Systèmes de numération pour les réels et pour les entiers : introduction illustrée et quelques exemples d'applications

Émilie Charlier

Département de mathématique, Université de Liège, Belgique

AFRIMATH, Abidjan
3 avril 2025

From numbers to words

Usually integers are represented by finite words while real numbers are represented by infinite words.

- In base 10: $148 \to 148$, $\frac{1}{3} \to 0.3333\cdots$, $\pi \to 3.141592\cdots$
- In base 2: $148 \to 10010100$, $\frac{1}{3} \to 0.01010101\cdots$, $\pi \to 11.001001000011\cdots$

The basic consideration is as follows: properties of numbers are translated into combinatorial properties of their representations.

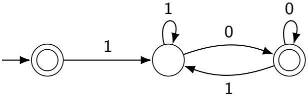
Recognizable sets of integers

A subset X of $\mathbb{N}$ is recognizable with respect to a given numeration system S, or S-recognizable, if the language

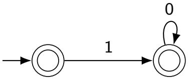
$$\{\mathrm{rep}_S(n) : n \in X\}$$

is regular, i.e., is accepted by a finite automaton.

- The set $2\mathbb{N}$ of even non-negative integers is 2-recognizable.

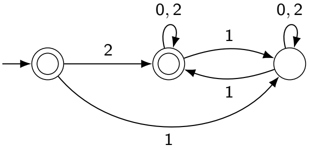


- The set $\{2^n : n \in \mathbb{N}\}$ of powers of 2 is 2-recognizable.



Changing the system

- The set $2\mathbb{N}$ of even non-negative integers is 3-recognizable.



In fact, the set $2\mathbb{N}$ is b-recognizable for all integer bases b.

- The set $\{2^n : n \in \mathbb{N}\}$ of powers of 2 is not 3-recognizable.

  This is a consequence of Cobham's theorem.

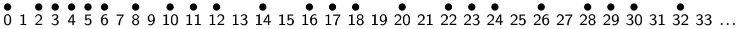
Cobham's theorem

Two integers k and ℓ are multiplicatively independent if $k^m = \ell^n$ and $m, n \in \mathbb{N}$ implies $m = n = 0$.

Theorem (Cobham 1969)

Let b and b' be multiplicatively independent integer bases. If a subset of $\mathbb{N}$ is simultaneously b-recognizable and b'-recognizable, then it is a finite union of arithmetic progressions (possibly finite).

$2\mathbb{N} \cup (3\mathbb{N} + 2) \cup \{3\}$



From words to numbers

On the other hand, infinite words may also represent sets of numbers:
the characteristic sequence of a subset of $\mathbb{N}$ is a binary infinite word.

- The set $2\mathbb{N}$ gives the periodic infinite word $10101010\cdots$
- The set $\{2^n : n \in \mathbb{N}\}$ gives the aperiodic infinite word $01101000100000001000\cdots$

Exercise: Show that the characteristic sequence of a subset of $\mathbb{N}$ is ultimately periodic, that is, of the form $uvvv\cdots$, if and only if it is a finite union of arithmetic progressions (possibly finite).

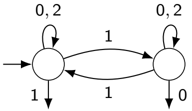
$2\mathbb{N} \cup (3\mathbb{N}+2) \cup \{3\}$

For this reason, we also talk about ultimately periodic sets of integers.

Linking recognizable sets and automatic sequences

For an integer base $b \geq 2$, a subset X of $\mathbb{N}$ is b-recognizable if and only if its characteristic sequence is *b-automatic*: there exists a DFAO that on input $\mathrm{rep}_b(n)$ ouputs 1 if $n \in X$, and outputs 0 otherwise.

