Decidability through first-order logic and regular sequences

Émilie Charlier

Département de Mathématique, Université de Liège

Dyadisc 3 - Amiens - 2019 July 3-5

Plan of the talk

I. b-automatic sequences and b-recognizable sets

II. Characterizing b-recognizable sets via logic

III. Applications to decidability questions for automatic sequences

IV. Enumeration: counting b-definable properties of b-automatic sequences is b-regular

V. Logic and non-standard numeration systems

I. Automatic sequences and recognizable sets

Integer bases

Let $b \geq 2$ be an integer (the integer base).

A natural number n is represented by the finite word

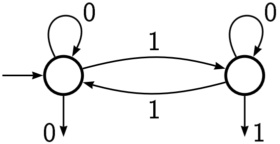
$$\mathrm{rep}_b(n) = c_\ell \cdots c_1 c_0$$

over the alphabet $A_b = \{0, 1, \ldots, b-1\}$ obtained from the greedy algorithm:

$$n = \sum_{i=0}^{\ell} c_i b^i.$$

b-automatic sequences

Take $b = 2$ and consider the following DFAO:



For each n, the DFAO reads $\mathrm{rep}_2(n)$ and outputs 0 or 1 according to the last state that is reached.

We obtain the Thue-Morse sequence

$$0110100110010110100101100110 1001 \cdots$$

b-automatic sequences

A sequence $x\colon \mathbb{N}^d \to \mathbb{N}$ is said to be b-automatic if there exists a DFAO with input alphabet A_b such that for each $\mathbf{n} \in \mathbb{N}^d$, $x(\mathbf{n})$ is the symbol outputted by the DFAO after reading $\mathrm{rep}_b(\mathbf{n})$.

Two remarks:

- A b-automatic sequence can take only finitely many values.
- We can work in any dimension d:

$$\mathrm{rep}_2 \begin{bmatrix} 5 \\ 3 \\ 10 \end{bmatrix} = \begin{bmatrix} 101 \\ 11 \\ 1010 \end{bmatrix}^{0} = \begin{bmatrix} 0101 \\ 0011 \\ 1010 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

b-recognizable sets of integers

A set $X \subseteq \mathbb{N}^d$ is b-recognizable if the language

$$\mathrm{rep}_b(X) = \{\mathrm{rep}_b(\mathbf{n}) \colon \mathbf{n} \in X\}$$

is regular.

It is equivalent to say that its characteristic sequence $\chi_X \colon \mathbb{N}^d \to \{0, 1\}$ is b-automatic: there exists a DFAO that on input $\mathrm{rep}_b(\mathbf{n})$ ouputs 1 if $\mathbf{n} \in X$, and outputs 0 otherwise.

The set of evil numbers $\{0, 3, 5, 6, 9, 10, 12, 15, 17, 18, 20, 23, \ldots\}$, i.e. the natural numbers having an even number of 1 in base 2, is 2-recognizable. Its characteristic sequence is the Thue-Morse sequence.

Cobham-Semenov theorem

Semi-linear sets of $\mathbb{N}^d$ are finite unions of sets of the form

$$\mathbf{p}_0 + \mathbf{p}_1\,\mathbb{N} + \cdots + \mathbf{p}_\ell\,\mathbb{N}$$

where $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_\ell \in \mathbb{N}^d$.

Theorem (Cobham 1969, Semenov 1977)

Let b and b' be multiplicatively independent integer bases.
If a subset of $\mathbb{N}^d$ is simultaneously b-recognizable and b'-recognizable, then it is semi-linear.

Alternative definitions of b-recognizable sets

There exist several equivalent definitions of b-recognizable sets of integers using

- logic
- b-uniform morphisms
- finiteness of the b-kernel
- algebraic formal series
- recognizable/rational formal series

See the survey of Bruyère-Hansel-Michaux-Villemaire.

II. Characterizing b-recognizable sets via logic

Definable sets

Theorem (Büchi 1960, Bruyère 1985)

A subset X of $\mathbb{N}^d$ is b-recognizable iff it is b-definable.

