Regularity of the language of greedy numeration systems.

Savinien Kreczman

23 May 2023

Greedy numeration systems

Definition

Greedy numeration system :

- Sequence $(U_n)_{n\in\mathbb{N}}$ satisfying

$$(i)\ U_0 = 1 \quad (ii)\ U \text{ is increasing}$$

$$(iii)\ \exists C \in \mathbb{N},\ \forall n \in \mathbb{N},\ \frac{U_{n+1}}{U_n} \leq C.$$

Greedy numeration systems

Definition

Greedy numeration system :

- Sequence $(U_n)_{n\in\mathbb{N}}$ satisfying

$$(i)\ U_0 = 1 \quad (ii)\ U \text{ is increasing}$$

$$(iii)\ \exists C \in \mathbb{N},\ \forall n \in \mathbb{N},\ \frac{U_{n+1}}{U_n} \leq C.$$

- Map $\mathrm{rep} : \mathbb{N} \to \{0, \ldots, C-1\}^*$ given by the *greedy algorithm* :
  Given $n \in \mathbb{N}$, let M be such that $U_M \leq n < U_{M+1}$, set $r_M = n$. Then, if r_i is defined, set $a_i = \lfloor \frac{r_i}{U_i} \rfloor$ and $r_{i-1} = r_i - U_i a_i$.
  $\mathrm{rep}_U(n)$ is the word $a_M \ldots a_0$.

Some examples

- Sequence $U = (1, 10, 100, 1000, \ldots)$: usual decimal numeration system.

Some examples

- Sequence $U = (1, 10, 100, 1000, \ldots)$: usual decimal numeration system.
- Sequence $F = (1, 2, 3, 5, 8, 13, 21, \ldots)$: *Zeckendorf* numeration system.

Some examples

- Sequence $U = (1, 10, 100, 1000, \ldots)$: usual decimal numeration system.
- Sequence $F = (1, 2, 3, 5, 8, 13, 21, \ldots)$: *Zeckendorf* numeration system.
  Representation of 19 in this system : $19 = 13 + 5 + 1$, so $\text{rep}_F(19) = 101001$.

Language of the system

$$\mathrm{rep}_F(19) \quad = \quad 101001$$

Language of the system

$$\begin{array}{rcr} & & \vdots \\ \text{rep}_F(19) & = & 101001 \\ \text{rep}_F(20) & = & 101010 \\ \text{rep}_F(21) & = & 1000000 \\ & & \vdots \end{array}$$

Language of the system

$$\begin{array}{rcr} & & \vdots \\ \mathrm{rep}_F(19) & = & 101001 \\ \mathrm{rep}_F(20) & = & 101010 \\ \mathrm{rep}_F(21) & = & 1000000 \\ & & \vdots \end{array}$$

Definition

$\mathrm{rep}_U(\mathbb{N})$ is the *language* L_U of the numeration U.

$$L_F = \{0,1\}^* \setminus (0\{0,1\}^* \cup \{0,1\}^* 11 \{0,1\}^*)$$

Language of the system

$$U_n = 2^n \longrightarrow L_U = \{0,1\}^* \setminus 0\{0,1\}^*$$
$$U_n = F_n \longrightarrow L_U = \{0,1\}^* \setminus (0\{0,1\}^* \cup \{0,1\}^* 11 \{0,1\}^*)$$

Language of the system

$U_n = 2^n \longrightarrow L_U = \{0,1\}^* \setminus 0\{0,1\}^*$
$U_n = F_n \longrightarrow L_U = \{0,1\}^* \setminus (0\{0,1\}^* \cup \{0,1\}^* 11 \{0,1\}^*)$
$U_n = n^2 \longrightarrow L_U$ is not regular.

Language of the system

$U_n = 2^n \longrightarrow L_U = \{0,1\}^* \setminus 0\{0,1\}^*$
$U_n = F_n \longrightarrow L_U = \{0,1\}^* \setminus (0\{0,1\}^* \cup \{0,1\}^*11\{0,1\}^*)$
$U_n = n^2 \longrightarrow L_U$ is not regular.

