Parry's 1960 theorem and some applications in combinatorics on words

Manon Stipulanti

Workshop *Random Number Representations, Stochastic Processes, and Quantum Transport*
Aalborg, Denmark
May 20–21, 2024

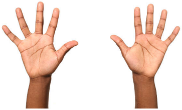
Numeration systems

set of rules to represent (natural) numbers

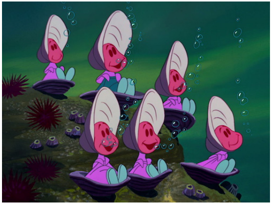
base 2

base 6

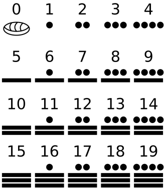
base 10



base 12



base 20



Formal definition

Definition

NS = an increasing sequence $U = (U(n))_{n\geq 0}$ of integers with $U(0) = 1$

- $U(n) = k^n$ for an integer $k \geq 1$

$$1024 = 2 \cdot (20)^2 + 11 \cdot 20 + 4 \cdot 1$$

$$17 = 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2$$

$$= 2 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2$$

- factorial $U(n) = n!$

  $463 = 3 \cdot 5! + 4 \cdot 4! + 1 \cdot 3! + 0 \cdot 2! + 1 \cdot 1! + 0 \cdot 0!$

  (unbounded alphabet)

- Zeckendorf/Fibonacci $U = 1, 2, 3, 5, 8, 13, 21, \ldots$

  $U(0) = 1 \quad U(1) = 2 \quad U(n+2) = U(n+1) + U(n)$

  $20 = 1 \cdot 13 + 0 \cdot 8 + 1 \cdot 5 + 0 \cdot 3 + 1 \cdot 2 + 0 \cdot 1$

Existence & uniqueness

Let $U = (U(n))_{n\geq 0}$ be a NS.

Euclidean algorithm $\rightsquigarrow n = \sum_{i=0}^{k} d_i U(i) \quad d_i \geq 0$

$$\mathrm{rep}_U(n) = d_k d_{k-1} \cdots d_0$$

Example: $U = 1, 2, 3, 5, 8, 13, 21, \ldots$ $\qquad \mathrm{rep}_U(20) = 101010$

$$\begin{array}{r|l} 20 & 1 \cdot 13 + 7 \\ 7 & 0 \cdot 8 + 7 \\ 7 & 1 \cdot 5 + 2 \\ 2 & 0 \cdot 3 + 2 \\ 2 & 1 \cdot 2 + 0 \\ 0 & 0 \cdot 1 \end{array}$$

Theorem (Fraenkel, 1983)

Existence: Any $n \geq 0$ can be written as $\sum_{i=0}^{k} d_i U(i)$.

Uniqueness: iff $d_i U(i) + d_{i-1} U(i-1) + \cdots + d_0 U(0) < U(i+1)\ \forall\, i \geq 0$ (greedy condition).

Real bases

- $\beta \in \mathbb{R}_{>1}$
- $x \in \mathbb{R} \cap [0,1] \rightsquigarrow x = \sum_{j=1}^{+\infty} c_j \beta^{-j} \quad c_j \in \mathbb{N}$
- greedy condition: $c_j \beta^{-j} + c_{j+1}\beta^{-j-1} + \cdots < \beta^{-j+1} \quad \forall\, j \geq 1$
- β-expansion: $d_\beta(x) = c_1 c_2 \cdots$
- special case: $d_\beta(1)$

Example: golden ratio φ

$$\varphi^2 = \varphi + 1$$
$$\Rightarrow 1 = \frac{1}{\varphi} + \frac{1}{\varphi^2}$$
$$\rightsquigarrow d_\varphi(1) = 110^\omega$$

Quasi-greedy β-expansions

$d_\beta(1) = t_1 t_2 \cdots \rightsquigarrow d_\beta^*(1)$

	$d_\beta(1)$	$d_\beta^*(1)$
“finite” case	$t_1 \cdots t_{m-1} \underbrace{t_m}_{\neq 0} 0^\omega$	$(t_1 \cdots t_{m-1}(t_m - 1))^\omega$
infinite case	$t_1 t_2 \cdots$	$t_1 t_2 \cdots$

Examples:

β	$d_\beta(1)$	$d_\beta^*(1)$
golden ratio φ	110^ω	$(10)^\omega$
φ^2	21^ω	21^ω

Parry's 1960 theorem

$D_\beta = \{d_\beta(x) \mid x \in [0,1)\}$ $\qquad S_\beta =$ closure of D_β

Theorem (Parry, 1960)

Let $\beta \in \mathbb{R}_{>1}$.
Let $\mathbf{s}$ be an infinite sequence over $\mathbb{N}$.
Then

$$\begin{aligned} \mathbf{s} \in D_\beta &\quad \text{iff} \quad \sigma^k(\mathbf{s}) < d^*_\beta(1) \quad \forall\, k \geq 0 \\ \mathbf{s} \in S_\beta &\quad \text{iff} \quad \sigma^k(\mathbf{s}) \leq d^*_\beta(1) \quad \forall\, k \geq 0 \end{aligned}$$

Purely combinatorial condition: compare $\mathbf{s}$ to $d^*_\beta(1)$ using the lex. order to check whether $\exists\, x \in [0,1)$ s.t. $\mathbf{s} = d_\beta(x)$

Example: golden ratio φ

	$(100)^\omega$	$101001000\cdots$	$\cdots 11 \cdots$
$\leq d^*_\varphi(1) = (10)^\omega$	✓	✓	✗

Parry numbers

Definition

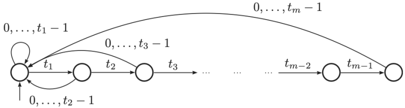
$\beta \in \mathbb{R}_{>1}$ is a Parry number if $d_\beta(1)$ is ultimately periodic.

Examples: integers, golden ratio φ, φ^2

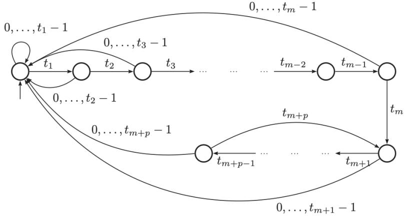
Corollary: $\beta \in \mathbb{R}_{>1}$ Parry number

num. lang. = labels of path in

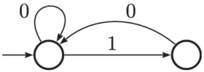
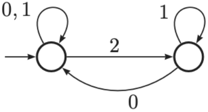
$$d_\beta(1) = t_1 \cdots t_m 0^\omega$$



$$d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+k})^\omega$$



Examples:

β	golden ratio φ	φ^2
$d^*_\beta(1)$	$(10)^\omega$	21^ω
		
in **s**	11 forbidden	

Linking both worlds

$U = (U(n))_{n\geq 0}$ $\qquad$ $\beta \in \mathbb{R}_{>1}$

$$\mathbb{N} \quad \xleftrightarrow{\text{(Bertrand, 1989)}} \quad \mathbb{R}$$

Definition

Let U be a NS.
Let L be the numeration language of U.
Then U is Bertrand if

$$\forall\, w,\ w \in L \iff w0 \in L.$$

Examples: integer bases, Fibonacci/Zeckendorf

Bertrand's 1989 theorem

recall from Parry's theorem...

