Weakly-unambiguous Parikh automata and their link with holonomic series

Alin Bostan, Arnaud Carayol, Florent Koechlin, Cyril Nicaud

LIGM UMR 8049 CNRS

May 2020, 12th

Link between languages and combinatorics

$$L(x) = \sum_{w \in L} x^{|w|} = \sum_{n \in \mathbb{N}} \ell_n x^n$$

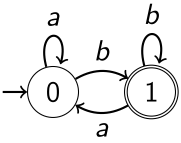
ℓ_n : number of words of length n

Formal languages		Generating series
L	$\longrightarrow$	$L(x)$

Link between languages and combinatorics

$$L(x) = \sum_{w \in L} x^{|w|} = \sum_{n \in \mathbb{N}} \ell_n x^n \qquad \ell_n : \text{number of words of length } n$$

Formal languages		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $L(x) = P(x)/Q(x)$



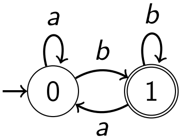
$$\begin{cases} q_0(x) = xq_0(x) + xq_1(x) \\ q_1(x) = 1 + xq_1(x) + xq_0(x) \end{cases}$$

$$L(x) = \frac{x}{1-2x}$$

Link between languages and combinatorics

$$L(x_1, \ldots, x_r) = \sum_{w \in L} x_1^{|w|_{a_1}} \ldots x_r^{|w|_{a_r}} \qquad \Sigma = \{a_1, \ldots, a_r\}$$

Formal languages		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $L(x) = P(x)/Q(x)$



$$\begin{cases} q_0(x_a, x_b) = x_a q_0(x_a, x_b) + x_b q_1(x_a, x_b) \\ q_1(x_a, x_b) = 1 + x_b q_1(x_a, x_b) + x_a q_0(x_a, x_b) \end{cases}$$

$$L(x_a, x_b) = \frac{x_b}{1-(x_a+x_b)}$$

Link between languages and combinatorics

$$L(x) = \sum_{w \in L} x^{|w|} = \sum_{n \in \mathbb{N}} \ell_n x^n$$

ℓ_n : number of words of length n

Formal languages		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $L(x) = P(x)/Q(x)$
Unambiguous context-free	$\longrightarrow$	algebraic $P(x, L(x)) = 0$

$$\begin{cases} S \to aSB \mid \varepsilon \\ B \to cB \mid bS \end{cases} \qquad \begin{cases} S(x) = xS(x)B(x) + 1 \\ B(x) = xB(x) + xS(x) \end{cases}$$

$$x^2 S(x)^2 - (1 - x)S(x) + 1 - x = 0$$

Link between languages and combinatorics

$$L(x_1, \ldots, x_r) = \sum_{w \in L} x_1^{|w|_{a_1}} \ldots x_r^{|w|_{a_r}}$$

$$\Sigma = \{a_1, \ldots, a_r\}$$

Formal languages		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $L(x) = P(x)/Q(x)$
Unambiguous context-free	$\longrightarrow$	algebraic $P(x, L(x)) = 0$

$$\begin{cases} S \to aSB \mid \varepsilon \\ B \to cB \mid bS \end{cases}$$

$$\begin{cases} S(\vec{x}) = x_a S(\vec{x}) B(\vec{x}) + 1 \\ B(\vec{x}) = x_c B(\vec{x}) + x_b S(\vec{x}) \end{cases}$$

$$x_a x_b S(x_a, x_b, x_c)^2 - (1 - x_c) S(x_a, x_b, x_c) + 1 - x_c = 0$$

Link between languages and combinatorics

$$L(x_1, \ldots, x_r) = \sum_{w \in L} x_1^{|w|_{a_1}} \ldots x_r^{|w|_{a_r}} \qquad \Sigma = \{a_1, \ldots, a_r\}$$

Formal languages		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $L(x) = P(x)/Q(x)$
ambiguous context-free	$\longrightarrow$	?

$$\begin{cases} S \rightarrow aSB \mid \varepsilon \\ B \rightarrow cB \mid bS \end{cases} \qquad \begin{cases} S(\vec{x}) = x_a S(\vec{x}) B(\vec{x}) + 1 \\ B(\vec{x}) = x_c B(\vec{x}) + x_b S(\vec{x}) \end{cases}$$

$$x_a x_b S(x_a, x_b, x_c)^2 - (1 - x_c) S(x_a, x_b, x_c) + 1 - x_c = 0$$

Link between languages and combinatorics

$$L(x) = \sum_{w \in L} x^{|w|} = \sum_{n \in \mathbb{N}} \ell_n x^n$$

ℓ_n : number of words of length n

Formal languages		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $L(x) = P(x)/Q(x)$
Unambiguous context-free	$\longrightarrow$	algebraic $P(x, L(x)) = 0$

$$\frac{1 - 2x + 225x^2}{(1 - 25x)(625x^2 + 14x + 1)} = 1+9x+49x^2+\dots$$ [Bousquet-Mélou 08]

$$G(x) = 1+2x+11x^2+\dots$$ [Bostan & Kauers 10, Drmota & Banderier 13]

Analytic criteria for inherent ambiguity

Theorem (Chomsky and Schützenberger 63)

The generating series of an unambiguous context-free language is algebraic.

Contraposition

If the generating series of a context-free language is not algebraic, then it is inherently ambiguous.

Detailed Example

Example (Flajolet 87)

$\mathcal{D} = \{a^{n_1} b\, a^{n_2} b \ldots a^{n_k} b \;:\; k \in \mathbb{N}^*,\; n_1 = 1 \text{ and } \exists j < k, n_{j+1} \neq 2n_j\}$ is inherently ambiguous.

