Automata and $\mathrm{FO}[\mathbb{N}, <, \mathrm{mod}]$-definable integer relations

Arthur MILCHIOR

IRIF
Université Paris Diderot, France
LACL, UPEC, Créteil, France

17 juin 2016
Séminaire automate

The main result

Theorem

It is decidable in linear time whether $R \subseteq \mathbb{N}^d$, accepted by a minimal automaton in base $b \geq 2$ is accepted by an automaton in base 1.

Outline:

Introduction

Definitions

The FO[<,mod]-definable sets

Similar problems

Two tools

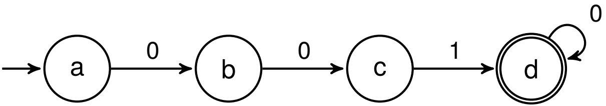
Representation of d-tuples of integers.

	Base 2	Base 1
7	1110	1111111_
8	0001	11111111
$(n, n+1)$	$(1,0)^*(0,1)\{(0,0)+(1,1)\}^*$	$(1,1)^*(\_,1)$
$2\mathbb{N}$	$0\{0,1\}^*$	$(11)^*$

Exponential explosion: Example $\{2^i\}$

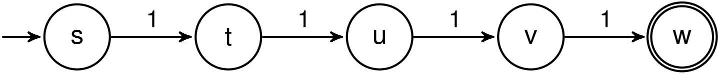
Let $R_i = \{2^i\}$.
Minimal automaton accepting R_i in base 2 has $i + 2$ states.

R_2 in base 2



Minimal automaton accepting R_i in base 1 has $2^i + 1$ states.

R_2 in base 1

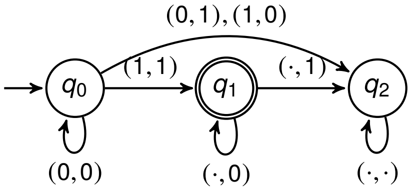


Known results relating automata and logics

Logic	Automaton in base	
FO$[+, V_b]$	$b \geq 2$	Büchi 60, Bruyère 85
FO$[+]$	all $b \geq 2$	Cobham 69, Semenov 77 (Michaux-Villemaire 96)
FO$[<, \text{mod}]$	1	Straubing 91
FO$[+1, \text{mod}]$		

$V_b(n)$: greatest power of b dividing n.

Example: $V_2(000101001110) = 0001$.



Dimension 1

Definition

A state $R \subseteq \mathbb{N}$ is ultimately periodic if there exists a threshold $t \in \mathbb{N}$, and a period p such that for all $n \geq t$, $(n \in R) \Longleftrightarrow (n + p \in R)$.

It is equivalent to state that a set $R \subseteq \mathbb{N}$ is:

–FO$[+]$-definable,

–FO$[<, \mathrm{mod}]$-definable,

–ultimately periodic.

Example

$$R = 3\mathbb{N} \cup \{4\} = \{0, 3, 4, 6, 9, \dots\}$$

R admits threshold $t = 5$ and periodicity $p = 3$. It is defined by:

$$\phi(x) = \{x = 4 \vee x \equiv 0 \mod 3\}.$$

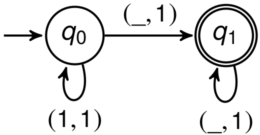
Characterization of FO$[<, \mathrm{mod}]$

Definition (Regular set)

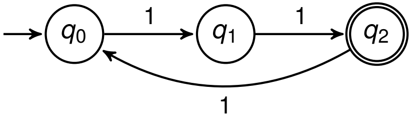
A regular set is a set accepted in base 1.

Theorem (Straubing 91)

A set is FO$[<, \mathrm{mod}]$*-definable if and only if it is regular.*



(a) $x_0 < x_1$



(b) $3\mathbb{N} + 2$

Quantifier elimination

Theorem (From Cooper 72, mentioned in Smoryński 91)

The logic $\mathrm{FO}[=_{\mathbb{N}}, +_{\mathbb{N}}, <, \mathrm{mod}]$ *admits quantifier elimination.*

