Permutation groups and the Morse-Hedlund Theorem

Émilie Charlier
joint work with Svetlana Puzynina and Luca Zamboni

Département de Mathématique, Université de Liège

Worskshop "Ergodic Theory & Symbolic Dynamics"
London, September 14, 2017

Factor complexity of infinite words

The Fibonacci word

$$f = 0100101001001010010100100101001 0\ldots$$

is the fixed point of the morphism $0 \mapsto 01$ and $1 \mapsto 0$.

Factors of length n:

1	0,1
2	00,01,10
3	001,010,100,101
4	0010,0100,0101,1001,1010
$\vdots$	

It can be shown that there are exactly $n+1$ factors of length n in f.

Notation

- Alphabet: finite non-empty set, usually denoted by A.
- Word over A:
  $x = x_0 x_1 x_2 \cdots$ (infinite word)
  $x = x_0 x_1 \cdots x_{n-1}$ (finite word of length $|x| = n$).
- A^n is the set of all words of length n over A.
- Factor u of an infinite word w: $u = x_i \ldots x_{i+j}$ for some $i, j \in \mathbb{N}$.
- $\mathrm{Fac}_n(x)$ is the set of the factors of x of length n.

Factor complexity

The factor complexity of an infinite word x is the function $p_x\colon \mathbb{N} \to \mathbb{N}$ which counts the number of factors of length n of x:

$$\forall n \in \mathbb{N},\ p_x(n) = \left|\operatorname{Fac}_n(x)\right|.$$

Some properties:

- $\forall n \in \mathbb{N},\ p_x(n) \leq |A|^n$.
- $p_x(n)$ is a non-decreasing function.

Some more examples

- The (binary) Champernowne word

$$c = 0\ 1\ 10\ 11\ 100\ 101\ 110\ 111\ 1000 \cdots$$

has maximal factor complexity 2^n.

- The Thue-Morse word is the fixed point of the morphism $0 \mapsto 01, 1 \mapsto 10$ beginning with 0:

$$t = 0110100110010110 \cdots$$

We have $p_t(3) = 6$: no factors $000, 111$.
The factor complexity of Thue-Morse is computed in [Brlek 1987].

Complexity and periodicity

- Purely periodic word: $x = v^\omega = vvv\cdots$
- Ultimately periodic word: $x = uv^\omega = uvvv\cdots$
- Aperiodic means not ultimately periodic.

Theorem (Hedlund-Morse 1940, first part)

An infinite word x is aperiodic iff $\forall n \in \mathbb{N},\ p_x(n) \geq n+1$.

Sturmian words and balance

- An infinite word over A is C-balanced if for all factors u, v of the same length and for each $a \in A$, we have $\big||u|_a - |v|_a\big| \leq C$.

Theorem (Hedlund-Morse 1940, second part)

An infinite word x is such that $\forall n \in \mathbb{N},\ p_x(n) = n + 1$ iff it is binary, aperiodic and 1-balanced.

- Aperiodic binary infinite word of minimal complexity are called Sturmian words.
- We have already seen that the Fibonacci word is Sturmian.

Several generalizations of Morse-Hedlund

Other complexity functions, and their links with periodicity.

- Abelian complexity, which counts the number of abelian classes of words of each length n occurring in x: [Coven-Hedlund 1973], [Richomme-Saari-Zamboni 2011].
- Palindrome complexity, which counts the number of palindromes of each length n occurring in x: [Allouche-Baake-Cassaigne-Damanik 2003].
- Cyclic complexity, which counts the number of conjugacy classes of factors of each length n occurring in x: [Cassaigne-Fici-Sciortino-Zamboni 2017].
- Maximal pattern complexity: [Kamae-Zamboni 2002].

Several generalizations of Morse-Hedlund

Higher dimensions:

- Nivat conjecture: Any 2-dimensional word having at most mn rectangular blocks of size $m \times n$ must be periodic.
- It is known that the converse is not true.
- [Durand-Rigo 2013], in which they re-interpret the notion of periodicity in terms of Presburger arithmetic.

Our contribution

- New notion of complexity by group actions.
- Encompass most complexity functions studied so far.

Abelian complexity

- Two finite words are abelian equivalent if they contain the same numbers of occurrences of each letter: $00111 \sim_{ab} 01101$.
- The abelian complexity function $a_x(n)$ counts the number of abelian classes of words of length n occurring in x.

