Automatic sequences in rational base numeration systems (and even more)

Joint work Michel Rigo (ULiège)

Manon Stipulanti (ULiège)

FNRS Postdoc Researcher

ULiège Discrete Math. Seminar
Liège (Belgium)
March 31, 2021

How it started

Eric Rowland presented our paper

E. Rowland and M. Stipulanti, Avoiding 5/4-powers on the alphabet of nonnegative integers, *Electron. J. Combin.* **27** (2020), Paper 3.42, 39 pp.

at the *One World Combinatorics on Words Seminar.*
He (notably) talked about the word

$$w_{3/2} = 001102100112001103100113001102100114001103 1\cdots ,$$

which is the lexicographically least word avoiding $\frac{3}{2}$-powers on the alphabet of non-negative integers. Then Michel asked:

Is this word $\frac{3}{2}$-regular?

So we started to look at automaticity and regularity in rational bases.

The plan

Define:

- Abstract numeration systems
- Labeled trees and periodic labeled signatures
- Rational base numeration systems
- Automatic sequences

Two main results:

- A version of Cobham's theorem
- "Factor complexity" in trees

Abstract numeration systems

Introduced by Pierre Lecomte and Michel Rigo in 2001, an abstract numeration system (ANS for short) is a triple $S = (L, A, <)$ where

- $(A, <)$ is a totally ordered (finite) alphabet
- L is an infinite language over A.

We say that L is the numeration language.

The map $\mathrm{rep}_S : \mathbb{N} \to L$ is the one-to-one correspondence mapping $n \in \mathbb{N}$ onto the $(n+1)$st word in the radix ordered language L.
This word is called the S-representation of n.
(The S-representation of 0 is the first word in L.)

The inverse map is denoted by $\mathrm{val}_S : L \to \mathbb{N}$. For any word w in L, $\mathrm{val}_S(w)$ is its S-(numerical) value.

Remark: Nothing is assumed on the language L, i.e., it might well be neither regular nor prefix-closed.

Examples of ANS

- Integer base numeration systems:
  Let $k \geq 2$ be an integer. We let $A_k = \{0, 1, \ldots, k-1\}$ and
  $$L_k = \{\varepsilon\} \cup \{1, \ldots, k-1\}\{0, \ldots, k-1\}^*.$$
  Then $S = (L_k, A_k, <)$ with $0 < 1 < \cdots < k-1$ is an ANS.
- Fibonacci (or Zeckendorff) numeration system:

Let $S = (L, \{0, 1\}, <)$ with $0 < 1$ and L containing words avoiding 11.
In radix order: $L = \{\varepsilon, 1, 10, 100, 101, 1000, 1001, 1010, \ldots\}$

S-representation	S-value
ε	0
1	1
10	2
100	3
101	4
1000	5

A "non-standard" ANS

$S = (a^*b^*, \{a, b\}, <)$ with $a < b$
$a^*b^* = \{\varepsilon, a, b, aa, ab, bb, aaa, aab, abb, \ldots\}$ in radix order

S-representation	S-value
ε	0
a	1
b	2
aa	3
ab	4
bb	5
aaa	6
aab	7

Position of a word in a^*b^*: $\mathrm{val}_S(a^p b^q) = \frac{(p+q)(p+q+1)}{2} + q$.
For instance, $\mathrm{val}_S(ab) = \frac{2\cdot 3}{2} + 1 = 4$.

From ANS to trees

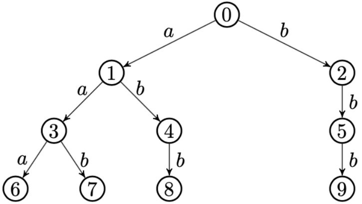
Prefix-closed languages define labeled trees.

Let $S = (L, A, <)$ be an ANS where L is prefix-closed.
We define the tree $T(L)$ as follows.

- The set of nodes of $T(L)$ is L.
- If w and wd are words in L with $d \in A$, then there is an edge from w to wd with label d in $T(L)$.

The children of a node are ordered by the labels of the letters in the ordered alphabet A.
Nodes are enumerated by breadth-first traversal.

Example: $S = (a^*b^*, \{a, b\}, <)$ with $a < b$
$a^*b^* = \{\varepsilon, a, b, aa, ab, bb, aaa, \ldots\}$

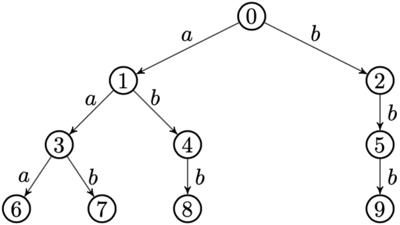


Let T be a labeled tree.

The signature of T is the sequence of the degrees of the nodes visited by the breadth-first traversal of T.

The labeling of T is the sequence of the labels of the edges visited by the breadth-first traversal of T.

Example:



Signature:
$2, 2, 1, 2, 1, 1, 2, 1, 1, 1, 2, \ldots$

Labeling:
$a, b, a, b, b, a, b, b, b, a, b, \ldots$

Remark: Sometimes it is convenient to consider i-trees: the root is assumed to be a child of itself.

Particular ANS: rational bases

Let p and q be two relatively prime integers with $p > q > 1$.