$\beta \in \mathbb{R}_{>1}$

$S_\beta = \{\mathbf{s} \mid \sigma^k(\mathbf{s}) \leq d^*_\beta(1) \; \forall\, k \geq 0\}$ $\quad S'_\beta = \{\mathbf{s} \mid \sigma^k(\mathbf{s}) \leq d_\beta(1) \; \forall\, k \geq 0\}$

Theorem (Bertrand, 1989 + Charlier, Cisternino, S., 2022)

Let $U = (U(n))_{n\geq 0}$ be a NS.
Let L be the numeration language of U.
Then U is Bertrand iff one of the following cases occurs

- Case 1: $U(n) = n + 1 \quad \forall\, n \geq 0$
- Case 2: $\exists\, \beta \in \mathbb{R}_{>1}$ s.t. $L = \text{Fac}(S_\beta)$
- Case 3: $\exists\, \beta \in \mathbb{R}_{>1}$ s.t. $L = \text{Fac}(S'_\beta)$

Theorem (Bertrand, 1989 + Charlier, Cisternino, S., 2022)

Let $U = (U(n))_{n\geq 0}$ be a NS.
Let L be the numeration language of U.

Then U is Bertrand iff one of the following cases occurs

Case 1: $U(n) = n+1 \, \forall \, n \geq 0$ Case 2: $\exists \, \beta \in \mathbb{R}_{>1}$ s.t. $L = \mathrm{Fac}(S_\beta)$ Case 3: $\exists \, \beta \in \mathbb{R}_{>1}$ s.t. $L = \mathrm{Fac}(S'_\beta)$

In Case 2/Case 3,

- unique β
- for $d^*_\beta(1) = a_1 a_2 \cdots / d_\beta(1) = a_1 a_2 \cdots$

$$U(i) = a_1 U(i-1) + a_2 U(i-2) + \cdots + a_i U(0) + 1 \quad \forall \, i \geq 0$$

- dominant root: $\lim_{i\to+\infty} \frac{U(i+1)}{U(i)} = \beta$

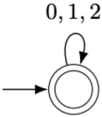
Remarks:

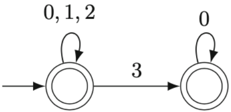
- Case 2 $\rightsquigarrow$ Bertrand, 1989
- New: when $d_\beta(1) \neq d^*_\beta(1) \rightsquigarrow 2$ Bertrand NSs

Canonical and non-canonical Bertrand

Example: $\beta = 3$

canonical Bertrand NS	non-canonical Bertrand NS
$d_3^*(1) = 2^\omega$	$d_3(1) = 30^\omega$
$U_1(0) = 1$ $U_1(n+1) = 2U_1(n) + \cdots + 2U_1(0) + 1$ $U_1 = 1, 3, 9, 27, 81, 243, \ldots$ $U_1(n) = 3^n$	$U_2(0) = 1$ $U_2(n+1) = 3U_2(n) + 1$ $U_2 = 1, 4, 13, 40, 121, 364, \ldots$
$L = \{\varepsilon\} \cup \{1,2\}\{0,1,2\}^*$	$L = \{0,1,2\}^* \cup \{0,1,2\}^* 30^*$

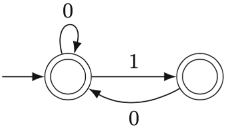


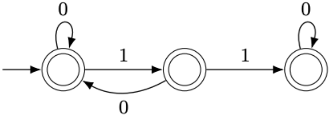


Canonical and non-canonical Bertrand

Example: $\beta = \varphi$ golden ratio

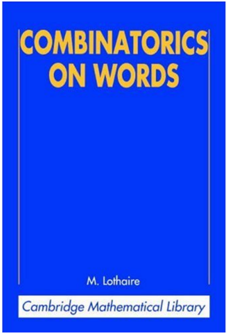
canonical Bertrand NS	non-canonical Bertrand NS
$d^*_\varphi(1) = (10)^\omega$	$d_\varphi(1) = 110^\omega$
$U_1(0) = 1 \quad U_1(1) = 2$	$U_2(0) = 1 \quad U_2(1) = 2$
$U_1(n+2) = U_1(n+1) + U_1(n)$	$U_2(n+1) = U_2(n) + U_2(n-1) + 1$
$U_1 = 1, 2, 3, 5, 8, 13, 21, \ldots$	$U_2 = 1, 2, 4, 7, 12, 20, 33, \ldots$
$L = \{\varepsilon\} \cup 1\{0, 01\}^*$	$L = \{0, 10\}^*\{\varepsilon, 1\} \cup \{0, 10\}^*110^*$



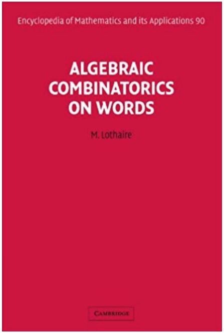


Combinatorics on words

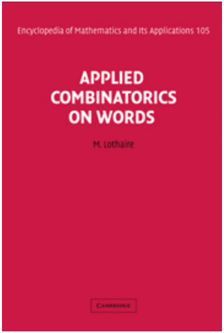
- the study of sequences of symbols
  words letters
- Lothaire's books
  collective work, different sets of authors



1983
overview



2002
overview



2005
applications

- relatively new area of discrete mathematics
- initiated by the Norwegian mathematician Axel Thue in 1906

(1863–1922)

- some topics of interest
  - combinatorial structure of words
  - regularities and patterns in words
  - important classes/families of words
    e.g., balanced, (a)periodic, automatic, regular, Lyndon, Sturmian, Arnoux-Rauzy, episturmian, de Bruijn
  - equations on words

An example

Thue $\rightsquigarrow$ square-free words

- A square is a non-empty word of the form xx.

  Example: in English $\underbrace{\texttt{cous}}\underbrace{\texttt{cous}}$ $\underbrace{\texttt{mur}}\underbrace{\texttt{mur}}$

- A word is square-free if is does not contain any squares.

  Example: `word` is square-free
  `repetition` is not

Question: are there infinite square-free words?

