The Zimin word and words with a Lyndon Orbit

Caius Wojcik - ICJ - Lyon 1

School for young researchers, CIRM

4-8 March 2019

Notations :

- $\mathcal{A}$: a set, called the alphabet.
- $\mathcal{A}^+$: set of non-empty finite words over $\mathcal{A}$.
- $|u|$: length of a finite word u.
- $\mathbb{P}_n(x)$: prefix of length $n \geq 1$ of an infinite word x.
- $T : x_0x_1x_2\ldots \longmapsto x_1x_2x_3\ldots$ the shift on infinite words.
- $T^n(x)$: the n-th suffix of the infinite word x.
- $[a, b]$: integer interval for $a \leq b$ and $a, b \in \mathbb{N}$.
- $\Omega(x)$: Dynamical orbit, or subshift of an infinite word x. Topological closure of the set of suffixes of x, in $\mathcal{A}^{\mathbb{N}}$ endowed with product topology associated with discrete topology on $\mathcal{A}$.

We are interested in the differences between suffixes of an infinite word and general elements of its dynamical orbit.

Definition (Zimin patterns)

We define the Zimin patterns over the alphabet $\mathcal{A}_X = \{x_1, x_2, \dots\}$ *as*

$$Z_1 = x_1 \quad \text{and} \quad Z_{n+1} = Z_n x_{n+1} Z_n, \quad \text{for } n \geq 1.$$

$$Z_2 = x_1x_2x_1, \quad Z_3 = x_1x_2x_1x_3x_1x_2x_1,$$
$$Z_4 = x_1x_2x_1x_3x_1x_2x_1x_4x_1x_2x_1x_3x_1x_2x_1.$$

- The Zimin patterns are palindromic,
- Over a finite alphabet, the Zimin patterns are unavoidable :

For any infinite word x over an alphabet $\mathcal{A}$, and any $n \geq 1$, there exists a morphism $\psi : \mathcal{A}_X \longrightarrow \mathcal{A}^+$ such that $\psi(Z_n)$ is a factor of x.

Definition (Zimin word)

We define the Zimin word over the alphabet $\mathcal{A}_X = \{x_1, x_2, \ldots\}$ *as*

$$Z = \lim Z_n.$$

Equivalently :

- $\prod_{n\geq 1} x_{val_2(n)+1}$ where val_2 is the 2-adic valuation,
- $Z = \varphi(Z)$, where φ is the morphism defined by $\varphi(x_i) = x_1 x_{i+1}$ for $i \geq 1$ on $\mathcal{A}_X$.

Define the sequences $(U_n)_{n\geq 1}$ and $(V_n)_{n\geq 1}$ with $U_1 = V_1 = x_1$ and for $n \geq 1$:

$$U_{n+1} = x_{n+1} U_1 U_2 \dots U_n \quad \text{and} \quad V_{n+1} = V_n V_{n-1} \dots V_1 x_{n+1}.$$

Example : $U_2 = x_2x_1$, $U_3 = x_3x_1x_2x_1$, $U_4 = x_4x_1x_2x_1x_3x_1x_2x_1$.
We have :

- $\widetilde{U_n} = V_n$,
- $|Z_n| = 2^n - 1$, and $|U_n| = |V_n| = 2^{n-1}$,
- $U_{n+1} = x_{n+1}Z_n$, and $V_{n+1} = Z_n x_{n+1}$,
- Equivalently, $Z_n = U_1 U_2 \dots U_n = V_n V_{n-1} \dots V_1$.

Also :

$$Z = \prod_{i\geq 1} U_i.$$

For $A = \{i_1 < \ldots < i_k\} \subset \mathbb{N}^*$, set

$$U_A = U_{i_1} U_{i_2} \ldots U_{i_k} = \overrightarrow{\prod_{i \in A}} U_i .$$

For $B = \{j_1 < \ldots < j_k\} \subset \mathbb{N}^*$, set

$$V_B = V_{j_k} V_{j_{k-1}} \ldots V_{j_1} = \overleftarrow{\prod_{j \in B}} V_j .$$

- The strict suffixes of U_n are exactly of the form U_A where $\emptyset \neq A \subset [1, n[$.
- For all $A \subset [1, n[$, $Z_{n-1} = V_{[1,n[\setminus A} U_A$ and $U_n = x_n V_{[1,n[\setminus A} U_A$.

Between two numbers of same 2-adic valuation there must be a number with a higher 2-adic valuation.
For u a factor of Z, let

$$k(u) = \max\{k \geq 1 \mid x_k \text{ appears in } u\}.$$

The letter $x_{k(u)}$ appears only once in u.

