Regularity of positional numeration systems without a dominant root

Émilie Charlier
Joint work with Savinien Kreczman

Département de mathématique, Université de Liège, Belgique

Workshop on Uniform Distribution of Sequences, Vienne
23 avril 2025

Positional numeration systems

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

A natural number x is represented by the finite word

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

obtained from the greedy algorithm:

$$x = \sum_{n=0}^{\ell-1} a_n U_n.$$

A description of the numeration language

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

strongly depends on the base sequence U.

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?

Linear systems

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

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

A way to see this is:

- Note that, for each n, U_n is the number of words of length n in L_U.
- This implies that the formal series
$$S = \sum_{n \geq 0} U_n X^n$$
is $\mathbb{Z}$-rational, i.e., $S = \frac{P}{Q}$ for polynomials $P, Q \in \mathbb{Z}[X]$ with $Q(0) = 1$.
- This, in turn, implies that the sequence $(U_n)_{n \geq 0}$ satisfies a linear recurrence relation over $\mathbb{Z}$.

Hollander's study for dominant root systems

This question was studied by Hollander in 1998 in the case of positional systems $U = (U_n)_{n\geq 0}$ with a dominant root, i.e., such that the limit $\lim_{n\to\infty} \frac{U_n}{U_{n-1}}$ exists and is greater than 1.

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

Proposition (Hollander 1998)

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

Some intuition for these two arguments

Proposition (Hollander 1998)

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

- The implication
$$L \text{ regular} \implies \mathrm{Max}(L) \text{ regular}$$
is true for any language. See [Shallit 1994] for a nice short proof.
- The other direction does not work in general.
- The point here is to prove that the converse is true for numeration languages. This is due to the following property of numeration languages:
$$w \in L_U \iff \text{for all } s \in \mathrm{Suff}(w),\ s \leq_{\mathrm{lex}} \mathrm{rep}_U(U_{|s|} - 1).$$

Proposition (Hollander 1998)

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

Here, the key observation is:

- If $d_\beta(1)$ is infinite, then
$$\lim_{n\to\infty} \mathrm{rep}_U(U_n - 1) = d_\beta(1).$$
- If $d_\beta(1) = d_1 \cdots d_\ell$ with $d_\ell \neq 0$, then for all lengths L and all large enough indices n, there exists $j \geq 0$ such that
$$\mathrm{Pref}_L(\mathrm{rep}_U(U_n - 1)) = \mathrm{Pref}_L(\mathbf{w}_j)$$
where $\mathbf{w}_j = (d_1 \cdots d_{\ell-1}(d_\ell - 1))^j d_1 \cdots d_\ell 0^\omega$.

We will refer to these words $\mathbf{w}_j$ as the "intermediate" β-representations of 1.

β-polynomials

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

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

is called a β-polynomial.
For q, r minimal, we get the canonical β-polynomial, simply denoted P_β.

Note that in the case where $d_\beta(1) = d_1 \cdots d_\ell$ with $d_\ell \neq 0$, then as $d_\beta^*(1) = (d_1 \cdots d_{\ell-1}(d_\ell - 1))^\omega$, we simply get

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

Theorem (Hollander 1998)

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

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

Getting rid of the dominant root condition

First step: Exploit the positiveness of the generating series.

- If L_U is regular, then the series $\sum_{n\geq 0} U_n X^n$ is $\mathbb{N}$-rational (and not just $\mathbb{Z}$-rational).
- By a result of Soittola from 1976, we know that if a $\mathbb{Z}$-rational series has nonnegative coefficients and a dominating eigenvalue, then it is $\mathbb{N}$-rational.
- Conversely, we can derive from another result of Soittola that if a series $\sum_{i\geq 0} s_n X^n$ is $\mathbb{N}$-rational, then there exists some $p \geq 1$ such that for each $i \in \{0, \dots, p-1\}$, the limit

$$\lim_{n\to+\infty} \frac{s_{pn-i}}{s_{pn-i-1}}$$

exists.
- Consequently, if L_U is regular, then we can associate a p-tuple of real numbers $(\beta_0, \dots, \beta_{p-1})$ where for each i,

$$\beta_i := \lim_{n\to+\infty} \frac{U_{pn-i}}{U_{pn-i-1}}.$$

Alternate bases

Second step: Introduce alternate bases and link them with maximal words of each length in L_U.

- For a tuple $B = (\beta_0, \ldots, \beta_{p-1})$, we consider representations of real numbers of the form
$$x = \frac{a_0}{\beta_0} + \frac{a_1}{\beta_0\beta_1} + \frac{a_2}{\beta_0\beta_1\beta_2} + \cdots$$
where $\beta_n := \beta_{n \bmod p}$ for all $n \geq 0$.
- We use a greedy algorithm to define greedy B-expansions of real numbers $d_B(x)$.
- We define the quasi-greedy B-expansion of 1 as $d_B^*(1) = \lim_{x \to 1^-} d_B(x)$.
- We get a Parry-kind characterization of allowable sequences: a sequence $a_0 a_1 a_2 \cdots$ is the B-expansion of a real number in $[0, 1)$ if and only if for all $n \geq 0$, one has $a_n a_{n+1} a_{n+2} \cdots <_{\text{lex}} d^*_{S^n(B)}(1)$ where $S^n(B)$ is the shifted base $(\beta_n, \beta_{n+1}, \ldots)$.