1 letter	$000\cdots$	✗
2 letters	0 1 0 ?	✗
4 letters	03121 01213 01321 01231 01321 $\cdots$	(Thue, 1906)
3 letters	012021012102 $\cdots$	(Thue, 1912)

$\rightsquigarrow$ squares are avoidable on ≥ 3 letters

Application 1: Generalized Pascal's triangles

Pascal's triangle

$\mathrm{P}\colon (m,k) \in \mathbb{N}\times\mathbb{N} \mapsto \binom{m}{k} \in \mathbb{N}$

$\binom{m}{k}$	k=0	1	2	3	4	5	6	7	⋯
m = 0	1	0	0	0	0	0	0	0	
1	1	1	0	0	0	0	0	0	
2	1	2	1	0	0	0	0	0	
3	1	3	3	1	0	0	0	0	
4	1	4	6	4	1	0	0	0	
5	1	5	**10**	**10**	5	1	0	0	
6	1	6	15	**20**	15	6	1	0	
7	1	7	21	35	35	21	7	1	
⋮									⋱

Binomial coefficients

$$\binom{m}{k} = \frac{m!}{(m-k)!\,k!}$$

Pascal's rule

$$\binom{m}{k} = \binom{m-1}{k} + \binom{m-1}{k-1}$$

A specific construction

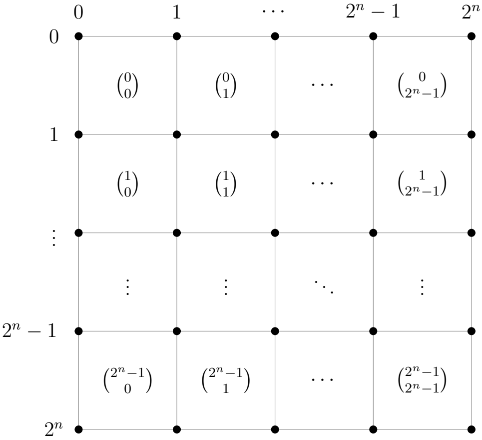
- Grid: first 2^n rows and columns of the Pascal's triangle

$$\left(\binom{m}{k}\right)_{0\leq m,k<2^n}$$

- Color each square in
  - white if $\binom{m}{k} \equiv 0 \bmod 2$
  - black if $\binom{m}{k} \equiv 1 \bmod 2$

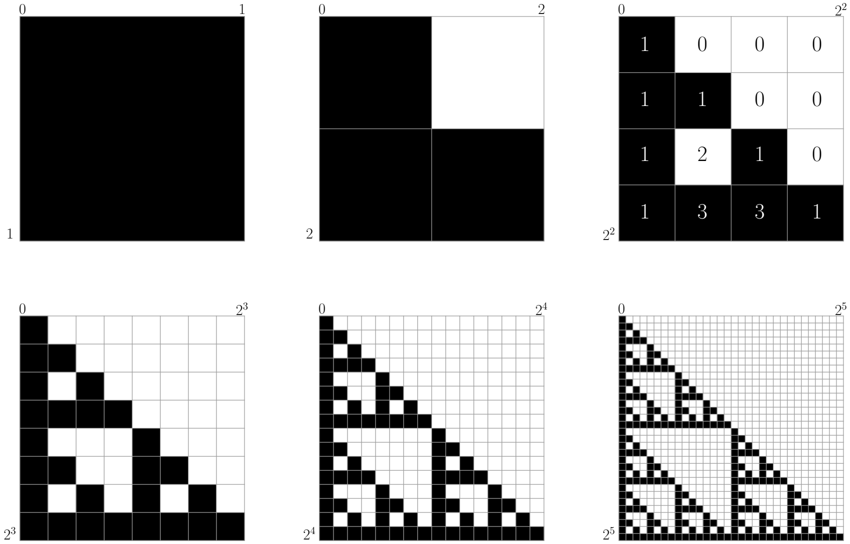
- Normalize by a homothety of ratio $1/2^n$

$\rightsquigarrow$ sequence of compact sets in $[0,1]^2$

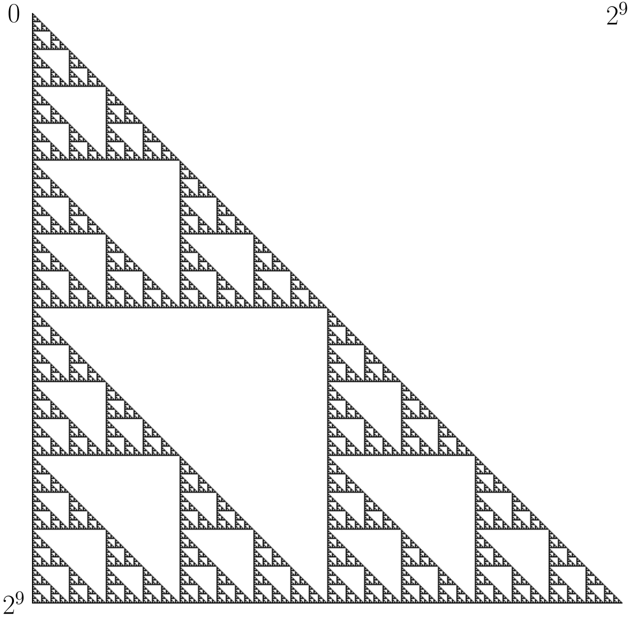


$\mathbb{N}^2 \cap [0, 2^n]^2$

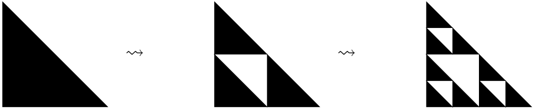
The first six elements

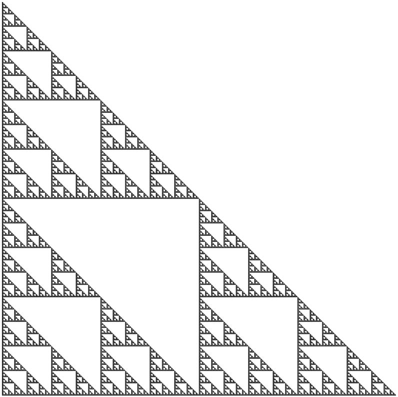


The tenth element



Sierpiński's triangle





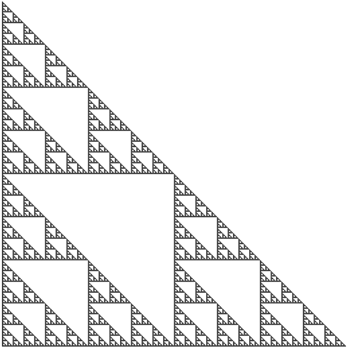
A folklore result

Theorem

The sequence of compact sets for

$$\left(\binom{m}{k} \bmod 2\right)_{0\leq m,k<2^n}$$

converges to the Sierpiński's triangle (w.r.t. the Hausdorff distance).