Question

What conditions on U are necessary or sufficient for the regularity of L_U?

Language of the system

$$U_n = 2^n \longrightarrow L_U = \{0,1\}^* \setminus 0\{0,1\}^*$$
$$U_n = F_n \longrightarrow L_U = \{0,1\}^* \setminus (0\{0,1\}^* \cup \{0,1\}^* 11\{0,1\}^*)$$
$U_n = n^2 \longrightarrow L_U$ is not regular.

Question

What conditions on U are necessary or sufficient for the regularity of L_U?

Theorem (Shallit, 1994)

If L_U is regular, then U is a linear recurrent sequence.

Language of the system

$U_n = 2^n \longrightarrow L_U = \{0,1\}^* \setminus 0\{0,1\}^*$
$U_n = F_n \longrightarrow L_U = \{0,1\}^* \setminus (0\{0,1\}^* \cup \{0,1\}^* 11\{0,1\}^*)$
$U_n = n^2 \longrightarrow L_U$ is not regular.

Question

What conditions on U are necessary or sufficient for the regularity of L_U?

Theorem (Shallit, 1994)

If L_U is regular, then U is a linear recurrent sequence.

We further restrict ourselves to the case of the *dominant root condition*, where

$$\lim_{n\to\infty} \frac{U_{n+1}}{U_n} = \beta > 1$$

Maximal words

Definition

$$\text{Maxlg}\, L = \{\text{rep}_U(U_n - 1) | n \in \mathbb{N}.\}$$

This set contains the greatest word of each length for the lexicographic order.

Maximal words

Definition

$$\operatorname{Maxlg} L = \{\operatorname{rep}_U(U_n - 1) | n \in \mathbb{N}.\}$$

This set contains the greatest word of each length for the lexicographic order.

Proposition

L_U is regular iff $\operatorname{Maxlg} L_U$ is regular.

$\operatorname{Maxlg} L_U$ is slender, so it is regular iff it can be expressed as a finite union of languages xy^*z.

Maximal words

Definition

$$\operatorname{Maxlg} L = \{\operatorname{rep}_U(U_n - 1) | n \in \mathbb{N}.\}$$

This set contains the greatest word of each length for the lexicographic order.

Proposition

L_U is regular iff $\operatorname{Maxlg} L_U$ *is regular.*

$\operatorname{Maxlg} L_U$ is slender, so it is regular iff it can be expressed as a finite union of languages xy^*z.
Example: the language for Zeckendorf numeration is

$$L_F = \{\varepsilon, 1, 10, 100, 101, 1000, 1001, 1010, 10000, \dots\}$$

so

$$\operatorname{Maxlg} L_F = \{\varepsilon, 1, 10, 101, 1010, \dots\}$$

β-representation

Write $1 = \sum_{i=1}^{\infty} \frac{d_i}{\beta^i}$ with a greedy algorithm. The infinite word $d_1 d_2 \ldots$ is the β-*representation* of 1, written $d_\beta(1)$.

β-representation

Write $1 = \sum_{i=1}^{\infty} \frac{d_i}{\beta^i}$ with a greedy algorithm. The infinite word $d_1 d_2 \dots$ is the *β-representation* of 1, written $d_\beta(1)$.
If the β-representation of 1 ends with a tail of zeroes, we say that it is *finite* and we omit this tail. We write $d_\beta(1) = d_1 \dots d_l$.
This representation gives information on the prefixes of words in $\mathrm{MaxIg}\, L_U$.