For a word $w = w_\ell w_{\ell-1} \cdots w_0 \in A_p^*$, the value of w in base $\frac{p}{q}$ is the rational number

$$\mathrm{val}_{\frac{p}{q}}(w) = \sum_{i=0}^{\ell} \frac{w_i}{q} \left(\frac{p}{q}\right)^i.$$

Example: $p = 3$, $q = 2$ $\qquad A_3 = \{0, 1, 2\}$

$w \in \{0,1,2\}^*$	$\mathrm{val}_{\frac{3}{2}}(w)$
$\varepsilon, 0$	0
1	$\frac{1}{2} \cdot \left(\frac{3}{2}\right)^0 = \frac{1}{2}$
2	$\frac{2}{2} \cdot \left(\frac{3}{2}\right)^0 = 1$
10	$\frac{1}{2} \cdot \left(\frac{3}{2}\right)^1 + \frac{0}{2} \cdot \left(\frac{3}{2}\right)^0 = \frac{3}{4}$
21	$\frac{2}{2} \cdot \left(\frac{3}{2}\right)^1 + \frac{1}{2} \cdot \left(\frac{3}{2}\right)^0 = 2$

Remark: $\mathrm{val}_{\frac{p}{q}}(w)$ is a not always an integer.

Definition: A word $w \in A_p^*$ is a representation of an integer $n \geq 0$ in base $\frac{p}{q}$ if $\mathrm{val}_{\frac{p}{q}}(w) = n$.

Theorem (Akiyama, Frougny, Sakarovitch, 2008)

Representations in rational bases are unique up to leading zeroes.

In base $\frac{p}{q}$:
$\mathrm{rep}_{\frac{p}{q}}(n)$ denotes the representation of n that does not start with 0.
By convention, the representation of 0 is the empty word ε.
The numeration language is the set

$$L_{\frac{p}{q}} = \left\{ \mathrm{rep}_{\frac{p}{q}}(n) \mid n \geq 0 \right\}.$$

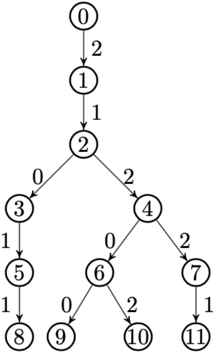
Example: In base $\frac{3}{2}$:

$$L_{\frac{3}{2}} = \{\varepsilon, 2, 21, 210, 212, 2101, 2120, 2122, \ldots\}$$

Properties:

- For all $u, v \in A_p^*$, $\mathrm{val}_{\frac{p}{q}}(uv) = \mathrm{val}_{\frac{p}{q}}(u) \left(\frac{p}{q}\right)^{|v|} + \mathrm{val}_{\frac{p}{q}}(v)$.
- $m < n$ if and only if $\mathrm{rep}_{\frac{p}{q}}(m) < \mathrm{rep}_{\frac{p}{q}}(n)$ for the radix order.
- $L_{\frac{p}{q}} \subseteq A_p^*$ is not regular.
- $L_{\frac{p}{q}} \subseteq A_p^*$ is prefix-closed.

Example: base $\frac{3}{2}$



Signature: $2, 1, 2, 1, 2, 1, \ldots$

(i-tree: if we add an edge of label 0 onto the root)

Labeling: $0, 2, 1, 0, 2, 1, 0, 2, 1, \ldots$

nth node: $\mathrm{rep}_{\frac{3}{2}}(n)$

(breadth-first traversal)

Edges: $n \xrightarrow{a \in A_3} m \iff m = \frac{3}{2} \cdot n + \frac{a}{2}$.

ANS given by periodic labeled signatures

A labeled signature is an infinite sequence $(w_n)_{n\geq 0}$ of finite words providing

- a signature $(|w_n|)_{n\geq 0}$ and
- a consistent labeling (made of the sequence of letters of $(w_n)_{n\geq 0}$)

of a tree.

The canonical breadth-first traversal of this tree produces an ANS.

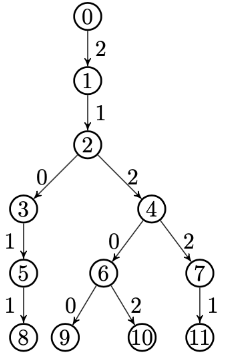
Example:

Labeled signature: $(02, 1)^\omega$

Signature: $\underbrace{|02|}_{=2}, \underbrace{|1|}_{=1}, \underbrace{|02|}_{=2}, \underbrace{|1|}_{=1}, \ldots$

Labeling: $0, 2, 1, 0, 2, 1, 0, 2, 1, \ldots$

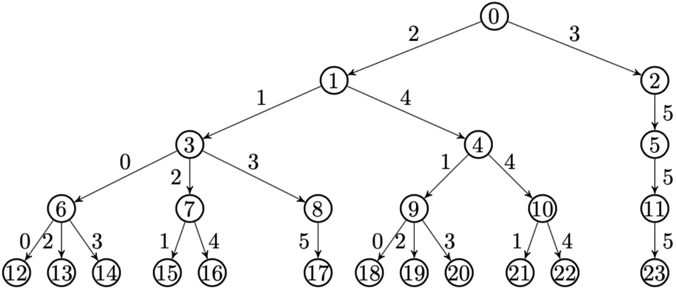
Base $\frac{3}{2}$



Example: Rational bases

p/q	corresp. labeled sign.
$3/2$	$(02, 1)^\omega$
$5/2$	$(024, 13)^\omega$
$7/3$	$(036, 25, 14)^\omega$
$11/4$	$(048, 159, 26(10), 37)^\omega$

Example: i-tree associated with the labeled signature $(023, 14, 5)^\omega$



Automatic sequences

Definition: Let $S = (L, A, <)$ be an ANS and let B be a finite alphabet. An infinite word $\mathbf{x} = x_0 x_1 x_2 \cdots \in B^{\mathbb{N}}$ is S-automatic if there exists a deterministic finite automaton with output (DFAO for short) $\mathcal{A} = (Q, q_0, A, \delta, \mu : Q \to B)$ such that

$$x_n = \mu(\delta(q_0, \mathrm{rep}_S(n))) \quad \forall\, n \geq 0.$$

We read most significant digits first (not a restriction).

