Complément de mémoire : Reading group

Reading of chapter 2 "Substitutions, arithmetic and finite automata: an introduction" in Pytheas Fogg N., Substitutions in dynamics, arithmetics, and combinatorics. Berlin : Springer, 2002.

Manon Stipulanti

University of Liège
Faculty of Sciences
Department of mathematics
Professor : M. Rigo

23 February 2015

Table of contents

Table of contents

Introduction

This chapter consists in giving examples of sequences that are pure morphic or morphic words. Moreover these words come from fixed points of morphisms of constant length $\ell \in \{2, 3\}$. The author wants to show how these morphic words can also be defined by an automatic property of the ℓ-adic development of the integers and how some of them are answers to some problem.

In order to do so, the author introduces four famous sequences:

- Thue–Morse,
- Rudin–Shapiro,
- Baum–Sweet,
- Cantor.

Table of contents

Definition

The Thue–Morse sequence $\mathbf{u} = (u_n)_{n\in\mathbb{N}}$ is defined as the (unique) fixed point beginning by a of the Morse morphism σ defined on the alphabet $\{a, b\}$ by $\sigma(a) = ab$, $\sigma(b) = ba$, i.e.

$$\mathbf{u} = abbabaabbaababbabaababbaabbabaab \cdots .$$

If we denote a by 0 and b by 1, it is easy to see that the Thue–Morse sequence verifies the following combinatorial properties.

Properties

(i) We have $u_0 = 0$ and, for all $n \in \mathbb{N}$, $u_{2n} = u_n$ and $u_{2n+1} = 1 - u_n$.

(ii) For all $k, n \in \mathbb{N}$, at position of the form $k2^n$ of the sequence appears $\sigma^n(0)$ if $u_k = 0$ and $\sigma^n(1)$ if $u_k = 1$.

If we define, for all $r \in \mathbb{N}$,

$$U_r = \sigma^r(a) \text{ and } V_r = \sigma^r(b),$$

we can prove by induction that these sequences of words over $\{a, b\}$ are uniquely defined by the following relations:

$$\begin{cases} U_0 &= a \\ V_0 &= b \end{cases} \text{ and } \begin{cases} U_{r+1} &= U_r V_r \\ V_{r+1} &= V_r U_r. \end{cases}$$

Remark

If we define the exchange morphism E on $\{a, b\}$ by $E(a) = b$, $E(b) = a$, then we have $V_r = E(U_r)$ for all $r \geq 0$.

Another definition of the Thue–Morse sequence is then given by the following result.

Proposition

We have $\mathbf{u} = \lim_{r \to +\infty} U_r$.

Proof: It is clear by definition.

$\square$

Remark

The other fixed point of σ begining by b is equal to $\lim_{r \to +\infty} V_r$.

Consequence

In order to build the Thue–Morse sequence, we can concatenate rules of the form $U_{r+1} = U_r V_r$. In other words, each block is obtained by concatenating the previous block and its opposite. The first few blocks are

$$\begin{array}{l} 0 \\ 01 \\ 0110 \\ 01101001 \\ 0110100110010110 \\ \vdots \end{array}$$

Remember the following result from chapter 1.

Proposition 1.3.1.

A sequence $\mathbf{w}$ is k-automatic in direct reading if and only if $\mathbf{w}$ is the image by a letter-to-letter projection of a fixed point of a morphism of constant length k.

In particular, if a sequence $\mathbf{w}$ is a fixed point of a k-uniform morphism, then $\mathbf{w}$ can be defined by a property of the k-adic development of the integers. Moreover, this property is simple enough to be recognisable by a finite automaton.

In our case, we define the two following subsets of $\mathbb{N}$:

- $\mathbb{N}_a$ the set of integers n such that $u_n = a$;
- $\mathbb{N}_b$ the set of integers n such that $u_n = b$.

If we observe the first letters of the Thue–Morse sequence, we see that

$$\mathbb{N}_a = \{0, 3, 5, 6, 9, 10, 12, 15, \ldots\},$$
$$\mathbb{N}_b = \{1, 2, 4, 7, 8, 11, 13, 14, \ldots\}$$

because

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$\mathbf{u} =$	a	b	b	a	b	a	a	b	b	a	a	b	a	b	b	a	$\cdots$

It is clear that $\mathbb{N}_a \cap \mathbb{N}_b = \emptyset$ and $\mathbb{N}_a \cup \mathbb{N}_b = \mathbb{N}$, so that $\mathbb{N}_a$ and $\mathbb{N}_b$ form a partition of $\mathbb{N}$.

