Normal sequences, sequences of complexity $2n+1$ and continued fractions algorithms

Gilles Didier

Institut de Mathématiques de Marseille
Aix-Marseille Université

Seqbio 2017

Lille - November 7th 2017

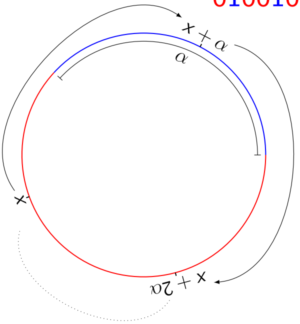
Sturmian sequences

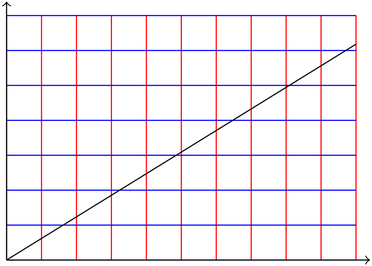
sequence u = infinite word over a finite alphabet
complexity : $\mathrm{p}_u(n)$ = the number of distinct words of length n in u

If $\mathrm{p}_u(n) \leq n$ then u is ultimately periodic

A sequence u is Sturmian if $\mathrm{p}_u(n) = n + 1$

0100101001001010010100100101001001...





Generating Sturmian sequences

Two morphisms : $\sigma_0 : \begin{array}{ccc} 0 & \to & 0 \\ 1 & \to & 10 \end{array}$ and $\sigma_1 : \begin{array}{ccc} 0 & \to & 01 \\ 1 & \to & 1 \end{array}$

$$\begin{array}{ccccccccc} u & = & 0 & 1 & 0 & 0 & 1 & \dots \\ \sigma_0(u) & = & 0 & 10 & 0 & 0 & 10 & \dots \end{array}$$

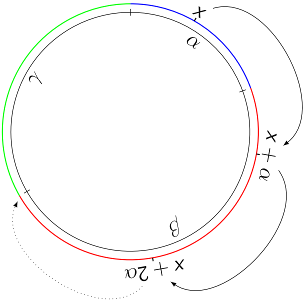
$$\lim_{n\to\infty} \sigma_0^{a_1}\sigma_1^{a_2}\dots\sigma_0^{a_{2n}}\sigma_1^{a_{2n+1}}(0)$$

is Sturmian with frequencies $(\alpha, 1-\alpha)$ where $\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots}}}$

Since letters of (ultimately) periodic sequences have rational frequencies, Sturmian sequences are the simplest class of two-letters sequences in which we can observe all non-trivial frequencies.

Three-letters sequences with arbitrary frequencies (α, β, γ)

Such sequences u have at least complexity $\mathrm{p}_u(n) = 2n + 1$



We already know

- Codings of rotations over three intervals
- Arnoux-Rauzy sequences
- Cassaigne sequences

All the sequences just above

- can be generated by applying morphisms following some multidimensional continued fraction representation of (α, β, γ)
- fulfill a combinatoric property : they are *normal*

((normal) bi-) Special words

w is *right-special* if $u = ...wx...wy...$ for two letters $x \neq y$; $w\genfrac{}{}{0pt}{}{x}{y}$

w is *left-special* if $u = ...aw...bw...$ for two letters $a \neq b$; $\genfrac{}{}{0pt}{}{a}{b}w$

w is *bi-special* if it is both left- and right-special

If w is bi-special $\genfrac{}{}{0pt}{}{a}{b}w$ and $w\genfrac{}{}{0pt}{}{x}{y}$ then it may be either :

