# On the abelian complexity of infinite words

**Idrissae KABORE**

Université Nazi BONI, Bobo Dioulasso (Burkina Faso)

**WORDS 2025**-Nancy, July, 2, 2025

# Outline

- Introduction
- Definitions, notations
- Classical complexity
- Abelian complexity
- Abelian complexity of Thue-Morse words, Tribonacci
- Abelian comlexity and uniform recurrence
- Link between abelian complexity and uniform frequency of letters
- Abelian comlexity and equi-frequency of letters in binary words
- Two questions

# Definitions, notations

- $\mathcal{A}_d = \{a_1, a_2, \cdots, a_d\}$, $d$-ary alphabet.
- $\mathcal{A}_d^*$: set of finite words on $\mathcal{A}_d$
- $\mathcal{A}_d^\omega$: set of infinite words on $\mathcal{A}_d$
- Let $u$ be a finite word on $\mathcal{A}_d$
  $\chi(u) = (|u|_{a_1}, |u|_{a_2}, \cdots, |u|_{a_d})$ is the Parikh vector of $u$.

# Definitions, notations

Let $u \in \mathcal{A}^\omega$

- $\mathcal{F}_n(u)$: set of factors of $u$ of length $n$.
- $\mathcal{F}(u)$: set of all factors of $u$.
- $u$ is aperiodic if $u$ is not ultimately periodic.
- $u$ is recurrent if any factor of $u$ appears infinitely often.
- $u$ is uniformly recurrent if for all $n \in \mathbb{N}$, there exists $N$ such that any factor of $u$ of length $N$ contains any factor of $u$ of length $n$.

# Definitions, notations

A morphism is a map $f : \mathcal{A}^* \to \mathcal{A}^*$ such that $f(uv) = f(u)f(v)$, for all $u, v \in \mathcal{A}^*$.

If, for a letter $a \in \mathcal{A}$, $f(a) = au$ with $u \neq \varepsilon$, the morphism has a unique fixed point beginning with $a$, which is the infinite word $\lim_{n\to\infty} f^n(a)$.

# Classical complexity

**Definition**

*Given an infinite word $u$, the factor complexity of $u$ is the map defined by $\rho_u(n) = \#\mathcal{F}_n(u)$ for all $n \in \mathbb{N}$.*

# Classical complexity

**Definition**

*Given an infinite word $u$, the factor complexity of $u$ is the map defined by $\rho_u(n) = \#\mathcal{F}_n(u)$ for all $n \in \mathbb{N}$.*

It counts the number of distinct factors of $u$ of given length.
Morse and Hedlund used factor complexity function to characterize aperiodic infinite words and define Sturmian words as follows:

**Theorem (Morse Hedlund, 1938)**

*Let $u$ be an infinite word. The following statements are equivalent:*

1. *$u$ is aperiodic.*
2. *There exists $n$ such that $\rho_u(n) \leq n$.*
3. *$(\rho_u(n))_n$ is bounded.*

**Definition**

*An infinite word $u$ is called Sturmian if its complexity satisfies $\rho_u(n) = n + 1$ for all $n \in \mathbb{N}$.*

Sturmian words constitute the class of non-ultimately periodic infinite words whose complexity is minimal.

Inspired by the notion of factor complexity, other notions of complexity have been developed. One such notion is the Abelian complexity of infinite words, the topic of this talk.

# Abelian complexity

Two words $u$ and $v$ in $\mathcal{A}^*$ are said to be abelian equivalent, and we note $u \sim_{\text{ab}} v$, if $|u|_a = |v|_a$ for all $a \in \mathcal{A}$,
$\sim_{\text{ab}}$ define an equivalence relation on $\mathcal{A}^*$.

**Definition**

*The abelian complexity of an infinite word $u$ is the function defined by*

$$\rho_u^{ab}(n) = \#\mathcal{F}_n(u)/\sim_{ab} = \#\left\{\chi\left(v\right) : v \in \mathcal{F}_n\left(u\right)\right\}.$$

# Abelian complexity

Let $u$ be an infinite word, $n$ be an integer and $a$ be a letter. Define, $\mathbf{m}_a(n) = \min\left\{|w|_a : w \in \mathcal{F}_n(u)\right\}$ and $\mathbf{M}_a(n) = \max\left\{|w|_a : w \in \mathcal{F}_n(u)\right\}$.

