From palindromes to Sturmian and dendric shifts in memory of Aldo de Luca

Dominique Perrin

June 28, 2019

Outline

- Sturmian words
- The palindromization map
- Sturmian and episturmian shifts
- Extension to the free group
- S-adic representations
- Dendric shifts

Sturmian words

Sturmian words are easy to define. They are the infinite words on a binary alphabet $A = \{a, b\}$ having $n+1$ distinct factors of length n for every $n \geq 0$.
The best known example of Sturmian word is the infinite Fibonacci word

$$x = abaababaabaab\cdots$$

which is the unique fixed point of the Fibonacci morphism $\varphi : a \to ab, b \to a$. The words $F_n = \varphi^n(a)$ for $n \geq 0$ are prefixes of x and are called the (finite) Fibonacci words.

Special words

Since there is one more factor of length $n+1$ than of length n, a Sturmian word has, for every n, exactly one factor u of length n having two left extensions au and bu. Such a factor is called left-special (one defines symmetrically the right-special factors).
For example, the Fibonacci words F_n are left-special with respect to the infinite Fibonacci word x. Indeed, this is true for $n=0$ since aa and ba are factors of x and next, one has

$$aF_{n+1} = \varphi(bF_n), \quad abF_{n+1} = \varphi(aF_n)$$

showing by induction that all F_n are left-special.

Standard words

The different left-special factors of a Sturmian word are necessarily prefixes of one another. A Sturmian (infinite) word is called standard if its left-special factors are its prefixes (or equivalently if all its prefixes are left-special).

For example, the infinite Fibonacci word is standard. Indeed, all finite Fibonacci words are among its prefixes and thus all its prefixes are left-special.

Palindromes

Aldo invented the fundamental notion of iterated palindromic closure. It is the following transformation Pal on words. For a word w, denote by $w^{(+)}$ the shortest palindrome having w as a prefix. For example, one has $(aaba)^{(+)} = aabaa$ and $(abaa)^{(+)} = abaaba$.
We then define Pal as the unique map from A^* to A^* such that $\mathrm{Pal}(\varepsilon) = \varepsilon$ and next, for $w \in A^*$ and $a \in A$,

$$\mathrm{Pal}(wa) = (\mathrm{Pal}(w)a)^{(+)}.$$

For example, $\mathrm{Pal}(aba) = abaaba$.
Since $\mathrm{Pal}(w)$ is a prefix of $\mathrm{Pal}(wa)$, the map Pal extends to the set A^ω of infinite words on A.

The Palindromization Theorem

The following result appears in a fundamental paper by Aldo on Sturmian words.

Theorem (de Luca, 1997)

Let A be a binary alphabet. The map Pal is a bijection from the set of words in A^ω with an infinite number of occurrences of each letter onto the set of standard Sturmian words.

The word $y \in A^\omega$ is called the directive word of $\mathrm{Pal}(y)$. For example, the word $y = (ab)^\omega$ is the directive word of the Fibonacci infinite word.

Justin's Formula

The Palindromization Theorem is closely related to another important statement, known as Justin's Formula.
For a letter $a \in A$, let L_a be the transformation on words consisting in placing an a before every letter $b \neq a$ and let $u \mapsto L_u$ be its extension to a morphism. Symmetrically, R_a places an a after each $b \neq a$.
Then Justin's Formula states that for every $u, v \in A^*$

$$\mathrm{Pal}(uv) = L_u(\mathrm{Pal}(v))\,\mathrm{Pal}(u).$$

or symmetrically

$$\mathrm{Pal}(uv) = \mathrm{Pal}(u)R_u(\mathrm{Pal}(v)).$$

Consequently for $u \in A^*$ and $v \in A^\omega$,

$$\mathrm{Pal}(uv) = L_u(\mathrm{Pal}(v))$$

It is easy to deduce from Justin's Formula that $x = \mathrm{Pal}((ab)^\omega)$ is the Fibonacci infinite word. Indeed, since $(ab)^\omega = (ab)(ab)^\omega$, one obtains by Justin's Formula, with $u = ab$ and $v = (ab)^\omega$

$$x = L_{ab}(x).$$

Since $L_{ab} = \varphi^2$, where φ is the Fibonacci morphism, we conclude that x is the Fibonacci infinite word.