β-representation

Write $1 = \sum_{i=1}^{\infty} \frac{d_i}{\beta^i}$ with a greedy algorithm. The infinite word $d_1 d_2 \ldots$ is the *β-representation* of 1, written $d_\beta(1)$.
If the β-representation of 1 ends with a tail of zeroes, we say that it is *finite* and we omit this tail. We write $d_\beta(1) = d_1 \ldots d_l$.
This representation gives information on the prefixes of words in $\mathrm{MaxIg}\, L_U$.
Example : with β equal to the golden ratio ϕ, we have $1 = \frac{1}{\phi} + \frac{1}{\phi^2}$. The ϕ-representation of 1 is 11.

β-polynomial

If $d_\beta(1) = d_1 \ldots d_l$, we set

$$b(x) = x^l - d_1 x^{l-1} - d_2 x^{l-2} - \ldots - d_l.$$

This is the *canonical β-polynomial*.

β-polynomial

If $d_\beta(1) = d_1 \dots d_l$, we set

$$b(x) = x^l - d_1 x^{l-1} - d_2 x^{l-2} - \dots - d_l.$$

This is the *canonical β-polynomial*.
The set of *extended β-polynomials* is the set

$$\{b(x)x^k(1 + x^l + x^{2l} + \dots + x^{(a-1)l}) | k \in \mathbb{N}, a \in \mathbb{N}_0\}.$$

Canonical and extended β-polynomials can also be defined when $d_\beta(1)$ is infinite periodic.

β-polynomial

If $d_\beta(1) = d_1 \ldots d_l$, we set

$$b(x) = x^l - d_1 x^{l-1} - d_2 x^{l-2} - \ldots - d_l.$$

This is the *canonical β-polynomial*.
The set of *extended β-polynomials* is the set

$$\{b(x)x^k(1 + x^l + x^{2l} + \ldots + x^{(a-1)l})|k \in \mathbb{N}, a \in \mathbb{N}_0\}.$$

Canonical and extended β-polynomials can also be defined when $d_\beta(1)$ is infinite periodic.
Example : with β equal to the golden ratio ϕ, we have $b(x) = x^2 - x - 1$. All the polynomials

$$(x^2 - x - 1)x^k(1 + x^2 + \ldots + x^{2a-2})$$

are extended ϕ-polynomials.

Theorem (Hollander 1998)

<table>
<tr><th>$d_\beta(1)$ infinite periodic</th><th>$d_\beta(1) = d_1 \cdots d_l$</th><th>$d_\beta(1)$ infinite aperiodic</th></tr>
<tr><td rowspan="2">U doesn't follow an extended β-polynomial</td><td>U doesn't follow an extended β-polynomial multiplied by a divisor of $x^l - 1$</td><td rowspan="3"></td></tr>
<tr><td>U follows an extended β-polynomial multiplied by a divisor of $x^l - 1$</td></tr>
<tr><td>U follows an extended β-polynomial</td><td>U follows an extended β-polynomial</td></tr>
</table>

Regular language

Non-regular language

A central gray area

When $d_\beta(1)$ is finite and U satisfies a β-polynomial times a divisor of $x^l - 1$ without satisfying a β-polynomial, regularity may depend on the initial conditions.

A central gray area

When $d_\beta(1)$ is finite and U satisfies a β-polynomial times a divisor of $x^l - 1$ without satisfying a β-polynomial, regularity may depend on the initial conditions.

Example

Let $\beta = 3$, U be the sequence satisfying $(x-3)(x-1)$ with initial conditions $(1, 4)$

. Then the lexicographically greatest words at each length are given below.

n	$\operatorname{Maxlg} U \cap \Sigma^n$
1	3
2	30
3	300
4	3000
5	30000
6	300000

A central gray area

When $d_\beta(1)$ is finite and U satisfies a β-polynomial times a divisor of $x^l - 1$ without satisfying a β-polynomial, regularity may depend on the initial conditions.