For example, the DFAO



generates the periodic sequence

$$1010101010\cdots$$

when reading 3-representations of integers, which corresponds to the subset of even non-negative integers

$$\{0, 2, 4, 6, 8, \ldots\}.$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 10.$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 10.$$

To get the Thue-Morse word, apply those rules iteratively from 0:

$$0$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 10.$$

  To get the Thue-Morse word, apply those rules iteratively from 0:

$$\underline{0}1$$

Automatic sequences

A sequence $f : \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 10.$$

To get the Thue-Morse word, apply those rules iteratively from 0:

$$0\underline{1}10$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

  $$0 \mapsto 01$$
  $$1 \mapsto 10.$$

  To get the Thue-Morse word, apply those rules iteratively from 0:

  $$01\underline{1}010$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 10.$$

To get the Thue-Morse word, apply those rules iteratively from 0:

$$011\underline{0}1001$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 10.$$

  To get the Thue-Morse word, apply those rules iteratively from 0:

$$0110\underline{1}00110$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$

$$1 \mapsto 10.$$

To get the Thue-Morse word, apply those rules iteratively from 0:

$$01101\underline{0}011001$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 10.$$

To get the Thue-Morse word, apply those rules iteratively from 0:

$$011010\underline{0}1100101\cdots$$

Automatic sequences

A sequence $f\colon \mathbb{N} \to B$ is called automatic with respect to a numeration system S, or *S-automatic*, if there exists a DFA0 $\mathcal{A} = (Q, q_0, \delta, A, \tau, B)$ such that

$$\forall n \in \mathbb{N}, \quad f(n) = \tau(\delta(q_0, \mathrm{rep}_S(n)))$$

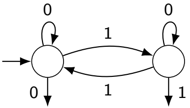
- The Thue-Morse word $01101001100101\cdots$ is a fixed point of the substitution

  $$0 \mapsto 01$$
  $$1 \mapsto 10.$$

  To get the Thue-Morse word, apply those rules iteratively from 0:

  $$011010\underline{0}1100101\cdots$$

  This infinite word is 2-automatic since it is generated by the DFAO

  when reading integers in base 2.

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$
\begin{aligned}
0 &\mapsto 01 \\
1 &\mapsto 0.
\end{aligned}
$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$0$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$\underline{0}1$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$0\underline{1}0$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$01\underline{0}01$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$\begin{aligned} 0 &\mapsto 01 \\ 1 &\mapsto 0. \end{aligned}$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$010\underline{0}101$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$0100\underline{1}010$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$01001\underline{0}1001$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$\begin{aligned} 0 &\mapsto 01 \\ 1 &\mapsto 0. \end{aligned}$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$010010\underline{1}0010$$

- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$0100101\underline{0}01001\cdots$$

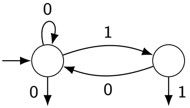
- The Fibonacci sequence $0100101001001\cdots$ is the fixed point of the substitution

$$0 \mapsto 01$$
$$1 \mapsto 0.$$

To get the Fibonacci word, apply those rules iteratively from 0:

$$0100101\underline{0}01001\cdots$$

The Fibonacci sequence $0100101001001\cdots$ is generated by the DFAO



when reading the Zeckendorf representations of the integers.

A range of numeration systems

- Unary representations

  A natural number n is represented by the finite word $\mathrm{rep}_1(n) = a^n$ where a is any fixed symbol.

  Exercise: Show that the 1-recognizable subsets of $\mathbb{N}$ are exactly the ultimately periodic sets.

- Binary representations

$\cdots$	16	8	4	2	1	
$\cdots$	a_4	a_3	a_2	a_1	a_0	
						0
					1	1
				1	0	2
				1	1	3
			1	0	0	4
			1	0	1	5
			1	1	0	6
			1	1	1	7
		1	0	0	0	8

We have $n = \sum_{i=0}^{\ell-1} a_i 2^i$ with $a_{\ell-1} = 1$, and we write $\mathrm{rep}_2(n) = a_{\ell-1} \cdots a_0$.

- Integer base representations

  Let $b \geq 2$ be an integer. A natural number n is represented by the finite word $\mathrm{rep}_b(n) = a_{\ell-1} \cdots a_0$ obtained from the greedy algorithm:

$$n = \sum_{i=0}^{\ell-1} a_i b^i.$$

  The greedy algorithm only imposes to have a nonzero leading digit $a_{\ell-1}$.