Episturmian words

The beauty of all this is that it carries on to larger alphabets, as shown by Droubay, Justin and Pirillo (2001).
Indeed, a generalization of Sturmian words can be defined on arbitrary finite alphabets as follows. An infinite word x on the alphabet A is called episturmian if

1. its set of factors is closed under reversal and if,
2. for each $n \geq 1$, there is exactly one left-special factor u of length n, that is, with more than one extension au for $a \in A$ which is a factor of x.

The Tribonacci word

For example, the Tribonacci word

$$x = abacaba\cdots$$

which is the unique fixed point of the Tribonacci morphism $\tau : a \to ab, b \to ac, c \to a$ is episturmian.

The unique left special factor of length n is the prefix of length n of x. As for the Fibonacci word, this is easy to verify by induction since $T_n = \tau^n(a)$ satisfies

$$aT_{n+1} = \tau(cT_n), \quad abT_{n+1} = \tau(aT_n), \quad acT_{n+1} = \tau(bT_n).$$

Standard episturmian words

Like for Stumian words, an episturmian word is standard if its left-special factors are its prefixes.

The Tribonacci word is standard.

The Palindromization Theorem (and Justin's Formula) hold for any alphabet, replacing 'Sturmian' by 'episturmian'.

The word $y \in A^\omega$ is again called the directive word of $x = \mathrm{Pal}(y)$.

For example, the directive word of the Tribonacci word is $y = (abc)^\omega$. This can be verified again easily using Justin's Formula. Indeed, the word $x = \mathrm{Pal}((abc)^\omega)$ satisfies

$$x = L_{abc}(x)$$

But $L_{abc} = \tau^3$ where τ is the Tribonacci morphism. Indeed,

$$L_{abc}(a) = L_{ab}(ca) = L_a(bcba) = abacaba$$

while

$$\tau^3(a) = \tau^2(ab) = \tau(abac) = abacaba$$

and similarly for b, c.

Strict episturmian words

An episturmian word x is strict if the unique left-special factor u of length n of x is for every n such that au is a factor of x for every $a \in A$.

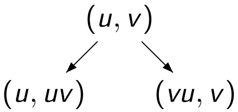
For example, the Tribonacci word is strict. Strict episturmian words are also called *Arnoux-Rauzy words* after the initial paper of Arnoux and Rauzy.

As a complement to the Palidromization Theorem, the word $\mathrm{Pal}(y)$ is a strict episturmian word if every letter appears infinitely often in y.

Droubay, Justin and Pirillo give due credit to Aldo by calling *Al* the following condition for an infinite word x: if v is a prefix of x, then $v^{(+)}$ is also a prefix of x. It is clear that an infinite word satisfies *Al* if and only if $x = \mathrm{Pal}(y)$ for some $y \in A^\omega$ and thus if and only if it is episturmian.

Central words

A formulation of the Palindromization Theorem in terms of finite words has been given by Aldo by using the notion of central word.
Central words are closely related to finite standard words, that are the basic bricks for constructing standard Sturmiam words, in the sense that every standard Sturmian word is the limit of a sequence of finite standard words (see Lothaire(2002)). For instance, the finite Fibonacci words are standard. Finite standard words can be defined directly as the words appearing in the pairs (u, v) obtained starting with (a, b) and applying iteratively one of the so-called Rauzy rules



A finite word w over the alphabet $\{a, b\}$ is central if wab (or equivalently wba) is a standard word. Aldo also proved that *the map* Pal *is a bijection from* $\{a, b\}^*$ *onto the set of central words*.

In relation with this result (and thus also with Theorem 1), a very deep connection between palindromes and periods in finite words was discovered by Aldo and Filippo Mignosi. They proved that *a word* w *in* $\{a, b\}^*$ *is central if and only if, for some relatively prime natural numbers* p, q *,* w *has two periods* p *and* q *and is of length* $|w| = p + q - 2$*.* This also shows that central words correspond to the extremal case of the famous Fine and Wilf periodicity lemma.

Aldo's work on Sturmian words has recently been put in perspective in the very nice book of Christophe Reutenauer, where its connections with number theory (in particular with continued fractions) are presented.

A palindromization map for the free group

The map $u \mapsto R_u$ extends to an automorphism of the free group $FG(A)$ on the alphabet A. The palindromization map $\mathrm{Pal} : A^* \to A^*$ has been extended to the free group $FG(A)$ by Kassel and Reutenauer (2008) for a binary alphabet.

