# A Decision Problem for Ultimately Periodic Sets in Non-Standard Numeration Systems

Emilie Charlier, Michel Rigo (University of Liège)

http://www.discmath.ulg.ac.be/

Journées Montoises, Mons, August 2008

Let's start with classical $k$-ary numeration system, $k \geq 2$:

$$n = \sum_{i=0}^{\ell} d_i k^i, \quad \mathrm{rep}_k(n) = d_\ell \cdots d_0 \in \{0, \dots, k-1\}^*$$

**Definition**

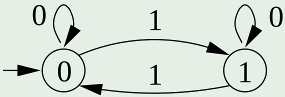
A set $X \subseteq \mathbb{N}$ is *$k$-recognizable*, if the language

$$\mathrm{rep}_k(X) = \{\mathrm{rep}_k(x) \mid x \in X\}$$

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

## EXAMPLES OF $k$-RECOGNIZABLE SETS

- In base 2, the set of even integers: $\text{rep}_2(2\mathbb{N}) = \{0,1\}^*0$.
- In base 2, the set of powers of 2: $\text{rep}_2(X) = 1\,0^*$.
- In base 2, the "*Thue-Morse set*" is 2-recognizable, i.e., $\{n \in \mathbb{N} \mid \mathbf{S}(\text{rep}_2(n)) \equiv 0 \pmod 2\}$.




- Given a $k$-automatic sequence $(x_n)_{n\geq 0}$ over an alphabet $\Sigma$, then for all $\sigma \in \Sigma$, the set $\{i \in \mathbb{N} \mid x_i = \sigma\}$ is $k$-recognizable.

## SOME NATURAL QUESTIONS, WELL-KNOWN ANSWERS...

- Characterization(s) of the $k$-recognizable sets
- Does the property depend on the choice of the base $k$ ?
- If so, are there (infinite) sets that are recognizable for all bases ?

$k, \ell$ are *multiplicatively independent* if $k^m = \ell^n \Rightarrow m = n = 0$, i.e., $\log k / \log \ell$ is irrational.

| 2 | 3 | 5 | 6 | 7 | 10 | 11 | $\cdots$ |
|---|---|---|---|---|---|---|---|
| 4 | 9 | 25 | 36 | 49 | 100 | 121 | $\cdots$ |
| 8 | 27 | 125 | 216 | 343 | 1000 | 1331 | $\cdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | |

### PROPOSITION (EASY)

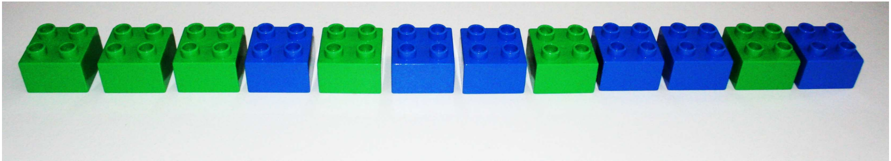
Let $k, \ell \geq 2$ be multiplicatively dependent.
$X \subset \mathbb{N}$ is $k$-recognizable IFF $X$ is $\ell$-recognizable.

### COBHAM'S THEOREM (1969)

Let $k, \ell \geq 2$ be two multiplicatively independent integers.

If $X \subseteq \mathbb{N}$ is both $k$- and $\ell$-recognizable,
then $X$ is ultimately periodic (finite union of A. P.).

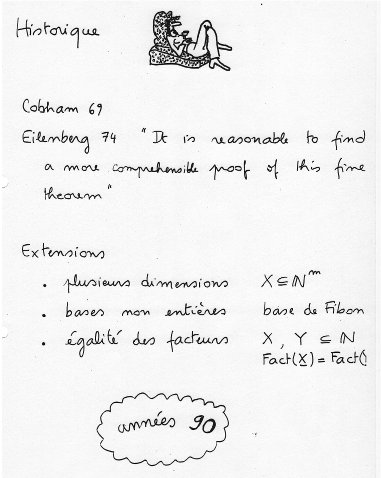
Index: 3, Period: 3



Véronique Bruyère promoted Cobham's result...

Véronique Bruyère
Juin 96.

Encore un exposé sur le Théorème de Cobham ...