- *weak* : both aw and bw are not right-special
  e.g., only awx and bwy occur in u
- *normal* : only one word among aw and bw is right-special
- *strong* : both $aw\genfrac{}{}{0pt}{}{x}{y}$ and $bw\genfrac{}{}{0pt}{}{x}{y}$

u is *normal* if all its bi-special words are normal

Return words

A *return word* over w in u is word of u which starts with w and ends just before the next occurrence of w in u

$$u = \dots \overbrace{w\underbrace{\dots\dots}_{\text{no } w}}^{\text{return word}}\overbrace{w\underbrace{\dots}_{\text{no } w}}^{\text{return word}}\overbrace{w\underbrace{\dots\dots\dots}_{\text{no } w}}^{\text{return word}}\overbrace{w\underbrace{\dots\dots}_{\text{no } w}}^{\text{return word}}\dots$$

If the set of return words R over w is finite : $R = \{r_1, r_2, \dots, r_p\}$, the *morphism associated to R* is the map $\rho : \{1, 2, \dots, p\} \to \mathcal{A}^+$ such that $\rho(i) = r_i$ for all $1 \leq i \leq p$.

In this case, up to a finite prefix, there exists a sequence v over alphabet $\{1, 2, \dots, p\}$ such that $u = \rho(v)$

Normal sequences and return words

Let u be a normal sequence.

Theorem

There exists a nonnegative integer S, which will be referred to as the increment of u, s. t. for all $n > 0$, $\mathrm{p}_u(n) = Sn + |\mathcal{A}| - S$.

Let u be a normal sequence of increment S.

Theorem

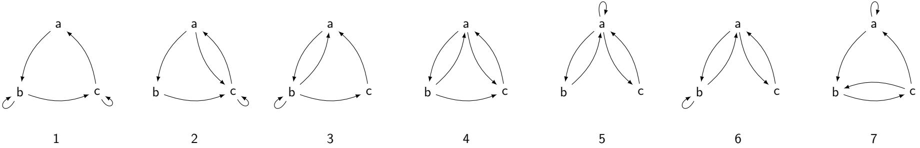
For all words w of u, there are exactly $S + 1$ return words over w.

Theorem

Let w be a word of u and ρ the morphism of its return words, up to a finite prefix, we have $u = \rho(v)$ where v is a normal sequence of increment S over an alphabet of cardinality $S + 1$.

Let ρ be a morphism of return words of any normal sequence of increment S and u a normal sequence of increment S with complexity $Sn + 1$ then $\rho(u)$ is a normal sequence of increment S.

Sequences of complexity $2n + 1$



Proposition

The graph of words of length 1 *of a sequence* u *of complexity* $2n + 1$ *is, up to renaming its letters, one of those displayed above.*
The letter frequencies $\mathrm{f}_u(a)$, $\mathrm{f}_u(b)$, $\mathrm{f}_u(c)$ *satisfy :*

- *for graph 1,* $\mathrm{f}_u(a) < \min\{\mathrm{f}_u(b), \mathrm{f}_u(c)\}$,
- *for graph 2,* $\mathrm{f}_u(b) < \mathrm{f}_u(a) < \mathrm{f}_u(c)$,
- *for graph 3,* $\mathrm{f}_u(c) < \mathrm{f}_u(a) < \mathrm{f}_u(b)$,
- *for graphs 4 and 5,* $\mathrm{f}_u(a) > \max\{\mathrm{f}_u(b), \mathrm{f}_u(c)\}$,
- *for graph 6,* $\mathrm{f}_u(c) < \mathrm{f}_u(a) < \mathrm{f}_u(b) + \mathrm{f}_u(c)$,
- *for graph 7,* $\mathrm{f}_u(b) = \mathrm{f}_u(c)$.