For all $n \in \mathbb{N}$, we denote by $S_2(n)$ the sum (with carry) of the digits in the dyadic development of n, i.e.

$$S_2(n) = \sum_{i \geq 0} n_i \quad \text{if } n = \sum_{i \geq 0} n_i 2^i$$

with $n_i \in \{0, 1\}$ for all $i \geq 0$. For example, we have

n	0	1	2	3	4	5	6	7	$\cdots$
$(n)_2$	0	1	10	11	100	101	110	111	$\cdots$
$S_2(n)$	0	1	1	2	1	2	2	3	$\cdots$

if we denote by $(n)_2$ the dyadic development of n.

We can show the following result by induction.

Proposition

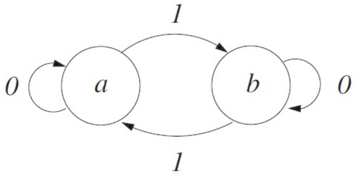
We have

$$\mathbb{N}_a = \{n \in \mathbb{N} \mid S_2(n) \text{ is even}\},$$
$$\mathbb{N}_b = \{n \in \mathbb{N} \mid S_2(n) \text{ is odd}\}.$$

We thus obtained the wanted property of the dyadic development of integers. Thanks to that simple property, we can construct an automaton recognising the Thue–Morse sequence.

Automaton

The following 2-automaton with initial state a and exit map id given by $\mathrm{id}(a) = a$ and $\mathrm{id}(b) = b$ recognises the Thue-Morse sequence (in direct reading).



Indeed, we have

$$u_n = a \Leftrightarrow n \in \mathbb{N}_a \Leftrightarrow S_2(n) \text{ is even} \Leftrightarrow (n)_2 \text{ has an even number of 1's,}$$
$$u_n = b \Leftrightarrow n \in \mathbb{N}_b \Leftrightarrow S_2(n) \text{ is odd} \Leftrightarrow (n)_2 \text{ has an odd number of 1's.}$$

The Prouhet–Tarry–Escott problem

The Prouhet–Tarry–Escott problem

Given the positive integers q and r, find an infinite number of sequences of q^r numbers that can be cut in q sets of q^{r-1} elements such that, for any $k < r$, the sum of all the k-th powers of the elements of each set is the same.

The solution to this problem was given by Prouhet in 1851. If $q = 2$, a solution to this problem is given by the Thue–Morse sequence as we will see.

By induction, we can easily prove the following lemma.

Lemma

Let r be a nonnegative integer. Then exactly half of the integers

$$0, 1, \dots, 2^{r+1} - 1,$$

namely 2^r integers, have a dyadic development containing an even number of 1's.

Corollary

For all nonnegative integer r, we have

$$\#\{n \in \mathbb{N}_a \mid n < 2^{r+1}\} = 2^r = \#\{n \in \mathbb{N}_b \mid n < 2^{r+1}\}.$$

Proposition

For all $k, r \in \mathbb{N}$ such that $k < r$, we have

$$\sum_{\substack{n \in \mathbb{N}_a \\ n < 2^r}} n^k = \sum_{\substack{n \in \mathbb{N}_b \\ n < 2^r}} n^k.$$

Proof: First of all, define

$$A_r = \{n \in \mathbb{N}_a \mid n < 2^r\},$$
$$B_r = \{n \in \mathbb{N}_b \mid n < 2^r\}.$$

For all $k < r$, we have

$$\sum_{\substack{n \in \mathbb{N}_a \\ n < 2^r}} n^k = \sum_{\substack{n \in \mathbb{N}_b \\ n < 2^r}} n^k \Leftrightarrow \sum_{n \in A_r} n^k = \sum_{n \in B_r} n^k$$

$$\Leftrightarrow \sum_{n=0}^{2^r-1} (-1)^{S_2(n)} n^k = 0.$$

Now define the polynomial $F_r \in \mathbb{Z}[X]$ by

$$F_r(X) = \sum_{n=0}^{2^r-1} (-1)^{S_2(n)} X^n.$$

By induction on r, we can show that

$$F_r(X) = \prod_{k=0}^{r-1} \left(1 - X^{2^k}\right).$$

Since $1 - X^n = (1 - X)(1 + X + X^2 + \cdots + X^{n-1})$ for all $n \in \mathbb{N}_0$, we actually have

$$F_r(X) = (1 - X)^r G_r(X)$$

with $G_r \in \mathbb{Z}[X]$. It follows that

$$(D^k F_r(X))(1) = 0$$

for all $k < r$, i.e.

$$(D^k F_r(X))(1) = \sum_{n=k}^{2^r-1} (-1)^{S_2(n)} n(n-1) \cdots (n-k+1) = 0$$

for all $k < r$.