[C-Cisternino 2021]

Lemma

Let $U = (U_n)_{n \geq 0}$ be a positional numeration system with an associated alternate base $(\beta_0, \ldots, \beta_{p-1})$. For all $i \in \{0, \ldots, p-1\}$, all lengths ℓ and all large enough indices n, there exists j such that

$$\mathrm{Pref}_\ell(\mathrm{rep}_U(U_{pn-i} - 1)) = \mathrm{Pref}_\ell(\mathbf{w}_{i,j})$$

where the infinite words $\mathbf{w}_{i,j}$ are $(\beta_i, \ldots, \beta_{i+p-1})$-representations of 1 which are "intermediate" between the greedy and the quasi-greedy one.

Proposition

Let U be a positional numeration system with a regular numeration language L_U, and let $(\beta_0, \ldots, \beta_{p-1})$ be an associated alternate base.
Then for each $i \in \{0, \ldots, p-1\}$, the quasi-greedy expansion $d^*_{S^i(B)}(1)$ is ultimately periodic.

Such alternate bases are called Parry.

[C-Kreczman 2025+]

Intermediate representations of 1

Let $U = (U_n)_{n\geq 0}$ be defined by $U_{n+10} = 16U_{n+5} - 9U_n$ for $n \geq 0$ and the following initial conditions.

n	0	1	2	3	4	5	6	7	8	9
U_n	1	2	3	6	10	19	29	48	96	151

Then for $i \in \{0, \dots, 4\}$, the limits

$$\beta_i := \lim_{n\to+\infty} \frac{U_{5n-i}}{U_{5n-i-1}}$$

exist, and can be effectively computed.

Set $B = (\beta_0, \dots, \beta_4)$.

We get the following greedy and quasi-greedy $S^i(B)$-expansions of 1:

i	$d_{S^i(B)}(1)$
0	1110^ω
1	$11000(10000)^\omega$
2	20^ω
3	110^ω
4	110^ω

i	$d^*_{S^i(B)}(1)$
0	$(11010)^\omega$
1	$11000(10000)^\omega$
2	$1(10110)^\omega$
3	$(10110)^\omega$
4	$1011000(10000)^\omega$

For $i = 0$, the intermediate representations are given by

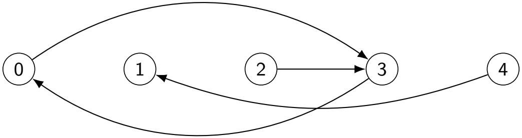
$$\mathbf{w}_{0,1} = 110 \cdot 110^\omega$$

$$\mathbf{w}_{0,2} = 110 \cdot 10 \cdot 1110^\omega$$

$$\mathbf{w}_{0,3} = 110 \cdot 10 \cdot 110 \cdot 110^\omega$$

$$\vdots$$

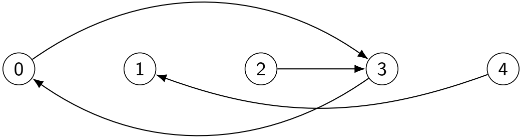
We encode the possible interactions of the remainders $i \in \{0, \ldots, p-1\}$ in a graph G:



Third step: Suppose that U is a positional numeration system with an associated Parry alternate base $(\beta_0, \ldots, \beta_{p-1})$. Then study the regularity of the sub-languages

$$\begin{aligned} L_{U,i} &:= \{w \in \mathrm{Max}(L_U) : |w| \equiv -i \pmod{p}\} \\ &= \{\mathrm{rep}_U(U_{pn-i} - 1) : n \geq 1\} \end{aligned}$$

by analyzing all possible interactions between remainders as encoded in the graph G.



Suppose that $d_{S^i(B)}(1) = t_{i,1} \cdots t_{i,q} \left(t_{i,q+1} \cdots t_{i,q+r}\right)^{\omega}$. Then we define a sequence $\Delta_{i,q,r}$ by

$$(\Delta_{i,q,r})_n = \left(U_{pn-i} - \sum_{s=1}^{q+r} t_{i,s} U_{pn-i-s}\right) - \left(U_{pn-i-r} - \sum_{s=1}^{q} t_{i,s} U_{pn-i-r-s}\right).$$

Case 1

Suppose that i has no outgoing edge in the graph G, i.e., $d_{S^i(B)}(1)$ is infinite.
The language $L_{U,i}$ is regular if and only if there exist some q, r such that the sequence $\Delta_{i,q,rp}$ is ultimately zero.