  Thus, the set of all greedy representations is

$$\{1, \ldots, b-1\}\{0, \cdots, b-1\}^* \cup \{\varepsilon\}.$$

- Zeckendorf representations

Let $F = (F_i)_{i \geq 0} = (1, 2, 3, 5, 8, \dots)$ be the sequence obtained from the rules:

$$F_0 = 1,\ F_1 = 2 \text{ and } F_{i+2} = F_{i+1} + F_i \text{ for } i \geq 0.$$

Again, we can use the greedy algorithm in order to produce a sequence of digits $a_{\ell-1} \cdots a_0$ such that $n = \sum_{i=0}^{\ell-1} a_i F_i$:

$\cdots$	8	5	3	2	1	
$\cdots$	a_4	a_3	a_2	a_1	a_0	n
						0
					1	1
				1	0	2
			1	0	0	3
			1	0	1	4
		1	0	0	0	5
		1	0	0	1	6
		1	0	1	0	7
	1	0	0	0	0	8

In addition to having a nonzero leading digit $a_{\ell-1}$, the greedy algorithm imposes that the valid representations do not contain two consecutive 1's.

The set of all greedy representations is

$$1\{0, 01\}^* \cup \{\varepsilon\}.$$

[Zeckendorf 1972]

- Positional representations

Let $U = (U_i)_{i\geq 0}$ be a base sequence, that is, an increasing sequence of integers such that $U_0 = 1$ and the quotients $\frac{U_{i+1}}{U_i}$ are bounded.

A natural number n is represented by the finite word

$$\mathrm{rep}_U(n) = a_{\ell-1} \cdots a_0$$

obtained from the greedy algorithm:

$$n = \sum_{i=0}^{\ell-1} a_i U_i.$$

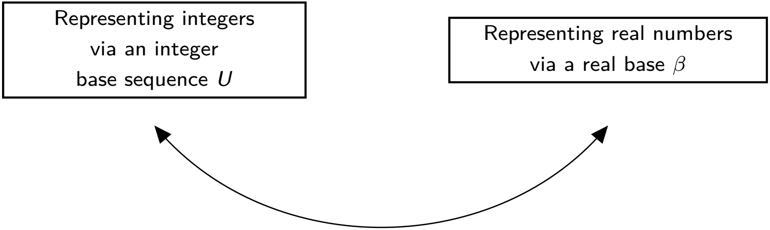
A description of the numeration language

$$L_U = 0^* \{\mathrm{rep}_U(n) : n \in \mathbb{N}\}$$

strongly depends on the base sequence U.

Given such a system U, other choices of representations could be made, such as the lazy algorithm for instance.

Knuth 1981, Fraenkel 1985


Representing integers via an integer base sequence U
Representing real numbers via a real base β

Which link?

Representing real numbers in base 3

Any $x \in [0,1)$ can be decomposed in a unique way as

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}$$

where $a_i \in \{0,1,2\}$ and $a_i a_{i+1} a_{i+2} \cdots \neq 2^\omega$ for all i.

We write $d_3(x) = a_1 a_2 a_3 \cdots$.

Define $D_3 = \{d_3(x) : x \in [0,1)\}$.

The topological closure of D_3 is called the 3-shift:

$$S_3 = \{\mathbf{w} \in \{0,1,2\}^\omega : \mathrm{Fac}(\mathbf{w}) \subseteq \mathrm{Fac}(D_3)\} = \{0,1,2\}^\omega.$$

Straightforward but crucial observation: $\mathrm{Fac}(S_3) = L_3$.

Representing real numbers in base φ

Let $\varphi = \frac{1+\sqrt{5}}{2}$ (the golden mean).

Any $x \in [0,1)$ can be decomposed in a unique way as

$$x = \sum_{i=1}^{\infty} \frac{a_i}{\varphi^i}$$

where $a_i \in \{0,1\}$, $a_i a_{i+1} \neq 11$ and $a_i a_{i+1} a_{i+2} \cdots \neq (10)^\omega$ for all i.

We write $d_\varphi(x) = a_1 a_2 a_3 \cdots$.

Define $D_\varphi = \{d_\varphi(x) : x \in [0,1)\}$.

The topological closure of D_φ is called the φ-shift:

$$S_\varphi = \{\mathbf{w} \in \{0,1\}^\omega : \mathrm{Fac}(\mathbf{w}) \subseteq \mathrm{Fac}(D_\varphi)\} = \{0,1\}^\omega \setminus \{0,1\}^* 11 \{0,1\}^\omega.$$

Straightforward but crucial observation: $\mathrm{Fac}(S_\varphi) = \mathcal{N}_F$.