Definitions:

- ϵ-fattening of a subset $S \subset \mathbb{R}^2$

$$[S]_\epsilon = \bigcup_{x\in S} B(x,\epsilon)$$

- $(\mathcal{H}(\mathbb{R}^2), d_h)$ complete space of compact subsets $\neq \emptyset$ of $\mathbb{R}^2$ with the Hausdorff distance d_h

$$d_h(S,S') = \inf\{\epsilon \in \mathbb{R}_{>0} \mid S \subset [S']_\epsilon \quad \text{and} \quad S' \subset [S]_\epsilon\}$$

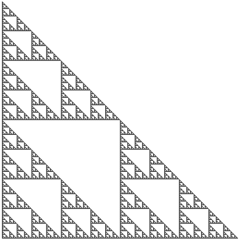
Theorem (von Haeseler, Peitgen, and Skordev, 1992)

Let p be a prime and $s > 0$.
The sequence of compact sets corresponding to

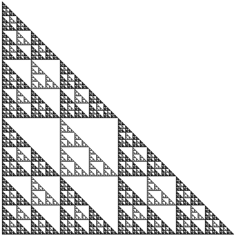
$$\left(\binom{m}{k} \bmod p^s\right)_{0 \leq m,k < p^n}$$

converges when n tends to infinity (w.r.t. the Hausdorff distance).

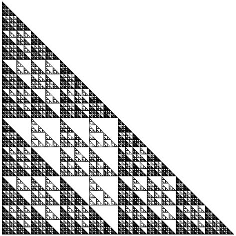
$p = 2,\ s = 1$



$p = 2,\ s = 2$



$p = 2,\ s = 3$



Binomial coefficient of finite words

integers $\rightsquigarrow$ words

Definition

Let u, v be finite words.
The binomial coefficient $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u ("scattered" subword).

Example: $u = 101001$ $\qquad$ $v = 101$

101001	**10**100**1**	
101**0**0**1**	**1**010**01**	$\Rightarrow \binom{101001}{101} = 6$
10**1**0**01**	10**1**0**01**	

Generalization of binomial coefficients over $\mathbb{N}$:

$$\text{letter } a: \quad \binom{a^m}{a^k} = \binom{\overbrace{a\cdots a}^{m \text{ times}}}{\underbrace{a\cdots a}_{k \text{ times}}} = \binom{m}{k} \quad \forall\, m, k \in \mathbb{N}$$

Generalized Pascal's triangle P_2 in base 2

$P_2 \colon (m,k) \in \mathbb{N} \times \mathbb{N} \mapsto \binom{\text{rep}_2(m)}{\text{rep}_2(k)} \in \mathbb{N}$

	$\binom{\text{rep}_2(m)}{\text{rep}_2(k)}$	$\text{rep}_2(k)$								
		ε	1	10	11	100	101	110	111	$\cdots$
	ε	1	0	0	0	0	0	0	0	
	1	1	1	0	0	0	0	0	0	
	10	1	1	1	0	0	0	0	0	
	11	1	**2**	0	**1**	0	0	0	0	
$\text{rep}_2(m)$	100	1	1	2	0	1	0	0	0	
	101	1	2	1	1	0	1	0	0	
	110	1	2	2	1	0	0	1	0	
	111	1	3	0	**3**	0	0	0	1	
	$\vdots$									$\ddots$

Rule (not local)

$$\binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

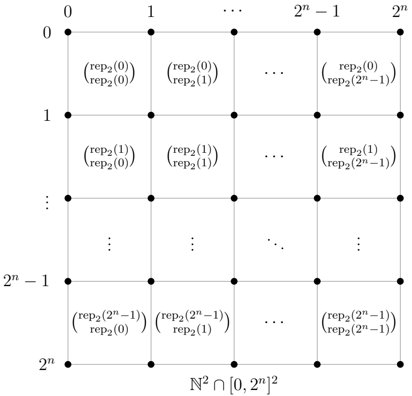
$\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)}$		$\mathrm{rep}_2(k)$								
		ε	**1**	10	**11**	100	101	110	**111**	$\cdots$
	ε	**1**	0	0	0	0	0	0	0	
	1	**1**	**1**	0	0	0	0	0	0	
	10	1	1	1	0	0	0	0	0	
	11	**1**	**2**	0	**1**	0	0	0	0	
$\mathrm{rep}_2(m)$	100	1	1	2	0	1	0	0	0	
	101	1	2	1	1	0	1	0	0	
	110	1	2	2	1	0	0	1	0	
	111	**1**	**3**	0	**3**	0	0	0	**1**	
	$\vdots$									$\ddots$

the usual Pascal's triangle

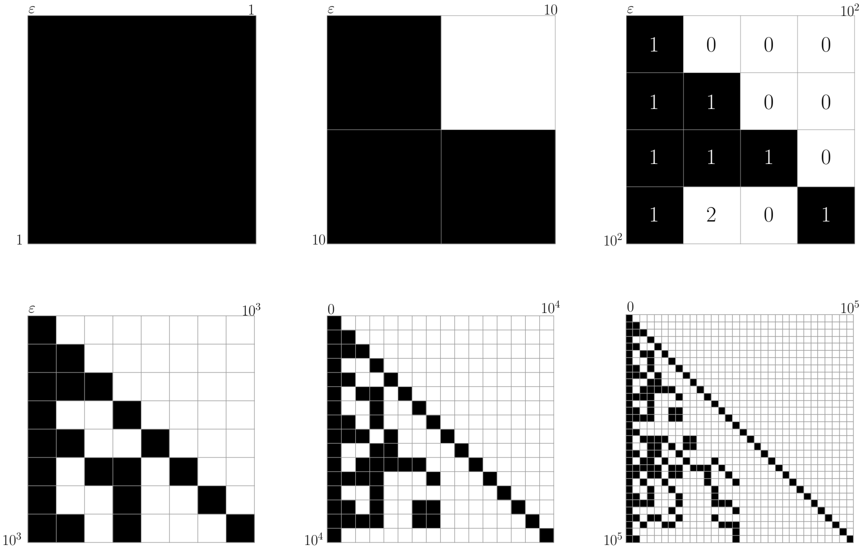
Same construction

- Grid: first 2^n rows and columns of P_2
- Color each square in
  - white if $\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \equiv 0 \bmod 2$
  - black if $\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \equiv 1 \bmod 2$
- Normalize by a homothety of ratio $1/2^n$

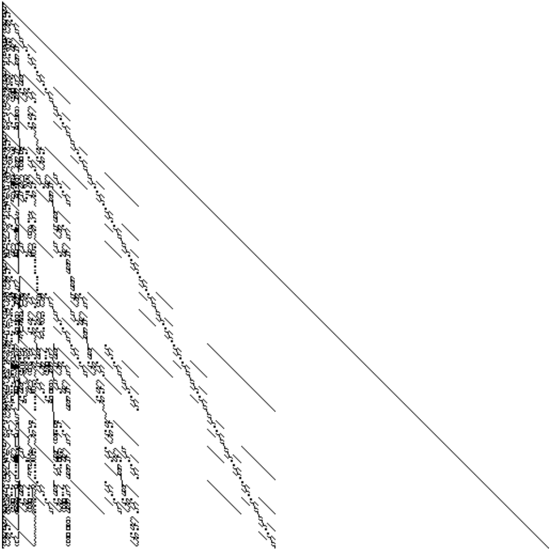
$\rightsquigarrow$ sequence of compact sets in $[0,1]^2$