Since $n(n-1)\cdots(n-k+1)=0$ for $n \in \{0,...,k-1\}$, we have

$$\sum_{n=0}^{2^r-1}(-1)^{S_2(n)}n(n-1)\cdots(n-k+1)=0$$

for all $k<r$. Now, by induction on $k \in \{0,...,r-1\}$, we show that

$$\sum_{n=0}^{2^r-1}(-1)^{S_2(n)}n^k=0.$$

Thanks to the previous lemma, the case holds for $k=0$. Suppose the property is true for $k \in \{0,...,\ell-1\}$ with $\ell<r$.

We will now show it still holds for $k = \ell$. We know that

$$\sum_{n=0}^{2^r-1} (-1)^{S_2(n)} \underbrace{n(n-1)\cdots(n-\ell+1)}_{=n^\ell + \sum_{j=0}^{\ell-1} \alpha_j n^j \text{ with } \alpha_j \in \mathbb{Z}\, \forall j} = 0$$

$$\Rightarrow \sum_{n=0}^{2^r-1} (-1)^{S_2(n)} n^\ell + \sum_{j=0}^{\ell-1} \alpha_j \underbrace{\sum_{n=0}^{2^r-1} (-1)^{S_2(n)} n^j}_{=0 \text{ by induction}} = 0$$

$$\Rightarrow \sum_{n=0}^{2^r-1} (-1)^{S_2(n)} n^\ell = 0$$

which is what we wanted to prove.

$\square$

Proposition

For all $k, r, n_0 \in \mathbb{N}$ such that $k < r$, we have

$$\sum_{\substack{n \in n_0 + \mathbb{N}_a \\ n < n_0 + 2^r}} n^k = \sum_{\substack{n \in n_0 + \mathbb{N}_b \\ n < n_0 + 2^r}} n^k. \tag{1}$$

<u>Proof:</u> For $\alpha \in \{a, b\}$, we have, by using the binomial theorem,

$$\sum_{\substack{n \in n_0 + \mathbb{N}_\alpha \\ n < n_0 + 2^r}} n^k = \sum_{\substack{n \in \mathbb{N}_\alpha \\ n < 2^r}} (n_0 + n)^k = \sum_{i=0}^{k} \binom{k}{i} n_0^{k-i} \sum_{\substack{n \in \mathbb{N}_\alpha \\ n < 2^r}} n^i.$$

Thanks to the previous result, we actually proved (1).

$\square$

Let r be a positive integer. We have to find an infinite number of sequences of 2^r numbers that can be cut into two sets of 2^{r-1} elements such that, for any $k < r$, the sum of all the k-th powers of the elements of each set is the same. The Thue–Morse sequence answers the question because:

- As $\#A_r = \#B_r = 2^{r-1}$, we know that, for all $n_0 \in \mathbb{N}$, the sets

  $$A_{r,n_0} = \{n \in n_0 + \mathbb{N}_a \mid n < 2^r\}, B_{r,n_0} = \{n \in n_0 + \mathbb{N}_b \mid n < 2^r\}$$

  both have 2^{r-1} elements.
- As n_0 varies into $\mathbb{N}$, we actually have an infinite number of sequences.
- For a fixed value of n_0, each sequence has 2^r elements that are equally partitioned into the two sets A_{r,n_0} and B_{r,n_0}.
- For a fixed value of n_0, each sequence verifies

  $$\sum_{\substack{n \in n_0 + \mathbb{N}_a \\ n < n_0 + 2^r}} n^k = \sum_{\substack{n \in n_0 + \mathbb{N}_b \\ n < n_0 + 2^r}} n^k.$$

$\Rightarrow$ We have a solution of Prouhet's problem in the case $q = 2$.

A statistical property

By a theorem of Fréchet, any monotone function f can be decomposed as $f = f_1 + f_2 + f_3$ where

- f_1 is a monotone step-function,
- f_2 is a monotone function which is the integral of its derivative,
- f_3 is a monotone continuous function which has almost everywhere a derivative equal to zero.

Thanks to the statistical property of the Thue–Morse sequence we will show, Malher gave an explicit example of the Fréchet decomposition where $f_3 \neq 0$.

Statistical property

For all positive integers k, N, if

$$\gamma_N(k) = \frac{1}{N} \sum_{n<N} (-1)^{S_2(n)} (-1)^{S_2(n+k)},$$

then, for any k, the sequence $(\gamma_N(k))_{N>0}$ converges and its limit is non-zero for infinitely many k's.

$\Rightarrow$ The Thue–Morse sequence has positive correlations.