- $aab \notin \mathcal{D}$
- $abaabaaab \in \mathcal{D}$
- $abaabaaaab \notin \mathcal{D}$
- $ab\; a^2 b\; a^4 b \ldots a^{2^{k-1}} b \notin \mathcal{D}$

Detailed Example

Example (Flajolet 87)

$\mathcal{D} = \{a^{n_1} b\, a^{n_2} b \dots a^{n_k} b \;:\; k \in \mathbb{N}^*,\ n_1 = 1 \text{ and } \exists j < k, n_{j+1} \neq 2n_j\}$ is inherently ambiguous.

- By contradiction, suppose $\mathcal{D}$ is unambiguous. Then $D(x)$ is algebraic
- Aim: build from $D(x)$ a series that is not algebraic and use closure properties

Detailed Example

Example (Flajolet 87)

$\mathcal{D} = \{a^{n_1} b\, a^{n_2} b \dots a^{n_k} b \;:\; k \in \mathbb{N}^*,\; n_1 = 1 \text{ and } \exists j < k, n_{j+1} \neq 2n_j\}$ is inherently ambiguous.

- By contradiction, suppose $\mathcal{D}$ is unambiguous. Then $D(x)$ is algebraic
- Aim: build from $D(x)$ a series that is not algebraic and use closure properties
- $\mathcal{B} = ab(ab^*)^* \setminus \mathcal{D} = \{ab\; a^2 b\; a^4 b \dots a^{2^{k-1}} b \;:\; k \in \mathbb{N}^*\}$
- $B(x) = \frac{x^2}{1 - \frac{x}{1-x}} - D(x) = \text{algebraic}$
- So $B(x) = \sum_{k \geq 1} x^{2^k - 1 + k}$, which is lacunary
- So $B(x)$ is not algebraic. Contradiction

Remarks on this method

- Analytic criteria for solving some instances of an undecidable problem

- It can avoid technical proofs on automata based on pumping techniques.

- $L = \{a^n b^m c^p \ : \ n = m \text{ or } m = p\}$ is inherently ambiguous as a CF language yet $L(x) = \frac{2}{(1-x^2)(1-x)} - \frac{1}{1-x^3}$ is rational

- Specific about inherent ambiguity questions.
  $\rightarrow$ language of primitive words $\mathcal{L}_P$

  $$aabb \in \mathcal{L}_P, abab \notin \mathcal{L}_P$$

  CFL: open   not unambiguous CFL: [Peterson 96]

Hierarchy of languages and series

Language		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $Q(x)L(x) = P(x)$
$\not\supseteq$		$\not\supseteq$
Unambiguous context-free	$\longrightarrow$	algebraic $P(x, L(x)) = 0$
		$\not\supseteq$
?	$\longrightarrow$	holonomic $P(x, \partial_x) \cdot L(x) = 0$

Holonomic series in one variable (Stanley 80)

A series $f(x) = \sum_n a_n x^n$ is holonomic (or D-finite) if it satisfies a differential equation of the form:

$$P_k(x) f^{(k)}(x) + \ldots + P_0(x) f(x) = 0 \quad \text{with } P_i(x) \in \mathbb{Q}[x]$$

Equivalently a_n satisfies a linear recurrence of the form

$$p_r(n) a_{n+r} + \ldots + p_0(n) a_n = 0 \quad \text{with } p_i(n) \in \mathbb{Q}[n]$$

Closed by sum, product, composition with algebraic series, Hadamard product...

Example of holonomic series

- rational series $F = P/Q$: $(PQ)F' + (PQ' - P'Q)F = 0$
  $\rightarrow$ Linear recurrence with constant coefficients

- algebraic series (the proof is however not straightforward)
  $F(x) = \sqrt{1-x} := \sum \frac{4^{-n}}{1-2n}\binom{2n}{n}x^n$ satisfies $F^2 - 1 - x = 0$
  $2(1-x)F' - F = 0$
  $2(n+1)u_{n+1} - (2n+1)u_n = 0$

- $F(x) = e^x := \sum x^n/n!$ is holonomic but is not algebraic
  $F' - F = 0$
  $(n+1)u_{n+1} - u_n = 0$

Holonomic series in several variables (Lipshitz 89)

A series $f(x_1, \ldots, x_n)$ is holonomic (or D-finite) if it satisfies a system of partial derivative equations of the form:

$$\begin{cases} A_{1,r_1}(\vec{x})\ \partial_{x_1}^{r_1} f(\vec{x}) + \ldots + A_{1,1}(\vec{x})\ \partial_{x_1} f(\vec{x}) + A_{1,0}(\vec{x})\ f(\vec{x}) = 0 \\ \vdots \\ A_{n,r_n}(\vec{x})\ \partial_{x_n}^{r_n} f(\vec{x}) + \ldots + A_{n,1}(\vec{x})\ \partial_{x_n} f(\vec{x}) + A_{n,0}(\vec{x})\ f(\vec{x}) = 0 \end{cases}$$

with $A_{i,j}(\vec{x}) \in \mathbb{Q}[\vec{x}]$, and $\vec{x} = (x_1, \ldots, x_n)$.

We only use closure properties rather than the definition

Holonomic series in several variables

Theorem (Lipshitz 1988, 1989)

Holonomic series are closed under :

1. *arithmetic operations* $+, \times, -$
2. *specialization to 1, when it is well-defined: if* $f(x_1, \dots, x_n)$ *is holonomic, then* $f(x, 1, \dots, 1)$ *is holonomic too*
3. *Hadamard's product* $\odot$

$$f(x_1, \dots, x_n) = \sum_{\mathbf{i} \in \mathbb{N}^n} a(i_1, \dots, i_n) x_1^{i_1} \dots x_n^{i_n}$$

$$g(x_1, \dots, x_n) = \sum_{\mathbf{i} \in \mathbb{N}^n} b(i_1, \dots, i_n) x_1^{i_1} \dots x_n^{i_n}$$

$$f \odot g(x_1, \dots, x_n) = \sum_{\mathbf{i} \in \mathbb{N}^n} a(i_1, \dots, i_n) b(i_1, \dots, i_n) x_1^{i_1} \dots x_n^{i_n}$$