Historique

Cobham 69

Eilenberg 74 "It is reasonable to find a more comprehensible proof of this fine theorem"

Extensions

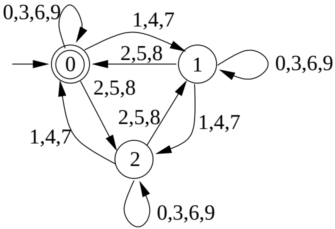
- plusieurs dimensions $X \subseteq \mathbb{N}^m$
- bases non entières base de Fibon
- égalité des facteurs $X, Y \subseteq \mathbb{N}$ $Fact(X) = Fact($

années 90


c'est pas
évident

## DIVISIBILITY CRITERIA

If $X$ is ultimately periodic, then $X$ is $k$-recognizable $\forall k \geq 2$.




## VARIOUS PROOF SIMPLIFICATIONS AND GENERALIZATIONS

G. Hansel, D. Perrin, F. Durand, V. Bruyère, F. Point, C. Michaux, R. Villemaire, A. Bès, J. Bell, J. Honkala, S. Fabre, C. Reutenauer, A.L. Semenov, L. Waxweiler, …

- V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and $p$-recognizable sets of integers, *Bull. Belg. Math. Soc.* **1** (1994).

## MAIN QUESTION FOR THIS TALK

Consider a $k$-recognizable set $X$, given by a DFA $\mathcal{A}_X$, decide whether or not $X$ is ultimately periodic ?

## THE ANSWER IS YES

J. Honkala, A decision method for the recognizability of sets defined by number systems, *Theoret. Inform. Appl.* **20** (1986).

Sketch :

- The number of states of $\mathcal{A}_X$ produces an upper bound on the possible maximal index and period for $X$.
- Consequently, there are finitely many candidates to check.
- For each pair $(i, p)$ of candidates, produce a DFA for all possible corresponding ultimately periodic sets and compare it with $\mathcal{A}_X$.

The question we are considering here was initially raised by J. Sakarovitch for abstract numeration systems

## Example (Appetizer)

Fibonacci system $F_{i+2} = F_{i+1} + F_i$, $F_0 = 1$, $F_1 = 2$
greedy expansion, $\ldots, 21, 13, 8, 5, 3, 2, 1$

| | | | | | |
|---|---|---|---|---|---|
| 1 | 1 | 8 | 10000 | 15 | 100010 |
| 2 | 10 | 9 | 10001 | 16 | 100100 |
| 3 | 100 | 10 | 10010 | 17 | 100101 |
| 4 | 101 | 11 | 10100 | 18 | 101000 |
| 5 | 1000 | 12 | 10101 | 19 | 101001 |
| 6 | 1001 | 13 | 100000 | 20 | 101010 |
| 7 | 1010 | 14 | 100001 | 21 | 1000000 |

The “pattern” 11 is forbidden, $A_F = \{0, 1\}$.

## EXAMPLE (CONTINUED)

We can define a $F$-recognizable set $X$ of integers:
$\mathrm{rep}_F(X) \subset \{0,1\}^*$ is regular.

## THE QUESTIONS BECOMES

$\longrightarrow$ Consider a $F$-recognizable set $X \subseteq \mathbb{N}$, given by a DFA $\mathcal{A}_X$, decide whether or not $X$ is ultimately periodic ?

First part (*upper bound on the period*) :

**“PSEUDO-RESULT”**

Let $X$ be ult. periodic with period $p_X$ ($X$ is $F$-recognizable).

Any DFA accepting $\text{rep}_F(X)$ has at least $f(p_X)$ states, where $f$ is increasing.

**“PSEUDO-COROLLARY”**

Let $X \subseteq \mathbb{N}$ be a $F$-recognizable set of integers s.t. $\text{rep}_F(X)$ is accepted by $\mathcal{A}_X$ with $k$ states.

If $X$ is ultimately periodic with period $p$, then

$$\boxed{f(p) \leq k} \quad \text{with} \begin{cases} k \text{ fixed} \\ f \text{ increasing.} \end{cases}$$

$\Rightarrow$ The number of candidates for $p$ is bounded from above.

## PROPOSITION (FIBONACCI)

Let $X$ be ultimately periodic with period $p_X$.
Any DFA accepting $\mathrm{rep}_F(X)$ has at least $p_X$ states.

