $\text{Fact}(x)$ such that, for every $n \in \mathbb{N}$, $u^n \in \text{Fact}(x)$. Then $|u| < k$.

Lemma

Let $x \in \mathbb{A}^{\mathbb{N}}$ with $\limsup Cl_x(n) < k$. Let u be a primitive word in $\text{Fact}(x)$ such that, for every $n \in \mathbb{N}$, $u^n \in \text{Fact}(x)$. Then $|u| < k$.

The word uuu contains the square $r(u)r(u)$ for every rotation of u .

An operator often used in symbolic dynamics is the **shift operator** T . That operator has many uses in combinatorics on words. It is defined on infinite words as follows :

$$\forall x \in \mathbb{A}^{\mathbb{N}}, x = x_0x_1x_2 \cdots, T(x) = x_1x_2x_3 \cdots$$

An operator often used in symbolic dynamics is the **shift operator** T . That operator has many uses in combinatorics on words. It is defined on infinite words as follows :

$$\forall x \in \mathbb{A}^{\mathbb{N}}, x = x_0x_1x_2 \cdots, T(x) = x_1x_2x_3 \cdots$$

For $x \in \mathbb{A}^{\mathbb{N}}$, we let $\Omega(x)$ denote the **shift orbit closure** of x , i.e. the closure in $\mathbb{A}^{\mathbb{N}}$ for the usual topology of the set

$$\{T^n(x) \mid n \in \mathbb{N}\}.$$

An operator often used in symbolic dynamics is the **shift operator** T . That operator has many uses in combinatorics on words. It is defined on infinite words as follows :

$$\forall x \in \mathbb{A}^{\mathbb{N}}, x = x_0x_1x_2 \cdots, T(x) = x_1x_2x_3 \cdots$$

For $x \in \mathbb{A}^{\mathbb{N}}$, we let $\Omega(x)$ denote the **shift orbit closure** of x , i.e. the closure in $\mathbb{A}^{\mathbb{N}}$ for the usual topology of the set

$$\{T^n(x) \mid n \in \mathbb{N}\}.$$

Theorem (Furstenberg)

For every infinite word x over a finite alphabet, $\Omega(x)$ contains at least one uniformly recurrent element.

Sketch of proof

- ▶ Starting from x , we construct a word y that does not contain powers of great order.

Sketch of proof

- ▶ Starting from x , we construct a word y that does not contain powers of great order.
- ▶ We consider a uniformly recurrent element in $\Omega(y)$ and derive a contradiction.

Closed complexity function

Definition

A subset S of \mathbb{N} is **syndetic** if it has bounded gaps, so if there exists d such that, for every integer n ,

$$S \cap [n, n + d] \neq \emptyset.$$

Closed complexity function

Definition

A subset S of \mathbb{N} is **syndetic** if it has bounded gaps, so if there exists d such that, for every integer n ,

$$S \cap [n, n + d] \neq \emptyset.$$

Theorem

Let $x \in \mathbb{A}^{\mathbb{N}}$ be such that there exist a positive integer d and a syndetic subset $S \subseteq \mathbb{N}$ with gaps smaller than d on which the closed complexity of x is bounded, i.e. there exists $k \in \mathbb{N}$ such that $Cl_x(n) < k$ for every $n \in S$. Then x is ultimately periodic.

Theorem (O. Parshina, P. 2020)

Let $x \in \mathbb{A}^{\mathbb{N}}$ be a right-infinite word over a finite alphabet \mathbb{A} . The following are equivalent :

1. x is aperiodic ;
2. $\liminf_{n \rightarrow +\infty} \text{Op}_x(n) = +\infty$.
3. There exists a syndetic set S such that $\text{Cl}_x(n)$ is bounded on S .

Thank you for your attention !