Table of contents

Definition

The Rudin–Shapiro sequence $\boldsymbol{\varepsilon} = (\varepsilon_n)_{n\in\mathbb{N}}$ over the alphabet $\{-1, 1\}$ is defined by the relations $\varepsilon_0 = 1$ and, for any nonnegative integer n,

$$\begin{cases} \varepsilon_{2n} &= \varepsilon_n \\ \varepsilon_{2n+1} &= (-1)^n \varepsilon_n. \end{cases}$$

We thus have

$$\boldsymbol{\varepsilon} = 111-111-11111-1-1-11-1\cdots.$$

The sequence $\boldsymbol{\varepsilon}$ can also be obtained as follows. Let σ be the morphism over $\{a, b, c, d\}$ defined by

$$\sigma(a) = ab,\ \sigma(b) = ac,\ \sigma(c) = db,\ \sigma(d) = dc.$$

Then σ is prolongable on a and thus has a unique infinite fixed point beginning by a that we will denote by $\mathbf{v} := \sigma^{\omega}(a)$. If we denote by ψ the coding given by

$$\psi(a) = 1 = \psi(b),\ \psi(c) = -1 = \psi(d),$$

we can prove by induction that $\boldsymbol{\varepsilon} = \psi(\mathbf{v})$.

By analogy with the fact that u_n gives the parity of the sum of the digits of the dyadic development of n, it is easy to verify by induction that ε_n gives the parity of the number of words 11 in the dyadic development of n.

Proposition

For any nonnegative integer n with a dyadic development $n = \sum_{i\geq 0} n_i 2^i$ with $n_i \in \{0,1\}$, we have

$$\varepsilon_n = (-1)^{\sum_{i\geq 0} n_i n_{i+1}}.$$

Indeed, the sum $\sum_{i\geq 0} n_i n_{i+1}$ exactly counts the number of 11's in $(n)_2$.

Another useful property of ε is the following one.

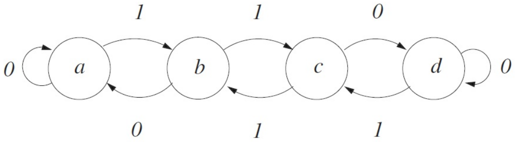
Property

For any nonnegative integers a, b and n such that $b < 2^n$, we have

$$\varepsilon_{2^{n+1}a+b} = \varepsilon_a \varepsilon_b.$$

Automaton

The following 2-automaton with initial state a and exit map ψ given by $\psi(a) = 1 = \psi(b)$ and $\psi(c) = -1 = \psi(d)$ recognises the Rudin–Shapiro sequence (in direct reading).



Indeed, we have

$$\varepsilon_n = 1 \Leftrightarrow \text{ the factor 11 appears an even number of times in } (n)_2,$$
$$\varepsilon_n = -1 \Leftrightarrow \text{ the factor 11 appears an odd number of times in } (n)_2.$$

The Salem problem

For any sequence $\boldsymbol{\alpha} = (\alpha_n)_{n\in\mathbb{N}} \in \{-1,1\}^{\mathbb{N}}$ and for all $N \in \mathbb{N}_0$, we have

$$\begin{aligned}
\int_0^1 \left| \sum_{n<N} \alpha_n \, e^{2i\pi n\theta} \right|^2 \, d\theta
&= \int_0^1 \sum_{n<N} \alpha_n \, e^{2i\pi n\theta} \, \overline{\sum_{m<N} \alpha_m \, e^{2i\pi m\theta}} \, d\theta \\
&= \sum_{n,m<N} \alpha_n \alpha_m \int_0^1 e^{2i\pi(n-m)\theta} \, d\theta
\end{aligned}$$

However, we have

$$\int_0^1 \mathrm{e}^{2i\pi(n-m)\theta}\ \mathrm{d}\theta = \begin{cases} 1 & \text{if } n = m, \\ 0 & \text{otherwise.} \end{cases}$$

Consequently, we get

$$\begin{aligned} \int_0^1 \left| \sum_{n<N} \alpha_n \,\mathrm{e}^{2i\pi n\theta} \right|^2 \ \mathrm{d}\theta &= \sum_{n<N} \underbrace{\alpha_n \alpha_n}_{=1} \\ &= N. \end{aligned}$$

We hence have

$$\sup_{\theta\in[0\,;1]}\left|\sum_{n<N}\alpha_n\,\mathrm{e}^{2i\pi n\theta}\right| \geq \left(\int_0^1\left|\sum_{n<N}\alpha_n\,\mathrm{e}^{2i\pi n\theta}\right|^2\,\mathrm{d}\theta\right)^{\frac{1}{2}} = \sqrt{N}$$

because, if $f \in \mathrm{L}^\infty([0\,;1])$, then we can show that $f \in \mathrm{L}^2([0\,;1])$ and $||f||_2 \leq ||f||_\infty$.