Crucial particular case: support series

Let $\mathcal{S} \subseteq \mathbb{N}^n$. The support series of $\mathcal{S}$ is

$$g(x_1, \dots, x_n) = \sum_{(i_1,\dots,i_n)\in\mathcal{S}} x_1^{i_1} \dots x_n^{i_n}$$

Let $f(x_1, \dots, x_n) = \sum_{(i_1,\dots,i_n)\in\mathbb{N}^n} a(i_1, \dots, i_n) x_1^{i_1} \dots x_n^{i_n}$. Then:

$$(f \odot g)(x_1, \dots, x_n) = \sum_{(i_1,\dots,i_n)\in\mathcal{S}} a(i_1, \dots, i_n) x_1^{i_1} \dots x_n^{i_n}$$

Example of Hadamard's product

Example

$\Omega_3 = \{w \in (a+b+c)^* \ : \ |w|_a \neq |w|_b \text{ or } |w|_b \neq |w|_c\}.$

- $abbca \in \Omega_3$, $abbcca \notin \Omega_3$.
- Ω_3 is context-free, inherently ambiguous as a CFL.

$$\Omega_3(x_a, x_b, x_c) = \underbrace{\frac{1}{1-(x_a+x_b+x_c)}}_{(a+b+c)^*} \odot \underbrace{\left(\frac{1}{(1-x_a)(1-x_b)(1-x_c)} - \frac{1}{1-x_a x_b x_c}\right)}_{|w|_a \neq |w|_b \text{ or } |w|_b \neq |w|_c}$$

$$= \frac{1}{1-(x_a+x_b+x_c)} - \frac{1}{1-(x_a+x_b+x_c)} \odot \frac{1}{1-x_a x_b x_c}$$

Example of Hadamard's product

Example

$\Omega_3 = \{w \in (a+b+c)^* \; : \; |w|_a \neq |w|_b \text{ or } |w|_b \neq |w|_c\}.$

$$\frac{1}{1-(x_a+x_b+x_c)} \odot \frac{1}{1-x_a x_b x_c} = [y_a^{-1} y_b^{-1} y_c^{-1}] \frac{1}{y_a y_b y_c} \frac{1}{1-(\frac{x_a}{y_a}+\frac{x_b}{y_b}+\frac{x_c}{y_c})} \frac{1}{1-y_a y_b y_c}$$

Mgfun [Chyzak] and gfun [Salvy and Zimmermann] give:

$$p_3(\vec{x})\partial_{x_a}^3 \Omega_3(\vec{x}) + p_2(\vec{x})\partial_{x_a}^2 \Omega_3(\vec{x}) + p_1(\vec{x})\partial_{x_a} \Omega_3(\vec{x}) + p_0(\vec{x})\Omega_3(\vec{x}) = 0$$

with $\|p_i\|_\infty \leq 7344$ and $\deg(p_i) \leq 9$.

Example of Hadamard's product

Example

$\Omega_3 = \{w \in (a+b+c)^* \ : \ |w|_a \neq |w|_b \text{ or } |w|_b \neq |w|_c\}.$

$$\frac{1}{1-(x_a+x_b+x_c)} \odot \frac{1}{1-x_a x_b x_c} = [y_a^{-1} y_b^{-1} y_c^{-1}] \frac{1}{y_a y_b y_c} \frac{1}{1-(\frac{x_a}{y_a}+\frac{x_b}{y_b}+\frac{x_c}{y_c})} \frac{1}{1-y_a y_b y_c}$$

Mgfun [Chyzak] and gfun [Salvy and Zimmermann] give:

$$p_3(\vec{x})\partial_{x_a}^3\Omega_3(\vec{x}) + p_2(\vec{x})\partial_{x_a}^2\Omega_3(\vec{x}) + p_1(\vec{x})\partial_{x_a}\Omega_3(\vec{x}) + p_0(\vec{x})\Omega_3(\vec{x}) = 0$$

with $\|p_i\|_\infty \leq 7344$ and $\deg(p_i) \leq 9$.

Remark (Flajolet 87)

$\Omega_3(x_a, x_b, x_c)$ is holonomic but not algebraic.

Previous attempts at a link with formal languages

- [Lipshitz 88] added linear constraints to the support of a holonomic series using a Hadamard product with a support series
- [Massazza 93] formalized the idea with (semi)linear constraints (Linear Constrained Languages)
- [Castiglione and Massazza 2017] RCM (Regular languages with semilinear Constraints and a (injective) Morphism)
  ex: $a^n b^m a^n b^m$

$\rightarrow$ not fully satisfactory from an automaton point of view.
Conjectured a link with deterministic Reversal Bounded Counter Machines.

Hierarchy of languages and series

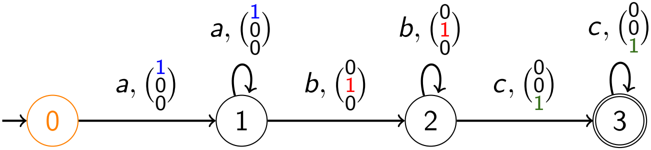
Language		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $Q(x)L(x) = P(x)$
$\rotatebox{90}{$\subsetneq$}$		$\rotatebox{90}{$\subsetneq$}$
Unambiguous context-free	$\longrightarrow$	algebraic $P(z, L(x)) = 0$
$\rotatebox{90}{$\subsetneq$}$		$\rotatebox{90}{$\subsetneq$}$
Weakly-unambiguous Pushdown PA	$\longrightarrow$	holonomic $P(x, \partial_x) \cdot L(x) = 0$

Hierarchy of languages and series

Language		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $Q(x)L(x) = P(x)$
$\subsetneq$		$\subsetneq$
Unambiguous context-free	$\longrightarrow$	algebraic $P(z, L(x)) = 0$
$\subsetneq$		$\subsetneq$
Weakly-unambiguous Pushdown PA	$\longrightarrow$	holonomic $P(x, \partial_x) \cdot L(x) = 0$

For the presentation we work with PA and not Pushdown PA.