Definable sets

Let $\mathcal{S}$ be a logical structure whose domain is S.
A set $X \subseteq S^d$ is definable in $\mathcal{S}$ if there exists a first-order formula $\varphi(x_1, \ldots, x_d)$ of $\mathcal{S}$ such that

$$X = \{(s_1, \ldots, s_d) \in S^d : \mathcal{S} \models \varphi(s_1, \ldots, s_d)\}.$$

A first-order formula is defined recursively from

- variables $x_1, x_2, x_3, \ldots$ describing elements of the domain S
- the equality $=$
- the relations and functions given in the structure $\mathcal{S}$
- the connectives $\neg, \vee, \wedge, \implies, \iff$
- the quantifiers $\forall, \exists$ on variables.

Presburger arithmetic $\langle \mathbb{N}, + \rangle$

$x \leq y$ is definable by $(\exists z)\ (x + z = y)$. Not true in $\langle \mathbb{Z}, + \rangle$.

$x = 0$ is definable by $x + x = x$. OK in $\langle \mathbb{Z}, + \rangle$.

$x = 1$ is definable by $x \neq 0 \wedge ((\forall y)\ (y = 0 \vee x \leq y))$. Not true in $\langle \mathbb{Z}, + \rangle$.

Inductively, $x = c$ is definable for every $c \in \mathbb{N}$.

The sets $a\mathbb{N} + b$ are definable: $a\mathbb{N} + b = \{x \colon (\exists y)\ (x = a \cdot y + b)\}$ where $a \cdot y$ stands for $y + y + \cdots y$ (a times).

In fact, a subset $X \subseteq \mathbb{N}$ is definable in $\langle \mathbb{N}, + \rangle$ iff it is a finite union of arithmetic progressions, or equivalently, ultimately periodic.

A subset $X \subseteq \mathbb{N}^d$ is definable in $\langle \mathbb{N}, + \rangle$ iff it is semi-linear.

b-definable sets

A set $X \subseteq \mathbb{N}^d$ is b-definable if it is definable in the structure $\langle \mathbb{N}, +, V_b \rangle$, where

- $+(x, y, z)$ is the ternary relation defined by $x + y = z$,
- $V_b(x)$ is the unary function defined as the largest power of b dividing x if $x \geq 1$ and $V_b(0) = 1$.

For example, the set $X = \{x \in \mathbb{N} \colon x \text{ is a power of } b\}$ is definable by $V_b(x) = x$.

It can be shown that the structures $\langle \mathbb{N}, +, V_b \rangle$ and $\langle \mathbb{N}, +, P_b \rangle$ are not equivalent, where $P_b(x)$ is 1 if x is a power of b and 0 otherwise.

The Büchi-Bruyère theorem

Theorem (Büchi 1960, Bruyère 1985)

A subset X of $\mathbb{N}^d$ is b-recognizable iff it is b-definable. Moreover, both directions are effective.

Sketch of the proof.

- From a DFA accepting $\mathrm{rep}_b(X)$, construct a first-order formula φ of the structure $\langle \mathbb{N}, +, V_b \rangle$ defining X, i.e. such that

$$X = \{(n_1, \ldots, n_d) \in \mathbb{N}^d : \varphi(n_1, \ldots, n_d) \text{ is true}\}.$$

- Conversely, given a first-order formula φ of the structure $\langle \mathbb{N}, +, V_b \rangle$ defining X, build a DFA accepting $\mathrm{rep}_b(X)$.
  This part is done by induction on the length of the formula φ.

Corollary

The first order theory of $\langle \mathbb{N}, +, V_b \rangle$ *is decidable*

Proof.

- We have to show that, given any closed first-order formula of $\langle \mathbb{N}, +, V_b \rangle$, we can decide whether it is true or false in $\mathbb{N}$.
- Since there is no constant in the structure, a closed formula of $\langle \mathbb{N}, +, V_b \rangle$ is necessarily of the form $\exists x \varphi(x)$ or $\forall x \varphi(x)$.
- The set

  $$X_\varphi = \{n \in \mathbb{N} \colon \langle \mathbb{N}, +, V_b \rangle \models \varphi(n)\}$$

  is b-definable, so it is b-recognizable by the Büchi-Bruyère theorem. This means that we can effectively construct a DFA accepting $\mathrm{rep}_b(X_\varphi)$.