The first six elements



The tenth element



Lines of different slopes: $1, 2, 4, 8, 16, \ldots$

The $(\star)$ condition

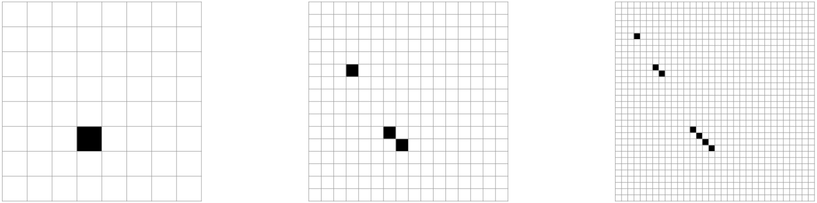
Definition

(u,v) satisfies $(\star)$ iff $u,v \neq \varepsilon$, $\binom{u}{v} \equiv 1 \bmod 2$, and $\binom{u}{v0} = 0 = \binom{u}{v1}$

Example: $(101, 11)$ satisfies $(\star)$ $\binom{101}{11} = 1$ $\binom{101}{110} = 0 = \binom{101}{111}$

Completion lemma: (u,v) satisfies $(\star) \Rightarrow (u0, v0), (u1, v1)$ satisfy $(\star)$

Example: $(101, 11)$ satisfies $(\star)$

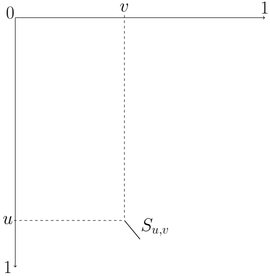


$\rightsquigarrow$ creation of a segment of slope 1

endpoint $(\frac{3}{8}, \frac{5}{8}) = (\frac{\mathrm{rep}_2(11)}{2^3}, \frac{\mathrm{rep}_2(101)}{2^3})$ length $\frac{\sqrt{2}}{2^3}$

Segments of slope 1

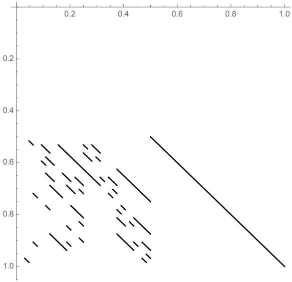
$(\star)$ describes lines of slope 1



(u, v) satisfying $(\star)$

$\rightsquigarrow$ closed segment $S_{u,v}$

- slope 1
- length $\frac{\sqrt{2}}{2^{|u|}}$
- origin $A_{u,v} = \left(\frac{\text{rep}_2(v)}{2^{|u|}}, \frac{\text{rep}_2(u)}{2^{|u|}}\right)$

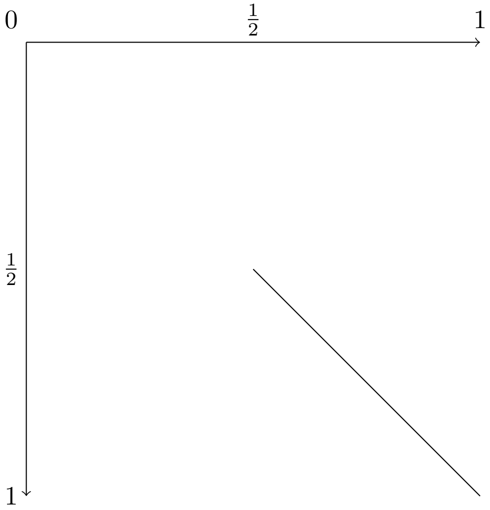


Definition: compact set $\subseteq$ those lines

$$\mathcal{A}_0 = \overline{\bigcup_{\substack{(u,v) \\ \text{satisfying } (\star)}} S_{u,v}} \subset [0,1]^2$$

Modifying the slope

with maps $c\colon (x,y) \mapsto (\frac{x}{2}, \frac{y}{2})$ and $h\colon (x,y) \mapsto (x, 2y)$



Example: $(1,1)$ satisfies $(\star)$

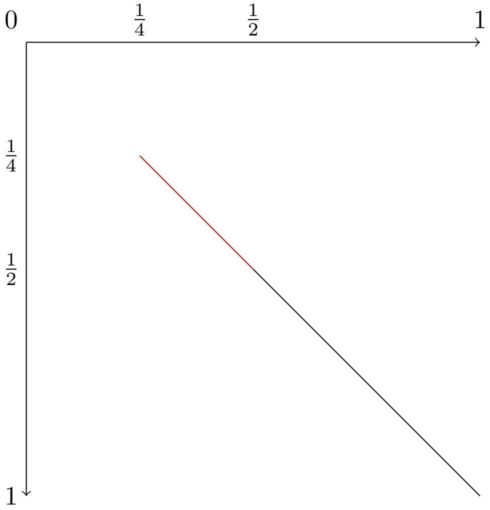
segment $S_{1,1}$
endpoint $(\frac{1}{2}, \frac{1}{2})$
length $\frac{\sqrt{2}}{2}$

Modifying the slope

with maps $c\colon (x,y) \mapsto (\frac{x}{2}, \frac{y}{2})$ and $h\colon (x,y) \mapsto (x, 2y)$

Example: $(1,1)$ satisfies $(\star)$

segment $S_{1,1}$
endpoint $(\frac{1}{2}, \frac{1}{2})$
length $\frac{\sqrt{2}}{2}$

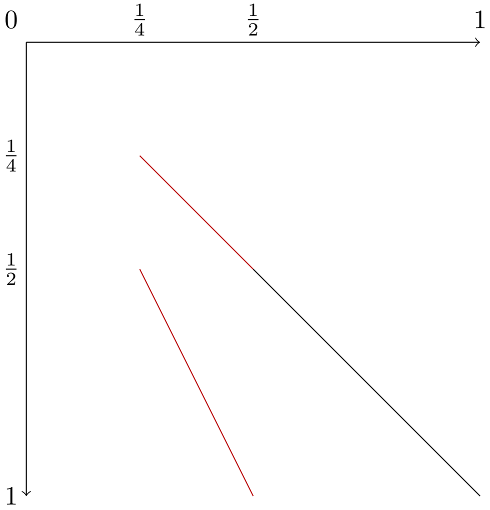


Modifying the slope

with maps $c\colon (x,y) \mapsto (\frac{x}{2}, \frac{y}{2})$ and $h\colon (x,y) \mapsto (x, 2y)$

Example: $(1,1)$ satisfies $(\star)$

segment $S_{1,1}$
endpoint $(\frac{1}{2}, \frac{1}{2})$
length $\frac{\sqrt{2}}{2}$