Representing real numbers via real bases $\beta > 1$

Let $\beta > 1$ be real number (called the base).

We may represent any $x \in [0, 1]$ by using the following greedy algorithm.

For all $i \geq 1$, let a_i be the greatest integer a such that

$$\sum_{j=1}^{i-1} \frac{a_j}{\beta^j} + \frac{a}{\beta^i} \leq x.$$

We get that

$$\sum_{i=1}^{\infty} \frac{a_i}{\beta^i} = x.$$

The infinite word $d_\beta(x) = a_1 a_2 \cdots$ is called the β-expansion of x.

Only finitely many digits are used, namely $0, 1, \ldots, \lfloor \beta \rfloor$.

[Rényi 1959]

The β-shift

For $\beta > 1$, we let $D_\beta = \{d_\beta(x) : x \in [0,1)\}$.

The β-shift is the topological closure of D_β:

$$S_\beta = \{\mathbf{w} \in \{0, \dots, \lceil \beta \rceil - 1\}^\omega : \mathrm{Fac}(\mathbf{w}) \subseteq \mathrm{Fac}(D_\beta)\}.$$

Parry's characterization of elements in the β-shift

In Parry's theorem, the β-expansion and the quasi-greedy β-expansion of 1 play crucial roles.

The quasi-greedy β-expansion of 1 is

$$d_\beta^*(1) = \lim_{x \to 1^-} d_\beta(x).$$

Combinatorial definition:

- If $d_\beta(1)$ does not end with a tail of zeros, then we simply have $d_\beta^*(1) = d_\beta(1)$.
- If $d_\beta(1) = d_1 \cdots d_\ell 0^\omega$ with $d_\ell \neq 0$, in which case we say that $d_\beta(1)$ is finite, then $d_\beta^*(1) = (d_1 \cdots d_{\ell-1}(d_\ell - 1))^\omega$.

Theorem (Parry 1960)

$$S_\beta = \{\mathbf{w} \in \{0, \ldots, \lceil \beta \rceil - 1\}^\omega : \forall i \geq 1,\ w_i w_{i+1} \cdots \leq_{\text{lex}} d_\beta^*(1)\}.$$

[Parry 1960]

Parry's descriptions of the 3-shift and the φ-shift

For $\beta = 3$, we get $d_3(1) = 30^\omega$ and $d_3^*(1) = 2^\omega$. So Parry's theorem gives

$$S_3 = \{w \in \{0,1,2\}^\omega : \forall i \geq 1,\ w_i w_{i+1} \cdots \leq_{\mathrm{lex}} 2^\omega\}.$$

For $\beta = \varphi$, we get $d_\varphi(1) = 110^\omega$ and $d_\varphi^*(1) = (10)^\omega$. So Parry's theorem gives

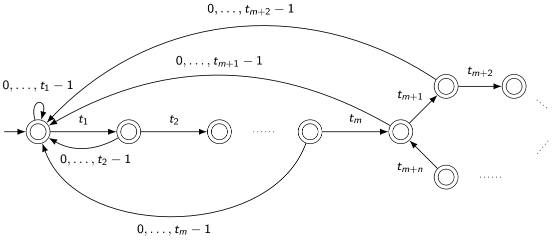
$$S_\varphi = \{w \in \{0,1\}^\omega : \forall i \geq 1,\ w_i w_{i+1} \cdots \leq_{\mathrm{lex}} (10)^\omega\}.$$

The β-shift S_β is called sofic if $\mathrm{Fac}(S_\beta)$ is a regular language.

As a consequence of Parry's characterization, we get:

Corollary

The β-shift is sofic if and only if $d^*_\beta(1)$ is an ultimately periodic word.



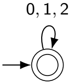
The Parry automaton associated with β where $d^*_\beta(1) = t_1 \dots t_m (t_{m+1} \cdots t_{m+n})^\omega$.

Such numbers and automata are named after Parry:

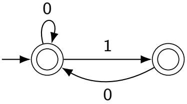
- A real base $\beta > 1$ is a called Parry number if $d^*_\beta(1)$ is an ultimately periodic word.
- The drawn automaton is called the Parry automaton associated with β

The Parry automata for 3, φ and φ^2

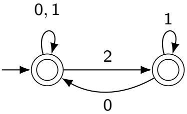
For $\beta = 3$, since $d_3^*(1) = 2^\omega$, we get



For $\beta = \varphi$, since $d_\varphi^*(1) = (10)^\omega$, we get



For $\beta = \varphi^2$, since $d_\varphi^*(1) = 21^\omega$, we get



Bertrand numeration systems

Let U be a positional numeration system.