In 1950, Salem asked the following question.

The Salem problem

Is it possible to find a sequence $\boldsymbol{\alpha} \in \{-1,1\}^{\mathbb{N}}$ such that there exists a constant $c > 0$ for which

$$\sqrt{N} \leq \sup_{\theta \in [0\,;1]} \left| \sum_{n<N} \alpha_n \, \mathrm{e}^{2i\pi n\theta} \right| \leq c\sqrt{N}$$

holds for any positive integer N?

The answer is "yes" and was given by Shapiro in 1951 and then Rudin in 1959. The answer they gave involves the Rudin–Shapiro sequence.

Proposition

For any nonnegative integer N, we have

$$\sup_{\theta\in[0\,;1]}\left|\sum_{n<N}\varepsilon_n\,\mathrm{e}^{2i\pi n\theta}\right|\leq(2+\sqrt{2})\sqrt{N}$$

where ε is the Rudin–Shapiro sequence.

I will not give the proof of this result, but I would like to insist on the fact that I had to take Allouche and Shallit's proof because it is easier to understand.

A statistical property

Proposition

For any positive integers k and N, we have

$$\left|\sum_{n<N} \varepsilon_n \varepsilon_{n+k}\right| < 2k + 4k \log_2\left(\frac{2N}{k}\right).$$

Again, I will not prove it and I did not understand Mauduit's proof.

Corollary

For any nonnegative integers k and N, define

$$\gamma_N(k) = \frac{1}{N} \sum_{n<N} \varepsilon_n \varepsilon_{n+k}.$$

Then, for any nonnegative integer k, the sequence $(\gamma_N(k))_{N>0}$ converges and

$$\lim_{N\to\infty} \gamma_N(k) = 0 \quad \text{for every } k \geq 1.$$

This result shows an unexpected behaviour.

Indeed, as the Thue–Morse sequence, the Rudin–Shapiro is defined by a very simple algorithm and its behaviour should be far from the "random" sequence. The possibility of predicting next terms of the sequence from a specific term should then be high. However, the latest corollary shows the contrary: the correlations are zero.

Table of contents

Definition

The Baum–Sweet sequence $\mathbf{f} = (f_n)_{n\in\mathbb{N}}$ with values in the alphabet $\mathbb{Z}_2 = \mathbb{Z}/2\,\mathbb{Z} = \{0, 1\}$ is defined by

$$f_n = \begin{cases} 1 & \text{if the dyadic development of } n \text{ contains} \\ & \text{no block of consecutive 0's of odd length,} \\ 0 & \text{otherwise} \end{cases}$$

for all $n \in \mathbb{N}$. We have

$$\mathbf{f} = 1101100101001001\cdots.$$

We can easily show by induction that the Baum–Sweet sequence verifies the following recurrent property.

Property

We have $f_0 = 1$ and, for any nonnegative integer n, we have

$$\begin{aligned} f_{2n+1} &= f_n, \\ f_{4n} &= f_n, \\ f_{4n+2} &= 0. \end{aligned}$$

The sequence $\mathbf{f}$ can also be obtained as follows. Let σ be the morphism over $\{a, b, c, d\}$ defined by

$$\sigma(a) = ab,\ \sigma(b) = cb,\ \sigma(c) = bd,\ \sigma(d) = dd.$$

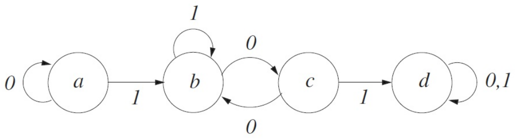
Then σ is prolongable on a and thus has a unique infinite fixed point beginning by a that we will denote by $\mathbf{v} := \sigma^\omega(a)$. If we denote by ψ the coding given by

$$\psi(a) = 1 = \psi(b),\ \psi(c) = 0 = \psi(d),$$

we can prove by induction that $\mathbf{f} = \psi(\mathbf{v})$.

Automaton

The following 2-automaton with initial state a and exit map ψ given by $\psi(a) = 1 = \psi(b)$ and $\psi(c) = 0 = \psi(d)$ recognises the Baum–Sweet sequence (in direct reading).



Indeed, we have

$$f_n = 1 \Leftrightarrow (n)_2 \text{ contains no block of 0's of odd length,}$$
$$f_n = 0 \Leftrightarrow (n)_2 \text{ contains at least one block of 0's of odd length.}$$

A solution to some problem

Remember the Lagrange's theorem stating that the continued fraction expansion of an irrational algebraic number $x \in \mathbb{R}$ is ultimately periodic if and only if x is quadratic.