For the Thue-Morse word $t = 0110100110010110\cdots$, we have

$$a_t(n) = \begin{cases} 2 & \text{if } n \text{ is odd} \\ 3 & \text{if } n \text{ is even} \end{cases}$$

We have $a_t(3) = 2$ since there are 2 abelian classes of factors of length 3:

$$\{001, 010, 100\} \text{ and } \{011, 101, 110\}.$$

Abelian complexity and periodicity

We clearly have the following implications:

$$\begin{aligned} \text{ultimate periodicity} &\Rightarrow \text{bounded factor complexity} \\ &\Rightarrow \text{bounded abelian complexity.} \end{aligned}$$

However, we have just seen that the converse is not true: the Thue-Morse word is aperiodic and its abelian complexity function is bounded by 3.

Theorem (Coven-Hedlund 1973, part 1)

An infinite word x is purely periodic iff $\exists n \geq 1$, $a_x(n) = 1$.

In particular, if x is aperiodic then $\forall n \geq 1$, $a_x(n) \geq 2$.
The converse is false: take $x = 01^\omega$.

Abelian complexity and balance

We clearly have the following implications:

$$\text{ultimate periodicity} \Rightarrow \text{bounded factor complexity}$$
$$\Rightarrow \text{bounded abelian complexity.}$$

Theorem (Coven-Hedlund 1973, part 2)

An infinite aperiodic word x is Sturmian iff $\forall n \geq 1$, $a_x(n) = 2$.

Theorem (Richomme-Saari-Zamboni 2011)

An infinite word has bounded abelian complexity iff it is C-balanced for some $C \geq 1$.

Cyclic complexity

- Two finite words u and v are conjugate if there exist words w_1, w_2 such that $u = w_1 w_2$ and $v = w_2 w_1$.
- The cyclic complexity function $c_x(n)$ counts the number of conjugacy classes of words of length n occurring in x.

For the Thue-Morse word $t = 0110100110010110\cdots$, we have $c_t(4) = 4$ since there are 4 conjugacy classes of factors of length 4:

$$\{0010, 0100\}$$
$$\{0110, 1001, 1100, 0011\}$$
$$\{0101, 1010\}$$
$$\{1011, 1101\}$$

Cyclic complexity, periodicity and Sturmian words

Theorem (Cassaigne-Fici-Sciortino-Zamboni 2014)

An infinite word is ultimately periodic iff it has bounded cyclic complexity.

One always has

$$a_x(n) \leq c_x(n) \leq p_x(n).$$

Hence $c_x(n) = 1$ for some $n \geq 1$ implies that x is purely periodic.

In [Cassaigne-Fici-Sciortino-Zamboni 2014] they consider $\liminf c_x(n)$:

- Sturmian words satisfy $\liminf c_x(n) = 2$.
- But this is not a characterization of Sturmian words since the period-doubling word also has $\liminf c_x(n) = 2$.

Generalization via group actions

- Let G be a subgroup of the symmetric group S_n: $G \leq S_n$.
- G acts on A^n by permuting the letters:

$$G \times A^n \to A^n, \ (g, u) \mapsto g * u = u_{g^{-1}(1)} u_{g^{-1}(2)} \cdots u_{g^{-1}(n)}.$$

- We write $u_1 \cdots u_n \overset{g}{\curvearrowright} u_{g^{-1}(1)} u_{g^{-1}(2)} \cdots u_{g^{-1}(n)}$.
- $0100 \overset{(1234)}{\curvearrowright} 0010$.
- $abcab \overset{(123)(45)}{\curvearrowright} cabba$.
- In particular $g * u \sim_{ab} u$.
- G-equivalence relation on A^n: for $u, v \in A^n$, $u \sim_G v$ if $\exists g \in G$, $g * u = v$.
- $u \sim_G v$ implies $u \sim_{ab} v$.

Complexity by actions of groups

- Now we consider a sequence of subgroups $\omega = (G_n)_{n\geq 1}$: for each $n \geq 1$, $G_n \leq S_n$.
- The group complexity $p_{\omega,x}(n)$ of x counts the number of G_n-classes of words of length n occurring in x.

For the Thue-Morse word $t = 0110100110010110\cdots$ and $G_4 = \langle (13), (24) \rangle$, we have $p_{\omega,t}(4) = 7$ while $p_t(4) = 10$.