Theorem (P., Reutenauer, 2019)

There exists a unique extension of Pal *to a map from* $FG(A)$ *to itself fixing every* $a \in A \cup A^{-1}$ *and satisfying*

$$\mathrm{Pal}(uv) = \mathrm{Pal}(u)R_u(\mathrm{Pal}(v)).$$

for every $u, v \in FG(A)$.

For example

$$\mathrm{Pal}(ab^{-1}) = aR_a(b^{-1}) = a(R_a(b))^{-1} = a(ab)^{-1} = b^{-1}$$

In particular, Pal is not anymore injective.

Elements with equal palindromization

The Braid group B_3 of braids on three strands is

$$B_3 = \langle \sigma_1, \sigma_2 \mid \sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2 \rangle.$$

Proposition (Kassel, Reutenauer)

The map $\beta : a \mapsto \sigma_1, b \mapsto \sigma_2^{-1}$ induces an isomorphism from the group generated by R_a, R_b onto B_3 and $\mathrm{Pal}(u) = \mathrm{Pal}(v)$ if and only if $\beta(u^{-1}v) \in \langle \sigma_1\sigma_2^{-1}\sigma_1^{-1} \rangle$.

For example, $\beta(b^{-1}) = \beta(ab^{-1})$ in agreement with

$$\begin{aligned} \beta(bab^{-1}) &= \sigma_2^{-1}\sigma_1\sigma_2 \\ &= \sigma_2^{-1}(\sigma_2\sigma_1\sigma_2\sigma_1^{-1}) = \sigma_1\sigma_2\sigma_1^{-1}. \end{aligned}$$

On more than two letters, we do not know the nature of the group generated by the R_a.

Sequential transducers

Justin's Formula $\mathrm{Pal}(uv) = \mathrm{Pal}(u)R_u(\mathrm{Pal}(v))$ can be seen as expressing that $\mathrm{Pal} : A^* \to A^*$ is realizable by a sequential transducer.

Proposition (P., Reutenauer)

The function Pal *is defined by the transducer on the set* $\mathrm{Aut}(FG(A))$ *with transition and output functions*

$$R_u \cdot a = R_{ua} \text{ and } R_u * a = R_u(a)$$

respectively, and with initial state $i = R_\varepsilon$.

Cocycles

Justin's Formula is also related to the notion of nonabelian group cohomology, as pointed out by Kassel and Reutenauer (2008). A function f from $FG(A)$ to itself, is a 1-cocycle with respect to a group morphism $\alpha : u \mapsto \alpha_u$ from $FG(A)$ to $\mathrm{Aut}(FG(A))$ if Equation

$$f(uv) = f(u)\alpha_u(f(v))$$

holds for all $u, v \in FG(A)$. Thus Pal, as a function from $FG(A)$ to itself, is a 1-cocycle with respect to the morphism R.

Trivial cocycles

A 1-cocycle is trivial if there is an element $x \in FG(A)$ such that

$$f(u) = x^{-1}\alpha_u(x)$$

for every $u \in FG(A)$. When A has two elements, one has

$$\mathrm{Pal}(u) = (ab)^{-1}R_u(ab)$$

Thus the 1-cocycle Pal is trivial.
This is not the case on more than two letters. Indeed, assume that $x \in FG(A)$ is such that $\mathrm{Pal}(u) = x^{-1}R_u(x)$ for all $u \in FG(A)$. One has then $xa = R_a(x)$ for every $a \in A$ and thus, by abelianization, $|x|_a + 1 = |x|$. This implies $|x|(\mathrm{Card}(A) - 1) = \mathrm{Card}(A)$ which is impossible for $\mathrm{Card}(A) \geq 3$ since $|x|$ is an integer.

Shift spaces

A shift space S is a set of infinite words on an alphabet A specified by a set of forbidden factors. It is thus defined uniquely by its set of factors $\mathcal{L}(S)$.
For example, the Golden mean shift is the set of infinite words on $A = \{a, b\}$ without factor bb.
A shift S is minimal if it does not contain properly any nonempty shft. It is equivalent to the property of for $\mathcal{L}(S)$ to be uniformly recurrent, that is such that for every $w \in \mathcal{L}(S)$ there is some $n \geq 1$ such that w is a factor of every word of $\mathcal{L}(S)$ of length n.