- The closed formula $\exists x \varphi(x)$ is true if $\text{rep}_b(X_\varphi)$ is nonempty, and false otherwise.
- As the emptiness of the language accepted by a DFA is decidable, we can decide if $\exists x \varphi(x)$ is true.
- The case $\forall x \varphi(x)$ reduces to the previous one since $\forall x \varphi(x)$ is logically equivalent to $\neg \exists x \neg \varphi(x)$. We can construct a DFA accepting the base-b representations of

  $$X_{\neg\varphi} = \mathbb{N} \setminus X_\varphi.$$

  The language it accepts is empty iff the formula $\forall x \varphi(x)$ is true.

III. Applications to decidability questions for automatic sequences

Corollary

If we can express a property $P(n)$ using quantifiers, logical operations, addition, subtraction, comparison, and elements of some b-automatic sequences, then $\exists n P(n)$, $\exists^{\infty} n P(n)$ and $\forall n P(n)$ are decidable.

In particular, what about the property $x(\mathbf{i}) = x(\mathbf{j})$?

If $x\colon \mathbb{N}^d \to \mathbb{N}$ is a b-automatic sequence then, for all letters a occurring in x, the subsets $x^{-1}(a)$ of $\mathbb{N}^d$ are b-recognizable.

Hence they are definable by some first-order formulae ψ_a of $\langle \mathbb{N}, +, V_b \rangle$ (by Büchi-Bruyère theorem): $\psi_a(\mathbf{n})$ is true iff $x(\mathbf{n}) = a$.

Therefore, we can express $x(\mathbf{i}) = x(\mathbf{j})$ by the first-order formula $\varphi(x_1, \ldots, x_d, y_1, \ldots, y_d)$ of $\langle \mathbb{N}, +, V_b \rangle$:

$$\varphi(\mathbf{i}, \mathbf{j}) \equiv \bigvee_a (\psi_a(\mathbf{i}) \wedge \psi_a(\mathbf{j})).$$

Applications

Consider the property of having an overlap.

A (unidimensional) sequence x has an overlap beginning at position i iff $(\exists \ell \geq 1)\ (\forall j \leq \ell)\ x(i+j) = x(i+\ell+j)$.

Now suppose that x is b-automatic.

Given a DFAO M_1 generating x, we first create an NFA M_2 that on input (i, ℓ) accepts if $(\exists j \leq \ell)\ x(i+j) \neq x(i+j+\ell)$.

To do this, M_2 guesses the base-b representation of j digit-by-digit, verifies that $j \leq \ell$, computes $i+j$ and $i+j+\ell$ on the fly, and accepts if $x(i+j) \neq x(i+j+\ell)$.

We now convert M_2 to a DFA M_3 using the subset construction, and inverse the final status of each state. Thus, M_3 accepts those pairs (i, ℓ) such that $(\forall j \leq \ell)\ x(i+j) = x(i+j+\ell)$.

Now we create an NFA M_4 that on input i guesses $\ell \geq 1$ and accepts iff M_3 accepts (i, ℓ).

As we can decide if M_4 accepts anything, we have obtained:

Proposition

It is decidable if a b-automatic sequence has an overlap.

Many decidability results for automatic sequences

- It is decidable whether a b-automatic sequence has k-powers (for a fixed k).
- It is decidable whether a b-automatic sequence is ultimately periodic.
- Given two b-automatic sequences x and y, it is decidable whether $\mathrm{Fac}(x) \subseteq \mathrm{Fac}(y)$.
- ...

What about deciding if a b-automatic sequence is Toeplitz?

The predicate

$$\forall n \, \exists p \geq 1 \, \forall \ell \; x(n) = x(n + \ell p)$$

is not a first order formula in $\langle \mathbb{N}, +, V_b \rangle$. Why? Is this property b-definable? What about the case where the periods p are restricted to powers of the base b?