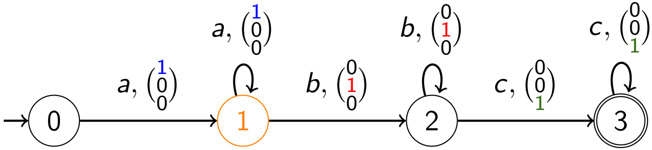
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

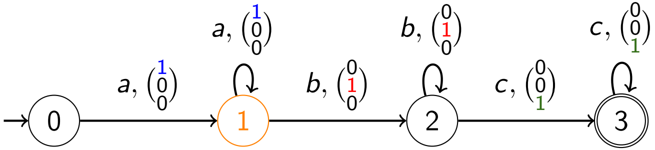
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

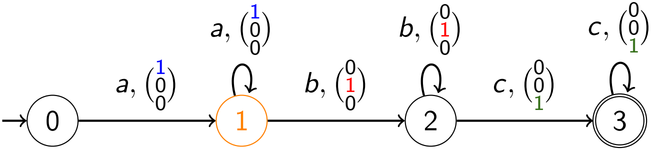
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$$

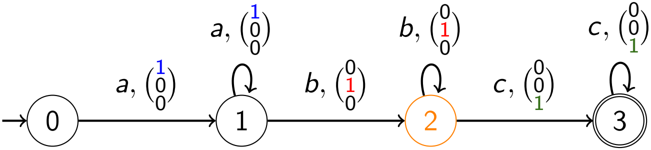
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$$

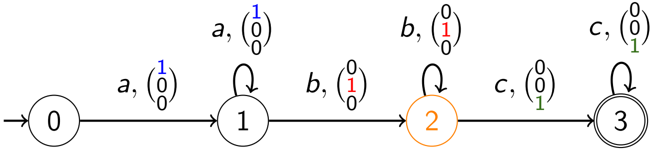
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}$$

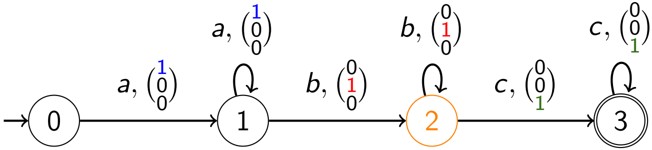
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}$$

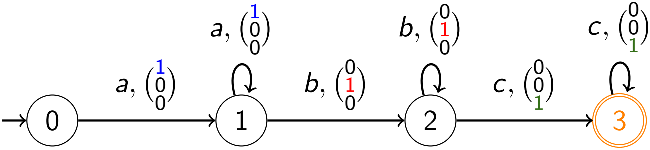
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix}$$

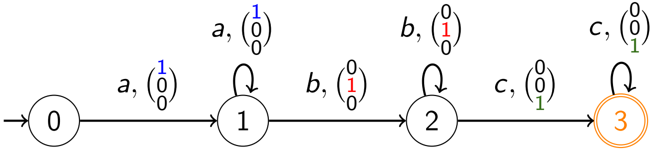
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n)\ :\ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}$$

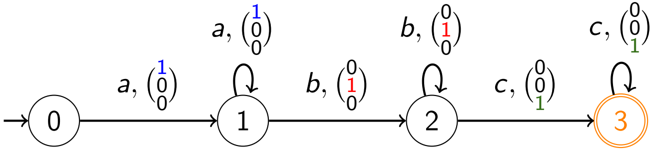
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 3 \\ 2 \end{pmatrix}$$

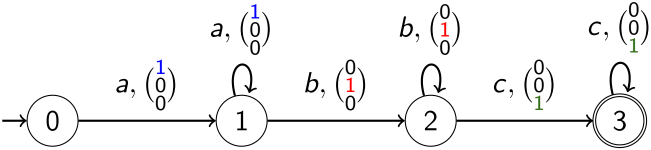
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \; : \; n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$$

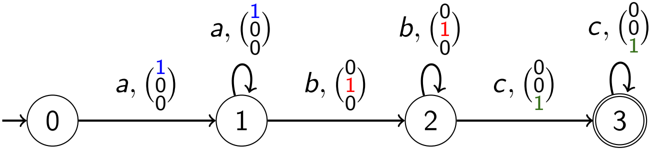
Parikh automata [Klaedtke and Rueß 03]



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

$$w = aaabbbccc \longrightarrow \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} \in C$$

$$w \in L(\mathcal{A})$$



$$C = \{(n, n, n) \ : \ n \in \mathbb{N}^*\}$$

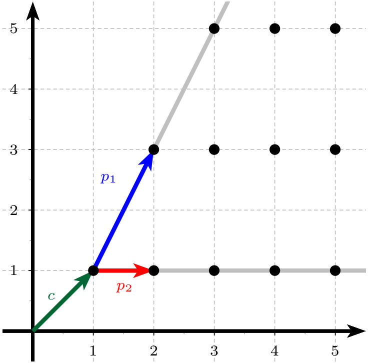
$$\ell = \{(a^n b^m c^p, \begin{pmatrix} n \\ m \\ p \end{pmatrix}) \ : \ n, m, p \in \mathbb{N}^*\}$$

$$L(\mathcal{A}) = \{a^n b^n c^n \ : \ n \in \mathbb{N}^*\}$$

Semilinear sets of $\mathbb{N}^d$ (Parikh 1966)

- Intuitively: boolean combinaison of linear (affine) inequalities defining subsets of $\mathbb{N}^d$
  $x_1 - x_2 = 0 \wedge x_2 - x_3 = 0 \rightarrow C = \{(n, n, n) : n \in \mathbb{N}\}$
- More generally, subsets defined by the Presburger arithmetic [Ginsburg and Spanier 66]
  $\Phi(x_1, x_2) := \exists x, x_1 - 3x = 0 \wedge 1 + 2x_1 - x_2 = 0$
  $\rightarrow \{(3n, 6n + 1) : n \in \mathbb{N}\}$