**Lemma (Intermidiate Vectors Lemma)**

*Let $u$ be an infinite word and $n$ be an integer. Then, for any letter $a$, we have*

$$\forall m \in \left[\mathbf{m}_a(n), \mathbf{M}_a(n)\right], \ \exists v \in \mathcal{F}_n\left(u\right), \ |v|_a = m.$$

# Abelian complexity

**Lemma (first difference of $\mathbf{m}_a$ and $\mathbf{M}_a$)**

*Let $u$ be an infinite word. Then, for any letter $a$ and every $n \in \mathbb{N}$, we have:*

- $\mathbf{m}_a(n+1) - \mathbf{m}_a(n) \in \{0, 1\}$
- $\mathbf{M}_a(n+1) - \mathbf{M}_a(n) \in \{0, 1\}$.

# Abelian complexity

**Lemma (Max-min formula of $\rho_u^{ab}$)**

*Let $u$ be an infinite binary word. Then, for all $n \in \mathbb{N}$:*

- $\mathbf{m}_a(n) + \mathbf{M}_b(n) = \mathbf{M}_a(n) + \mathbf{m}_b(n) = n$;
- $\rho_u^{ab}(n) = \mathbf{M}_a(n) - \mathbf{m}_a(n) + 1 = \mathbf{M}_b(n) - \mathbf{m}_b(n) + 1.$

**Proof.**

For any $w \in \mathcal{F}_n(u)$, $\chi(w)$ is in the set
$\{(\mathbf{m}_a(n), n - \mathbf{m}_a(n))\,; (\mathbf{m}_a(n) + 1, n - \mathbf{m}_a(n) - 1)\,; \cdots ;$
$\cdots ; (\mathbf{M}_a(n), n - \mathbf{M}_a(n))\}$.
By Intermidiate Vectors Lemma, all these vectors are attained.
Therefore, $\mathbf{M}_b(n) = n - \mathbf{m}_a(n)$ and $\mathbf{m}_b(n) = n - \mathbf{M}_a(n)$. The results follow. $\square$

# Abelian complexity

**Lemma (first difference of $\rho_u^{ab}$ )**

*Let $u$ be an infinite binary word. Then*

$$\forall n \in \mathbb{N} : \rho_u^{ab}(n+1) - \rho_u^{ab}(n) \in \{-1,\, 0,\, +1\}\,.$$

**Proof.**

We have $\rho_u^{ab}(n) = \mathbf{M}_a(n) - \mathbf{m}_a(n) + 1$. Since, by Lemma of first difference of $\mathbf{m}_a$ and $\mathbf{M}_a$, $\mathbf{m}_a(n+1) - \mathbf{m}_a(n) \in \{0,\, 1\}$ and $\mathbf{M}_a(n+1) - \mathbf{M}_a(n) \in \{0,\, 1\}$, it follows that

$$\rho_u^{ab}(n+1) - \rho_u^{ab}(n) = (\mathbf{M}_a(n+1) - \mathbf{M}_a(n)) - (\mathbf{m}_a(n+1) - \mathbf{m}_a(n)) \in \{-1,\, 0,\, +1\}\,.$$

$\square$

# Abelian complexity

It is known that

**Theorem (Richomme, Saari, Zamboni, 2009)**

*For all infinite word $u$ over $\mathcal{A}_d$, and for all $n \geq 0$,*

$$1 \leq \rho_u^{ab}(n) \leq \binom{n+d-1}{d-1}$$

*In particular, the abelian complexity is bounded by $O(n^d)$.*

## Abelian complexity: periodic words and Sturmian words

Like the classical complexity, the abelian complexity intervenes in characterization of some classes of infinite words: infinite periodic words and Sturmian words.

**Theorem (Coven, Hedlund, 1973)**

*An infinite word $u$ is periodic if and only if $\rho_u^{ab}(n) = 1$ for some $n \geq 1$.*

# Abelian complexity: periodic words and Sturmian words

Like the classical complexity, the abelian complexity intervenes in characterization of some classes of infinite words: infinite periodic words and Sturmian words.