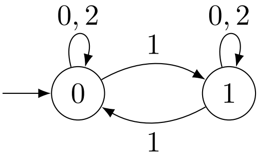
Remark: We talk about...

- k-automatic seq. in the base-k numeration system $(L_k, A_k, <)$
- $\frac{p}{q}$-automatic seq. in the base-$\frac{p}{q}$ numeration system $(L_{\frac{p}{q}}, A_p, <)$.

Toy example: the sum-of-digits in base $\frac{3}{2}$

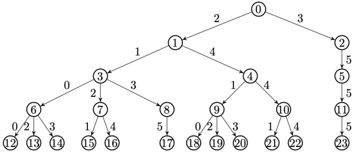
n	0	1	2	3	4	5	6	7
$\mathrm{rep}_{\frac{3}{2}}(n)$	ε	2	21	210	212	2101	2120	2122
$s(n)$	0	2	3	3	5	4	5	7
$t(n) = s(n) \bmod 2$	0	0	1	1	1	0	1	1

The sequence $\mathbf{t}$ is $\frac{3}{2}$-automatic as it is generated by the following DFAO when reading base-$\frac{3}{2}$ representations:



Another example

Periodic labeled signature $(023, 14, 5)^\omega$ producing the i-tree



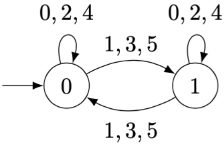
The first few words in the corresponding ANS S:

$\varepsilon, 2, 3, 21, 24, 35, 210, 212,$
$213, 241, 244, 355, \ldots$

The sum-of-digits in S modulo 2

n	0	1	2	3	4	5	6	7
$\text{rep}_S(n)$	ε	2	3	21	24	35	210	212
s.o.d	0	2	3	3	6	8	3	5
s.o.d mod2	0	0	1	1	0	0	1	1

is S-automatic since it is generated by the DFAO



Cobham's theorem

Theorem (Cobham, 1972)

A sequence is k-automatic if and only if it is the image under a coding of a fixed point of a k-uniform morphism.

Many generalizations exist.

Goal: Generalization to our context of S-automatic sequences for ANS built on tree languages with a purely periodic labeled signatures.

Needed: alternate fixed points and block substitutions.

Alternating morphisms

The Kolakoski–Oldenburger word is the unique word $\mathbf{k}$ over $\{1,2\}$ starting with 2 and satisfying $\Delta(\mathbf{k}) = \mathbf{k}$ where Δ is the run-length encoding map

$$\mathbf{k} = 2211212212211\cdots . \qquad \text{[OEIS, A000002]}$$

How to build it?
"Write what you read": each term of $\mathbf{k}$ generates a run of one or two future terms.

2	starting point
$\underline{2}2$	the first letter 2 generates a run of "22"
$22\underline{11}$	the second letter 2 generates a run of "11"
$2211\underline{2}$	the first letter 1 generates a run of "2"
$22112\underline{1}$	the second letter 1 generates a run of "1"
$221121\underline{22}$	and so on and so forth...

It is a well-known and challenging object of study in CoW.
<u>Conjecture</u>: The density of 1 in $\mathbf{k}$ is $\frac{1}{2}$. (Still open.)

Alternate fixed point

Alternative definition (Culik, Karhumäki, Lepistö, 1992): $\mathbf{k} = k_0k_1k_2\cdots$ can be obtained by periodically iterating two morphisms

$$\mathbf{k} = h_0(k_0)h_1(k_1)\cdots h_0(k_{2n})h_1(k_{2n+1})\cdots$$

where

$$h_0 : \begin{cases} 1 \mapsto 2 \\ 2 \mapsto 22 \end{cases} \quad \text{and} \quad h_1 : \begin{cases} 1 \mapsto 1 \\ 2 \mapsto 11. \end{cases}$$

$\mathbf{k}$	2	2	1	1	2	1	2	2	1
(h_0, h_1)	h_0	h_1	h_0	h_1	h_0	h_1	h_0	h_1	h_0
$(h_0, h_1)(\mathbf{k}) = \mathbf{k}$	22	11	2	1	22	1	22	11	2

We say that $\mathbf{k}$ is an alternate fixed point of (h_0, h_1).

Definition

Let $r \geq 1$ be an integer and let $f_0, \ldots, f_{r-1}$ be morphisms over an alphabet A.
An word $\mathbf{w} = w_0w_1\cdots \in A^\omega$ is an alternate fixed point of $(f_0, \ldots, f_{r-1})$ if $\mathbf{w} = f_0(w_0)f_1(w_1)\cdots f_{r-1}(w_{r-1})f_0(w_r)\cdots f_{i \bmod r}(w_i)\cdots$.