Semilinear sets of $\mathbb{N}^d$ (Parikh 66)



Semilinear = Finite union of linear sets $\vec{c} + P^*$ where
$P = \{p_1, \ldots, p_r\}$ and $P^* = \{\lambda_1 p_1 + \ldots + \lambda_r p_r \ : \ \lambda_i \in \mathbb{N}\}$

Semilinear sets of $\mathbb{N}^d$ (Parikh 66)

Theorem (Eilenberg and Schützenberger 69, Ito 69)

If C is semilinear, then $C(x_1, \ldots, x_d) = \sum_{\vec{v} \in C} x_1^{v_1} \ldots x_d^{v_d}$, its support series, is rational.

If $C = \bigcup_{i=1}^{m} \vec{c_i} + P_i^*$ is an unambiguous description of C:

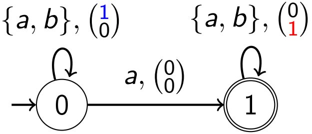
$$C(x_1, \ldots, x_d) = \sum_{i=1}^{m} \frac{\vec{x}^{c_i}}{\prod_{p \in P_i} (1 - \vec{x}^{p})}$$

Remark

In the sequel we will deal with holonomic series of the form $f \odot C$, where C is the support series of a semilinear set.

Weakly-unambiguous Parikh automaton

Weakly-unambiguous: at most one accepting run for every word.



$$C = \{(n, n) \ : \ n \in \mathbb{N}\}$$

$L(\mathcal{A}) = \{w_1 a w_2 \ : \ |w_1| = |w_2|, w_1, w_2 \in \Sigma^*\}$ with $\Sigma = \{a, b\}$.

$\neq$ [Cadilhac, Finkel and McKenzie 13] Unambiguous constraint automata

Weakly-unambiguous Parikh automata

- PA coincide with the class of Reversal Bounded Counter Machines [Klaedtke and Rueß 03]
- Deterministic versions do not coincide.

- Weakly-unambiguous PA coincide with the class of unambiguous RBCM...
  ...and the class of RCM languages!

- Weakly-unambiguous PA are closed under intersection, and left quotient with words.
- Closure under union? Complement? Still open.

Weakly-unambiguous Parikh automata

- PA coincide with the class of Reversal Bounded Counter Machines [Klaedtke and Rueß 03]
- Deterministic versions do not coincide.

- Weakly-unambiguous PA coincide with the class of unambiguous RBCM...
  ...and the class of RCM languages!

- Weakly-unambiguous PA are closed under intersection, and left quotient with words.
- Closure under union? Complement? Still open.

- Languages recognized by weakly-unambiguous PA have holonomic generating series

Weighted generating series of a PA

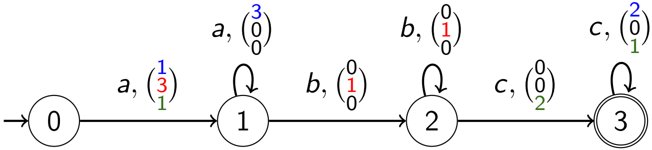
Definition (Generating series of the runs of a PA)

$$q(x, y_1, \ldots, y_d) = \sum_{n, i_1, \ldots, i_d} q_{n, i_1, \ldots, i_d} \, x^n y_1^{i_1} \ldots y_d^{i_d}$$

where $q_{n, i_1, \ldots, i_d}$ denotes the number of runs from q to a final state, labelled by (w, v) with $|w| = n$ and $v = (i_1, \ldots, i_d)$.

The generating series of these runs are classically rational.

Example



$$\begin{cases} q_0(x, y_1, y_2, y_3) = x y_1 y_2^3 y_3 q_1(x, y_1, y_2, y_3) \\ q_1(x, y_1, y_2, y_3) = x y_1^3 q_1(x, y_1, y_2, y_3) + x y_2 q_2(x, y_1, y_2, y_3) \\ q_2(x, y_1, y_2, y_3) = x y_2 q_2(x, y_1, y_2, y_3) + x y_3^2 q_3(x, y_1, y_2, y_3) \\ q_3(x, y_1, y_2, y_3) = x y_1^2 y_3 q_3(x, y_1, y_2, y_3) + 1 \end{cases}$$

Weakly-unambiguous PA have holonomic series

Proposition

The generating series of a language recognized by a weakly-unambiguous Parikh Automaton is holonomic.

- $q_I(x, y_1, \ldots, y_d)$ counts every run of the automaton from q_I to a final state. It is rational
- $C(y_1, \ldots, y_d) = \sum_{(i_1,\ldots,i_d)\in C} y_1^{i_1} \ldots y_d^{i_d}$ support series of the semilinear set C, which is rational
- $A(x, y_1, \ldots, y_d) := q_I(x, y_1, \ldots, y_d) \odot \frac{1}{1-x} C(y_1, \ldots, y_d)$ counts the accepting runs of the automaton, sorted by length and vector value. It is holonomic

Weakly-unambiguous PA have holonomic series

Proposition

The generating series of a language recognized by a weakly-unambiguous Parikh Automaton is holonomic.

- $q_I(x, y_1, \ldots, y_d)$ counts every run of the automaton from q_I to a final state. It is rational
- $C(y_1, \ldots, y_d) = \sum_{(i_1,\ldots,i_d)\in C} y_1^{i_1} \ldots y_d^{i_d}$ support series of the semilinear set C, which is rational
- $A(x, y_1, \ldots, y_d) := q_I(x, y_1, \ldots, y_d) \odot \frac{1}{1-x} C(y_1, \ldots, y_d)$ counts the accepting runs of the automaton, sorted by length and vector value. It is holonomic
- $A(x, 1, \ldots, 1)$ counts the accepting runs of the automaton, sorted by length. It is holonomic
- By weak-unambiguity, $L(x) = A(x, 1, \ldots, 1)$.