Two desirable properties of the numeration language $L_U = 0^*\mathrm{rep}_U(\mathbb{N})$ are:

- L_U is prefix-closed if all prefixes of words in L_U also belong to L_U.
- L_U is prolongable if for all w in L_U, the word $w0$ also belongs to L_U.

We say that U is a Bertrand numeration system if L_U is both prefix-closed and prolongable.

Equivalently: $\forall w \in A_U^*,\ w \in L_U \iff w0 \in L_U$.

[Bertrand-Mathis 1989]
[Bruyère & Hansel 1997]

Canonical Bertrand systems associated with a real base β

For a real number $\beta > 1$, define

$$U_i = a_1 U_{i-1} + a_2 U_{i-2} + \cdots + a_i U_0 + 1, \quad \forall i \geq 0$$

where $(a_i)_{i\geq 1}$ is given by $d_\beta^*(1)$.

The so-obtained sequence $U = (U_i)_{i\geq 0}$ defines a positional numeration system for representing integers.

This numeration system is Bertrand, and it has β as a dominant root, meaning that

$$\lim_{i\to\infty} \frac{U_{i+1}}{U_i} = \beta.$$

Moreover, we have the language equality

$$L_U = \mathrm{Fac}(S_\beta).$$

Thanks to Parry's characterization, we see that

$$L_U \text{ is regular} \iff \beta \text{ is a Parry number.}$$

[Bertrand-Mathis 1989]

Canonical Bertrand systems associated with 3, φ and φ^2

For $\beta = 3$, since $d_3^*(1) = 2^\omega$, we get $U_i = 2U_{i-1} + 2U_{i-2} + \cdots + 2U_0 + 1$.

This gives $U_0 = 1$, $U_1 = 2U_0 + 1 = 3$, $U_2 = 2U_1 + 2U_0 + 1 = 9$,
$U_3 = 2U_2 + 2U_1 + 2U_0 + 1 = 27$...

For $\beta = \varphi$, since $d_\varphi^*(1) = (10)^\omega$, we get

$$U_i = \begin{cases} U_{i-1} + U_{i-3} + \cdots + U_1 + 1, & \text{if } i \equiv 0 \pmod 2; \\ U_{i-1} + U_{i-3} + \cdots + U_0 + 1, & \text{if } i \equiv 1 \pmod 2. \end{cases}$$

This gives $U_0 = 1$, $U_1 = U_0 + 1 = 2$, $U_2 = U_1 + 1 = 3$, $U_3 = U_2 + U_0 + 1 = 5$,
$U_4 = U_3 + U_1 + 1 = 8$...

For $\beta = \varphi^2$, since $d_{\varphi^2}^*(1) = 21^\omega$, we get $U_i = 2U_{i-1} + U_{i-2} + \cdots + U_0 + 1$.

This gives $U_0 = 1$, $U_1 = 2U_0 + 1 = 3$, $U_2 = 2U_1 + U_0 + 1 = 8$,
$U_3 = 2U_2 + U_1 + U_0 + 1 = 21$...

Non-Bertrand systems

Define

$$U_i = a_1 U_{i-1} + a_2 U_{i-2} + \cdots + a_i U_0 + 1, \quad \forall i \geq 0$$

with the sequence of coefficients given by

$$(a_i)_{i \geq 1} = 10110^\omega.$$

This system is again linked with the Golden ratio φ since $\frac{1}{\varphi} + \frac{1}{\varphi^3} + \frac{1}{\varphi^4} = 1$. It has φ as a dominant root : $\lim_{i \to \infty} \frac{U_{i+1}}{U_i} = \varphi$.

We have

$$U_0 = 1, \quad U_1 = U_0 + 1 = 2, \quad U_2 = U_1 + 1 = 3, \quad U_3 = U_2 + U_0 + 1 = 5,$$
$$U_i = U_{i-1} + U_{i-3} + U_{i-4} + 1, \quad i \geq 4$$

so that $U = (1, 2, 3, 5, 9, 15, 24, 39, \ldots)$.