• Idea of the proof. The sequence is *purely periodic* mod $p_X$.
Indeed, $F_{n+2} = F_{n+1} + F_n$ and $F_n = F_{n+2} - F_{n+1}$

### $X = (8\mathbb{N} + 3) \cup (8\mathbb{N} + 5) \cup \{0, 1, 16, 20, 88\}$

| $\cdots$101725505321 | 101725505321 | 101725505321 | |
|---|---|---|---|
| | | 000000010000 | 0 |
| *some* | | 000000000001 | 1 |
| *special* | *with* | 000000000010 | 2 |
| *words* | *enough* | 000000000100 | 3 |
| | *leading* | 000000000101 | 4 |
| *complete* | | 000000001000 | 5 |
| *residue* | | 000000001001 | 6 |
| *set* | | 000000001010 | 7 |

Enough leading zeroes to :

- start a new period for $(F_n \bmod 8)_{n\geq 0}$
- ensure to keep greedy representation when concatenating with something: 10100101.000000000100
- be in the periodic part of $X$

$$X = (8\mathbb{N}+3) \cup (8\mathbb{N}+5) \cup \{0, 1, 16, 20, 88\}$$

| $\cdots$101725505321 | 101725505321 | 101725505321 | |
|---|---|---|---|
| | | 000000010000 | 0 |
| *some* | | 000000000001 | 1 |
| *special* | *with* | 000000000010 | 2 |
| *words* | *enough* | 000000000100 | 3 |
| | *leading* | 000000000101 | 4 |
| *complete* | | 000000001000 | 5 |
| *residue* | | 000000001001 | 6 |
| *set* | | 000000001010 | 7 |

$w^{-1}L = \{u \mid wu \in L\} \leftrightarrow$ states of minimal automaton of $L$

**$X = (8\mathbb{N}+3) \cup (8\mathbb{N}+5) \cup \{0, 1, 16, 20, 88\}$**

| $\cdots$ 101725505321 | 101725505321 | 101725505321 |
|---|---|---|
| | | 000000010000 |
| | | 000000000001 |
| | | 000000000010 |
| | | 000000000100 |
| | | 000000000101 |
| | | 000000001000 |
| | | 000000001001 |
| | | 000000001010 |

$10000^{-1}\,\mathrm{rep}_F(X)$, $1^{-1}\,\mathrm{rep}_F(X)$, $10^{-1}\,\mathrm{rep}_F(X)$, $100^{-1}\,\mathrm{rep}_F(X)$, $101^{-1}\,\mathrm{rep}_F(X)$, $1000^{-1}\,\mathrm{rep}_F(X)$, $1001^{-1}\,\mathrm{rep}_F(X)$ and $1010^{-1}\,\mathrm{rep}_F(X)$ are pairwise distinct !

$w^{-1}L = \{u \mid wu \in L\} \leftrightarrow$ states of minimal automaton of $L$

**$X = (8\mathbb{N}+3) \cup (8\mathbb{N}+5) \cup \{0, 1, 16, 20, 88\}$**

| | | |
|---|---|---|
| $\cdots$101725505321 | 101725505321 | 101725505321 |
| | | 000000010000 |
| | | 000000000001 |
| | | 000000000010 |
| | | 000000000100 |
| | | 000000000101 |
| | | 000000001000 |
| | | 000000001001 |
| | | 000000001010 |

$10000^{-1}\,\mathrm{rep}_F(X)$, $1^{-1}\,\mathrm{rep}_F(X)$, $10^{-1}\,\mathrm{rep}_F(X)$, $100^{-1}\,\mathrm{rep}_F(X)$, $101^{-1}\,\mathrm{rep}_F(X)$, $1000^{-1}\,\mathrm{rep}_F(X)$, $1001^{-1}\,\mathrm{rep}_F(X)$ and $1010^{-1}\,\mathrm{rep}_F(X)$ are pairwise distinct !

$w^{-1}L = \{u \mid wu \in L\} \leftrightarrow$ states of minimal automaton of $L$

**$X = (8\mathbb{N}+3) \cup (8\mathbb{N}+5) \cup \{0, 1, 16, 20, 88\}$**

| $\cdots$101725505321 | 101725505321 | 101725505321 |
|---|---|---|
| | | 000000010000 |
| | | 000000000001 |
| | | 000000000010 |
| | 10000 | 000000000100 |
| | | 000000000101 |
| | 10000 | 000000001000 |
| | | 000000001001 |
| | | 000000001010 |

$10000^{-1}\,\mathrm{rep}_F(X)$, $1^{-1}\,\mathrm{rep}_F(X)$, $10^{-1}\,\mathrm{rep}_F(X)$, $100^{-1}\,\mathrm{rep}_F(X)$, $101^{-1}\,\mathrm{rep}_F(X)$, $1000^{-1}\,\mathrm{rep}_F(X)$, $1001^{-1}\,\mathrm{rep}_F(X)$ and $1010^{-1}\,\mathrm{rep}_F(X)$ are pairwise distinct !