Block substitution

Alternative definition (Dekking, 1980): Since

$$\mathbf{k} = \underbrace{h_0(2)h_1(2)}_{=g(22)}\ \underbrace{h_0(1)h_1(1)}_{=g(11)}\ \underbrace{h_0(2)h_1(1)}_{=g(21)}\ \underbrace{h_0(2)h_1(2)}_{=g(22)}\ \underbrace{h_0(1)h_1(2)}_{=g(12)}\cdots$$
$$= 2211\ \ 21\ \ 221\ \ 2211\ \ 211\cdots$$

$\mathbf{k}$ is also a fixed point of the 2-block substitution

$$g : \begin{cases} 11 \mapsto h_0(1)h_1(1) = 21 \\ 12 \mapsto h_0(1)h_1(2) = 211 \\ 21 \mapsto h_0(2)h_1(1) = 221 \\ 22 \mapsto h_0(2)h_1(2) = 2211. \end{cases}$$

Remark: Lengths of images under g are not all equal.

Definition

An r-block substitution $g : A^r \to A^*$ maps a word $w_0 \cdots w_{rn-1} \in A^*$ to

$$g(w_0 \cdots w_{r-1})g(w_r \cdots w_{2r-1}) \cdots g(w_{r(n-1)} \cdots w_{rn-1}).$$

If the length of the word is not a multiple of r, then the suffix of the word is ignored under the action of g.
An infinite word $\mathbf{w} = w_0 w_1 \cdots \in A^\omega$ is a fixed point of the r-block substitution $g : A^r \to A^*$ if

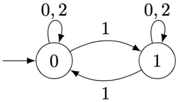
$$\mathbf{w} = g(w_0 \cdots w_{r-1})g(w_r \cdots w_{2r-1}) \cdots .$$

Proposition

If an infinite word over A is an alternate fixed point of $(f_0, \ldots, f_{r-1})$, then it is a fixed point of an r-block substitution.

Proof: For every of length-r word $a_0 \cdots a_{r-1}$, define the r-block substitution $g : A^r \to A^*$ by $g(a_0 \cdots a_{r-1}) = f_0(a_0) \cdots f_{r-1}(a_{r-1})$.

Goal: From a DFAO, build alternating morphisms.



ANS built on the purely periodic labeled signature $(w_0, w_1, \ldots, w_{r-1})^\omega$

Base $\frac{3}{2}$ $\quad (02, 1)^\omega$

Sum-of-digits mod 2

$\mathbf{t} = 0011101111101 \cdots$ $\frac{3}{2}$-automatic

$\mathbf{x}$ automatic produced by a DFA $\mathcal{A} = (Q, q_0, A, \delta)$

$$f_0 : \begin{cases} 0 \mapsto \delta(0,0)\delta(0,2) = 00 \\ 1 \mapsto \delta(1,0)\delta(1,2) = 11 \end{cases}$$

$$f_1 : \begin{cases} 0 \mapsto \delta(0,1) = 1 \\ 1 \mapsto \delta(1,1) = 0 \end{cases}$$

$f_i : Q \to Q^{|w_i|}$:
$q \mapsto \delta(q, w_{i,0}) \cdots \delta(q, w_{i,|w_i|-1})$
$\forall i \in \{0, \ldots, r-1\}$

$f_0(0) = 00$

$f_0(0) f_1(0) = 00\ 1$

$f_0(0) f_1(0) f_0(1) = 00\ 1\ 11$

$f_0(0) f_1(0) f_0(1) f_1(1) f_0(1) = 00\ 1\ 11\ 0\ 11\ 1$

$\mathbf{x}$ is the alternate fixed point of $(f_0, \ldots, f_{r-1})$ starting with q_0

$\mathbf{t}$ alternate fixed point of (f_0, f_1)

Proposition (Rigo, S., 2021)

Let $r \geq 1$ be an integer and let A be a finite alphabet of digits.
Let $w_0, \ldots, w_{r-1}$ be r non-empty words in $\mathsf{inc}(A^*)$.
Consider the language $L(\mathsf{s})$ of the i-tree generated by the purely periodic signature $\mathsf{s} = (w_0, w_1, \ldots, w_{r-1})^\omega$ and the corresponding ANS $S = (L(\mathsf{s}), A, <)$.
Let $\mathcal{A} = (Q, q_0, A, \delta)$ be a DFA.
For $i \in \{0, \ldots, r-1\}$, we define the r morphisms from Q^* to itself by

$$f_i : Q \to Q^{|w_i|}, q \mapsto \delta(q, w_{i,0}) \cdots \delta(q, w_{i,|w_i|-1}),$$

where $w_{i,j}$ denotes the jth letter of w_i.
The alternate fixed point $\mathbf{x} = x_0 x_1 \cdots$ of $(f_0, \ldots, f_{r-1})$ starting with q_0 is the sequence of states reached in $\mathcal{A}$ when reading the words of $L(\mathsf{s})$ in increasing radix order, i.e., for all $n \geq 0$, $x_n = \delta(q_0, \mathrm{rep}_S(n))$.

Alternate fixed points to automatic sequences

Goal: From alternating morphisms, build a DFAO.