This system is not Bertrand since for example, $1100, 11000 \in L_U$ but $11, 110, 110000 \notin L_U$, showing that L_U is neither prefix-closed nor prolongable.

In fact, we have

$$U_{i+2} = \begin{cases} U_{i+1} + U_i, & \text{if } i \equiv 2, 3 \pmod 4; \\ U_{i+1} + U_i + 1, & \text{if } i \equiv 0, 1 \pmod 4. \end{cases}$$

The canonical Bertrand system U associated with β has the property that

$$\mathrm{rep}_U(U_i - 1) = \mathrm{Pref}_i(d^*_\beta(1)), \quad \text{for all } i \geq 0.$$

Proposition (Hollander 1998)

Let U be a positional numeration system such that $\frac{U_{i+1}}{U_i} = \beta > 1$.

- *If $d_\beta(1) = d^*_\beta(1)$ is not finite, then*

$$\lim_{i\to\infty} \mathrm{rep}_U(U_i - 1) = d^*_\beta(1).$$

- *If $d_\beta(1) = d_1 \cdots d_\ell 0^\omega$ with $d_\ell \neq 0$, then for all $n \geq 0$ and all large enough i, there exists $k \geq 0$ such that*

$$\mathrm{Pref}_n(\mathrm{rep}_U(U_i - 1)) = \mathrm{Pref}_n((d_1 \cdots d_{\ell-1}(d_\ell - 1))^k d_1 \cdots d_\ell 0^\omega).$$

[Hollander 1998]

The Zeckendorf system $F = (1, 2, 3, 5, 8, 13, 21, 34, \dots)$, which is the canonical Bertrand system associated with φ satisfies

$$\mathrm{rep}_F(1) = 1,\ \mathrm{rep}_F(2) = 10,\ \mathrm{rep}_F(4) = 101,\ \mathrm{rep}_F(7) = 1010,\ \mathrm{rep}_F(11) = 10101, \dots$$

that is

$$\mathrm{rep}_F(F_i - 1) = \mathrm{Pref}_i(d_\varphi^*(1)) = \mathrm{Pref}_i((10)^\omega).$$

The non-Bertrand system $U = (1, 2, 3, 5, 9, 15, 24, 39, \dots)$ we've seen before (still with the dominant root φ) is such that

$$\mathrm{rep}_U(1) = 1,\ \mathrm{rep}_U(2) = 10,\ \mathrm{rep}_U(4) = 101,\ \mathrm{rep}_U(8) = 1100,\ \mathrm{rep}_U(14) = 11000, \dots$$

that is

$$\mathrm{rep}_U(U_i - 1) = \begin{cases} \mathrm{Pref}_i((10)^\omega), & \text{if } i \equiv 0, 1 \pmod 4; \\ \mathrm{Pref}_i(110^\omega), & \text{if } i \equiv 2, 3 \pmod 4. \end{cases}$$

A characterization of Bertrand numeration systems

Proposition (C., Cisternino & Stipulanti 2022)

Let U be a positional numeration system such that $\lim_{i\to\infty} \frac{U_{i+1}}{U_i} = \beta > 1$.

If $\lim_{i\to\infty} \mathrm{rep}_U(U_i - 1)$ exists, then it is either $d^*_\beta(1)$ or $d_\beta(1)$.

Theorem (C., Cisternino & Stipulanti 2022)

A positional numeration system U is Bertrand if and only if one of the following conditions is satisfied.

1. *We have $\mathrm{rep}_U(U_i - 1) = \mathrm{Pref}_i(10^\omega)$ for all $i \geq 0$.*
2. *There exists $\beta > 1$ such that $\mathrm{rep}_U(U_i - 1) = \mathrm{Pref}_i(d^*_\beta(1))$ for all $i \geq 0$.*
3. *There exists $\beta > 1$ such that $\mathrm{rep}_U(U_i - 1) = \mathrm{Pref}_i(d_\beta(1))$ for all $i \geq 0$.*

[C., Cisternino & Stipulanti 2022]

Regularity of L_U

A fundamental question is the following:

- Given a positional system U, can we decide if the numeration language L_U is regular?
- And even more precisely, can characterize those systems U giving rise to a regular numeration language L_U?