$w^{-1}L = \{u \mid wu \in L\} \leftrightarrow$ states of minimal automaton of $L$

$X = (8\mathbb{N}+3) \cup (8\mathbb{N}+5) \cup \{0, 1, 16, 20, 88\}$

| $\cdots$ 101725505321 | 101725505321 | 101725505321 |
|---|---|---|
| | | 000000010000 |
| | | 000000000001 |
| | | 000000000010 |
| | 10000 | 000000000100 |
| | | 000000000101 |
| | 10000 | 000000001000 |
| | | 000000001001 |
| | | 000000001010 |

$10000^{-1}\mathrm{rep}_F(X)$, $1^{-1}\mathrm{rep}_F(X)$, $10^{-1}\mathrm{rep}_F(X)$, $100^{-1}\mathrm{rep}_F(X)$, $101^{-1}\mathrm{rep}_F(X)$, $1000^{-1}\mathrm{rep}_F(X)$, $1001^{-1}\mathrm{rep}_F(X)$ and $1010^{-1}\mathrm{rep}_F(X)$ are pairwise distinct !

$w^{-1}L = \{u \mid wu \in L\} \leftrightarrow$ states of minimal automaton of $L$

$X = (8\mathbb{N}+3) \cup (8\mathbb{N}+5) \cup \{0, 1, 16, 20, 88\}$

| | | |
|---|---|---|
| $\cdots$101725505321 | 101725505321 | 101725505321 |
| | | 000000010000 |
| | 10 | 000000000001 |
| | | 000000000010 |
| | 10000/10 | 000000000100 |
| | | 000000000101 |
| | 10000 | 000000001000 |
| | | 000000001001 |
| | | 000000001010 |

$10000^{-1}\,\mathrm{rep}_F(X)$, $1^{-1}\,\mathrm{rep}_F(X)$, $10^{-1}\,\mathrm{rep}_F(X)$, $100^{-1}\,\mathrm{rep}_F(X)$, $101^{-1}\,\mathrm{rep}_F(X)$, $1000^{-1}\,\mathrm{rep}_F(X)$, $1001^{-1}\,\mathrm{rep}_F(X)$ and $1010^{-1}\,\mathrm{rep}_F(X)$ are pairwise distinct !

### Lemma – definition of (minimal) period

Let $X$ be an ult. periodic set of period $p_X$ and preperiod $a_X$.

Let $i, j \geq a_X$. If $i \not\equiv j \mod p_X$ then $\exists t < p_X$ such that either

- $i + t \in X$ and $j + t \notin X$, or
- $i + t \notin X$ and $j + t \in X$.

### DEFINITION

A *numeration system* is an increasing sequence $U = (U_i)_{i \geq 0}$ of integers s.t. $U_0 = 1$ and $C_U := \sup_{i \geq 0} \lceil U_{i+1}/U_i \rceil$ is finite. $A_U = \{0, \ldots, C_U - 1\}$.

The *greedy $U$-representation* of a positive integer $n$ is the unique finite word $\text{rep}_U(n) = w_\ell \cdots w_0$ over $A_U$ satisfying

$$n = \sum_{i=0}^{\ell} w_i U_i, \; w_\ell \neq 0 \text{ and } \sum_{i=0}^{t} w_i U_i < U_{t+1}, \; \forall t = 0, \ldots, \ell.$$

$$\text{val}_U(x_\ell \cdots x_0) = \sum_{i=0}^{\ell} x_i U_i, \quad \forall x_\ell \cdots x_0 \in A_U^*.$$

$\rightarrow$ *$U$-recognizable* set $X$ of integers: $\text{rep}_U(X)$ is regular.

**DEFINITION**

A numeration system $U = (U_i)_{i\geq 0}$ is said to be *linear*, if $U$ satisfies a homogenous linear recurrence relation. For all $i \geq 0$,

$$U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i$$

for some $k \geq 1$, $a_1, \ldots, a_k \in \mathbb{Z}$ and $a_k \neq 0$.