A negative result by Schaeffer

If x is an arbitrary b-automatic sequence, then the predicate

$$\text{“}x[i, i+2n-1] \text{ is an abelian square”}$$

is not expressible in the logical theory $\langle \mathbb{N}, +, V_b \rangle$.

Complexity issues

In the worst case, we have a tower of exponentials:

$$2^{2^{\cdot^{\cdot^{\cdot^{2^n}}}}}$$

where n is the number of states of the given DFAO and the height of the tower is the number of alternating quantifiers if the first-order predicate.

This procedure was implemented by Mousavi, giving birth the Walnut software.

In practice, Goc, Henshall, Mousavi, Shallit and others were able to run their programs in order to prove (and/or reprove) many results about b-automatic sequences.

IV. Enumeration: counting b-definable properties of b-automatic sequences is b-regular

In fact, what we showed is

Proposition

Let $x \colon \mathbb{N} \to \mathbb{N}$ *be a* b*-automatic sequence and let* $y \colon \mathbb{N} \to \mathbb{N}$ *be defined as* $y(i) = 1$ *if* x *has an overlap at position* i*, and* $y(i) = 0$ *otherwise. Then* y *is* b*-automatic.*

In the same vein, we can prove that counting b-definable properties of a b-automatic sequence give rise to a b-regular sequence.

b-regular sequences

Let K be a commutative semiring. A sequence $x\colon\ \mathbb{N}^d \to K$ is (K, b)-regular if there exist

- an integer $m \geq 1$
- vectors $\lambda \in K^{1\times m}$ and $\gamma \in K^{m\times 1}$
- a morphism of monoids $\mu\colon\ ((A_b)^d)^* \to K^{m\times m}$

such that

$$\forall \mathbf{n} \in \mathbb{N}^d, \quad x(\mathbf{n}) = \lambda\mu\big(\operatorname{rep}_b(\mathbf{n})\big)\gamma.$$

The triple (λ, μ, γ) is called a linear representation of x and m is its dimension.

A useful result

Theorem

For any b-definable subset X of $\mathbb{N}^{d+1}$, the sequence $a\colon \mathbb{N}^d \to \mathbb{N} \cup \{\infty\}$ defined by

$$a(n_1, \ldots, n_d) = \mathrm{Card}\{m \in \mathbb{N} \colon (n_1, \ldots, n_d, m) \in X\}$$

is $(\mathbb{N} \cup \{\infty\}, b)$-regular. If moreover $a(\mathbb{N}^d) \subseteq \mathbb{N}$, then a is $(\mathbb{N}, b)$-regular.

Application to the factor complexity

Corollary

For any b-automatic sequence $x \colon \mathbb{N} \to \mathbb{N}$, the factor complexity of x is $(\mathbb{N}, b)$-regular.

- Let $x \colon \mathbb{N} \to \mathbb{N}$ be a b-automatic sequence.
- For all $n \in \mathbb{N}$, let $p_x(n)$ denote the number of length-n factors of x.
- Then $p_x(n) = \#\{i \in \mathbb{N} \colon \forall j < i,\ x[j, j+n-1] \neq x[i, i+n-1]\}$.
- Consider $X = \{(i, n) \in \mathbb{N}^2 \colon \forall j < i,\ x[j, j+n-1] \neq x[i, i+n-1]\}$.
- Since x is b-automatic, the set X is b-definable.
- By choice of X, we have $p_x(n) = \#\{i \in \mathbb{N} \colon (i, n) \in X\}$.
- From the previous theorem, x is $(\mathbb{N}, b)$-regular.

An open problem

What about the counting the number of rectangular factors of size (m, n) in a bidimensional b-automatic sequence? Is the corresponding bidimensional sequence $(\mathbb{N}, b)$-regular?