$(f_0, \dots, f_{r-1})$ $f_i : A^* \to A^*$ ℓ_i-unif.
f_0 is prolongable on a

ANS S built on the purely periodic labeled signature $(w_0, \dots, w_{r-1})^\omega$

$w_0 = 0 \cdots (\ell_0 - 1)$
$w_1 = \ell_0(\ell_0 + 1) \cdots (\ell_0 + \ell_1 - 1)$
$\vdots$
$w_{r-1} = \left(\sum_{j<r-1} \ell_j\right) \cdots \left(\sum_{j<r} \ell_j - 1\right)$

$\mathcal{A} = (A, a, B, \delta)$ with

$B = \{0, \dots, \sum_{j<r} \ell_j - 1\}$
$\delta(b, i) = [f_{j_i}(b)]_{t_i}$ for $b \in A, i \in B$ such that
$i = \sum_{k \le j-1} \ell_k + t$ with $j \ge 0, 0 \le t < \ell_j$

The alternate fixed point of $(f_0, \dots, f_{r-1})$ starting with a is S-automatic (produced by $\mathcal{A}$)

$$f_0 : \begin{cases} 0 \mapsto 00 \\ 1 \mapsto 11 \end{cases} \quad f_1 : \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 0 \end{cases}$$

ANS built on $(\underbrace{01}_{\ell_0=2}, \underbrace{2}_{\ell_1=1})^\omega$

$\mathcal{A} = (\{0,1\}, 0, \{0,1,2\}, \delta)$

Write i in $[0,2)$, $[2,3)$

$0 = 0 + 0 \quad j = 0$

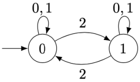
$\delta(0, 0) = [f_0(0)]_0 = 0 \quad \delta(1, 0) = [f_0(1)]_0 = 1$

$1 = 0 + 1 \quad j = 0$

$\delta(0, 1) = [f_0(0)]_1 = 0 \quad \delta(1, 1) = [f_0(1)]_1 = 1$

$2 = 2 + 0 \quad j = 1$

$\delta(0, 2) = [f_1(0)]_0 = 1 \quad \delta(1, 2) = [f_1(1)]_0 = 0$



The alternate fixed point of (f_0, f_1) is automatic. Up to a coding of the ANS, it is equal to $\mathbf{t}$ ($\frac{3}{2}$-automatic).

Proposition (Rigo, S., 2021)

Let $r \geq 1$ be an integer and let A be a finite alphabet.
For all $i \in \{0, \ldots, r-1\}$, let $f_i : A^* \to A^*$ be a ℓ_i-uniform morphism such that f_0 is prolongable on some $a \in A$.
Consider the language $L(\mathsf{s})$ of the i-tree generated by the purely periodic labeled signature

$$\mathsf{s} = \left(0 \cdots (\ell_0 - 1), \ell_0(\ell_0+1) \cdots (\ell_0 + \ell_1 - 1), \ldots, \left(\sum_{j<r-1} \ell_j\right) \cdots \left(\sum_{j<r} \ell_j - 1\right)\right)^\omega$$

and the corresponding ANS $S = (L(\mathsf{s}), B, <)$.
Let $\mathcal{A} = (A, a, B, \delta)$ be the DFA where $B = \{0, \ldots, \sum_{j<r} \ell_j - 1\}$ and its transition function $\delta : A \times B \to A$ is defined as follows: for all $i \in B$, $\exists! j_i \geq 0$ and $\exists! t_i \geq 0$ such that $i = \sum_{k \leq j_i - 1} \ell_k + t_i$ with $t_i < \ell_{j_i}$, so we set $\delta(b, i) = [f_{j_i}(b)]_{t_i}$ $\forall b \in A$.
Then the alternate fixed point $\mathbf{x} = x_0 x_1 \cdots$ of $(f_0, \ldots, f_{r-1})$ starting with a is the sequence of the states reached in $\mathcal{A}$ when reading the words of $L(\mathsf{s})$ by increasing radix order, i.e., for all $n \geq 0$, $x_n = \delta(a, \mathrm{rep}_S(n))$.

Theorem (Rigo, S., 2021)

Let A, B be two finite alphabets. An infinite word over B is the image under a coding $g : A \to B$ of an alternate fixed point of uniform morphisms (not necessarily of the same length) over A if and only if it is S-automatic for an ANS built on a tree language with a purely periodic labeled signature.

Proof: It follows from the previous two propositions.

Particular case in rational bases

Corollary (Rigo, S., 2021)

If a sequence is $\frac{p}{q}$-automatic, then it is the image under a coding of a fixed point of a q-block substitution whose images all have length p.

Example: The sum-of-digits mod 2 $\mathbf{t} = 0011101111101\cdots$ is $\frac{3}{2}$-automatic.

$$f_0 : \begin{cases} 0 \mapsto 00 \\ 1 \mapsto 11 \end{cases} \qquad f_1 : \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 0 \end{cases}$$

Then $\mathbf{t}$ is also a fixed point of the 2-block substitution

$$g : \begin{cases} 00 \mapsto f_0(0)f_1(0) = 001 \\ 01 \mapsto f_0(0)f_1(1) = 000 \\ 10 \mapsto f_0(1)f_1(0) = 111 \\ 11 \mapsto f_0(1)f_1(1) = 110. \end{cases}$$

Observe that images under g all have length 3.

Non-morphic sequences

Theorem (Cobham, 1972) rewritten

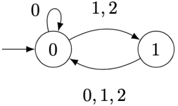
An automatic sequence in an integer base is a morphic word, i.e., the image, under a coding, of a fixed point of a prolongable morphism.