We have six singleton classes of length 4:

$$[0010], [0100], [0101], [1010], [1011], [1101]$$

and one class of order 4:

$$[0110 \overset{(13)(24)}{\curvearrowright} 1001 \overset{(24)}{\curvearrowright} 1100 \overset{(13)}{\curvearrowright} 0011].$$

Group actions: generalization of factor, abelian and cyclic complexities

Each choice of sequence $\omega = (G_n)_{n\geq 1}$ defines a unique complexity which reflects a different combinatorial property of an infinite word.

As particular cases, we recover

- factor complexity: if $\omega = (Id_n)_{n\geq 1}$ then $p_{\omega,x}(n) = p_x(n)$
- abelian complexity: if $\omega = (S_n)_{n\geq 1}$ then $p_{\omega,x}(n) = a_x(n)$
- cyclic complexity: if $\omega =< (12\cdots n) >_{n\geq 1}$ then $p_{\omega,x}(n) = c_x(n)$.

The quantity $\varepsilon(G)$

- For $G \leq S_n$ and $i \in \{1, 2, \ldots, n\}$, the G-orbit of i is

$$G(i) = \{g(i) \mid g \in G\}.$$

- The number of distinct G-orbits is denoted

$$\varepsilon(G) = \left|\{G(i) \mid i \in \{1, 2, \ldots, n\}\}\right|.$$

- For $n = 6$ and $G =< (13), (256) >$, we have $\varepsilon(G) = 3$:

$$123456.$$

- If $G = Id$, then $\varepsilon(G) = n$.
- If G contains an n-cycle, then $\varepsilon(G) = 1$.

Complexity by group actions: $\varepsilon(G)$

- For $G \leq S_n$, $\varepsilon(G)$ is the number of G-orbits of $\{1, \ldots, n\}$.

Example (The Klein group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$)

First take $G = \{\mathrm{id}, (12), (34), (12)(34)\}$.
Then the G-orbits are $\{1, 2\}$ and $\{3, 4\}$, hence $\varepsilon(G) = 2$.

Second, consider $G' = \{\mathrm{id}, (12)(34), (13)(24), (14)(23)\}$.
Then the only G'-orbit is $\{1, 2, 3, 4\}$, hence $\varepsilon(G') = 1$.

- This shows an interesting phenomenon: the quantity $\varepsilon(G)$ depends on the embedding of G into S_n.

Generalisation of the Morse-Hedlund theorem

Theorem 1 (Charlier-Puzynina-Zamboni 2017)

Let x be an infinite aperiodic word, $\omega = (G_n)_{n\geq 1}$, $G_n \leq S_n$.

- Then $\forall n \geq 1,\ p_{\omega,x}(n) \geq \varepsilon(G_n) + 1$.
- If $\forall n \geq 1,\ p_{\omega,x}(n) = \varepsilon(G_n) + 1$ then x is Sturmian.

Corollary

An infinite aperiodic word is Sturmian iff there exists $\omega = (G_n)_{n\geq 1}$, $G_n \leq S_n$ such that $\forall n \geq 1,\ p_{\omega,x}(n) = \varepsilon(G_n) + 1$.

Sketch of the proof

Theorem 1, second part

Let x be an infinite aperiodic word, $\omega = (G_n)_{n\geq 1}$, $G_n \leq S_n$.
If $\forall n \geq 1$, $p_{\omega,x}(n) = \varepsilon(G_n) + 1$ then x is Sturmian.

- Since $\varepsilon(G_1) = 1$, then $p_{\omega,x}(1) = 2$, and hence x is binary.
- Suppose that x is not Sturmian, that is, not 1-balanced.
- Key lemma: $\exists n \geq 2$, a Sturmian word y and a bispecial factor $u \in \{0,1\}^{n-2}$ of y s.t. $\mathrm{Fac}_n(x) = \mathrm{Fac}_n(y) \cup \{0u0, 1u1\}$.
- u is a bispecial factor of y means that $u0$, $u1$, $0u$, $1u$ are factors of y.
- Since y is Sturmian, exactly one of $0u0$ and $1u1$ is a factor of y, hence $p_{\omega,x}(n) \geq p_{\omega,y}(n) + 1$.
- Apply first part of the theorem to y to get $p_{\omega,x}(n) \geq p_{\omega,y}(n) + 1 \geq \epsilon(G_n) + 2$, a contradiction.