Example

$$\exists x.(x = y \wedge x = 0) \vee (x + 4 \leq y \wedge x \equiv 0 \mod 2).$$

equivalent to

$$\begin{aligned}
&\bigvee_{i=-2}^{2}\{(\quad i = y \wedge \quad i = 0) \vee (\quad i + 4 \leq y \wedge \quad i \equiv 0 \mod 2)\}\vee \\
&\bigvee_{i=-6}^{-2}\{(y+i = y \wedge y+i = 0) \vee (y+i + 4 \leq y \wedge y+i \equiv 0 \mod 2)\}\vee \\
&\bigvee_{i=-2}^{2}\{(y+i = y \wedge y+i = 0) \vee (y+i + 4 \leq y \wedge y+i \equiv 0 \mod 2)\}.
\end{aligned}$$

Theorem (Peladeau 92)

The class $\mathcal{R}$ of regular set is maximal such that $\mathrm{FO}[\mathcal{R}, (P_a)_{a \in A}]$ *only defines regular languages.*

Example

$$\phi = \exists m. m \times 2 = \text{last} \wedge \forall y. P_b(y) \Longleftrightarrow y = m$$

Satisfied by:

a	a	b	a	a
0	1	2	3	4
		m		

Not satisfied by:

a	a	a	a
0	1	2	3

The formula ϕ defines the non-regular language $\{a^n b a^n \mid n \in \mathbb{N}\}$.
Hence $\times 2$ is not $\mathrm{FO}[<, \mathrm{mod}]$-definable.

Theorem

The class $\mathcal{R}$ of regular set is the maximal fragment of $\mathrm{FO}[+]$ *such that the satisfiability of* $\exists\mathrm{MSO}[<,\mathcal{R}]$ *is decidable.*

This result does not hold for logics which are not fragment of $\mathrm{FO}[+]$.

The main result

Input: a minimal automaton $\mathcal{A}$ in base $b \geq 2$.

Theorem (Decision algorithm)

It is decidable whether R accepted by $\mathcal{A}$ *is* $\mathrm{FO}[<, \mathrm{mod}]$*-definable is decidable in linear time.*

Theorem (Construction algorithm)

If R is $\mathrm{FO}[<, \mathrm{mod}]$*-definable, an existential* $\mathrm{FO}[<, \mathrm{mod}]$*-formula defining R can be computed in time* $O\left(n^3 \log(n)\right)$.

Outline

Similar problem for FO[+]

Theorem

It is decidable whether $R \subseteq \mathbb{N}^d$ accepted by $\mathcal{A}$ is FO[+]*-definable is decidable.*

dimension d	*time complexity*	
1		*(Honkala 86)*
any	*3EXP*	*(Muchnik 91)*
any	*polynomial*	*(Leroux 06)*
1	*quasi-linear*	*(Marsault-Sakarovitch 13)*

Honkala's algorithm

Theorem (Honkala 86)

It is decidable whether $R \subseteq \mathbb{N}$ accepted by $\mathcal{A}$ is $\mathrm{FO}[+]$*-definable is decidable.*

Lemma

The threshold t and the period p is at most b^n where n is the number of states of $\mathcal{A}$.

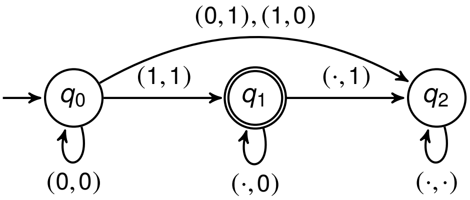
Algorithm

(1) Runs over all sets $R \subseteq \mathbb{N}$ with threshold and period less than b^n.

(2) Generates the minimal automaton $\mathcal{A}_R$ in base b accepting R.

(3) Accepts if $\mathcal{A}_R = \mathcal{A}$.

Honkala's example



If this automaton accepts an ultimately periodic set, the threshold and period are at most $2^3 = 8$.

Proposition

This methods allows to recognize any class C of languages such that:

–there exists a function s from language to integers,

–for all language $L \in C$, the minimal automaton accepting L has at least $s(L)$ states,

–the set $\{L \mid s(L)\}$ is finite and computable.

For C the ultimately periodic sets, take $s(R) = \max(t, p)$.

Muchnik's algorithm

Theorem (Muchnik 91)

There exists $\phi_d \in \mathrm{FO}[+, R]$ which states "$R \subseteq \mathbb{N}^d$ is $\mathrm{FO}[+]$ – definable".

Corollary (Muchnik 91)

It is decidable whether $R \subseteq \mathbb{N}^d$ accepted by $\mathcal{A}$ is $\mathrm{FO}[+]$-definable is decidable in 4EXP-time.