Example

Let $\beta = 3$, U be the sequence satisfying $(x-3)(x-1)$ with initial conditions $(1, 4)$ and V be the one satisfying the same polynomial with initial conditions $(1, 2)$. Then the lexicographically greatest words at each length are given below.

n	Maxlg $U \cap \Sigma^n$	Maxlg $V \cap \Sigma^n$	n	Maxlg $V \cap \Sigma^n$
1	3	1	7	2222120
2	30	20	8	22222111
3	300	211	9	222222110
4	3000	2210	10	2222222101
5	30000	22201	11	22222222100
6	300000	222200	12	222222222020

Taking a good look at suffixes

Look at the suffixes of the words in the nonregular case.

n	$\text{Maxlg } V \cap \Sigma^n$		
1	1		
2	20		
3	211		
4	2210		
5	22201		
6	222200		
7	2222120		
8	22222111		
9	222222110		

Taking a good look at suffixes

Look at the suffixes of the words in the nonregular case.

n	$\mathrm{Maxlg}\ V \cap \Sigma^n$	last 4 digits
1	1	1
2	20	20
3	211	211
4	2210	2210
5	22201	2201
6	222200	2200
7	2222120	2120
8	22222111	2111
9	222222110	2110

Taking a good look at suffixes

Look at the suffixes of the words in the nonregular case.

n	Maxlg $V \cap \Sigma^n$	last 4 digits	value
1	1	1	1
2	20	20	4
3	211	211	13
4	2210	2210	40
5	22201	2201	39
6	222200	2200	38
7	2222120	2120	37
8	22222111	2111	36
9	222222110	2110	35

The operator Δ_P

If $P(x) = x^l - d_1 x^{l-1} - d_2 x^{l-2} - \ldots - d_l$, $\Delta_P(U)$ is a sequence defined by

$$(\Delta_P(U))_n = U_{n+l} - d_1 U_{n+l-1} - d_2 U_{n+l-2} - \ldots - d_l U_n.$$

In particular, U satisfies the recurrence relation associated with P if and only if $\Delta_P(U) = 0$.

Idea

Assume that $d_\beta(1) = d_1 \cdots d_l$, $b(x)$ is the canonical β-polynomial and that the first $l-1$ digits of $rep(U_m - k)$ are $d_1 \cdots d_{l-1}$. Then, the next remainder is

$$U_m - k - \sum_{i=1}^{l-1} d_i U_{m-i}$$

Idea

Assume that $d_\beta(1) = d_1 \cdots d_l$, $b(x)$ is the canonical β-polynomial and that the first $l-1$ digits of $rep(U_m - k)$ are $d_1 \cdots d_{l-1}$. Then, the next remainder is

$$U_m - k - \sum_{i=1}^{l-1} d_i U_{m-i}$$

which is

$$d_l U_{m-l} + \Delta_b(U)_{m-l} - k$$

Idea (2)

Depending on the value of $\Delta_b(U)_{m-l}$, two behaviours may occur :

$U_m - k$

Idea (2)

Depending on the value of $\Delta_b(U)_{m-l}$, two behaviours may occur :

$U_m - k$

$\Delta_b(U)_{m-l} \geq k$ → digit d_l

remainder $\Delta_b(U)_{m-l} - k$

Idea (2)

Depending on the value of $\Delta_b(U)_{m-l}$, two behaviours may occur :

$U_m - k$

- $\Delta_b(U)_{m-l} < k$ → digit $d_l - 1$, remainder $U_{m-l} + \Delta_b(U)_{m-l} - k$
- $\Delta_b(U)_{m-l} \geq k$ → digit d_l, remainder $\Delta_b(U)_{m-l} - k$

Idea (2)

Depending on the value of $\Delta_b(U)_{m-l}$, two behaviours may occur :

$U_m - k$

- $\Delta_b(U)_{m-l} < k$ $\longrightarrow$ digit $d_l - 1$, remainder $U_{m-l} + \Delta_b(U)_{m-l} - k$
- $\Delta_b(U)_{m-l} \geq k$ $\longrightarrow$ digit d_l, remainder $\Delta_b(U)_{m-l} - k$

In the upper case, this is repeated as long as possible with a new value of k

$\longrightarrow$ we may understand regularity with the values of $\Delta_b(U)$

Idea (3.1)