Theorem

Every factor u of Z writes uniquely in the form

$$u = U_A x_{k(u)} V_B$$

where $A, B \subset [1, k(u)[$.

Let u and v two factors of Z such that $k(u) < k(v)$, set :

$$u = u_{A_1} x_{k(u)} v_{B_1} \quad \text{and} \quad v = u_{A_2} x_{k(v)} v_{B_2}$$

- uv is a factor of Z

$$\iff k(u) \notin A_2 \text{ and } A_2 \cap [1, k(u)[= [1, k(u)[\setminus B_1$$

  and in this case :

$$uv = u_{A_1 \cup \{k(u)\} \cup (A_2 \cap]k(u), k(v)[)} x_{k(v)} v_{B_2}$$

- u is a suffix of v

$$\iff k(u) \in B_2 \text{ and } B_2 \cap [1, k(u)[= B_1.$$

Corollary

Let u, v, w be three factors of Z with $k(u) < k(v) < k(w)$, then

- *If uv and vw are factors of Z, then uvw is a factor of Z,*
- *If uw and vw are factors of Z, then u is a suffix of v.*

Lemma

Let $m \geq 1$, written $m = \sum_{i=0}^{N} b_i 2^i$ in base 2. Then :

$$\widetilde{\mathbb{P}}_m(Z) = \prod_{i=0}^{N}{}^{\uparrow} U_{i+1}^{b_i} \quad \textit{and} \quad T^m(Z) = \prod_{i\geq 0}^{N}{}^{\uparrow} U_{i+1}^{1-b_i}.$$

For

$$\gamma = \sum_{i=1}^{+\infty} b_i 2^i$$

a 2-adic number not an integer, this allows us to define the infinite products

$$\widetilde{\mathbb{P}}_\gamma(Z) = \prod_{i=0}^{+\infty}{}^{\uparrow} U_{i+1}^{b_i} \quad \text{and} \quad T^\gamma(Z) = \prod_{i=0}^{+\infty}{}^{\uparrow} U_{i+1}^{1-b_i}$$

If γ is not an integer, then we have the reciprocity :

$$T^{\gamma}(Z) = \widetilde{\mathbb{P}}_{\overline{\gamma}}(Z)$$

where $\gamma \longmapsto \overline{\gamma} = -1 - \gamma$ is an involution on the set of 2-adic numbers that are not integers.

Let $\Omega(Z)$ be the set of infinite words sharing the same factors as Z.

A bi-infinite word is a Zimin orbit if it shares the same factor as Z.

Theorem

Every Y in $\Omega(Z)$ writes uniquely in the form

$$Y = \widetilde{\mathbb{P}}_{\gamma}(Z)$$

for a 2-adic number not a natural number.

Moreover, every Zimin orbit write uniquely in the form

$$\widetilde{T^{\overline{\gamma}}(Z)} \cdot T^{\gamma}(Z)$$

for a 2-adic number γ not an integer.

- We have obtained a similar statement for the Ostrowski number system and Sturmian words.
- There is no letter a such that aZ shares the same factors as Z.

An extension of an infinite word x is a word of the form ux for a word u sharing the same factors as x. If u is a letter, we call it an extension by a letter.

Corollary

An element of $\Omega(Z)$ is a suffix of Z if and only if it admits only a finite number of extensions.

The Doubling-Period word :

Defined as $D = \psi(Z)$ where $\psi(x_n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$.

Also

$$D = \sigma(D) = 01000101010001000100010\ldots$$

where $\sigma : 0 \mapsto 01, 1 \mapsto 00$.
With $\varphi : x_i \mapsto x_1 x_{i+1}$ for $i \geq 1$ and $\varphi(Z) = Z$, the diagram

$$\begin{array}{ccc} \mathcal{A}_X^+ & \xrightarrow{\varphi} & \mathcal{A}_X^+ \\ \downarrow{\scriptstyle \psi} & & \downarrow{\scriptstyle \psi} \\ \{0,1\}^+ & \xrightarrow[\sigma]{} & \{0,1\}^+ \end{array}$$

commutes.

Solving the equation $\psi(u) = w$:

- If $|w| \geq 2^k + 2^{k-1}$, then the positions in u of every letter x_ℓ with $1 \leq \ell \leq k$ are uniquely determined. This lower bound is optimal.
- If u can be chosen so that the letters $x_{k(u)-2}$, $x_{k(u)-1}$ and $x_{k(u)}$ appear in this order, then solutions to $\psi(u) = w$ differ by only one letter.