Normal sequences of complexity $2n+1$

Such sequences have increment 2 thus always 3 return words

The *incidence matrix* of morphism $\sigma : \mathcal{B} \to \mathcal{A}$ is the $\mathcal{A} \times \mathcal{B}$ matrix $M^{(\sigma)}$ where entry $M^{(\sigma)}_{a,x}$ is the number of occurrences of a in $\sigma(x)$

$$\sigma : \begin{array}{ccc} a & \to & ab \\ b & \to & b \end{array} \quad , \quad M^{(\sigma)} = \begin{array}{c} \\ a \\ b \end{array} \begin{array}{c} \begin{array}{cc} \sigma(a) & \sigma(b) \end{array} \\ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \end{array}$$

If $u = \sigma(v)$ letter frequencies of u and v are related through $M^{(\sigma)}$

Theorem

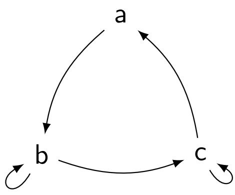
Incidence matrices of morphisms associated to return words over letters of normal sequences of complexity $2n+1$ with graph 1 to 6 are unimodular (i.e., have determinant 1 or -1).
Those with graph 7 are not inversible.

Normal sequences of complexity $2n+1$

If u is a normal sequence of complexity $2n+1$ with graph 1,

- its return words over a are one of the following sets
  - ab^kc^ℓ, $ab^kc^{\ell+1}$, $ab^{k+1}c^\ell$,
  - ab^kc^ℓ, $ab^kc^{\ell+1}$, $ab^{k+1}c^{\ell+1}$,
  - $ab^{k+1}c^\ell$, $ab^{k+1}c^{\ell+1}$, ab^kc^ℓ,
  - $ab^{k+1}c^\ell$, $ab^{k+1}c^{\ell+1}$, $ab^kc^{\ell+1}$;
- its return words over b are
  - b, $bc^\ell a$, $bc^{\ell+1}a$;
- its return words over c are
  - c, cab^k, cab^{k+1} ;

where $k = \left\lfloor \frac{\mathrm{f}_u(b)}{\mathrm{f}_u(a)} \right\rfloor$ and $\ell = \left\lfloor \frac{\mathrm{f}_u(c)}{\mathrm{f}_u(a)} \right\rfloor$.



The incidence matrix of the morphism associated to $\{ab^kc^\ell, ab^kc^{\ell+1}, ab^{k+1}c^\ell\}$ is

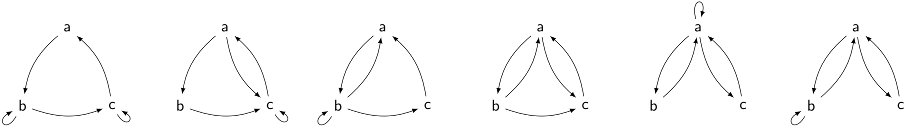
$$\begin{pmatrix} 1 & 1 & 1 \\ k & k & k+1 \\ \ell & \ell+1 & \ell \end{pmatrix}$$

Normal sequences of complexity $2n+1$

Theorem

Let (α, β, γ) be a frequency vector of rational independent entries. The set of normal sequences of complexity $2n+1$ with letter frequencies α, β and γ is uncountable.

$$\alpha < \beta < \gamma$$



1
$$\begin{aligned} f_u(a) &= \alpha \\ f_u(b) &= \beta \\ f_u(c) &= \gamma \end{aligned}$$

2
$$\begin{aligned} f_u(a) &= \beta \\ f_u(b) &= \alpha \\ f_u(c) &= \gamma \end{aligned}$$

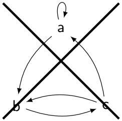
3
$$\begin{aligned} f_u(a) &= \beta \\ f_u(b) &= \gamma \\ f_u(c) &= \alpha \end{aligned}$$

4
$$\begin{aligned} f_u(a) &= \gamma \\ f_u(b) &= \alpha \\ f_u(c) &= \beta \end{aligned}$$

5
$$\begin{aligned} f_u(a) &= \gamma \\ f_u(b) &= \alpha \\ f_u(c) &= \beta \end{aligned}$$

6
$$\begin{aligned} f_u(a) &= \beta \\ f_u(b) &= \gamma \\ f_u(c) &= \alpha \end{aligned}$$

α, β, γ rationally independent $\Rightarrow$



7