Sturmian shifts

By a Sturmian set, we mean the set of factors of a strict episturmian word.
A Sturmian shift (also called Arnoux-Rauzy shift) is the shift space with set of factors a Sturmian set.
For example, the Fibonacci shift is the shift generated by the Fibonacci word.

Extension graph

Let S be a shift on the alphabet A. For $w \in \mathcal{L}(S)$, denote by $LE(w)$ (resp. $RE(w)$) the set of letters a such that $aw \in \mathcal{L}(S)$ (resp. $wa \in \mathcal{L}(S)$) and by $BE(w)$ the set of pairs $(a, b) \in A \times A$ such that $awb \in \mathcal{L}(S)$. The extension graph of w is the undirected graph $\mathcal{E}(w)$ with vertices the disjoint union of $LE(w)$ and $RE(w)$ and edges the elements of $BE(S)$.

Dendric shifts

A shift S is said to be dendric (or equivalently $\mathcal{L}(S)$ is a tree set) if $\mathcal{E}(w)$ is a tree for every $w \in \mathcal{L}(S)$.
The family of dendric shifts contains the Sturmian shifts as well as the interval exchange shifts.

Example

The Fibonacci shift is dendric. The graphs $\mathcal{E}(\varepsilon)$ and $\mathcal{E}(a)$ are shown below.

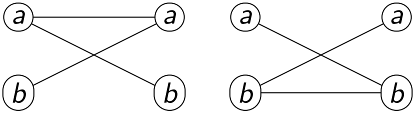


Figure : The graphs $\mathcal{E}(\varepsilon)$ and $\mathcal{E}(a)$ for the Fibonacci shift.

The coding of the Fibonacci word by nonoverlapping blocks of length 2

$$yxzzyxzzyyx\cdots$$

(with $x = aa, y = ab, z = ba$) is also dendric.

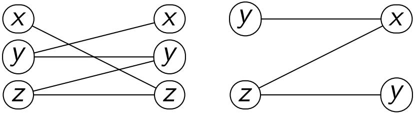


Figure : The graphs $\mathcal{E}(\varepsilon)$ and $\mathcal{E}(y)$.

Return words

For $w \in \mathcal{L}(S)$, a return word to w is a word $u \in \mathcal{L}(S)$ such that $wu \in \mathcal{L}(S)$ and that wu has exactly two occurrences of w, one as a prefix and one as a suffix. We denote by $\mathcal{R}_S(w)$ the set of return words to w in S. Thus a return word indicates the word to be read before the next occurrence of w in a left to right scan. A symmetric notion of left return word is obtained replacing wu by uw in the definition.

Let for example S be the golden mean shift. Then $\mathcal{R}_S(a) = \{a, ba\}$ and $\mathcal{R}_S(b) = a^+b$.

Return words in Sturmian shifts

Proposition (Justin, Vuillon, 2000)

Let s be a standard strict episturmian word over A, let $\Delta = a_0 a_1 \cdots$ be its directive word, and let (u_n) be its sequence of palindrome prefixes.

(i) *The return words to u_n are the words $R_{a_0 \cdots a_{n-1}}(a)$ for $a \in A$.*

(ii) *For each factor u of s, there exist a word z and an integer n such that the return words to u are the words zyz^{-1}, where y ranges over the return words to u_n.*

As a consequence, the group generated by every $\mathcal{R}_S(w)$ is, for any Sturmian shift S, equal to the free group on A.

Example

Let $s = abaabab \cdots$ be the Fibonacci word. Then $u_1 = a$, $u_2 = aba$ and $u_3 = abaaba$. $\mathcal{R}_S(u_3) = \{aba, baaba\} = \{R_{aba}(a), \mathcal{R}_{aba}(b)\}$ and $\mathcal{R}_S(aa) = \{baa, babaa\} = z\mathcal{R}_S(u_3)z^{-1}$ with $z = ba$.

Return Theorem

The following result generalizes what we have seen for Sturmian shifts.

Theorem (BDDLPRR[a],2015)

[a] Berthé, De Felice, Dolce, Leroy, Perrin, Reutenauer, Rindone

Let S be a dendric and minimal shift on the alphabet A. For any $w \in \mathcal{L}(S)$, the set $\mathcal{R}_S(w)$ is a basis of the free group $FG(A)$.