What we need for making comparisons :

**LEMMA**

Let $a, b$ be nonnegative integers and $U = (U_i)_{i \geq 0}$ be a linear numeration system. The language

$$\mathrm{val}_U^{-1}(a\mathbb{N} + b) = \{w \in A_U^* \mid \mathrm{val}_U(w) \in a\mathbb{N} + b\} \subset A_U^*$$

is regular.

In particular, if $\mathbb{N}$ is $U$-recognizable then a DFA accepting $\mathrm{rep}_U(a\mathbb{N} + b)$ can be obtained efficiently and any ultimately periodic set is $U$-recognizable.

**REMARK (SHALLIT ’94)**

If $\mathbb{N}$ is $U$-recognizable, then $U$ is linear.

A technical hypothesis :

$$\lim_{i\to+\infty} U_{i+1} - U_i = +\infty. \tag{1}$$

Most systems are built on an exponential sequence $(U_i)_{i\geq 0}$

**LEMMA**

*Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying (1). For all $j$, there exists $L$ such that for all $\ell \geq L$,*

$$10^{\ell-|\mathrm{rep}_U(t)|}\,\mathrm{rep}_U(t),\ t = 0, \ldots, U_j - 1$$

*are greedy $U$-representations. Otherwise stated, if $w$ is a greedy $U$-representation, then for $r$ large enough, $10^r w$ is also a greedy $U$-representation.*

$N_U(m) \in \{1, \dots, m\}$ = the number of values that are taken infinitely often by $(U_i \mod m)_{i \geq 0}$.

**Proposition**

Let $U = (U_i)_{i \geq 0}$ be a numeration system satisfying (1).
If $X \subseteq \mathbb{N}$ is an ult. periodic $U$-recognizable set of period $p_X$, then any DFA accepting $\text{rep}_U(X)$ has at least $N_U(p_X)$ states.

Upper bound on the period:

**COROLLARY**

Let $U = (U_i)_{i \geq 0}$ be a numeration system satisfying (1). Assume that

$$\lim_{m \to +\infty} N_U(m) = +\infty.$$

Then the period of an ultimately periodic set $X \subseteq \mathbb{N}$ such that $\mathrm{rep}_U(X)$ is accepted by a DFA with $d$ states is bounded by the smallest integer $s_0$ such that for all $m \geq s_0$, $N_U(m) > d$.

## LEMMA

*If $U = (U_i)_{i\geq 0}$ is a linear numeration system satisfying a recurrence relation of order $k$ like*

$$U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i$$

*with $a_k = \pm 1$, then $\lim_{m\to+\infty} N_U(m) = +\infty$.*

## IDEA OF THE PROOF

| 1 | 2 | 3 | 5 | 8 | 13 | 21 | 34 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 0 | | | | | | $N_U(3) = 3$ |
| 1 | 2 | 3 | | | | | | $N_U(4) \geq 3$ |
| 1 | 2 | 3 | 5 | 8 | | | | $N_U(10) \geq 5$ |
| 1 | 2 | 3 | 5 | 8 | 13 | 21 | | $N_U(30) \geq 7$ |

$$N_U(m) \geq \log_\tau m$$

Upper bound on the index:

For a sequence $(U_i)_{i\geq 0}$ of integers, if $(U_i \mod m)_{i\geq 0}$ is ultimately periodic, we denote its (minimal) index by $\iota_U(m)$.

### PROPOSITION

Let $U = (U_i)_{i\geq 0}$ be a linear numeration system.

Let $X \subseteq \mathbb{N}$ be an ult. periodic $U$-recognizable set of period $p_X$ and index $a_X$.

Then any DFA accepting $\text{rep}_U(X)$ has at least $|\text{rep}_U(a_X - 1)| - \iota_U(p_X)$ states.

If $p_x$ is bounded and $a_x$ is increasing, then the number of states is increasing.

**THEOREM**

*Let $U = (U_i)_{i\geq 0}$ be a linear numeration system such that $\mathbb{N}$ is $U$-recognizable and satisfying a recurrence relation of order $k$ with $a_k = \pm 1$ and condition (1).*

*It is decidable whether or not a $U$-recognizable set is ultimately periodic.*

## REMARK

If $\gcd(a_1, \ldots, a_k) = g \geq 2$, for all $n \geq 1$ and for all $i$ large enough, we have $U_i \equiv 0 \mod g^n$ and assumption about $N_U(m)$ does not hold !