Theorem

- *D has exactly two extensions by a letter,*
- *every element of $\Omega(D)$ distinct from D has exactly one extension by a letter,*
- *For $y \in \Omega(D)$ with $y \neq D$, the following are equivalent :*
    - *There exists $Y \in \Omega(Z)$ with $\psi(Y) = y$,*
    - *D is not a suffix of y,*

A such Y is then uniquely determined.

The bi-infinite orbits of the Doubling-Period word are exactly :

$$\widetilde{D} \cdot 0 \cdot D$$

$$\widetilde{D} \cdot 1 \cdot D$$

$$\psi\left(\widetilde{T^{\overline{\gamma}}(Z)} \cdot T^{\gamma}(Z)\right)$$

for γ a 2-adic number not an integer.

This results helped for specific validations of the conjecture :
An infinite word x over an alphabet $\mathcal{A}$ is ultimately-periodic if and only if for every coloring on $\mathcal{A}^+$ there exist a suffix y of x admitting a factorisation $y = u_1 u_2 u_3 \cdots$ in finite words such that the family

$$(u_{n_1} u_{n_2} \cdots u_{n_k})_{k \geq 1, n_1 < n_2 < \ldots < n_k}$$

is monochromatic.

A finite word is unbordered if it has no strict prefix that is also one of its suffix.

Definition

A Lyndon word u over an alphabet $\mathcal{A}$ is an unbordered word for which there exists a total order on $\mathcal{A}$ such that u is strictly smaller than every of its suffixes.

Examples : 011, 00110111, 0001000101, 0010010110111.

Remark : The unbordered condition is independent of the choice of the total order on $\mathcal{A}$.

Every finite word over a totally ordered alphabet writes uniquely as a decreasing product of Lyndon words.
Lyndon words have multiple applications, for example they describe basis for free Lie Algebra with the use of necklace polynomials...

Lyndon words are linked to the Zimin patterns, for example with Shirshov's theorem :

Let x be an infinite word over a finite and totally ordered alphabet $\mathcal{A}$, such thatfor every factor u of x, there exists $p \geq 1$ such that u^p is not a factor of x. Then for all $n \geq 2$, there exists a morphism $\psi : \mathcal{A}_X \longrightarrow \mathcal{A}^+$ such that $\psi(Z_n)$ is a factor of x with

- For all $i = 1 \ldots n$, $\psi(V_i)$ is a Lyndon word,
- We have $\psi(V_1) > \psi(V_2) > \ldots > \psi(V_n)$
- For all non-trivial permutations σ on $[1, n]$, $\psi(V_{\sigma(n)})\psi(V_{\sigma(n-1)}) \cdots \psi(V_{\sigma(1)}) < \psi(Z_n)$

Definition

An infinite word x is called (infinite) Lyndon if there exists a total order on $\mathcal{A}$ such that

$$x < T^n(x) \quad \text{for all } n \geq 1.$$

Examples :

- $\prod_{i \geq 1} 01^i$
- $0c_\alpha$ and $1c_\alpha$ where c_α is a characteristic Sturmian word.

An infinite Lyndon word is not ultimately-periodic and admits no factorisation in terms of its prefixes.

Lyndon words have known multiple generalisations :

- By E. Charlier (Nyldon words),
- By G. Fici (Universal Lyndon words),
- By C. Reutenauer (Generalisation of Galois words).

We introduce a new variation.

Definition

An infinite word x over an alphabet $\mathcal{A}$ is said to have a Lyndon orbit if every of its suffix is Lyndon.

Remark : The total order on $\mathcal{A}$ may vary.
Trivial example : The word $L = x_1 x_2 x_3 x_4 \ldots$.
The pattern $xyxx$ cannot appear in a word with a Lyndon orbit.

Theorem

A word with a Lyndon orbit must be defined on an infinite alphabet.

proof : For x a word with a Lyndon orbit over a finite alphabet,

- By finiteness of the number of total order on $\mathcal{A}$, we can classify the suffixes of x with respect to the order for wich they are Lyndon,
- This implies by the Lyndon property that x has at most a finite number m of elements in $\Omega(x)$ that are not suffixes of x,
- This implies that for every $\lambda \geq 1$ there exists a suffix of x with at most m factors of length λ,
- If we take $\lambda \geq m$, by the Morse-Hedlund theorem we obtain that x is ultimately-periodic, contradiction.

Theorem

If x has a Lyndon orbit, then every element of $\Omega(x)$ has a Lyndon orbit.

Proof : Let $y \in \Omega(x)$

- Take a sequence $(u_n)_{n\geq 1}$ such that $\lim T^{u_n}(x) = y$, and for $n \geq 1$ let $\leq_n$ be a total order on $\mathcal{A}$ such that $T^{u_n}(x)$ is Lyndon,
- For $i \geq 1$, let F_i be the set of the first i letters appearing in y. Will necessary be an infinite sequence after the proof,
- Construct with the countable axiom of choice, with the non-emptyness of a countable projective limit of finite sets, a total order on $\mathcal{A}$, that is a sub-projective-limit of the sequence $(\leq_n)_{n\geq 1}$,
- Check that y is Lyndon for this total order with the use of the definitions of F_i.