Polynomial : $(x-3)(x-1)$,

Idea (3.1)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,4)$
Sequence of U's : $1, 4, 13, 40, 121, \ldots$
$\Delta_b(U)$: $1, 1, 1, \ldots$

Idea (3.1)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,4)$
Sequence of U's : $1, 4, 13, 40, 121, \ldots$
$\Delta_b(U) : 1, 1, 1, \ldots$

$$40 - 1$$

Idea (3.1)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,4)$
Sequence of U's : $1, 4, 13, 40, 121, \ldots$
$\Delta_b(U)$: $1, 1, 1, \ldots$

$$40 - 1 \xrightarrow{3} 0$$

Idea (3.1)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,4)$
Sequence of U's : $1, 4, 13, 40, 121, \ldots$
$\Delta_b(U)$: $1, 1, 1, \ldots$

$$40-1 \xrightarrow{\quad 3 \quad} 0 \xrightarrow{\quad 0 \quad} 0 \xrightarrow{\quad 0 \quad} \cdots$$

Idea (3.1)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,4)$
Sequence of U's : $1, 4, 13, 40, 121, \ldots$
$\Delta_b(U)$: $1, 1, 1, \ldots$

$$13-1 \xrightarrow{3} 0 \xrightarrow{0} \cdots$$

$$40-1 \xrightarrow{3} 0 \xrightarrow{0} 0 \xrightarrow{0} \cdots$$

$$121-1 \xrightarrow{3} 0 \xrightarrow{0} 0 \xrightarrow{0} 0 \xrightarrow{0} \cdots$$

Idea (3.2)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,3)$
Sequence of U's : $1, 3, 9, 27, 81, \ldots$
$\Delta_b(U)$: $0, 0, 0, \ldots$

Idea (3.2)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,3)$
Sequence of U's : $1, 3, 9, 27, 81, \ldots$
$\Delta_b(U)$: $0, 0, 0, \ldots$

$$27 - 1$$

Idea (3.2)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,3)$
Sequence of U's : $1, 3, 9, 27, 81, \ldots$
$\Delta_b(U)$: $0, 0, 0, \ldots$

$$27 - 1 \xrightarrow{2} 9 - 1$$

Idea (3.2)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,3)$
Sequence of U's : $1, 3, 9, 27, 81, \ldots$
$\Delta_b(U)$: $0, 0, 0, \ldots$

$$9-1$$

$$27-1 \xrightarrow{2} 9-1$$

Idea (3.2)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,3)$
Sequence of U's : $1, 3, 9, 27, 81, \ldots$
$\Delta_b(U)$: $0, 0, 0, \ldots$

$$9-1$$

$$27-1 \xrightarrow{2} 9-1$$

$$81-1 \xrightarrow{2} 27-1 \xrightarrow{2} 9-1$$

Idea (3.2)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,3)$
Sequence of U's : $1, 3, 9, 27, 81, \ldots$
$\Delta_b(U)$: $0, 0, 0, \ldots$

$$9-1 \xrightarrow{2} 3-1 \xrightarrow{2} \cdots$$

$$27-1 \xrightarrow{2} 9-1 \xrightarrow{2} 3-1 \xrightarrow{2} \cdots$$

$$81-1 \xrightarrow{2} 27-1 \xrightarrow{2} 9-1 \xrightarrow{2} 3-1 \xrightarrow{2} \cdots$$

Idea (3.3)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,2)$
Sequence of U's : $1, 2, 5, 14, 41, 122\ldots$
$\Delta_b(U)$: $-1, -1, -1, \ldots$