However we know nothing about the expansion of nonquadratic irrational algebraic numbers.

The problem

Is there any algebraic number of degree at least 3 with bounded continued fraction expansion, i.e. such that the partial quotients in the expansion are bounded?

If we replace $\mathbb{R}$ by the field $\mathbb{Z}_2[[X^{-1}]]$ of formal power series in X^{-1} over $\mathbb{Z}_2$, the analogous problem was solved in 1976 by Baum and Sweet. Indeed, they gave an example of an algebraic element of degree 3 with a bounded continued fraction expansion, i.e. with partial quotients in $\mathbb{Z}_2[X]$ of bounded degree.

Proposition

Let $F(X) = \sum_{n\geq 0} f_n X^{-n}$ be the formal power series with coefficients given by the Baum–Sweet sequence. Then,

(1) F is an algebraic element of degree 3 over $\mathbb{Z}_2[X]$, i.e. there exists a nontrivial polynomial P of degree 3 with coefficients in $\mathbb{Z}_2[X]$ such that $P(F) = 0$.

(2) The continued fraction expansion of F is bounded and consists of elements of the set $\{1, X, X+1, X^2, X^2+1\}$.

Proof: I will show the first part, not the second one. We have

$$\begin{aligned}
F(X) &= \sum_{n\geq 0} f_n X^{-n} \\
&= \sum_{2n+1\geq 0} \underbrace{f_{2n+1}}_{=f_n} X^{-(2n+1)} + \sum_{4n\geq 0} \underbrace{f_{4n}}_{=f_n} X^{-4n} + \sum_{4n+2\geq 0} \underbrace{f_{4n+2}}_{=0} X^{-(4n+2)} \\
&= X^{-1} \sum_{n\geq 0} f_n X^{-2n} + \sum_{n\geq 0} f_n X^{-4n} \\
&= X^{-1} F(X^2) + F(X^4).
\end{aligned}$$

If $A(X) = \sum_{n\geq 0} a_n X^n$ is a formal power series with $a_n \in \mathbb{Z}_2$ for all $n \geq 0$, we have $A(X^2) = A(X)^2$. Consequently, we get

$$\begin{aligned}
F(X) = X^{-1}F(X)^2 + F(X)^4 &\Rightarrow XF(X) = F(X)^2 + XF(X)^4 \\
&\Rightarrow X = F(X) + XF(X)^3 \Rightarrow XF(X)^3 + F(X) + X = 0.
\end{aligned}$$

If we take the polynomial $P(Y) = X^3 Y + Y + X$, we get

$$P(F) = 0$$

and P is a nontrivial polynomial of degree 3 with coefficients in $\mathbb{Z}_2[X]$.

$\square$

Table of contents

Definition

The Cantor sequence $\mathbf{c} = (c_n)_{n\in\mathbb{N}}$ is defined as the (unique) fixed point beginning by a of the Cantor morphism τ defined on the alphabet $\{a, b\}$ by $\tau(a) = aba$, $\tau(b) = bbb$, i.e.

$$\mathbf{c} = ababbbababbbbbbbbbababbbaba \cdots .$$

Let's define the two following subsets of $\mathbb{N}$:

- $\mathbb{C}_a$ the set of integers n such that $c_n = a$;
- $\mathbb{C}_b$ the set of integers n such that $c_n = b$.

If we observe the first letters of the Cantor sequence, we see that

$$\mathbb{C}_a = \{0, 2, 6, 8, ...\},$$
$$\mathbb{C}_b = \{1, 3, 4, 5, 7, 9, 10, 11, 12, 13, 14, 15, ...\}$$

because

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$\mathbf{c} =$	a	b	a	b	b	b	a	b	a	b	b	b	b	b	b	b	$\cdots$

It is clear that $\mathbb{C}_a \cap \mathbb{C}_b = \emptyset$ and $\mathbb{C}_a \cup \mathbb{C}_b = \mathbb{N}$, so that $\mathbb{C}_a$ and $\mathbb{C}_b$ form a partition of $\mathbb{N}$.

Then we can show the following result by induction.