Inherent weak-ambiguity

Proposition

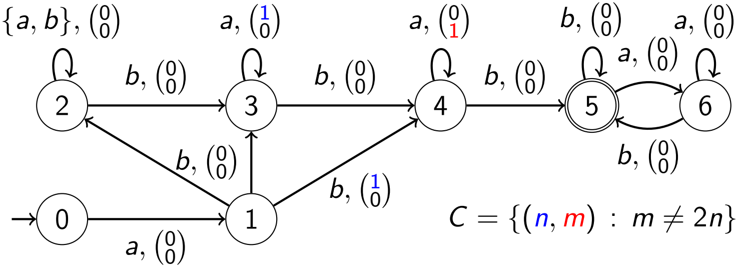
The generating series of a language recognized by a weakly-unambiguous Parikh Automaton is holonomic.

Contraposition

If the generating series of a language recognized by a PA is not holonomic, then it is inherently weakly-ambiguous as a PA language.

Example

$\mathcal{D} = \{a^{n_1} b\, a^{n_2} b \dots a^{n_k} b \ : \ k \in \mathbb{N}^*,\ n_1 = 1 \text{ and } \exists j < k, n_{j+1} \neq 2n_j\}$ is inherently weakly-ambiguous as a PA language.



Ambiguous automaton: *ababab* has two accepting runs.

From $D(x)$ we built a lacunary series. Lacunary series are not holonomic.

Criteria for non holonomy

Theorem (Stanley 1980)

Let $f(x) = \sum a_n x^n$ *:*

- *If* f *has an infinite number of singularities,* f *is not holonomic.*
- *If* a_n *does not satisfy a linear recurrence with polynomial coefficients, then* f *is not holonomic.*

Example $(B(x) = \sum_{k\geq 1} x^{2^k-1+k})$

$2^{k+1} - 1 + k + 1 - (2^k - 1 + k) \to \infty$ incompatible with any

$$p_r(n)a_{n+r} + \ldots + p_0(n)a_n = 0 \quad \text{with } p_i(n) \in \mathbb{Q}[n]$$

Limits of the method

Inherent weak-ambiguity is undecidable, by Greibach's theorem, using undecidability of universality of PA [Klaedtke and Rueß 03]

The series criterium may fail. There exist inherently weakly-ambiguous PA languages having holonomic series.

Inherently weakly-ambiguous language with algebraic series

Proposition

$\mathcal{L}_{even} = \{a^{n_1} b a^{m_1} b \dots a^{n_k} b a^{m_k} b \ : \ k \in \mathbb{N}^*, \exists i \in [1, k], n_i = m_i\}$ is inherently weakly-ambiguous as a PA.

- $aaabaab\, aabaab\, abaab \in \mathcal{L}_{even}$
- It is deterministic context-free $\Rightarrow$ algebraic generating series
- The proof uses Ramsey's theorem, and is very specific to this language. It shows inherent ambiguity for a wider family of automata.

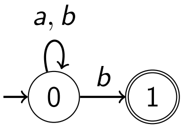
An algorithmic consequence of holonomy

Holonomy of the generating series has algorithmic consequences

$\rightarrow$ It has already been used for standard unambiguous finite automata!

1. Present the case of the inclusion problem for unambiguous finite automata
2. Show how the same general ideas apply to weakly-unambiguous PA.

Inclusion separation problem



Proposition (Stearns and Hunt 85)

Given two unambiguous finite automata $\mathcal{A}$ and $\mathcal{B}$ such that

$$L(\mathcal{B}) \subsetneq L(\mathcal{A})$$

Then there is a small witness word $w \in L(\mathcal{A}) \setminus L(\mathcal{B})$ such that

$$|w| < |Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$$

Sketch of the proof

- $L_{\mathcal{A}}(x) = \sum_n a_n x^n$ generating series of $L(\mathcal{A})$
- $L_{\mathcal{B}}(x) = \sum_n b_n x^n$ generating series of $L(\mathcal{B})$.
- $G(x) = L_{\mathcal{B}}(x) - L_{\mathcal{A}}(x)$ rational, degrees at most $r \leq |Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$
- Then $g_n = b_n - a_n$ satisfies:

$$\forall n \geq r,\ c_r g_n = c_{r-1} g_{n-1} + \ldots + c_0 g_{n-r}$$

Sketch of the proof

- $L_{\mathcal{A}}(x) = \sum_n a_n x^n$ generating series of $L(\mathcal{A})$
- $L_{\mathcal{B}}(x) = \sum_n b_n x^n$ generating series of $L(\mathcal{B})$.
- $G(x) = L_{\mathcal{B}}(x) - L_{\mathcal{A}}(x)$ rational, degrees at most $r \leq |Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$
- Then $g_n = b_n - a_n$ satisfies:

$$\forall n \geq r,\ c_r g_n = c_{r-1} g_{n-1} + \ldots + c_0 g_{n-r}$$

- So if $a_n = b_n$ for every $n < r$, then $a_n = b_n$ for all n.
- As $L(\mathcal{A}) \subsetneq L(\mathcal{B})$, there exists $N < r$ such that $a_N < b_N$.

$\rightarrow$There is a small witness word of length $< |Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$ in $L(\mathcal{B}) \backslash L(\mathcal{A})$.

Inclusion problem

Input: two weakly-unambiguous Parikh automata $\mathcal{A}, \mathcal{B}$
Question: $L(\mathcal{A}) \subseteq L(\mathcal{B})$?