Consider the morphisms $g_0 : 0 \mapsto 01, 1 \mapsto 00$ and $g_1 : 0 \mapsto 1, 1 \mapsto 0$ yielding the 2-block substitution

$$h_2 : 00 \mapsto 011,\ 01 \mapsto 010,\ 10 \mapsto 001,\ 11 \mapsto 000$$

producing the word $\mathbf{F}_2 = 010\,011\,000\,011 \cdots$. It is $\frac{3}{2}$-automatic as it is generated by the following DFAO (built thanks to our proposition):

$$\begin{aligned}
\delta(0,0) &= [g_0(0)]_0 = 0 & \delta(1,0) &= [g_0(1)]_0 = 0\\
\delta(0,1) &= [g_0(0)]_1 = 1 & \delta(1,1) &= [g_0(1)]_1 = 0\\
\delta(0,2) &= [g_1(0)]_0 = 1 & \delta(1,2) &= [g_1(1)]_0 = 0
\end{aligned}$$



Property (Lepistö, 1993)

The word $\mathbf{F}_2$ is neither purely morphic nor morphic.

Decorated trees (not only for Christmas)

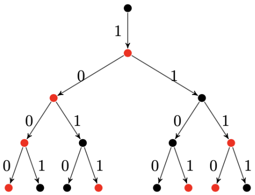
ANS $S = (L, A, <)$ where $L = \{w_0 < w_1 < \cdots\}$ is prefix-closed producing a tree $T(L)$.
We now add an extra information on every node.

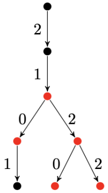
Definition: Let $\mathbf{x} = x_0 x_1 \cdots \in B^{\mathbb{N}}$ where B is a finite alphabet.
A decoration of $T(L)$ by $\mathbf{x}$ is the map $L \to B : w_n \mapsto x_n$.

Example: (0 means black, 1 means red)

Base 2
Decoration:
Thue-Morse
$0110100110010110\cdots$

Base $\frac{3}{2}$
Decoration:
Sum-of-digits modulo 2
$0011101111011011\cdots$



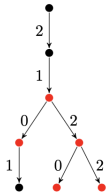


Factors of trees (subtrees)

Definition: Let T be a labeled tree.
The domain $\mathrm{dom}(T)$ of T is the set of labels of paths from the root to its nodes.

Example:

The domain of the tree on the right is $\{\varepsilon, 21, 210, 212, 2101, 2120, 2122\}$.



Definition: Let $w \in L$ and let $h \geq 0$.
We let $T[w, h]$ denote the factor of height h rooted at w of $T(L)$.
The prefix of height h of $T(L)$ is the factor $T[\varepsilon, h]$.

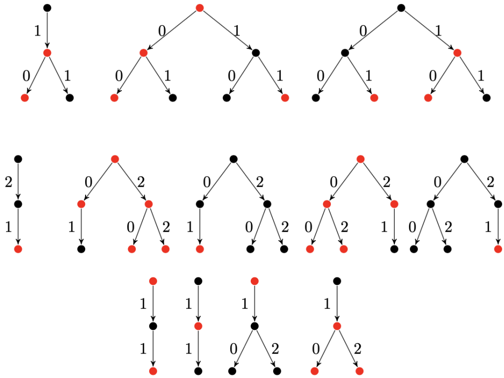
Example: The previous tree is a factor of height 4 of $T(L_{\frac{3}{2}})$. It is its prefix of height 4.

Definition: Two factors $T[w,h]$ and $T[w',h]$ of the same height are equal if they have the same domain and the same decorations.
We let $F_h = \{T[w,h] \mid w \in L\}$ denote the set of factors of height h occurring in $T(L)$.

F_2:

Base 2
Decoration:
Thue-Morse
$0110100110010110\cdots$

Base $\frac{3}{2}$
Decoration:
Sum-of-digits modulo 2
$0011101111011011\cdots$

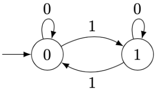


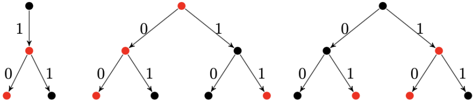
Automatic decorations of trees

Theorem (Rigo, S., 2021)

A sequence $\mathbf{x}$ is k-automatic if and only if, in the labeled tree $T(L_k)$ decorated by $\mathbf{x}$, there exists a height $h \geq 0$ such that $\#F_h = \#F_{h+1}$.

Example: Base 2, decoration being Thue-Morse $0110100110010110\cdots$
Each factor $T[w,h]$ (not the prefix) is determined by $\delta(q_0,w)$: it is a full binary tree, and the decorations are determined by $\tau(\delta(q_0,wu))$ with $u \in A_2^{\leq h}$. Therefore $\#F_h \leq \#Q + 1$.





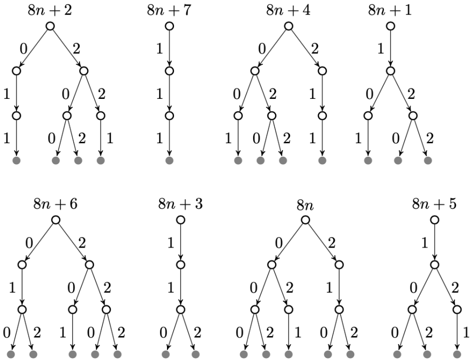
Theorem (Rigo, S., 2021)

Let $S = (L, A, <)$ be an ANS built on a prefix-closed regular language L. A sequence $\mathbf{x}$ is S-automatic if and only if, in the labeled tree $T(L)$ decorated by $\mathbf{x}$, there exists a height $h \geq 0$ such that $\#F_h = \#F_{h+1}$.