**Theorem (Coven, Hedlund, 1973)**

*An infinite word $u$ is periodic if and only if $\rho_u^{ab}(n) = 1$ for some $n \geq 1$.*

**Theorem (Coven, Hedlund, 1973)**

*An infinite word $u$ is Sturmian if and only if, $\rho_u^{ab}(n) = 2$, for all $n \geq 1$.*

Inspired by this characterization of Sturmian words, Rauzy asked whatether there exist aperiodic words on a ternary alphabet such that $\rho_u^{ab}(n) = 3$ for all $n \geq 1$.
Due to Richomme et al. the two folowing results provide positive answers:

# Abelian complexity: Sturmian words and following

**Theorem (Richomme, Saari, Zamboni, 2009)**

*For all aperiodic balanced word $u$ on a ternary alphabet, $\rho_u^{ab}(n) = 3$, for all $n \geq 1$.*

**Theorem (Richomme, Saari, Zamboni, 2009)**

*Let $u$ be an aperiodic binary word on $\{a, b\}$. then $\rho_{\sigma(u)}^{ab}(n) = 3$, for all $n \geq 1$ where $\sigma$ denotes the substitution defined by $\sigma(a) = abc$ and $\sigma(b) = acb$.*

So, Richomme et al. posed the question: "Does there exist a recurrent word over a $4$-ary alphabet with exactly $4$ Abelian factors of each length?"
Currie and Rampersad provide a negative answer:

**Theorem (Currie, Rampersad, 2011)**

*Let $d \geq 4$ be an integer. There is no recurrent word over an $d$-ary alphabet with exactly $d$ Abelian factors of each length $\geq 1$.*

# Abelian complexity: binary Thue-Morse word

The binary Thue-Morse word $\mathbf{t}_2$ is the fixed point from $a$ of the Thue-Morse substition $\mu_2 : a \mapsto ab, \ b \mapsto ba$.

**Theorem (Cassaigne, Richomme, Saari, Zamboni, 2011)**

*The abelian complexity of $\mathbf{t}_2$ is given by:*

$$\rho_{\mathbf{t}_2}^{ab}(n) = \begin{cases} 3 & \textit{if } n \textit{ is even, } n \geq 2 \\ 2 & \textit{if } n \textit{ is odd} \end{cases} .$$

# Abelian complexity: Thue-Morse words

The binary Thue-Morse word $\mathbf{t}_3$ is the fixed point from $a$ of the substition $\mu_3 : a \mapsto abc, b \mapsto bca, c \mapsto cab$.

**Theorem (Kaboré, Kientéga, 2017)**

*The abelian complexity of $\mathbf{t}_3$ is given by:*

$$\rho^{ab}_{\mathbf{t}_3}(n) = \begin{cases} 1 & \textit{if} \quad n = 0 \\ 3 & \textit{if} \quad n = 1 \\ 7 & \textit{if} \quad n = 3k, \ k \geq 1 \\ 6 & \textit{otherwise} \end{cases}$$

Chen and Wen (2019) determine completely the abelian complexity of generalized Thue-Morse word $\mathbf{t}_d$ ($d \geq 2$) and find that the sequence $\left(\rho^{ab}_{\mathbf{t}_d}(n)\right)_n$ is ultimately periodic with period $d$.

# Abelian complexity: Tribonacci word

The Tribonacci word $T$ is the fixed point from $a$ of the substitution $\tau : a \mapsto ab, b \mapsto ac, c \mapsto a$.

**Theorem (Richomme, Saari, Zamboni (2009), Turek (2013), Shallit (2021))**

$$\forall n \geq 1, \rho^{ab}_{\mathbf{t}_3}(n) \in \{3, 4, 5, 6, 7\}$$

*where eah element appears infinitely many times.*

And more generally, Shallit get the following result:

# Abelian complexity and automaticity

**Theorem (Shallit, 2021)**

*Let $u$ be a sequence that is automatic in some regular numeration system. Suppose that*

(a) *the Parikh vectors of length-$n$ prefixes of $u$ form a synchronized sequence; and*

(b) *the abelian complexity $\rho_u^{ab}$ is bounded above by a constant.*

*Then $\left(\rho_u^{ab}(n)\right)_{n\geq 0}$ is an automatic sequence and the DFAO computing it is effectively computable.*
*Furthermore, if condition (a) holds, then condition (b) can be tested algorithmically.*

# Abelian complexity: general results

Joint work with **Julien Cassaigne** (2016).

# Abelian complexity and uniform recurrence

**Theorem**

*Let $u$ be a uniformly recurrent binary word. Then, there exists a positive real number $\alpha < 1$ and a non-negative integer $n_0$ such that:*

$$\forall n \geq n_0,\ \rho_u^{ab}(n) \leq \alpha n.$$

# Abelian complexity and uniform recurrence

**Proof.**

Let $u$ be a uniformly recurrent word on $\{0, 1\}$.