Algorithm

(1) Transform ϕ_d into an automaton $\mathcal{A}'$ where R is encoded by $\mathcal{A}$.

(2) Accepts if $\mathcal{A}'$ accepts a non-empty language.

Muchnik's example

This method works for any class C of set of tuple of natural integers such that there exists a $\phi \in \mathrm{FO}[+, V_b, R]$ which holds if and only if $R \in C$.

Example

It is decidable whether an automaton accepts a subsemigroup of $(\mathbb{N}^d, +)$.

$$\forall \begin{array}{l} x_1, \ldots, x_d \\ y_1, \ldots, y_d \end{array} . \left\{ \begin{array}{l} (x_1, \ldots, x_d) \in R \wedge \\ (y_1, \ldots, y_d) \in R \end{array} \right\} \Longrightarrow (x_1 + y_1, \ldots, x_d + y_d) \in R$$

asserts that R is a subsemigroup of $(\mathbb{N}^d, +)$.

Leroux's algorithm

Theorem (Leroux 06)

It is decidable whether $R \subseteq \mathbb{N}^d$ accepted by $\mathcal{A}$ is $\mathrm{FO}[+]$*-definable is decidable in polynomial time.*

Theorem

It is decidable whether R accepted by $\mathcal{A}$ is $\mathrm{FO}[<, \mathrm{mod}]$*-definable is decidable in* 2EXP*-time.*

(1) Compute a polynomial-time $\mathrm{FO}[+]$-formula (Leroux 06),

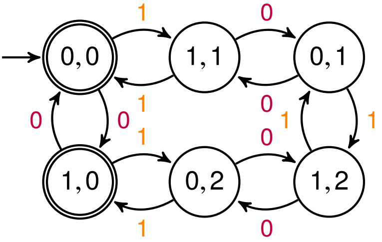
(2) Deciding whether this formula is equivalent to a $\mathrm{FO}[<, \mathrm{mod}]$-formula (Choffrut 08).

Marsault-Sakarovitch's algorithm

Theorem (Marsault-Sakarovitch 13)

It is decidable whether $R \subseteq \mathbb{N}$ accepted by $\mathcal{A}$ is FO[+]*-definable is decidable in linear time.*

Pascal automata - $3\mathbb{N}$



States:	$\{0,1\} \times \{0,1,2\}$
	(length, value)
Initial state:	$(0,0)$
Transition:	$\delta((l,v),a)$
	$= (l+1, v+2^l a)$
Final states:	$\{0,1\} \times \{0\}$.

$$(101 \cdot 1)_2 \equiv 101_2 + 1 \times 2^{|101|} \equiv 5 + 1 \times 2^3 \equiv 2 + 1 \times 2^1 \equiv 1 \mod 3.$$

General method

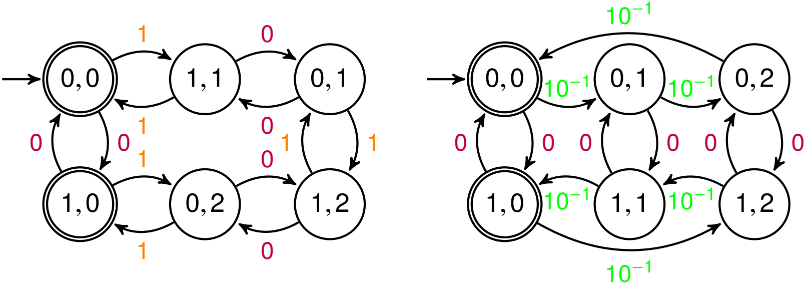
Proposition

Let $\mathbb{L}$ be a class of language such that:

1. *each language of $\mathbb{L}$ is accepted by an automaton of $\mathbb{A}$,*
2. *all automata of $\mathbb{A}$ accepts a language belonging to $\mathbb{L}$,*
3. *$\mathbb{A}$ is closed under quotient and*
4. *it is decidable in time $t(n)$ whether an automaton belongs to $\mathbb{A}$.*

It is decidable in time $t(n)$ whether a minimal automaton accepts a language of $\mathbb{L}$.