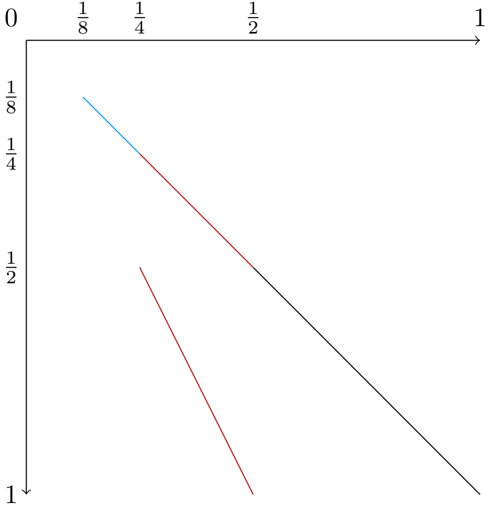


Modifying the slope

with maps $c\colon (x,y) \mapsto (\frac{x}{2}, \frac{y}{2})$ and $h\colon (x,y) \mapsto (x, 2y)$

Example: $(1,1)$ satisfies $(\star)$

segment $S_{1,1}$
endpoint $(\frac{1}{2}, \frac{1}{2})$
length $\frac{\sqrt{2}}{2}$

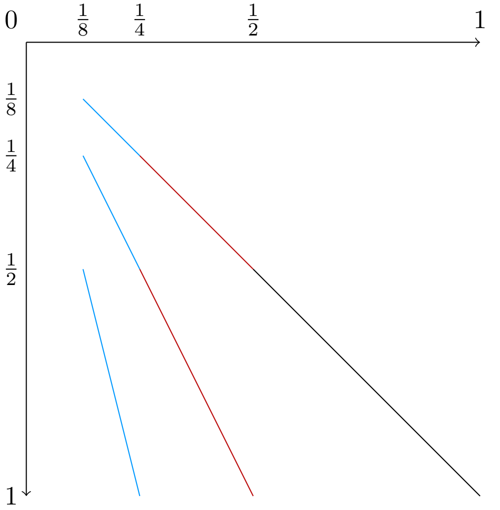


Modifying the slope

with maps $c\colon (x,y) \mapsto (\frac{x}{2}, \frac{y}{2})$ and $h\colon (x,y) \mapsto (x, 2y)$

Example: $(1,1)$ satisfies $(\star)$

segment $S_{1,1}$
endpoint $(\frac{1}{2}, \frac{1}{2})$
length $\frac{\sqrt{2}}{2}$

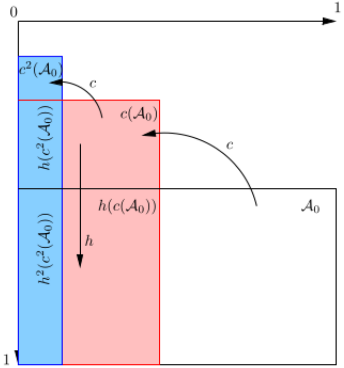


Definition: compact set $\subseteq$ lines of slopes 1, 2, 2^2, …, 2^n

$c\colon (x,y) \mapsto (\frac{x}{2}, \frac{y}{2})$

$h\colon (x,y) \mapsto (x, 2y)$

$$\mathcal{A}_n = \bigcup_{0 \leq j \leq i \leq n} h^j(c^i(\mathcal{A}_0))$$



$(\mathcal{A}_n)_{n\geq 0}$

increasingly nested

bounded union

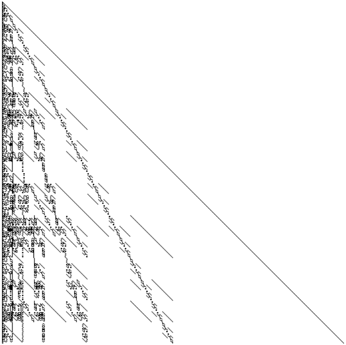
$\Rightarrow$ converges to the compact set

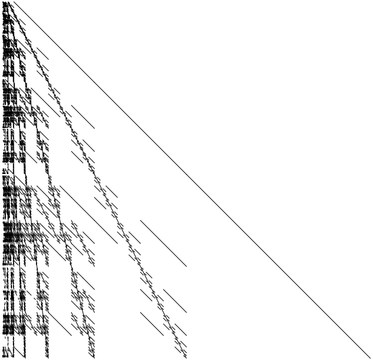
$$\mathcal{L} = \overline{\bigcup_{n\geq 0} \mathcal{A}_n}$$

(w.r.t. the Hausdorff distance)

Theorem (Leroy, Rigo, S., 2016)

The sequence of compact sets derived from P_2 converges to $\mathcal{L}$ (w.r.t. the Hausdorff distance).





($\star$) "simple" combinatorial characterization of $\mathcal{L}$

Extension modulo p

previously even/odd coefficients

Theorem (Lucas, 1878)

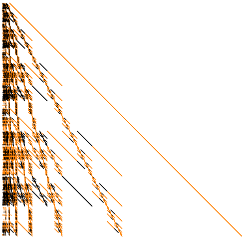
Let p be a prime number.

$$\left.\begin{array}{l} m = m_k p^k + \cdots + m_1 p + m_0 \\ n = n_k p^k + \cdots + n_1 p + n_0 \end{array}\right\} \Rightarrow \binom{m}{n} \equiv \prod_{i=0}^{k} \binom{m_i}{n_i} \bmod p$$

Theorem (Leroy, Rigo, S., 2016)

Let p be a prime and $0 < r < p$.
For binomial coeff. congruent to $r \bmod p$, the sequence derived from P_p converges to a compact set $\mathcal{L}_{p,r}$ (w.r.t. the Hausdorff distance).

Example: $\mathcal{L}_{3,1} \cup \mathcal{L}_{3,2}$



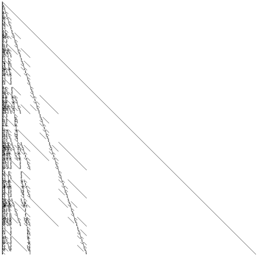
Generalization (S., 2019)

$\beta \in \mathbb{R}_{>1}$ Parry number
U canonical Bertrand NS
generalized Pascal's triangle P in U
compact: first $U(n)$ rows and col. of P

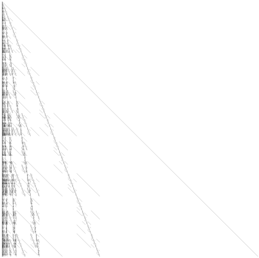
- $(\star')$ condition
- completion lemma with $(\star')$
- segments of slope 1
- compact set $\mathcal{A}_0 = \overline{\bigcup_{\substack{(u,v) \\ \text{satis.}(\star')}} S_{u,v}}$
- $c\colon (x,y) \mapsto (\frac{x}{\beta}, \frac{y}{\beta})$ $\quad h\colon (x,y) \mapsto (x, \beta y)$
- compact $\mathcal{A}_n = \bigcup_{0 \le j \le i \le n} h^j(c^i(\mathcal{A}_0)) \subseteq$ lines of slopes $1, \beta, \beta^2, \dots, \beta^n$
- $(\mathcal{A}_n)_{n \ge 0}$ converges to $\mathcal{L} = \overline{\bigcup_{n \ge 0} \mathcal{A}_n}$ (w.r.t. the Hausdorff dist.)
- modulo any prime p

Examples (mod 2)