- $u$ is constant. Then, $\rho_u^{ab}(n) = 1$ and the result holds,
- $u$ is not constant. Then

$$\exists N \in \mathbb{N},\ \forall w \in \mathcal{F}_N(u),\ 0, 1 \in \mathcal{F}(w)$$

So, $1 \leq |w|_1 \leq | \leq N - 1$and thus $\rho_u^{ab}(N) \leq N - 1$.
Let $n \in \mathbb{N}$ such that $\frac{n}{N} \geq 3$ and take $w \in \mathcal{F}_n(u)$. We can write:
$w = w_1 w_2 \cdots w_k w'$ where $k = \lfloor \frac{n}{N} \rfloor$, $|w_i| = N$ and $|w'| < N$. So,

$$k \leq |w|_1 \leq n - k$$

That is letter $1$ appears in $w$ at least $k$ times and at most $n - k$ times.
Therefore
$\rho_u^{ab}(n) \leq n - 2k + 1 \leq n - 2 \cdot \frac{n}{N} + 3 \leq n - \frac{n}{N} = n(1 - \frac{1}{N})$ .
To obtain the result it suffices to put $\alpha = 1 - \frac{1}{N}$ and $n_0 = 3N$ . $\square$

# Abelian complexity and uniform recurrence

**Theorem (Control formula of uniform recurrence by $\rho_u^{ab}$)**

*Let $u$ be a uniformly recurrent $d$-ary word. Then, there exist a positive real number $\alpha < 1$ and a non-negative integer $n_0$ such that:*

$$\forall n \geq n_0,\ \rho_u^{ab}(n) \leq \frac{\alpha n^{d-1}}{(d-1)!}.$$

# Link between abelian complexity and uniform frequency of letters

**Definition**

*We say that* $u$ *admits*

- *frequencies of letters if for any letter* $a$, *and any sequence* $(w_n)$ *of prefixes of* $u$ *such that* $\lim_{n\to\infty} |w_n| = \infty$, *then* $\lim_{n\to\infty} \frac{|w_n|_a}{|w_n|}$ *exists.*
- *uniform frequencies of letters if for any letter* $a$, *and any sequence* $(w_n)$ *of factors of* $u$ *such that* $\lim_{n\to\infty} |w_n| = \infty$, *then* $\lim_{n\to\infty} \frac{|w_n|_a}{|w_n|}$ *exists.*

# Link between abelian complexity and uniform frequency of letters

**Lemma**

*Let $u$ be an infinite word, $(w_i)$ be a sequence of factors of $u$ and $f$ be a real number in $(0,1)$ such that $\lim_{i\to\infty} |w_i| = \infty$ and $\lim_{i\to\infty} \frac{|w_i|_a}{|w_i|} = f$. Then*

$$\forall \delta > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ \exists v \in \mathcal{F}_n(u),\ \left| \frac{|v|_a}{|v|} - f \right| \leq \delta.$$

**Lemma**

*Let $u$ be an infinite word with a letter $a$ that does not admit frequency. Then, $\left(\frac{\mathbf{m}_a(n)}{n}\right)$ and $\left(\frac{\mathbf{M}_a(n)}{n}\right)$ converge and further $\lim \frac{\mathbf{m}_a(n)}{n} < \lim \frac{\mathbf{M}_a(n)}{n}$.*

# Link between abelian complexity and uniform frequency of letters

**Theorem**

*Let $u$ be an infinite binary word. Then $\rho_u^{ab}(n) = o(n)$ if and only if $u$ admits uniform frequencies of letters.*

The above theorem provides a characterization of binary words with uniform frequencies of letters. On a larger alphabet, we have a weaker result.

# Link between abelian complexity and uniform frequency of letters

**Proof.**

Let $u$ be an infinite binary word which does not admit uniform frequencies of letters and suppose by contradiction that $\rho_u^{ab}(n) = o(n)$. Then

$$\forall \delta > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ \rho_u^{ab}(n) \leq \delta n.$$

For every integer $n \geq n_0$, we have

$$\mathbf{M}_a(n) - \mathbf{m}_a(n) + 1 = \rho_u^{ab}(n) \leq \delta n.$$

Let $v$ be a factor of length $N$ sufficiently large. It follows

$$\mathbf{m}_a(n) \left\lfloor \frac{N}{n} \right\rfloor \leq |v|_a \leq \mathbf{M}_a(n) \left( \left\lfloor \frac{N}{n} \right\rfloor + 1 \right)$$

hence

# Link between abelian complexity and uniform frequency of letters

## Proof.

$$\mathbf{m}_a(n)\left(\frac{1}{n}-\frac{1}{N}\right) \leq \frac{|v|_a}{N} \leq \mathbf{M}_a(n)\left(\frac{1}{n}+\frac{1}{N}\right).$$

Since $u$ does not admit uniform frequencies of letters, there exists two distinct reals $f_1$ and $f_2$ and two sequences of factors $(v_k)$ and $(v'_k)$ satisfying $|v_k| = |v'_k| = k$ and

- $\lim_{k\to\infty} \frac{|v_k|_a}{k} = f_1$;
- $\lim_{k\to\infty} \frac{|v'_k|_a}{k} = f_2$.