V. Logic and non-standard numeration systems

Fibonacci representations

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

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

A natural number n is represented by the finite word

$$\mathrm{rep}_F(n) = c_\ell \cdots c_1 c_0$$

over the alphabet $A_F = \{0, 1\}$ obtained from the greedy algorithm:

$$n = \sum_{i=0}^{\ell} c_i F_i .$$

The greedy algorithm imposes, in addition to having a nonzero leading digit c_ℓ, that the valid representations do not contain two consecutive digits 1. The set of all possible representations is

$$\mathcal{L}_F = 1\{0, 01\}^* \cup \{\varepsilon\}.$$

U-systems

Let $U = (U_i)_{i\geq 0} = (1, 2, 3, 5, 8, \ldots)$ be a base sequence, that is, an increasing sequence of positive integers satisfying:

$$U_0 = 1 \quad \text{and} \quad C_U = \sup_{i\geq 0} \frac{U_{i+1}}{U_i} < +\infty.$$

A natural number n is represented by the finite word

$$\mathrm{rep}_U(n) = c_\ell \cdots c_1 c_0$$

over the alphabet $A_U = \{0, 1, \ldots, \lceil C_U \rceil - 1\}$ obtained from the greedy algorithm:

$$n = \sum_{i=0}^{\ell} c_i U_i.$$

In this case, we talk about U-automatic sequences and U-recognizable sets of integers.

A logical framework for positional numeration systems

Two problems:

- In general, $\mathbb{N}$ is not U-recognizable.
- The addition is not recognized by finite automaton.

Pisot systems

A Pisot number is an algebraic integer > 1 such that all of its Galois conjugates have absolute value < 1.

Working Hypothesis (WH) : U satifies a linear recurrence whose characteristic polynomial is the minimal polynomial of a Pisot number.

For such systems, Frougny showed that $\mathbb{N}$ and the addition are recognizable by finite automata.

A logical framework for Pisot systems

U-definable sets are subsets of $\mathbb{N}^d$ that are definable in the logical structure $\langle \mathbb{N}, +, V_U \rangle$, where

- $+(x, y, z)$ is the ternary relation defined by $x + y = z$,
- $V_U(x)$ is the unary function defined as the smallest U_i corresponding to a nonzero digit in $\text{rep}_U(x)$ if $x \geq 1$, and $V_U(0) = 1$.

Theorem (Bruyère-Hansel 1997)

Under WH, the U-recognizable sets of integers coincide with the U-definable sets of integers.

Corollary

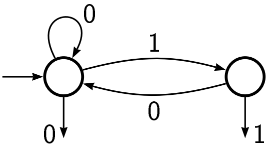
The first order theory of $\langle \mathbb{N}, +, V_U \rangle$ *is decidable*

This result implies that there exist algorithms to decide U-definable properties for U-automatic sequences.

As an application, one can prove (and reprove, or verify) many results about the Fibonacci infinite word

$$\mathbf{f} = 0100101001001010010100100101001 0\cdots$$

(which is the fixed point of $0 \mapsto 01,\ 1 \mapsto 0$).



Current work on enumeration (with Célia Cisternino and Manon Stipulanti)

What is a U-regular sequence? Several choices of definitions are possible.

In Manon Stipulanti's PhD thesis, it is proved that some sequence $S_\varphi \colon \mathbb{N} \to \mathbb{N}$ is *F-regular* by proving that there exist

- an integer $m \geq 1$
- vectors $\lambda \in K^{1\times m}$ and $\gamma \in K^{m\times 1}$
- a morphism of monoids $\mu \colon \{0, 01\}^* \to K^{m\times m}$

such that

$$\forall n \in \mathbb{N}, \quad S_\varphi(n) = \lambda \mu\big(0\,\mathrm{rep}_U(n)\big)\gamma$$

where, in order to compute $\mu\big(0\,\mathrm{rep}_U(n)\big)$, it is understood that $0\,\mathrm{rep}_U(n)$ is factored into blocks of 0 and 01.