A necessary condition is that the sequence $U = (U_i)_{i\geq 0}$ is linear, i.e., it must satisfy a linear recurrence relation with integer coefficients: there exist integers $c_1, \ldots, c_k$ such that

$$U_i = c_1 U_{i-1} + c_2 U_{i-2} \cdots + c_k U_{i-k}, \quad \text{for all } i \geq k.$$

The characteristic polynomial of the recurrence relation is

$$X^k - c_1 X^{k-1} - c_2 X^{k-2} - \cdots - c_k.$$

This question was studied by Hollander in the case of linear systems with a dominant root, i.e., such that the limit $\lim_{i\to\infty} \frac{U_{i+1}}{U_i}$ exists and is greater than 1.

A clever observation he made was that is sufficient to study the regularity of the language made of words of maximal length.

Proposition (Hollander 1998)

L_U is regular $\iff \mathrm{Max}(L_U) := \{\mathrm{rep}_U(U_i - 1) : i \geq 0\}$ is regular.

He also showed the following necessary condition:

Proposition (Hollander 1998)

If U has a dominant root $\beta > 1$ and if L_U is regular, then β is a Parry number.

In order to give Hollander's full statement, we need to introduce the notion of β-polynomials. Suppose that $d^*_\beta(1) = t_1 \dots t_m(t_{m+1} \cdots t_{m+n})^\omega$, then the polynomial

$$P_{\beta,m,n} = \left(X^{m+n} - \sum_{i=1}^{m+n} t_i X^{m+n-i}\right) - \left(X^m - \sum_{i=1}^{m} t_i X^{m-i}\right).$$

is called a β-polynomial.

For m, n minimal, we get the canonical β-polynomial, simply denoted P_β.

- If $d^*_\beta(1) = 21^\omega$, then $m = n = 1$ and

$$P_\beta = (X^2 - 2X - 1) - (X - 2) = X^2 - 3X + 1.$$

- If $d^*_\beta(1) = (10)^\omega$, then $m = 0$, $n = 2$ and

$$P_\beta = (X^2 - X - 0) - (X^0) = X^2 - X - 1.$$

In the case where $d_\beta(1) = d_1 \dots d_\ell 0^\omega$ is finite (with $d_\ell \neq 0$), it is easy to see that

$$P_\beta = X^\ell - \sum_{i=1}^{\ell} t_i X^{\ell-i}.$$

Theorem (Hollander 1998)

Let U be a linear numeration system with a dominant root $\beta > 1$.

- *If L_U is regular, then β is a Parry number.*
- *Case where $d_\beta(1) = d^*_\beta(1)$.*
  - *L_U is regular if and only if U satisfies a recurrence relation of characteristic polynomial $P_{\beta,m,n}$ for some m, n.*
- *Case where $d_\beta(1) = d_1 \ldots d_\ell 0^\omega$ with $d_\ell \neq 0$.*
  - *If U satisfies a recurrence relation of characteristic polynomial $P_{\beta,m,n}$ for some m, n, then L_U is regular.*
  - *If L_U is regular, then the base sequence U satisfies a recurrence relation of characteristic polynomial of the form $(X^\ell - 1)P_{\beta,m,n}$ for some m, n.*

β-integers and sturmian words

A real number $x \geq 0$ is a β-integer if its β-expansion is of the form

$$d_\beta(x) = a_{n-1} \cdots a_0.0^\omega \quad \text{with } n \in \mathbb{N}.$$

The set of all β-integers is denoted by $\mathbb{N}_\beta$.

A real number $x \geq 0$ is a β-integer if its β-expansion is of the form

$$d_\beta(x) = a_{n-1} \cdots a_0.0^\omega \quad \text{with } n \in \mathbb{N}.$$

The set of all β-integers is denoted by $\mathbb{N}_\beta$.

Lemma

This set is unbounded and discrete, i.e., it has no accumulation point in $\mathbb{R}$.

Proof

The β-expansion of a β-integer smaller than β^n is of the form $a_{m-1} \cdots a_0.0^\omega$ with $m \leq n$. Since $a_i < \beta$ for each i, there are only finitely many β-expansions having this property. □

β-integers and sturmian words

A real number $x \geq 0$ is a β-integer if its β-expansion is of the form

$$d_\beta(x) = a_{n-1} \cdots a_0.0^\omega \quad \text{with } n \in \mathbb{N}.$$

The set of all β-integers is denoted by $\mathbb{N}_\beta$.