Generalisation of the Morse-Hedlund theorem

Partial converse:

Theorem 2 (Charlier-Puzynina-Zamboni 2017)

Let x be a Sturmian word and $\omega = (G_n)_{n\geq 1}$, where G_n is an abelian subgroup of S_n. Then $\exists \omega' = (G'_n)_{n\geq 1}$, $G'_n \leq S_n$, such that $\forall n \geq 1$,

- G'_n is isomorphic to G_n
- $p_{\omega',x}(n) = \varepsilon(G'_n) + 1$.

As particular cases, we recover:

- Morse-Hedlund theorem: $\omega = (Id_n)_{n\geq 1}$, $p_{\omega,x}(n) = p_x(n)$, $\varepsilon(G_n) = n$.
- Abelian complexity: $\omega = (S_n)_{n\geq 1}$, $p_{\omega,x}(n) = a_x(n)$, $\varepsilon(G_n) = 1$.

We cannot always take $G' = G$

Theorem 2

Let x be a Sturmian word and $\omega = (G_n)_{n\geq 1}$, where G_n is an abelian subgroup of S_n. Then $\exists \omega' = (G'_n)_{n\geq 1}$, $G'_n \leq S_n$, such that $\forall n \geq 1$, G'_n is isomorphic to G_n and $p_{\omega',x}(n) = \varepsilon(G'_n) + 1$.

Consider the factors of length 4 of the Fibonacci word:

$$0010, 0100, 0101, 1001, 1010.$$

Let $G_4 = \langle(1234)\rangle$. Then $\varepsilon(G_4) = 1$ and $p_{\omega,f}(4) = 3 > \varepsilon(G_4) + 1$:

$$[0100 \overset{(1234)}{\curvearrowright} 0010], \quad [0101 \overset{(1234)}{\curvearrowright} 1010], \quad [1001].$$

But we can take $G'_4 = \langle(1324)\rangle$. Then $\varepsilon(G'_4) = 1$ and $p_{\omega',f}(4) = 2 = \varepsilon(G'_4) + 1$:

$$[0010 \overset{(1324)}{\curvearrowright} 0100], \quad [0101 \overset{(1324)}{\curvearrowright} 1001 \overset{(1324)}{\curvearrowright} 1010].$$

We cannot replace "isomorphic" by "conjugate"

Theorem 2

Let x be a Sturmian word and $\omega = (G_n)_{n \geq 1}$, where G_n is an abelian subgroup of S_n. Then $\exists \omega' = (G'_n)_{n \geq 1}$, $G'_n \leq S_n$, such that $\forall n \geq 1$, G'_n is isomorphic to G_n and $p_{\omega',x}(n) = \varepsilon(G'_n) + 1$.

Let $G =< (123)(456) >\leq S_6$. This is a cyclic subgroup of order 3. Then $\varepsilon(G) = 2$ and we can show that

$$\left| \mathrm{Fac}_6(f)/_{\sim_{G'}} \right| \geq 4$$

for each subgroup G' of S_6 which is conjugate to G.

Sketches of proof

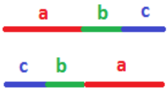
Theorem 2

Let x be a Sturmian word and $\omega = (G_n)_{n\geq 1}$, where G_n is an abelian subgroup of S_n. Then $\exists \omega' = (G'_n)_{n\geq 1}$, $G'_n \leq S_n$, such that $\forall n \geq 1$,

- G'_n is isomorphic to G_n
- $p_{\omega',x}(n) = \varepsilon(G'_n) + 1$.

First we prove Theorem 2 for an n-cycle.

abc-permutation [Pak-Redlich 2008]: The numbers $1, 2, \dots, n$ are divided into three subintervals of length a, b and c which are rearranged in the order c, b, a:



$$1, 2, \dots, n \mapsto c+b+1, c+b+2, \dots, n, c+1, c+2, \dots, c+b, 1, 2, \dots, c$$

Factors of length 6 in Fibonacci

Consider the *abc*-permutation with $a = 1$, $b = 2$, $c = 3$ on the lexicographic array of length 6.

0	0	1	0	0	1
0	0	1	0	1	0
0	1	0	0	1	0
0	1	0	1	0	0
1	0	0	1	0	0
1	0	0	1	0	1
1	0	1	0	0	1

This *abc*-permutation can be seen as a 6-cycle: (163524).