Corollary

If x had a Lyndon orbit, then every element of $\Omega(x)$ is defined over an infinite alphabet.

The word $\prod_{i\geq 1} x_i x_1^i$ has not a Lyndon orbit.

It is hard to construct examples of words with a Lyndon orbit that are recurrent. But the Zimin word Z is one of them.

Theorem

The Zimin word Z has a Lyndon orbit.

Recall that every $Y \in \Omega(Z)$ writes uniquely in the form

$$Y_M = \prod_{i \in M} U_i$$

for an infinite subset $M \subset \mathbb{N}^*$.
The set M is uniquely determined by the order of apparition $\leq_Y$ of letters in Y_M :

$$\begin{aligned} & x_{a_1} <_Y x_1 <_Y \dots <_Y x_{a_1-1} \\ <_Y & x_{a_2} <_Y x_{a_1+1} <_Y \dots <_Y x_{a_2-1} \\ <_Y & x_{a_3} <_Y x_{a_2+1} <_Y \dots \\ & \dots \\ x_{a_{i-1}-1} <_Y & x_{a_i} <_Y x_{a_{i-1}+1} <_Y \dots <_Y x_{a_i-1} \\ <_Y & x_{a_{i+1}} <_Y \dots \end{aligned}$$

Theorem

The Zimin word Z has a Lyndon orbit.

It is enough to show that every Y_M is Lyndon. But with the properties of the Zimin word, the word Y_M is Lyndon for the total order

$$x_{a_1} <_M x_{a_2} <_M x_{a_3} <_M \cdots <_M x_{a_i} <_M \cdots <_M x_1 <_M x_2 <_M \\ \cdots x_{a_1-1} <_M x_{a_1+1} <_M \cdots <_M x_{a_2-1} <_M x_{a_2+1} <_M \cdots$$

obtained by juxtaposition of M and M^c endowed with orders isomorphic to $\mathbb{N}$.

Moreover, an element $Y \in \Omega(Z)$ is a suffix of Z if and only if it Lyndon with respect to an order on $\mathcal{A}_X = \{x_1, x_2, x_3, \ldots\}$ isomorphic to $\mathbb{N}$.

- - -=== The end...? ===- - -

Bonus : About the Study of differences between suffix and general elements of the dynamical orbit.

Let $x = x_0 x_1 x_2 x_3 \dots$ be a non-ultimately periodic word over an alphabet $\mathcal{A}$. For u a factor of x, define the integers $A(u)$ and $B(u)$ such that

$$x_{A(u)} x_{A(u)+1} \cdots x_{B(u)-1}$$

is the first occurrence of u in x.

Due to non-ultimate periodicity, for all $k \geq 0$ there exist $n \geq 1$ such that

$$A(\mathbb{P}_n(T^k(x))) = k.$$

For u a factor of x, a decomposition $u = u_1 u_2 \dots u_n$ is called consecutive if

$$A(u) = A(u_1),\ B(u) = B(u_n)$$

and

$$B(u_i) = A(u_{i+1}) \text{ for all } i = 1 \dots n-1.$$

Definition

For u a factor of x define the consecutive length $L(u)$ of u as the maximal number of terms in a consecutive decomposition of u.

A decomposition of u is called maximal if it has $L(u)$ terms.

A factor u is called irreducible if $L(u) = 1$.

Proposition

A maximal consecutive decomposition is made of irreducible elements. Moreover, the consecutive length of k consecutive terms in a maximal decomposition equals k.

However the consecutive length has defaults :

Theorem

Let u and v be two consecutive factors of x, then

$$L(u) + L(v) \leq L(uv) \leq L(u) + L(v) + 1$$

It may happen that $L(uv) \neq L(u) + L(v)$ for two consecutive factors u and v.

For any suffix $y = T^k(x)$ of x, with $k \geq 0$ and any $\ell \geq 1$, there exists a prefix of y with consecutive length ℓ. There exists $n \geq 1$ such that

$$L(\mathbb{P}_n(T^k(x))) = \ell \quad \text{and} \quad A(\mathbb{P}_n(T^k(x))).$$

The consecutive length can be used to give a caracterisation of suffixes of x.

Theorem

An infinite word y is a suffix of x if and only if it admits only a finite number of irreducible prefixes.

− − − ====== THE END====== − − −

Thank you for your attention !