The rational bases case

Several extensions: in base $\frac{p}{q}$, except for the height-h prefix, each factor of height h is extended in exactly q ways to a factor of height $h+1$. To the first (leftmost) leaf of a factor of height h are attached children corresponding to one of the q words of the periodic signature.

Example: Base $\frac{3}{2}$ periodic labeled signature $(02,1)^\omega$



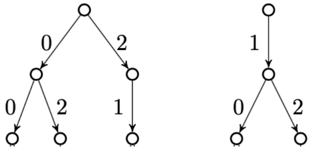
Factors appearing infinitely often

Definition: Let T be a labeled decorated tree and let $h \geq 0$.
We let $F_h^\infty \subseteq F_h$ denote the set of factors of height h occurring infinitely often in T.

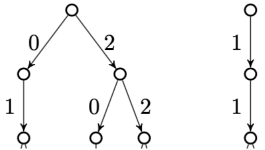
For any suitable letter a in the signature of T, we let $F_{h,a}^\infty \subseteq F_h^\infty$ denote the set of factors of height h occurring infinitely often in T such that the label of the edge between the first node on level $h-1$ and its first child is a. (Otherwise stated, the first word of length h in the domain of the factor ends with a.)

Example: Base $\frac{3}{2}$

$F_{2,0}^\infty$:

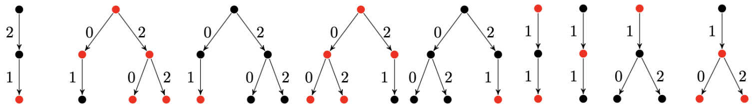


$F_{2,1}^\infty$:



Example: Base $\frac{3}{2}$

Assumptions: Factors of length 1 in $T(L_{\frac{3}{2}})$ can be extended as follows:

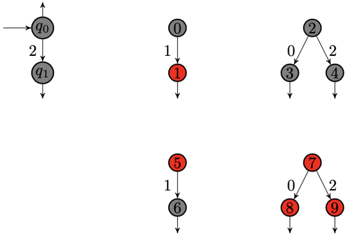


Assumptions:

- The first tree gives the prefix and occurs only once.
- The last eight trees of height 2 occur infinitely often.

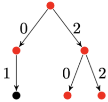
Build an NFA:

- States: $\{T[\varepsilon, 1]\} \cup F_1^{\infty}$
- Initial state: root of $T[\varepsilon, 1]$
- Final states: nodes of $T[\varepsilon, 1]$ and leaves in F_1^{∞}
- Transitions...



Example for transitions on state 7.
The corresponding tree has two extensions:

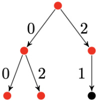
Extension 1



The tree hanging to the child 0 (resp. 2) of the root corresponds to state 5 (resp. 7).

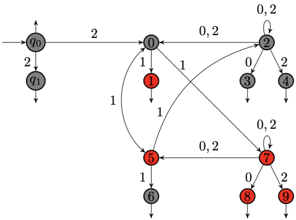
Transitions: $7 \xrightarrow{0} 5$ and $7 \xrightarrow{2} 7$

Extension 2



The tree hanging to the child 0 (resp. 2) of the root corresponds to state 7 (resp. 5).

Transitions: $7 \xrightarrow{0} 7$ and $7 \xrightarrow{2} 5$



Runs in the NFA for $w = 210 \in L_{\frac{3}{2}}$:

$q_0 \xrightarrow{2} q_1$ ✗

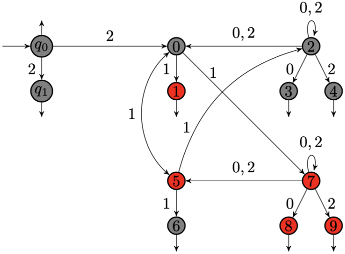
$q_0 \xrightarrow{2} 0 \xrightarrow{1} 1$ ✗

$q_0 \xrightarrow{2} 0 \xrightarrow{1} 5$ ✗

$q_0 \xrightarrow{2} 0 \xrightarrow{1} 7 \xrightarrow{0} 5$ ✗

$q_0 \xrightarrow{2} 0 \xrightarrow{1} 7 \xrightarrow{0} 7$ ✗

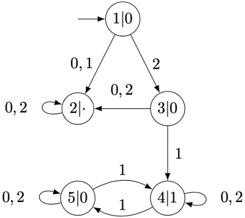
$q_0 \xrightarrow{2} 0 \xrightarrow{1} 7 \xrightarrow{0} 8$ ✓



Determinize the NFA (using the usual subset construction).

No output is set for state 2.

This DFA produces the sum-of-digits modulo 2 in base $\frac{3}{2}$ ($\frac{3}{2}$-automatic).