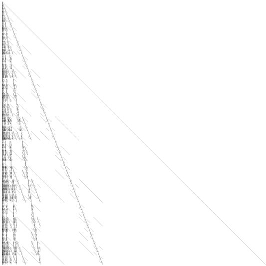
3



golden ratio φ

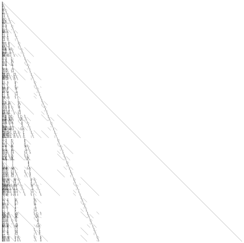


φ^2



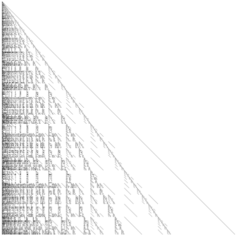
$\beta_1 \approx 2.47098$

dom. root of $X^4 - 2X^3 - X^2 - 1$



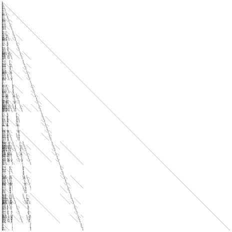
$\beta_2 \approx 1.38028$

dom. root of $X^4 - X^3 - 1$



$\beta_3 \approx 2.80399$

dom. root of $X^4 - 2X^3 - 2X^2 - 2$



Application 2: String attractors

Definition

Definition (Kempa, Prezza, 2018)

Let $w = w_0 w_1 \cdots w_{n-1}$ be a word.
A string attractor of w is a set $S \subseteq \{0, 1, \ldots, n-1\}$ s.t. every factor $x \neq \varepsilon$ of w has an occurrence in w that crosses an element of S.

Example: `alfalfa` $S = \{2, 3, 4\}$

length	factors	covered?
1	`a l f`	✓
2	`al lf fa`	✓
3	`alf lfa fal`	✓
4	`alfa lfal falf`	✓
5	`alfal lfalf falfa`	✓
6	`alfalf lfalfa`	✓
7	`alfalfa`	✓

no smaller S

Literature

- data compression (Kempa, Prezza, 2018)
- computational complexity theory
  NP-hard problem: find a smallest string attractor of a word
  (Kempa, Prezza, 2018)
- combinatorics on words
  - combinatorial properties, some families of infinite words
    (Mantaci, Restivo, Romana, Rosone, Sciortino, 2021)
  - automatic sequences (Schaeffer, Shallit, 2022+) (Gheeraert, Romana, S., 2024+)
  - Thue-Morse word
    (Kutsukake, Matsumoto, Nakashima, Inenaga, Bannai, Takeda, 2020) (Dolce, 2023)
  - Sturmian words
    (Mantaci, Restivo, Romana, Rosone, Sciortino, 2021) (Restivo, Romana, Sciortino, 2022)
  - episturmian words (Dvořáková, 2024)
  - Tribonacci word and k-bonacci-like words
    (Schaeffer, Shallit, 2022) (Gheeraert, Romana, S., 2023)
    (Cassaigne, Gheeraert, Restivo, Romana, Sciortino, S., 2024+)
  - Rote sequences (Dvořáková, Hendrychová, 2023+)
  - bi-infinite words (Béaur, Gheeraert, Hellouin de Menibus, 2024+)

A way to generate infinite words

Definition

A morphism is a map f s.t. $f(uv) = f(u)f(v)$ for all words u, v.
If $f(0) = 0x$ with $f^n(x) \neq \varepsilon \ \forall\, n$, the fixed point of f is

$$f^\omega(0) = \lim_{n \to +\infty} f^n(0).$$

Examples:

- $f: 0 \mapsto 01, 1 \mapsto 10$

$0 \quad f(0) = 01 \quad f^2(0) = 0110 \quad f^3(0) = 01101001 \quad f^4(0) = 0110100110010110$

Thue–Morse word $f^\omega(0) = 0110100110010110100101 10\cdots$

- $f: 0 \mapsto 01, 1 \mapsto 0$

$0 \quad f(0) = 01 \quad f^2(0) = 010 \quad f^3(0) = 01001 \quad f^4(0) = 01001010$

Fibonacci word $f^\omega(0) = 0100101001001010 01\cdots$

Starting point

Question: for $\mathbf{x}$, $\exists$ NS U s.t. s.a. of the prefixes of $\mathbf{x}$ may be described with U?

Example: Fibonacci word $f^\omega(0) = 010010100100101001\cdots$

n	length-n prefix of $f^\omega(0)$	string attractor
1	$\underline{0}$	$\{1\}$
2	$\underline{01}$	$\{1, 2\}$
3	$\underline{01}0$	$\{1, 2\}$
4	$0\underline{10}0$	$\{2, 3\}$
5	$0\underline{10}01$	$\{2, 3\}$
6	$0\underline{10}010$	$\{2, 3\}$
7	$01\underline{0}0\underline{1}01$	$\{3, 5\}$
$\vdots$	$\vdots$	$\vdots$
11	$01\underline{0}0\underline{1}010010$	$\{3, 5\}$
12	$0100\underline{1}01\underline{0}0100$	$\{5, 8\}$

Fibonacci/Zeckendorf NS

A family of words

Definition

Let $k \geq 2$ be an integer and set $c = (\underbrace{c_1}_{\geq 1}, \underbrace{c_2}_{\geq 0}, \ldots, \underbrace{c_{k-1}}_{\geq 0}, \underbrace{c_k}_{\geq 1})$.

$$f_c\colon \begin{cases} 0 \mapsto 0^{c_1} 1 \\ 1 \mapsto 0^{c_2} 2 \\ \vdots \\ k-2 \mapsto 0^{c_{k-1}}(k-1) \\ k-1 \mapsto 0^{c_k} \end{cases} \quad \text{and} \quad \mathbf{x}_c = f_c^{\omega}(0)$$

Example: $k = 2 \quad c = (1,1) \rightsquigarrow \mathbf{x}_c =$ Fibonacci word

Automaticity

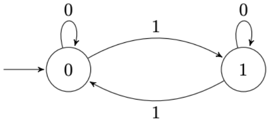
Definition

Let U be a NS.

$\mathbf{x} = (\mathbf{x}(n))_{n\geq 0}$ is U-automatic if $\exists$ DFAO $\mathcal{A}$ s.t. $\mathcal{A}$: $\mathrm{rep}_U(n) \rightsquigarrow \mathbf{x}(n)$.

Example: Thue–Morse $\mathbf{x} = 0110100110010110\cdots$ 2-automatic

n	0	1	2	3	4	5
$\mathrm{rep}_2(n)$	ε	1	10	11	100	101
$\mathbf{x}(n)$	0	1	1	0	1	0



Observation (Gheeraert, Romana, S., 2024+)

$c_0 \cdots c_{k-2}(c_{k-1} - 1)$ max. among its conjugates

$\Rightarrow \exists\, \beta$ simple-Parry s.t. $\mathbf{x}_c$ is U-automatic in the canonical NS U of β

Examples: φ golden ratio canonical NS = Fibonacci/Zeckendorf

- $k = 2$ $c = (1, 1)$ $\mathbf{x}_c = 0100101001\cdots$ Fib.-automatic
- $k = 4$ $c = (1, 0, 1, 1)$ $\mathbf{x}_c = 0120301001\cdots$ Fib.-automatic

Finding s.a.