Example

Let S be the Fibonacci shift. Then $\mathcal{R}_S(aa) = \{baa, babaa\}$ is a basis of the free group on $\{a, b\}$. Set indeed $x = baa$ and $y = babaa$. Then $a = xy^{-1}x$ and $b = yx^{-1}(xy^{-1}x)^{-1}$.

S-adic representation

Let $\mathcal{S}$ be a set of morphisms and $(\sigma_n)_{n\geq 0}$ be a sequence of morphisms in $\mathcal{S}$ with $\sigma_n : A_{n+1}^* \to A_n^*$ and $A_0 = A$. Let $\Delta = a_0 a_1 \cdots$ with $a_i \in A_i$. Assume that

$$u = \lim_{n\to\infty} \sigma_0 \sigma_1 \cdots \sigma_{n-1}(a_n)$$

exists and is an infinite word $u \in A^{\mathbb{N}}$. Let S be the shift such that $\mathcal{L}(S)$ is the set of factors of u. We call the sequence (σ_n) an *$\mathcal{S}$-adic representation* of S with directive sequence Δ.

Thus $\mathcal{S}$-adic representations are a generalization of the construction of shifts by iterating morphisms as shown more precisely below.

Example

Let $\varphi : A^* \to A^*$ be a morphism such that $u = \lim \varphi^n(a)$ exists and is an infinite word for some $a \in A$. Then $(\varphi, \varphi, \ldots)$ is an $\mathcal{S}$-adic representation of the shift S generated by u (that is such that $\mathcal{L}(S)$ is the set of factors of u) with directive word $aa\cdots$.

A sequence of morphisms $(\sigma_n)_{n\geq 0}$ is said to be *primitive* if for all $r \geq 0$ there exists $s > r$ such that all letters of A_r occur in all images $\sigma_r \cdots \sigma_{s-1}(a)$, $a \in A_s$. When all morphisms σ_n are equal, one finds the usual notion of primitive morphism.

A morphism $\alpha : A^* \to A^*$ is called elementary if it is a permutation of A or it is one of the morphisms $\alpha_{a,b}$, $\tilde{\alpha}_{a,b}$ defined for $a, b \in A$ with $a \neq b$ by

$$\alpha_{a,b}(c) = \begin{cases} ab & \text{if } c = a, \\ c & \text{otherwise} \end{cases} \quad \text{and} \quad \tilde{\alpha}_{a,b}(c) = \begin{cases} ba & \text{if } c = a, \\ c & \text{otherwise.} \end{cases}$$

Thus $\alpha_{a,b}$ places a letter b after each a and $\tilde{\alpha}_{a,b}$ places a letter b before each a. We denote by $\mathcal{S}_e$ the family of elementary morphisms. The elementary morphisms are automorphisms of the free group which are positive, that is, preserving A^*.

A morphism is called tame if it belongs to the submonoid generated by the elementary morphisms. Thus the tame morphisms form a finitely generated submonoid of the monoid of positive automorphisms of $FG(A)$ which is not itself finitely generated (Tan,Wen,Zhang, 2004).

Theorem (BDDLPRR, 2015)

Every minimal dendric shift has a primitive $\mathcal{S}_e$-adic representation.

The converse is not true and it is an open problem to characterize minimal dendric shifts by their $\mathcal{S}$-adic representation. Such a characterization exists for Sturmian shifts as recalled below.

Example

Any Sturmian shift has an $\mathcal{S}$-adic representation using the particular tame morphisms L_a for $a \in A$. Such a representation characterizes Sturmian shifts in the sense that a shift is Sturmian if and only if it has an $\mathcal{S}$-adic representation of the form (L_{a_n}) where every letter in A appears infinitely often in the directive word $\Delta = a_0 a_1 \cdots$. For example, the Fibonacci shift has directive word $(ab)^\omega$, as we have seen.

The proof of the Theorem uses the following result. A positive basis X of the free group $FG(A)$ is tame if there is a tame automorphism α of $FG(A)$ such that $X = \alpha(A)$.

Theorem

Let S be a minimal dendric shift. Any basis of the free group included in $\mathcal{L}(S)$ is tame.