Suppose that $d_{S^i(B)}(1) = t_{i,1} \cdots t_{i,\ell}$. Then we define a sequence Δ_i by

$$(\Delta_i)_n = U_{pn-i} - \sum_{s=1}^{\ell} t_{i,s} U_{pn-i-s}.$$

Case 2

Suppose that there is a path from i to some vertex with no outgoing edge in the graph G, i.e., there exist $i_1 = i, i_2, \ldots, i_k$ such that

- $d_{S^{i_j}(B)}(1) = t_{i_j,1} \cdots t_{i_j,\ell_{i_j}}$ for all $j \in \{1, \ldots, k-1\}$
- $d_{S^{i_k}(B)}(1) = d^*_{S^{i_k}(B)}(1)$
- $d^*_{S^i(B)}(1) = \left(\prod_{j=1}^{k-1}(t_{i_j,1} \cdots t_{i_j,\ell_{i_j}-1}(t_{i_j,\ell_{i_j}} - 1))\right) d_{S^{i_k}(B)}(1)$.

Suppose that the languages L_{U,i_j} are regular for all $j \in \{2, \ldots, k\}$.
The language $L_{U,i}$ is regular if and only if the sequence Δ_i is ultimately periodic.

Case 3

Suppose that i belongs to a cycle in the graph G, i.e., that there exists $i_1 = i, i_2, \dots, i_k$ such that

- $d_{S^{i_j}(B)}(1) = t_{i_j,1} \cdots t_{i_j,\ell_{i_j}}$ for all $j \in \{1, \dots, k\}$
- $d^*_{S^i(B)}(1) = \left(\prod_{j=1}^{k} (t_{i_j,1} \cdots t_{i_j,\ell_{i_j}-1}(t_{i_j,\ell_{i_j}} - 1))\right)^{\omega}$.

The languages $L_{U,i_1}, \dots, L_{U,i_k}$ are all regular if and only if

- the sequences $\Delta_{i_1}, \dots, \Delta_{i_k}$ are all ultimately periodic $j \in \{1, \dots, k\}$, with a common period M such that Mp is a multiple of $\ell_{i_1} + \cdots + \ell_{i_k}$
- for all $j \in \{1, \dots, k\}$, the M sums

$$(\Delta_{i_j})_n + (\Delta_{i_{j+1}})_{n-m_{j,1}} + \cdots + (\Delta_{i_{j+mk-1}})_{n-m_{j,mk-1}}$$

are all ultimately nonnegative, where $m = \frac{Mp}{\ell_{i_1}+\cdots+\ell_{i_k}}$ and $m_{j,h} = \frac{i_j+\ell_{i_j}+\cdots+\ell_{i_{j+h-1}}-i_{j+h}}{p}$.

(Here, the indices j in i_j are considered modulo k).

Case 4

Suppose that there is a path from i to a cycle in the graph G, i.e., that there exists $i_1 = i, i_2, \ldots, i_k, i_{k+1}, \ldots, i_{k+k'}$ such that

- $d_{S^{i_j}(B)}(1) = t_{i_j,1} \cdots t_{i_j,\ell_{i_j}}$ for all $j \in \{1, \ldots, k+k'\}$
- $d^*_{S^i(B)}(1) = \left(\prod_{j=1}^{k}(t_{i_j,1} \cdots t_{i_j,\ell_{i_j}-1}(t_{i_j,\ell_{i_j}} - 1))\right) \cdot \left(\prod_{j=k+1}^{k+k'}(t_{i_j,1} \cdots t_{i_j,\ell_{i_j}-1}(t_{i_j,\ell_{i_j}} - 1))\right)^{\omega}$.

Suppose that the languages L_{U,i_j} are regular for all $j \in \{2, \ldots, k+k'\}$.
The language $L_{U,i}$ is regular if and only if "a complicated condition holds, but nice enough so that it can be effectively checked".

Summing up

- All in all, we obtain a characterization of positional numeration systems with a regular numeration language.
- This characterization is effective.
- It also fills the gap in Hollander's result in the case of a simple Parry dominant root β.

  Indeed, suppose that $d_\beta(1) = d_1 \cdots d_\ell$ and if U satisfies a linear recurrence relation of characteristic polynomial $(X^\ell - 1)P_{\beta,q,r}$ for some q, r.

  Define the sequence $\Delta = (\Delta_n)_{n \geq \ell}$ by $\Delta_n = U_n - d_1 U_{n-1} - \cdots - d_\ell U_{n-\ell}$. In this situation, this sequence Δ is ultimately periodic with period $r = m\ell$, which is a multiple of ℓ.

  We then get that the numeration language L_U is regular if and only if the ℓ sums

$$\Delta_n + \Delta_{n-\ell} + \cdots + \Delta_{n-(m-1)\ell}$$

  are ultimately nonnegative.

Thank you!