Idea (3.3)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,2)$
Sequence of U's : $1, 2, 5, 14, 41, 122 \ldots$
$\Delta_b(U)$: $-1, -1, -1, \ldots$

$$41 - 1$$

Idea (3.3)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,2)$
Sequence of U's : $1, 2, 5, 14, 41, 122\ldots$
$\Delta_b(U)$: $-1, -1, -1, \ldots$

$$41-1 \xrightarrow{2} 14-2$$

Idea (3.3)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,2)$
Sequence of U's : $1, 2, 5, 14, 41, 122 \ldots$
$\Delta_b(U)$: $-1, -1, -1, \ldots$

$$14-1$$

$$41-1 \xrightarrow{2} 14-2$$

Idea (3.3)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,2)$
Sequence of U's : $1, 2, 5, 14, 41, 122\ldots$
$\Delta_b(U)$: $-1, -1, -1, \ldots$

$$14-1$$

$$41-1 \xrightarrow{2} 14-2$$

$$122-1 \xrightarrow{2} 41-2 \xrightarrow{2} 14-3$$

Idea (3.3)

Polynomial : $(x-3)(x-1)$, initial conditions $(1,2)$
Sequence of U's : $1, 2, 5, 14, 41, 122 \dots$
$\Delta_b(U)$: $-1, -1, -1, \dots$

$$
\begin{array}{ccccccc}
 & & & & 14-1 & \xrightarrow{2} & \cdots \\
 & & 41-1 & \xrightarrow{2} & 14-2 & \xrightarrow{2} & \cdots \\
122-1 & \xrightarrow{2} & 41-2 & \xrightarrow{2} & 14-3 & \xrightarrow{2} & \cdots
\end{array}
$$

The theorem

Theorem (Charlier, K., 2022)

Let U be a numeration system with dominant root β, with $d_\beta(1) = d_1 \cdots d_l$, such that U satisfies $b(x)x^k(x^{al} - 1)$. Then $(\Delta_b(U)_n)_{n\in\mathbb{N}}$ is eventually periodic with period al. With $\delta_1, \ldots, \delta_{al}$ this period, U is regular if and only if

$$\delta_j + \delta_{j+l} + \ldots + \delta_{j+(a-1)l}$$

is nonnegative for all j in $\{0, \ldots, l-1\}$.

Back to Hollander

If $d_\beta(1)$ is finite and U satisfies the extended β-polynomial $b(x)x^k(1 + x + \ldots + x^{(a-1)l})$, then $\Delta_b(U)$ satisfies $x^k(1 + x + \ldots + x^{(a-1)l})$, and

$$\delta_j + \delta_{j+l} + \ldots + \delta_{j+(a-1)l} = 0$$

due to this. We find the third of Hollander's items.

Simple criteria for new cases

Case where the polynomial is $b(x)(1 + x^l + \ldots + x^{(a-1)l})r(x)$ with $r(x)$ dividing $x^l - 1$:
If $r(x) = x - 1$, one can show that L_U is regular if and only if the final initial condition is larger than what it would have been without the $x - 1$.
This ties back to the example used by Hollander : if U satisfies $(x - 3)(x - 1)$ with initial conditions $(1, n)$, it is regular iff $n \geq 3$ (3 is the second term of the sequence satisfying $x - 3$ with initial condition 1).

Simple criteria for new cases

Case where the polynomial is $b(x)(1 + x^l + \ldots + x^{(a-1)l})r(x)$ with $r(x)$ dividing $x^l - 1$:
If $r(x)$ divides $1 + x + \ldots + x^{l-1}$: then $\Delta_b(U)$ satisfies $1 + x + \ldots + x^{al-1}$, so

$$\sum_{j=0}^{l-1} \delta_j + \delta_{j+l} + \ldots + \delta_{j+(a-1)l} = 0$$

Thus in order for all the required sums to be nonnegative, they must be zero, so $r(x) = 1$, with regularity in this case and only in this case.

Future work

Some questions remain open :

- Could the language still be context-free in the case where it is not regular?
- Can we more precisely describe the length of the common prefix between the elements of $\mathrm{Maxlg}\, L$ and $d_\beta(1)$?
- What about the case without a dominant root?

Thank you for your attention!