Lemma

This set is unbounded and discrete, i.e., it has no accumulation point in $\mathbb{R}$.

Proof

The β-expansion of a β-integer smaller than β^n is of the form $a_{m-1} \cdots a_0.0^\omega$ with $m \leq n$. Since $a_i < \beta$ for each i, there are only finitely many β-expansions having this property. □

Let $(x_k)_{k\in\mathbb{N}}$ be the increasing sequence of β-integers:

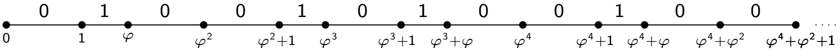
$$\mathbb{N}_\beta = \{x_k : k \in \mathbb{N}\}.$$

[Gazeau 1997]

Distances between consecutive β-integers

For $\beta = \varphi$, there are only two possible distances $\Delta_0 = 1$ and $\Delta_1 = \frac{1}{\varphi} = \varphi - 1$.

The distances $x_{k+1} - x_k$ between consecutive β-integers are coded by the Fibonacci word $0100101001001010\cdots$.



Theorem

The sequence $(x_{k+1} - x_k)_{k \geq 0}$ *of distances between consecutive* β*-integers takes only finitely many values if and only if the base* β *is a Parry number, in which case the corresponding infinite word is a fixed point of a primitive substitution.*

Let $d_\beta^*(1) = t_1 t_2 t_3 \cdots$. The possible distances are given by

$$\Delta_i = \mathrm{val}_\beta(0.t_{i+1} t_{i+2} t_{i+3} \cdots) = \sum_{k=1}^{\infty} \frac{t_{i+k}}{\beta^k}.$$

By letting $w_k = i$ if $x_{k+1} - x_k = \Delta_i$, the infinite word $\mathbf{w}_\beta = w_0 w_1 w_2 \cdots$ encodes the distances between β-integers.

If $d_\beta^*(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+n})^\omega$ for minimal m, n, then there are exactly $m + n$ distinct distances, and $\mathbf{w}_\beta$ is written over the alphabet $\{0, \ldots, m + n - 1\}$.

The infinite word $\mathbf{w}_\beta$ is the fixed point of the Parry substitution

$$\begin{aligned}
0 &\mapsto 0^{t_1} 1 \\
1 &\mapsto 0^{t_2} 2 \\
&\vdots \\
m{+}n{-}2 &\mapsto 0^{t_{m+n-1}} (m{+}n{-}1) \\
m{+}n{-}1 &\mapsto 0^{t_{m+n}} m.
\end{aligned}$$

Combinatorial properties of $\mathbf{w}_\beta$

The factor complexity of an infinite word $\mathbf{w}$ is the function $C(n)$ counting the number of factors of length n in $\mathbf{w}$.

Aperiodic words with factor complexity $C(n) = n + 1$ are called sturmian.

For example, the Fibonacci word $\mathbf{f} = 0100101001001010010100100101 0\cdots$ is sturmian:

$$\mathrm{Fac}_1(\mathbf{f}) = \{0, 1\}$$
$$\mathrm{Fac}_2(\mathbf{f}) = \{00, 01, 10\}$$
$$\mathrm{Fac}_3(\mathbf{f}) = \{001, 010, 100, 101\}$$
$$\mathrm{Fac}_4(\mathbf{f}) = \{0010, 0100, 0101, 1001, 1010\}$$

- $\mathbf{w}_\beta$ is sturmian if and only if β is a quadratic Parry number.
- In the case where $d_\beta(1)$ is finite, Arnoux-Rauzy words $\mathbf{w}_\beta$ are characterized in [Frougny, Masakova, Pelantova 2004].
- $\mathbf{w}_\beta$ with affine factor complexity $C(n) = an + b$ are characterized in [Bernat, Masakova, Pelantova 2007].

Current work:

- Regularity in the non dominant root case.
- Cantor real numeration systems, in particular, alternating real bases.
- In this context, generalized β-integers can be coded by words of other types, called S-adic words.

Thank you!

Merci !