The letter 0^{-1}

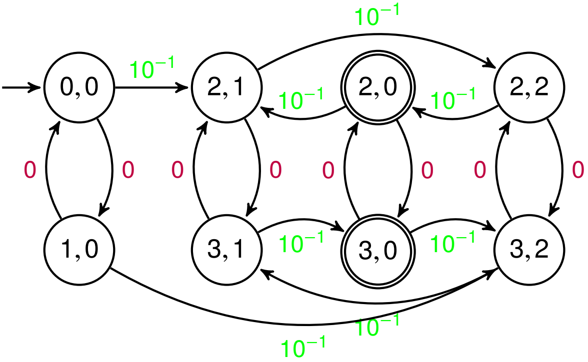


Lemma

Assuming m coprime with b, each state belongs to a 0*-cycle.*

The length of the (10^{-1})*-cycle is the periodicity.*

$3\mathbb{N}^{>0}$

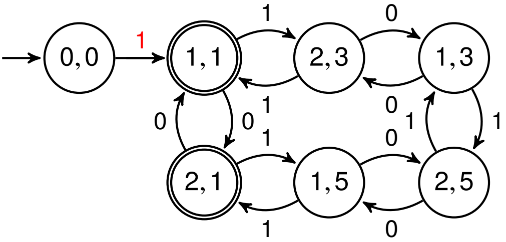


Outline

First step

Proposition (Leroux 06)

An automaton A accepts a set FO[+]*-definable if and only if A_q is* FO[+]*-definable for all q accessible in a step.*



$$\overline{\mathcal{A}_{0,0}}^{\mathbb{N}} = 1 + 2(\overline{\mathcal{A}_{1,1}}^{\mathbb{N}}) = 1 + 2(3\mathbb{N}) = 1 + 6\mathbb{N}$$

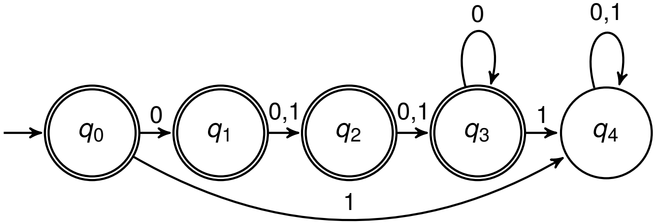
where $\overline{\mathcal{A}_{l,v}}^{\mathbb{N}}$ denotes the set of integers accepted when (l, v) is the initial state.

The initially-cyclic case

Corollary

An automaton A accepts a set $\mathrm{FO}[+]$*-definable if and only if A_q is* $\mathrm{FO}[+]$*-definable for each cyclic q.*

Also holds for $\mathrm{FO}[<, \mathrm{mod}]$.



(a) $\{0, 2, 4, 6\}$

$\mathcal{A}_{q_3}$ accepts $\{0\}$ and $\mathcal{A}_{q_4}$ accepts $\emptyset$:
$\mathcal{A}$ accepts a $\mathrm{FO}[<, \mathrm{mod}]$-definable set.

Left quotient of formulas

Proposition (Boudet, Comon 96)

Let $\phi(x_1, \ldots, x_d)$ *be a* FO$[+]$*-formula defining a set* $R \subseteq \mathbb{N}^d$*.*

Let $L \subseteq (\{0,1\}^d)^*$ *be the binary expansion of elements of* R*.*

Let $(a_1, \ldots, a_d) \in \{0,1\}^d$*.*

$(a_1, \ldots, a_d)^{-1}L$ *is defined by the* FO$[+]$*-formula:*

$$(a_1, \ldots, a_d)^{-1}\phi = \phi(a_1 + 2x_1, \ldots, a_d + 2x_d).$$

This result also holds for FO$[<, \mathrm{mod}]$.

$$(0,0)^{-1}(x_0 + 2 = x_1) \equiv (0 + 2x_0 + 2 = 0 + 2x_1) \equiv (x_0 + 1 = x_1)$$
$$(1,0)^{-1}(x_0 + 2 = x_1) \equiv (1 + 2x_0 + 2 = 0 + 2x_1) \equiv \text{false}$$

Conclusion

Can be tested with
`https://github.com/Arthur-Milchior/RegAut`

This algorithm also works for FO$[+1, \mathrm{mod}]$ and $\Sigma_0[=\mathbb{N}, <]$.

Open problems

Considering FO$[<]$.

Considering most-significant-digit first automata.

Applying similar method to real automata.