## EXAMPLES

- Honkala's integer bases: $U_{n+1} = k\, U_n$
- $U_{n+2} = 2U_{n+1} + 2U_n$

$$a, b, 2(a+b), 2(2a+3b), 4(3a+4b), 4(8a+11b), \ldots$$

Learn more about linear recurrent sequences mod $m$ . . .


- H.T. Engstrom, On sequences defined by linear recurrence relations, *Trans. Amer. Math. Soc.* **33** (1931).
- M. Ward, The characteristic number of a sequence of integers satisfying a linear recursion relation, *Trans. Amer. Math. Soc.* **33** (1931).
- M. Hall, An isomorphism between linear recurring sequences and algebraic rings, *Trans. Amer. Math. Soc.* **44** (1938).
- G. Rauzy, Relations de récurrence modulo $m$, Séminaire Delange-Pisot-Poitou, Th. Nombres **5**, (1963-1964).


To solve the case where $\gcd(a_1, \ldots, a_k) = 1$.

## DEFINITION

An *abstract numeration system* $S = (L, \Sigma, <)$ is given by an infinite regular language $L$ over a totally ordered alphabet $(\Sigma, <)$

## EXAMPLE

Consider the language $L = \{\varepsilon\} \cup \{a, ab\}^* \cup \{c, cd\}^*$ and the ordering $a < b < c < d$ of the alphabet.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | $\varepsilon$ | 5 | $cc$ | 10 | $ccc$ | 15 | $aaba$ | 20 | $ccdc$ |
| 1 | $a$ | 6 | $cd$ | 11 | $ccd$ | 16 | $abaa$ | 21 | $cdcc$ |
| 2 | $c$ | 7 | $aaa$ | 12 | $cdc$ | 17 | $abab$ | 22 | $cdcd$ |
| 3 | $aa$ | 8 | $aab$ | 13 | $aaaa$ | 18 | $cccc$ | 23 | $aaaaa$ |
| 4 | $ab$ | 9 | $aba$ | 14 | $aaab$ | 19 | $cccd$ | 24 | $aaaab$ |

One can define $S$-recognizable sets of integers.

**PROPOSITION (P. LECOMTE, M.R. '01)**

Let $S = (L, \Sigma, <)$ be an abstract numeration system built over an infinite regular language $L$ over $\Sigma$.

Any ultimately periodic set $X$ is $S$-recognizable and a DFA accepting $\mathrm{rep}_S(X)$ can be effectively obtained.

Upper bound on the period:

### PROPOSITION

Let $S = (L, \Sigma, <)$ be an abstract numeration system such that for all states $q$ of the trim minimal automaton $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ of $L$,

$$\lim_{j \to +\infty} \mathbf{u}_j(q) = +\infty$$

and $\mathbf{u}_j(q_{0,L}) > 0$ for all $j \geq 0$.

If $X \subseteq \mathbb{N}$ is an ult. periodic set of period $p_X$, then any DFA accepting $\mathrm{rep}_S(X)$ has at least $\lceil N_{\mathbf{v}}(p_X)/\#Q_L \rceil$ states where $\mathbf{v} = (\mathbf{v}_j(q_{0,L}))_{j \geq 0}$.

$N_{\mathbf{v}}(m) \to \infty$ ?

Upper bound on the index:

**PROPOSITION**

Let $S = (L, \Sigma, <)$ be an abstract numeration system.

If $X \subseteq \mathbb{N}$ is an ult. periodic set of period $p_X$ such that $\mathrm{rep}_S(X)$ is accepted by a DFA with $d$ states,

then the index $a_X$ of $X$ is bounded by a constant depending only on $d$ and $p_X$.

### THEOREM

*Let $S = (L, \Sigma, <)$ be an abstract numeration system such that for all states $q$ of the trim minimal automaton $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ of $L$*

$$\lim_{j \to \infty} \mathbf{u}_j(q) = +\infty$$

*and $\mathbf{u}_j(q_{0,L}) > 0$ for all $j \geq 0$. Assume moreover that $\mathbf{v} = (\mathbf{v}_i(q_{0,L}))_{i \geq 0}$ satisfies a linear recurrence relation with $a_k = \pm 1$.*

*It is decidable whether or not a $S$-recognizable set is ultimately periodic.*


c'est pas
évident