Theorem (Rigo, S., 2021)

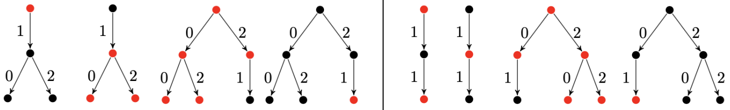
Let the tree $T(L_{\frac{p}{q}})$ be decorated by a sequence $\mathbf{x}$.
Suppose there exists some $h \geq 0$ such that $\#F^{\infty}_{h+1,a} \leq \#F^{\infty}_{h}$ for all $0 \leq j \leq q-1$ and all suitable letters $a \in A_p$.
Then $\mathbf{x}$ is $\frac{p}{q}$-automatic.

Automatic sequences to bounding $\#F^{\infty}_{h,a}$

Example: Base $\frac{3}{2}$

Decoration: the sum-of-digits mod 2 $\mathbf{t} = 001110111110\cdots$, $\frac{3}{2}$-automatic

$F^{\infty}_{2,0}$ $F^{\infty}_{2,1}$



$\#F^{\infty}_{2,0} \leq \#F^{\infty}_{1}$ $\#F^{\infty}_{2,1} \leq \#F^{\infty}_{1}$

Theorem (Rigo, S., 2021)

Let $\mathbf{x}$ be a $\frac{p}{q}$-automatic sequence generated by a DFAO $\mathcal{A} = (Q, q_0, A_p, \delta, \tau : A_p \to B)$ with the following property:

$\exists h \geq 0$ s.t. $\forall q \neq q' \in Q$ and $\forall w \in L_{\frac{p}{q}}$, $\exists u \in w^{-1}L_{\frac{p}{q}}$ with $|u| \leq h$ s. t. $\tau(\delta(q, u)) \neq \tau(\delta(q', u))$.

Then in the tree $T(L_{\frac{p}{q}})$ decorated by $\mathbf{x}$, we have $\#F^{\infty}_{h+1,a} \leq \#F^{\infty}_{h}$ for all $0 \leq j \leq q-1$ and all suitable letters $a \in A_p$.

References

- S. Akiyama, Ch. Frougny, and J. Sakarovitch, Powers of rationals modulo 1 and rational base number systems, *Israel J. Math.* **168** (2008), 53–91.
- J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, (2003).
- J. Berstel, L. Boasson, O. Carton, and I. Fagnot, Sturmian Trees, *Theoret. Comput. Sci.* **46** (2010), 443–478.
- V. Bruyère and G. Hansel, Bertrand numeration systems and recognizability, *Theoret. Comput. Sci.* **181** (1997), 17–43.
- É. Charlier, M. Le Gonidec, and M. Rigo, Representing real numbers in a generalized numeration system, *J. Comput. System Sci.* **77** (2011), no. 4, 743–759.
- A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164–192.
- K. Culik, J. Karhumäki, and A. Lepistö, Alternating iteration of morphisms and the Kolakovski sequence, in *Lindenmayer systems*, 93–106, Springer, Berlin, (1992).
- F. M.Dekking, Regularity and irregularity of sequences generated by automata, *Sém. Th. Nombres Bordeaux* 79–80 (1980), 901–910.
- J.-M. Dumont and A. Thomas, Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theoret. Comput. Sci.* **65** (1989), 153–169.
- T. Edgar, H. Olafson, and J. Van Alstine, Some combinatorics of rational base representations, preprint.
- J. Endrullis and D. Hendriks, On periodically iterated morphisms, *Proc. CSL-LICS'14* in Vienna (2014), 1–10.
- P. Lecomte and M. Rigo, Abstract numeration systems, Ch. 3, in *Combinatorics, Automata and Number Theory*, Encyclopedia Math. Appl. **135**, Cambridge University Press, (2010).

References

- A. Lepistö, On the power of periodic iteration of morphisms, ICALP 1993, 496–506, *Lect. Notes Comp. Sci* **700**, (1993).
- J. Peltomäki, A. Massuir, and M. Rigo, Automatic sequences based on Parry or Bertrand numeration systems, *Adv. Appl. Math.* **108** (2019), 11–30.
- V. Marsault, On $\frac{p}{q}$-recognisable sets, `arXiv:1801.08707`.
- V. Marsault, *Énumération et numération*, Ph.D. thesis, Télecom-Paristech, 2015.
- V. Marsault and J. Sakarovitch, On sets of numbers rationally represented in a rational base number system. Algebraic informatics, *Lect. Notes Comp. Sci.* **8080**, 89–100, Springer, Heidelberg, 2013.
- V. Marsault and J. Sakarovitch, Breadth-first serialisation of treesand rational languages, Developments in Language Theory - 18th International Conference, 2014, Ekaterinburg, Russia, August 26-29, 2014, *Lect. Notes Comp. Sci.* **8633**, 252–259.
- V. Marsault and J. Sakarovitch, Trees and languages with periodic signature, *Indagationes Mathematicae* **28** (2017), 221–246.
- V. Marsault and J. Sakarovitch, The signature of rational languages, *Theor. Comput. Sci.* **658** (2017), 216–234.
- J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés. Automata, languages and programming (Antwerp, 1984), 380–389, *Lect. Notes Comp. Sci.* **172**, Springer, Berlin, (1984).
- M. Rigo and A. Maes, More on generalized automatic sequences, *J. Autom. Lang. Comb.* **7** (2002), 351–376.
- M. Rigo, *Formal Languages, Automata and Numeration Systems*, ISTE–Wiley, (2014).
- ⋆ M. Rigo and M. Stipulanti, Automatic sequences: from rational bases to trees, `arxiv.2102.10828` preprint (2021).