It follows $\frac{\mathbf{m}_a(n)}{n} \leq f_i \leq \frac{\mathbf{M}_a(n)}{n}$, $i = 1, 2$. So, $|f_1 - f_2| \leq \frac{\mathbf{M}_a(n)-\mathbf{m}_a(n)}{n} \leq \delta$. As $\delta$ is arbitrarily small, $f_1 = f_2$. Contradiction!

Conversely, if $u$ admits uniform frequencies of letters, then both $\left(\frac{\mathbf{m}_a(n)}{n}\right)$ and $\left(\frac{\mathbf{M}_a(n)}{n}\right)$ converge to the same limit $f_a$. It follows that $\frac{\rho_u^{ab}(n)}{n} = \frac{\mathbf{M}_a(n)-\mathbf{m}_a(n)+1}{n} = o(1)$.

# Link between abelian complexity and uniform frequency of letters

**Theorem**

*Let $u$ be an infinite $d$-ary word. We have the following statements:*

1. *If $\rho_u^{ab}(n) = o(n)$ then $u$ admits uniform frequencies of letters.*
2. *If $u$ admits uniform frequencies of letters then $\rho_u^{ab}(n) = o\left(n^{d-1}\right)$.*

# Abelian complexity and equi-frequency of letters in infinite binary words

It is well known that the Thue-Morse word possesses uniform frequencies of letters and its abelian complexity satisfies $\rho^{ab}_{t_2}(2n) = 3$ and $\rho^{ab}_{t_2}(2n+1) = 2$ for all $n \geq 1$.

**Theorem**

*Let $u$ be an infinite binary word such that for all $n \geq 1$,*

$$\rho^{ab}_{u}(n) = \begin{cases} 3 & \textit{if } n \textit{ is even} \\ 2 & \textit{if } n \textit{ is odd} \end{cases}$$

*Then, $u$ admits uniform frequencies of letters and the letters have the same frequencies i.e., $\mathbf{f}_0(u) = \mathbf{f}_1(u) = \frac{1}{2}$.*

# Abelian complexity and equi-frequency of letters in infinite binary words

**Proof.**

Suppose

$$\forall n \geq 1,\ \rho_u^{ab}(n) = \begin{cases} 3 & \text{if } n \text{ is even} \\ 2 & \text{if } n \text{ is odd} \end{cases}$$

By induction, we check that

$$\forall n \geq 1, \forall w \in \mathcal{F}_{2n}(u), |w|_0 \in \{n-1,\ n,\ n+1\}.$$

□

Through the examples below we observe that the converse of the a bove Theorem does not hold.

• The image of any infinite binary aperiodic word $u$ by substitution $\Phi : 0 \mapsto 0101,\ 1 \mapsto 1100$ satisfies: $\rho_{\Phi(u)}^{ab}(n) \geq 3$ and $\mathbf{f}_0(\Phi(u)) = \mathbf{f}_1(\Phi(u)) = \frac{1}{2}$.

• The word $u = (000111)^{\omega}$ satisfies: $\rho_u^{ab}(3) = 4$ and $\mathbf{f}_0(u) = \mathbf{f}_1(u) = \frac{1}{2}$.

# Abelian complexity and equi-frequency of letters in infinite binary words

**Theorem**

*Let $u$ be an infinite binary word satisfying:*

1. $\rho_u^{ab}(n) = o(n)$,
2. $\exists n_0,\ \forall n \geq n_0,\ \rho_u^{ab}(n+1) \neq \rho_u^{ab}(n)$

*Then, $u$ admits uniform frequencies of letters and the frequencies are:* $\mathbf{f}_0(u) = \mathbf{f}_1(u) = \frac{1}{2}$.

# Two questions

- The Theorem on the equi-frequency of letters for any binary word having the same abelian complexity as the binary Thue-Morse word can be extended to the $d$-ary alphabet ($d \geq 3$)?
- As in the last Theorem, any $d$-ary word having the same abelian complexity of the $d$-ary Thue-Morse word $\mathbf{t}_d$ does admit equi-frequency of letters: $\mathbf{f}_a(u) = \frac{1}{d}$?

# Thank you for your attention!