Proposition

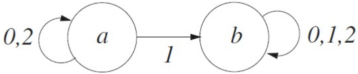
We have

$$\mathbb{C}_a = \{n \in \mathbb{N} \mid \text{if } n = \sum_{i \geq 0} n_i 3^i \text{ with } n_i \in \{0,1,2\}\ \forall i \geq 0, \\ \text{then } n_i \neq 1\ \forall i \geq 0\}$$

and

$$\mathbb{C}_b = \{n \in \mathbb{N} \mid \text{if } n = \sum_{i \geq 0} n_i 3^i \text{ with } n_i \in \{0,1,2\}\ \forall i \geq 0, \\ \text{then } \exists i \geq 0 \text{ such that } n_i = 1\}.$$

The following 3-automaton with initial state a and exit map id given by $\mathrm{id}(a) = a$ and $\mathrm{id}(b) = b$ recognises the Cantor sequence (in direct reading).



Indeed, we have

$$c_n = a \Leftrightarrow (n)_3 \text{ does not contain any 1's,}$$
$$c_n = b \Leftrightarrow (n)_3 \text{ contains at least one 1.}$$

Table of contents

A solution to some problem

After these four examples of sequences obtained as the image, under a coding, of a fixed point of a constant length morphism and their arithmetic description, the following question arises.

Problem

Is it possible to find a sequence that is a fixed point of a non-constant length morphism and can also be defined by some simple arithmetic property of the integers?

The answer is "yes" and is given by the Fibonacci sequence as we will see.

Definition

The Fibonacci sequence $\mathbf{f} = (f_n)_{n\in\mathbb{N}}$ is defined as the unique non-empty fixed point of the Fibonacci morphism φ defined on the alphabet $\{a, b\}$ by $\varphi(a) = ab$, $\varphi(b) = a$, i.e.

$$\mathbf{f} = ababbbababbbbbbbbababbbaba \cdots .$$

The Fibonacci numeration system

Let $(F_n)_{n\in\mathbb{N}}$ be the sequence of integers defined by the relations $F_0 = 1$, $F_1 = 2$ and, for any integer $n \geq 1$, $F_{n+1} = F_n + F_{n-1}$. This sequence of integers is called *the Fibonacci sequence of numbers*. We can show the following proposition.

Proposition

Every positive integer n can be written in a unique way as

$$n = \sum_{i=0}^{k} n_i F_i$$

with $n_k = 1$, $n_i \in \{0, 1\}$ and $n_i n_{i+1} = 0$ for any $i \in \{0, ..., k-1\}$. This numeration system is called the *Zeckendorff numeration system*.

We can thus define the Fibonacci expansion of an integer.

Definition

Let n be a positive integer. If

$$n = \sum_{i=0}^{k} n_i F_i$$

with $n_k = 1$, $n_i \in \{0,1\}$ and $n_i n_{i+1} = 0$ for any $i \in \{0, ..., k-1\}$, then we say that

$$\mathrm{Fib}(n) = n_k n_{k-1} ... n_0 \in \{0,1\}^{k+1}$$

is the *Fibonacci expansion of* n. If $n = 0$, we set $\mathrm{Fib}(0) = 0$.

Let's define the two following subsets of $\mathbb{N}$:

- $\mathbb{F}_a$ the set of integers n such that $f_n = a$;
- $\mathbb{F}_b$ the set of integers n such that $f_n = b$.

If we observe the first letters of the Fibonacci sequence, we see that

$$\mathbb{F}_a = \{0, 2, 3, 5, 7, 8, 10, 11, 13, 15, ...\},$$
$$\mathbb{F}_b = \{1, 4, 6, 9, 12, 14, ...\}$$

because

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$\mathbf{f} =$	a	b	a	a	b	a	b	a	a	b	a	a	b	a	b	a	$\cdots$

It is clear that $\mathbb{F}_a \cap \mathbb{F}_b = \emptyset$ and $\mathbb{F}_a \cup \mathbb{F}_b = \mathbb{N}$, so that $\mathbb{F}_a$ and $\mathbb{F}_b$ form a partition of $\mathbb{N}$.

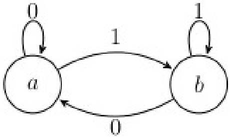
A solution to the problem

Proposition

We have

$$\mathbb{F}_a = \{n \in \mathbb{N} \mid \mathsf{Fib}(n) \in \{0,1\}^*0\},$$
$$\mathbb{F}_b = \{n \in \mathbb{N} \mid \mathsf{Fib}(n) \in \{0,1\}^*1\}.$$

Just as Fibonacci representation is the analogue of base-k representation, we can define the notion of *Fibonacci-automata* as the analogue of the more familiar notion of k-automata. The following Fibonacci-automaton with initial state a and exit map id given by $\mathrm{id}(a) = a$ and $\mathrm{id}(b) = b$ recognises the Fibonacci sequence (in direct reading).