- decidable for deterministic PA
- decidable for RCM [Castiglione and Massazza 17] (hence for weakly-unambiguous PA) without complexity bound
- undecidable for non-deterministic PA

$\rightarrow$Our contribution is to give explicit bounds in the weakly-unambiguous case

Essentially same ideas as regular case, however:

- $L_{\mathcal{A}}(x) = A(x, 1, \ldots, 1)$ where:

$$A(x, y_1, \ldots, y_d) := q_I(x, y_1, \ldots, y_d) \odot C(x, y_1, \ldots, y_d)$$

$\rightarrow$ Same problem with $L_{\mathcal{B}}(x)$

- Then $g_n = v_n - u_n$ satisfies a linear recurrence of the form

$$\forall n \geq r,\ c_r(n) g_n = c_{r-1}(n) g_{n-1} + \ldots + c_0(n) g_{n-r}$$

$G(x) = x^{1000} \ \rightarrow \ (1000 - n) g_n = 0$
$\rightarrow$ we need to go beyond r and the roots of c_r that are in $\mathbb{N}$

Inclusion separation for weakly-unambiguous automata?

- We want bounds on the polynomials and order of the recurrence of $G(x)$, depending on the size of the automata $\mathcal{A}$ and $\mathcal{B}$
- At each step (Hadamard product, $y = 1$, sum...), bound the size of the representation of the resulting holonomic series (holonomic series are represented by their system of differential equations)
- by a careful analysis of every operation:

Proposition

If $L(\mathcal{A}) \not\subseteq L(\mathcal{B})$, there exists a word $w \in L(\mathcal{B}) \setminus L(\mathcal{A})$ such that

$$|w| \leq 2^{2^{O(d^2 \log(dM))}}$$

where $d = d_{\mathcal{A}} + d_{\mathcal{B}}$, $M = |\mathcal{A}|\,|\mathcal{B}|\,\|\mathcal{A}\|_\infty\,\|\mathcal{B}\|_\infty$.

Consequence: inclusion problem

Input: two weakly-unambiguous Parikh automata $\mathcal{A}, \mathcal{B}$
Question: $L(\mathcal{A}) \subseteq L(\mathcal{B})$?

Proposition

We can decide in time $\leq 2^{2^{O(d^2 \log(dM))}}$ *whether* $L(\mathcal{A}) \subseteq L(\mathcal{B})$, *where* $d = d_{\mathcal{A}} + d_{\mathcal{B}}$, $M = |\mathcal{A}|\,|\mathcal{B}|\,\|\mathcal{A}\|_\infty\,\|\mathcal{B}\|_\infty$.

$\rightarrow$ dynamic programming approach to avoid an other exponential when enumerating every word of length less than the witness!

Conclusion

Language		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $Q(x)L(x) = P(x)$
$\subsetneq$		$\subsetneq$
Unambiguous context-free	$\longrightarrow$	algebraic $P(x, L(x)) = 0$
$\subsetneq$		$\subsetneq$
Weakly-unambiguous Pushdown PA	$\longrightarrow$	holonomic $P(x, \partial_x) \cdot L(x) = 0$

Conclusion

Language		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $Q(x)L(x) = P(x)$
$\subsetneq$		$\subsetneq$
Unambiguous context-free	$\longrightarrow$	algebraic $P(x, L(x)) = 0$
$\subsetneq$		$\subsetneq$
Weakly-unambiguous Pushdown PA	$\longrightarrow$	holonomic $P(x, \partial_x) \cdot L(x) = 0$

Remaining problems: closure under union, universality with a stack, implementation of algorithms...

Perspectives

Extension: larger classes with holonomic series?

Proposition (Bell and Chen 17)

Any holonomic series with coefficients in $\{0, 1\}$ *is the support series of a semilinear set.*

We are close to the limits of this approach
$\rightarrow$ need for new ideas to find other links between holonomic series and formal languages.

Extension: larger classes with holonomic series?

Proposition (Bell and Chen 17)

Any holonomic series with coefficients in $\{0, 1\}$ *is the support series of a semilinear set.*

We are close to the limits of this approach
$\rightarrow$ need for new ideas to find other links between holonomic series and formal languages.

Thank you!

Jason P. Bell and Shaoshi Chen, *Power series with coefficients from a finite set*, J. Comb. Theory, Ser. A **151** (2017), 241 – 253.

Alin Bostan and Manuel Kauers, *The complete generating function for Gessel walks is algebraic*, Proc. Amer. Math. Soc. **138** (2010), no. 9, 3063–3078, With an Appendix by Mark van Hoeij.

Mireille Bousquet-Mélou, *Rational and algebraic series in combinatorial enumeration*, International Congress of Mathematicians (ICM 2006), vol. 3, Eur. Math. Soc., Zürich, 2006, pp. 789–826.

Michaël Cadilhac, Alain Finkel, and Pierre McKenzie, *Affine Parikh automata*, RAIRO - Theor. Inf. and Applic. **46** (2012), no. 4, 511–545.

- ______, *Unambiguous constrained automata*, Int. J. Found. Comput. Sci. **24** (2013), no. 7, 1099–1116.
- Giusi Castiglione and Paolo Massazza, *On a class of languages with holonomic generating functions*, Theor. Comput. Sci. **658** (2017), 74–84.
- Louis Comtet, *Calcul pratique des coefficients de Taylor d'une fonction algébrique*, Enseignement Math. (2) **10** (1964), 267–270.
- Noam Chomsky and Marcel-Paul Schützenberger, *The algebraic theory of context-free languages*, Studies in Logic and the Foundations of Mathematics, vol. 35, Elsevier, 1963.
- Samuel Eilenberg and Marcel-Paul Schützenberger, *Rational sets in commutative monoids*, J. Algebra **13** (1969), no. 2, 173 – 191.