Hypotheses: $c_0 \cdots c_{k-2}(c_{k-1}-1)$ max. among its conjugates

$\Rightarrow$ correspoding simple-Parry β and canonical NS $U = (U(n))_{n\geq 0}$

$$S_{-1} = \emptyset \qquad S_n = \begin{cases} \{U(0), \dots, U(n)\} & 0 \leq n \leq k-1 \\ \{U(n-k+1), \dots, U(n)\} & n \geq k \end{cases}$$

Theorem (Gheeraert, Romana, S., 2024+)

- $m \in [U(n), L_n] \quad \Rightarrow \quad S_{n-1} \cup \{U(n)\}$ s.a. of $\mathbf{x}_c[0..m)$
- $m \in [\ell_n, L_n] \quad \Rightarrow \quad S_n$ s.a. of $\mathbf{x}_c[0..m)$

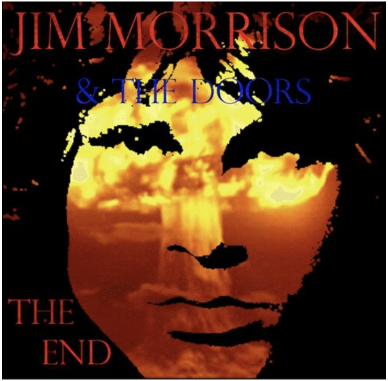
$$\ell_n = \begin{cases} U(n) & 0 \leq n \leq k-1 \\ U(n) + U(n-k+1) - U(n-k) - 1 & n \geq k \end{cases}$$

L_n = length of the longest common prefix between x_c and $(f_c^n(0))^\omega$

Corollary (Gheeraert, Romana, S., 2024+)

size of a smallest s.a. of $\mathbf{x}_c[0..m) \in [k, k+1] \quad \forall\, m \gg$

Conclusion



$\beta \in \mathbb{R}_{>1}$ Parry number
U canonical Bertrand NS

• generalized Pascal's and Sierpiński's triangles to words of U with a combinatorial description

• string attractors of U-automatic fixed points in terms of elements of U

References I

J.-P. Allouche and J. Shallit.
Automatic sequences. Theory, applications, generalizations.
Cambridge: Cambridge University Press, 2003.

J. Berstel and D. Perrin.
The origins of combinatorics on words.
Eur. J. Comb., 28(3):996–1022, 2007.

A. Bertrand-Mathis.
Comment écrire les nombres entiers dans une base qui n'est pas entière.
Acta Math. Hung., 54(3-4):237–241, 1989.

P. Béaur, F. Gheeraert, and B. Hellouin de Menibus.
String attractors and bi-infinite words.
Preprint available at https://arxiv.org/abs/2403.13449, 2024.

J. Cassaigne, F. Gheeraert, A. Restivo, G. Romana, M. Sciortino, and M. Stipulanti.
New string attractor-based complexities on infinite words.
Preprint available at https://hdl.handle.net/2268/309672, 2024.

É. Charlier, C. Cisternino, and M. Stipulanti.
A full characterization of Bertrand numeration systems.
In Developments in language theory. 26th international conference, DLT 2022, Tampa, FL, USA, May 9–13, 2022. Proceedings, pages 102–114. Cham: Springer, 2022.

F. Dolce.
String attractors for factors of the Thue-Morse word.
In Combinatorics on words. 14th international conference, WORDS 2023, Umeå, Sweden, June 12–16, 2023. Proceedings, pages 117–129. Cham: Springer, 2023.

References II

L. Dvořáková.
String attractors of episturmian sequences.
Theor. Comput. Sci., 986:5, 2024.
Id/No 114341.

L. Dvořáková and V. Hendrychová.
String attractors of Rote sequences.
Preprint available at https://arxiv.org/abs/2308.00850, 2023.

A. S. Fraenkel.
Systems of numeration.
Computer arithmetic, Proc. 6th Symp., Aarhus/Den. 37-42, 1983.

F. Gheeraert, G. Romana, and M. Stipulanti.
String attractors of fixed points of k-bonacci-like morphisms.
In Combinatorics on words. 14th international conference, WORDS 2023, Umeå, Sweden, June 12–16, 2023. Proceedings, pages 192–205. Cham: Springer, 2023.

F. Gheeraert, G. Romana, and M. Stipulanti.
String attractors of some simple-Parry automatic sequences.
Preprint available at https://arxiv.org/abs/2302.13647, 2024.

D. Kempa and N. Prezza.
At the roots of dictionary compression: string attractors.
In Proceedings of the 50th annual ACM SIGACT symposium on theory of computing, STOC '18, Los Angeles, CA, USA, June 25–29, 2018, pages 827–840. New York, NY: Association for Computing Machinery (ACM), 2018.

References III

K. Kutsukake, T. Matsumoto, Y. Nakashima, S. Inenaga, H. Bannai, and M. Takeda.
On repetitiveness measures of Thue-Morse words.
In String processing and information retrieval. 27th international symposium, SPIRE 2020, Orlando, FL, USA, October 13–15, 2020. Proceedings, pages 213–220. Cham: Springer, 2020.

J. Leroy, M. Rigo, and M. Stipulanti.
Generalized Pascal triangle for binomial coefficients of words.
Adv. Appl. Math., 80:24–47, 2016.

E. Lucas.
The theory of simply periodic functions.
Am. J. Math., 1:184–251, 1878.

S. Mantaci, A. Restivo, G. Romana, G. Rosone, and M. Sciortino.
A combinatorial view on string attractors.
Theor. Comput. Sci., 850:236–248, 2021.

W. Parry.
On the β-expansions of real numbers.
Acta Math. Acad. Sci. Hung., 11:401–416, 1960.

A. Restivo, G. Romana, and M. Sciortino.
String attractors and infinite words.
In Latin 2022: theoretical informatics. 15th Latin American symposium, Guanajuato, Mexico, November 7–11, 2022. Proceedings, pages 426–442. Cham: Springer, 2022.

References IV

M. Rigo and A. Maes.
More on generalized automatic sequences.
J. Autom. Lang. Comb., 7(3):351–376, 2002.

L. Schaeffer and J. Shallit.
String attractor for automatic sequences.
Preprint available at https://arxiv.org/abs/2012.06840, 2022.

M. Stipulanti.
Convergence of Pascal-like triangles in Parry-Bertrand numeration systems.
Theor. Comput. Sci., 758:42–60, 2019.

F. von Haeseler, H.-O. Peitgen, and G. Skordev.
Pascal's triangle, dynamical systems and attractors.
Ergodic Theory Dyn. Syst., 12(3):479–486, 1992.