Natural choices for U-regularity

A sequence $x\colon \mathbb{N}^d \to K$ is (K, U)-regular if there exist

- an integer $m \geq 1$
- vectors $\lambda \in K^{1\times m}$ and $\gamma \in K^{m\times 1}$
- a morphism of monoids $\mu\colon ((A_b)^d)^* \to K^{m\times m}$

such that

C1 $\forall \mathbf{n} \in \mathbb{N}^d, \quad x(\mathbf{n}) = \lambda\mu\big(\mathrm{rep}_U(\mathbf{n})\big)\gamma$

C2 $\forall w \in ((A_U)^d)^*, \quad x(\mathrm{val}_U(w)) = \lambda\mu(w)\gamma.$

Theorem (Cisternino-Charlier-Stipulanti)

Under WH, C1 $\iff$ C2.

Conjecture (analogue of the useful result)

Under WH, for any U-definable subset X of $\mathbb{N}^{d+1}$, the sequence $a\colon \mathbb{N}^d \to \mathbb{N} \cup \{\infty\}$ defined by

$$a(n_1, \ldots, n_d) = \mathrm{Card}\{m \in \mathbb{N} \colon (n_1, \ldots, n_d, m) \in X\}$$

is $(\mathbb{N} \cup \{\infty\}, U)$-regular. If moreover $a(\mathbb{N}^d) \subseteq \mathbb{N}$, then a is $(\mathbb{N}, U)$-regular.

Related works on real numbers

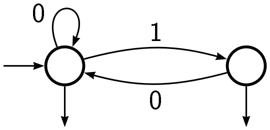
In general real numbers are represented by infinite words.

In this context, we consider Büchi automata. An infinite word is accepted when the corresponding path goes infinitely many times through an accepting state.

We talk about ω-languages and ω-regular languages.

β-recognizable and β-definable subsets of $\mathbb{R}^d$

- Notion of β-recognizability of subsets of $\mathbb{R}^d$, where $\beta > 1$ is a real base.
- For $\beta = \frac{1+\sqrt{5}}{2}$, the ω-language of the (quasi-greedy) β-representations of $[0, 1]$ is accepted by



- First order theory $\langle \mathbb{R}, +, \leq, \mathbb{Z}_\beta, X_\beta \rangle$ leading to a notion of β-definability.
- For β Pisot, β-recognizability coincide with β-definability.
- For β Pisot, the first order theory of $\langle \mathbb{R}, +, \leq, \mathbb{Z}_\beta, X_\beta \rangle$ is decidable.

Deciding topological properties

For β Pisot, the following properties of β-recognizable subsets X of $\mathbb{R}^d$ are decidable:

- X has a nonempty interior:

$$(\exists x \in X)\ (\exists \varepsilon > 0)\ (\forall y)\ (|x - y| < \varepsilon \implies y \in X).$$

- X is open:

$$(\forall x \in X)\ (\exists \varepsilon > 0)\ (\forall y)\ (|x - y| < \varepsilon \implies y \in X).$$

- X is closed: OK as $\mathbb{R}^d \setminus X$ is b-recognizable.
- ...

References

- Bruyère-Hansel-Michaux-Villemaire, Logic and p-recognizable sets of integers, Bull. Belg. Math. Soc. 1 (1994) 191–238.
- Bruyère-Hansel, Bertrand numeration systems and recognizability, TCS 181 (1997) 17–43.
- Charlier, First-Order Logic and Numeration Systems. Sequences, Groups, and Number Theory, Chapter 3, 89–141. Trends in Mathematics, Birkhaüser/Springer, Cham, 2018.
- Charlier-Leroy-Rigo, An analogue of Cobham's theorem for graph directed iterated function systems. Adv. in Math. 280 (2015) 86–12.
- Charlier-Rampersad-Shallit, Enumeration and decidable properties of automatic sequences. IJFCS 23(5) (2012) 1035–1066.
- Du-Mousavi-Rowland-Schaeffer-Shallit, Decision algorithms for Fibonacci-automatic words, II: Related sequences and avoidability, TCS 657 (2017) 146–162.
- Leroy-Rigo-Stipulanti, Counting the number of non-zero coefficients in rows of generalized Pascal triangles, Discrete Math. 340 (2017) 862–881.