- Philippe Flajolet, Stefan Gerhold, and Bruno Salvy, *On the non-holonomic character of logarithms, powers, and the nth prime function*, Electr. J. Comb. **11** (2005), no. 2.
- Philippe Flajolet, *Analytic models and ambiguity of context-free languages*, Theor. Comput. Sci. **49** (1987), no. 2, 283 – 309.
- Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, first ed., Cambridge University Press, 2009.
- Sheila Greibach, *A note on undecidable properties of formal languages*, Mathematical Systems Theory **2** (1968), no. 1, 1–6.
- Seymour Ginsburg and Edwin Spanier, *Semigroups, presburger formulas, and languages*, Pac. J. Math. **16** (1966), no. 2, 285–296.
- Seymour Ginsburg and Joseph Ullian, *Ambiguity in context free languages*, J. ACM **13** (1966), no. 1, 62–89.

- Oscar H. Ibarra, *Reversal-bounded multicounter machines and their decision problems*, J. ACM **25** (1978), no. 1, 116–133.
- Ryuichi Ito, *Every semilinear set is a finite union of disjoint linear sets*, J. Comput. Syst. Sci. **3** (1969), no. 2, 221–231.
- Manuel Kauers, *Bounds for D-finite closure properties*, Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation, (ISSAC 2014), ACM, New York, 2014, pp. 288–295.
- Felix Klaedtke and Harald Rueß, *Parikh automata and Monadic Second-Order logics with linear cardinality constraints*, Tech. Report 177, Freiburg University, 2002.

- ———, *Monadic second-order logics with cardinalities*, Automata, Languages and Programming, 30th International Colloquium, ICALP 2003, Lecture Notes in Computer Science, vol. 2719, Springer, 2003, pp. 681–696.
- Leonard Lipshitz, *The diagonal of a D-finite power series is D-finite*, J. Algebra **113** (1988), no. 2, 373 – 378.
- ———, *D-finite power series*, J. Algebra **122** (1989), no. 2, 353 – 373.
- Paolo Massazza, *Holonomic functions and their relation to linearly constrained languages*, ITA **27** (1993), no. 2, 149–161.
- ———, *On the conjecture $\mathcal{L}_{\mathrm{dfcm}} \subsetneq \mathrm{RCM}$*, Implementation and Application of Automata - 22nd International Conference, CIAA 2017, Lecture Notes in Computer Science, vol. 10329, Springer, 2017, pp. 175–187.

- ______, *On the generating functions of languages accepted by deterministic one-reversal counter machines*, Proceedings of the 19th Italian Conference on Theoretical Computer Science, (ICTCS 2018), CEUR Workshop Proceedings, vol. 2243, CEUR-WS.org, 2018, pp. 191–202.
- F. P. Ramsey, *On a Problem of Formal Logic*, Proc. London Math. Soc. (2) **30** (1929), no. 4, 264–286.
- Richard Edwin Stearns and Harry B. Hunt III, *On the equivalence and containment problems for unambiguous regular expressions, regular grammars and finite automata*, SIAM J. Comput. **14** (1985), no. 3, 598–611.
- Richard P. Stanley, *Differentiably finite power series*, Eur. J. Comb. **1** (1980), no. 2, 175 –188.

Conclusion

Language		Generating series
L	$\longrightarrow$	$L(x)$
Regular	$\longrightarrow$	rational $Q(x)L(x) = P(x)$
$\subsetneq$		$\subsetneq$
Unambiguous context-free	$\longrightarrow$	algebraic $P(x, L(x)) = 0$
$\subsetneq$		$\subsetneq$
Weakly-unambiguous Pushdown PA	$\longrightarrow$	holonomic $P(x, \partial_x) \cdot L(x) = 0$

Remark

Example

$\mathcal{D} = \{a^{n_1} b\, a^{n_2} b \dots a^{n_k} b \ :\ k \in \mathbb{N}^*,\ n_1 = 1 \text{ and } \exists j < k, n_{j+1} \neq 2n_j\}$ is inherently ambiguous as a PA language.

Consequence

Weakly-unambiguous PA are not closed under left quotient with regular languages.

$$\mathcal{D}_2 = \{c^j a^{n_1} b\, a^{n_2} b \dots a^{n_k} b \ :\ k \in \mathbb{N}^*, j < k,\ n_1 = 1 \wedge n_{j+1} \neq 2n_j\}$$

$$(c^*)^{-1}\mathcal{D}_2 \cap (a+b)^* = \mathcal{D}$$

Incomparable

$\{a^n b^m c^p \ : \ n = m \text{ or } m = p\}$ is

- inherently ambiguous as a CF language
- deterministic as a PA language

$\mathcal{L}_{even} = \{a^{n_1} b \dots a^{n_{2k}} b \ : \ k \in \mathbb{N}^*, \exists i \in [1, k], n_{2i-1} = n_{2i}\}$ is

- deterministic as a CF language
- inherently ambiguous as a PA language

Undecidability of inherent weak-ambiguity

General method [Greibach 68], by reducing the universality problem

$$L_1 = \Sigma_1^*?$$

$L = L_1 \# \Sigma^* \cup \Sigma_1^* \# \mathcal{D}$. Then:

$$L_1 = \Sigma_1^* \Leftrightarrow L \text{ is weakly-unambiguous}$$

- $\Rightarrow$ If $L_1 = \Sigma_1^*$, $L = \Sigma_1^* \# \Sigma^*$ is regular.
- $\Leftarrow$ By contraposition, let $y \notin L_1$. As $(y\#)^{-1}L = \mathcal{D}$ is not weakly-unambiguous, neither is L.

Inclusion separation for weakly-unambiguous automata?

- $\mathcal{A}$ given under the form $(\Sigma, Q, q_I, F, C, \Delta)$.
- C given under a unambiguous form $\cup_{i=1}^{p} c_i + P_i^*$

Inclusion separation for weakly-unambiguous automata?

- $\mathcal{A}$ given under the form $(\Sigma, Q, q_I, F, C, \Delta)$.
- C given under a unambiguous form $\cup_{i=1}^{p} c_i + P_i^*$
- $\|\mathcal{A}\|_\infty$ maximum coordinate of the vectors in the description of Δ and C
- $|\mathcal{A}| = |Q| + |\Delta| + p + \sum |P_i|$