For $G \leq S_n$, we say that $\sim_G$ is abelian transitive on x if $\forall u, v \in \mathrm{Fac}_n(x)$: $u \sim_{ab} v \Leftrightarrow u \sim_G v$.

Factors of length 6 in Fibonacci

Consider the *abc*-permutation with $a = 1$, $b = 2$, $c = 3$ on the lexicographic array of length 6.

0	0	1	0	0	1
0	0	1	0	1	0
0	1	0	0	1	0
0	1	0	1	0	0
1	0	0	1	0	0
1	0	0	1	0	1
1	0	1	0	0	1

This *abc*-permutation can be seen as a 6-cycle: (163524).

For $G \leq S_n$, we say that $\sim_G$ is abelian transitive on x if $\forall u, v \in \mathrm{Fac}_n(x)$: $u \sim_{ab} v \Leftrightarrow u \sim_G v$.

Factors of length 6 in Fibonacci

Consider the *abc*-permutation with $a = 1$, $b = 2$, $c = 3$ on the lexicographic array of length 6.

0	0	1	0	0	1
0	0	1	0	1	0
0	1	0	0	1	0
0	1	0	1	0	0
1	0	0	1	0	0
1	0	0	1	0	1
1	0	1	0	0	1

This *abc*-permutation can be seen as a 6-cycle: (163524).

For $G \leq S_n$, we say that $\sim_G$ is abelian transitive on x if $\forall u, v \in \mathrm{Fac}_n(x)$: $u \sim_{ab} v \Leftrightarrow u \sim_G v$.

(abc)-permutations

Lemma

Let x be a Sturmian word. Then for each $n \geq 1$ there exists an (a, b, c)-permutation on $\{1, 2, \ldots, n\}$ which is an n-cycle σ such that $\sim_{\langle \sigma \rangle}$ is abelian transitive on x.

Comments:

- We exhibit our (a, b, c)-permutation candidate.
- We show that it is actually an n-cycle [Pak, Redlich, 2008].
- We use lexicographic arrays for the proof of the abelian transitivity.
- In fact, we prove that $w_{(i+1)} = \sigma(w_{(i)})$ in each abelian class, where $w_{(i)}$ are ordered lexicographically.

A corollary

Corollary

If x is a Sturmian word then for each n there exists a cyclic group G_n generated by an n-cycle such that $\left|\mathrm{Fac}_n(x)/_{\sim_{G_n}}\right| = 2$.

In contrast, if we set $G_n = \langle(1, 2, \ldots, n)\rangle$ for each $n \geq 1$, then $\limsup p_{\omega,x}(n) = +\infty$, while $\liminf p_{\omega,x} = 2$.
[Cassaigne, Fici, Sciortino, Zamboni, 2015]

Theorem 2: construction for abelian groups

Theorem (Fundamental theorem of finite abelian groups)

Every finite abelian group G can be written as a direct product of cyclic groups $\mathbb{Z}/m_1\mathbb{Z} \times \mathbb{Z}/m_2\mathbb{Z} \times \cdots \times \mathbb{Z}/m_k\mathbb{Z}$ where the m_i are prime powers.

- The sequence $(m_1, m_2, \ldots, m_k)$ determines G up to isomorphism.
- The trace of G is given by $T(G) = m_1 + m_2 + \cdots + m_k$.

Proposition (Hoffman 1987)

If an abelian group G is embedded in S_n, then $T(G) \leq n$.

Open problem

Does Theorem 2 hold for non-abelian groups?

Question

Let x be a Sturmian word and $\omega = (G_n)_{n\geq 1}$, where $G_n \leq S_n$. Does there exist $\omega' = (G'_n)_{n\geq 1}$, $G'_n \leq S_n$, such that for all $n \geq 1$,

- G'_n is isomorphic to G_n
- $p_{\omega',x}(n) = \varepsilon(G'_n) + 1$.

Minimal complexity

complexity type	minimal complexity	words family
factor	n+1	Sturmian
abelian	2	Sturmian
cyclic	$\liminf = 2$	Sturmian+
group	$\varepsilon(G_n) + 1$	Sturmian
maximal pattern	2n+1	Sturmian+
arithmetical	linear	(asymptotically) Toeplitz

Arithmetical complexity: [Avgustinovich-Cassaigne-Frid 2006]