Indeed, we have

$$f_n = a \Leftrightarrow \mathrm{Fib}(n) \text{ ends with a } 0,$$
$$f_n = b \Leftrightarrow \mathrm{Fib}(n) \text{ ends with a } 1.$$

Table of contents

The problem

Problem

Let d and q be two integers greater or equal to 2. Is it possible to decide only from its q-adic development whether a given positive integer is divisible by d or not?

Let $\mathbf{v}$ be the periodic sequence over the alphabet $\{0, 1, ..., d-1\}$ given by

$$\mathbf{v} = 01\cdots(d-1)01\cdots(d-1)01\cdots(d-1)\cdots = (01\cdots(d-1))^{\omega}.$$

Actually, the sequence $\mathbf{v}$ codes the rests modulo d of all nonnegative integers. In order to solve this problem, a solution may be to find a q-automaton

(1) with d states: 0, 1, ..., $d-1$;

(2) with initial state 0;

(3) with exit map id where

$$\mathrm{id}(0) = 0,\ \mathrm{id}(1) = 1,\ ...,\ \mathrm{id}(d-1) = d-1;$$

(4) which recognises the sequence $\mathbf{v}$.

Indeed, thanks to that automaton, we will be able to decide whether a nonnegative integer is a multiple of d uniquely from its q-adic development:

- if the final state reached after reading the q-adic development of the integer (in direct reading) is 0, then the integer is divisible by d;
- if the final state reached after reading the q-adic development of the integer (in direct reading) is different from 0, then the integer is not divisible by d.

By the result 1.3.1. I recalled at the beginning of this talk, it is enough to find a morphism ρ of constant length q such that $\mathbf{v}$ is a fixed point of ρ. This can be done by cutting $\mathbf{v}$ into blocks of length q and rewriting $\mathbf{v}$ as

$$\mathbf{v} = \rho(0)\rho(1)\cdots\rho(d-1)\rho(0)\rho(1)\cdots\rho(d-1)\cdots .$$

Example

Consider the case where $q = 2$ and $d = 5$. We thus have

$$\mathbf{v} = 01234012340123401234012340123401234012340\cdots.$$

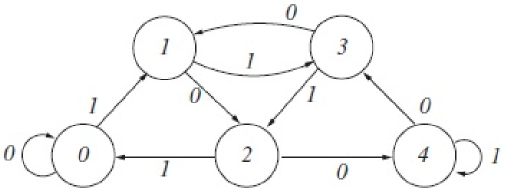
We take ρ as follows

$$\begin{aligned} &\rho(0) = 01, \quad \rho(1) = 23, \quad \rho(2) = 40, \\ &\rho(3) = 12, \quad \rho(4) = 34. \end{aligned}$$

We see that $\mathbf{v}$ is a fixed point of ρ. We are now able to show the final 2-automaton which was described above. Indeed, the proof of the result 1.3.1. indicates how to build the wanted 2-automaton: there exists an edge from a to b if b occurs in $\rho(a)$. If such an edge exists, it is labelled by i if b is the $(i+1)$-th letter of $\rho(a)$.

Example (continued)

We can thus build the following 2-automaton. With initial state 0 and exit map id, it recognises **v** (in direct reading).



We see that this 2-automaton answers the problem: if $n = \sum_{i=0}^{k} n_i 2^i$ with $n_i \in \{0, 1\}$ and $n_k = 1$, then we can feed the automaton with the word $n_k n_{k-1} \cdots n_1 n_0$. We reach a final state labelled by $r \in \{0, 1, 2, 3, 4\}$. If $r = 0$ (resp. $r \neq 0$), then n is divisible (resp. is not divisible) by 5 and we only decide this by using the dyadic development of n.

Thank you for listening!

Do you have any questions?

Table of contents

Bibliography

- Allouche J.-P., Shallit J., Automatic sequences : theory, applications, generalizations. Cambridge : Cambridge University Press, 2003.
- Baum L. E., Sweet M. M.. Continued fractions of algebraic power series in characteristic 2. *Annals of Mathematics*. 1976, 103, 593-610.
- Fei Du C., Mousavi H., Schaeffer L., Shallit J.. Decision algorithms for Fibonacci-automatic words, with applications to pattern avoidance. Available at <http://arxiv.org/abs/1406.0670> (consulted on 11th February 2015).
- Gallouët T., Herbin R., Mesure, intégration, probabilités. 2014. Available at <http://www.cmi.univ-mrs.fr/~gallouet/licence.d/mip-acor.pdf> (consulted on 18th February 2015).
- Pytheas Fogg N., Substitutions in dynamics, arithmetics, and combinatorics. Berlin : Springer, 2002.