Generating Functions and Ambiguity in Automata Theory

Florent Koechlin

Joint work with Alin Bostan, Arnaud Carayol and Cyril Nicaud

CNRS, LIPN, Université Sorbonne Paris Nord

Séminaire Flajolet
2025, June 5th

Reminder on Automata theory

Automata theory is interested in

- languages, i.e. formal sets of words over a given alphabet Σ.

  $(a+b)^* := \{w_1 \dots w_n \,:\, n \in \mathbb{N}, w_i \in \{a,b\}\}$
  $\{a^n b^n \,:\, n \in \mathbb{N}^*\}$

- described by finite structures

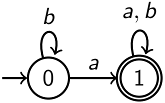
  automata, grammars, counter machines

- and the complexity of problems related to them

  emptyness, inclusion, universality, ...

Regular languages

Regular languages are the simplest languages in the Chomsky hierarchy. They are exactly the languages recognized by :

- Regular expressions :
  $\Sigma^* a \Sigma^*$, $(a+b)^* b$, $\Sigma^* a \Sigma^{r-1}, \ldots$
- (Deterministic) finite automata 

Accepting run of an automaton : labeled path from the initial state to a final state

Context-free languages

Context-free languages are the second-level class of languages in the Chomsky hierarchy. They are exactly the languages recognized by :

- Context-free grammars

$$S \to aSb \mid \varepsilon, \qquad S \to [S]S \mid \varepsilon, \qquad \begin{cases} S \to aSb \mid C \mid cc \\ C \to cC \mid c \end{cases}$$

- Non-deterministic pushdown automata

Regular languages $\subsetneq$ Context-free languages

$\{a^n b^n \mid n \in \mathbb{N}\}$ is context-free but not regular

Unambiguous context-free grammar

$S \to [S]S \mid \varepsilon$

Derivation

$S \Rightarrow [S]S \Rightarrow [[S]S]S \Rightarrow [[]S]S \Rightarrow [[][S]S]S \Rightarrow [[][]S]S \Rightarrow [[][]]S \Rightarrow [[][]]$

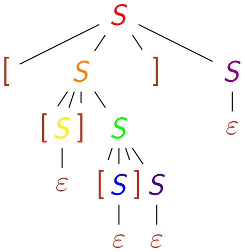
Unambiguous context-free grammar

$S \to [S]S \mid \varepsilon$

Derivation

$S \Rightarrow [S]S \Rightarrow [[S]S]S \Rightarrow [[]S]S \Rightarrow [[][S]S]S \Rightarrow [[][]S]S \Rightarrow [[][]]S \Rightarrow [[][]]$
$S \Rightarrow [S]S \Rightarrow [S] \Rightarrow [[S]S] \Rightarrow [[]S] \Rightarrow [[][S]S] \Rightarrow [[][S]] \Rightarrow [[][]]$

Derivation tree



Unambiguous context-free grammar

Every word in its language has exactly one derivation tree.

Unambiguous context-free languages

Relevant intermediate model between deterministic and non-deterministic context-free languages.

Unambiguous Context-free Languages $\subsetneq$ Context-free Languages

$\{a^n b^m c^p \mid n = m \text{ or } m = p\}$ is inherently ambiguous

Unambiguous context-free languages

Relevant intermediate model between deterministic and non-deterministic context-free languages.

Unambiguous Context-free Languages $\subsetneq$ Context-free Languages

$\{a^n b^m c^p \mid n = m \text{ or } m = p\}$ is inherently ambiguous

Finding inherently ambiguous languages is interesting. However:

- deciding whether a grammar is ambiguous is undecidable [Chomsky-Schützenberger'63]
- deciding whether a context-free language is inherently ambiguous is undecidable [Ginsburg-Ullian'66, Greibach'68]

Combinatorics of formal languages: generating functions

Generating function

Let L be a language, ℓ_n the number of words in L of length n:

$$L(x) = \sum_{n=0}^{+\infty} \ell_n x^n$$

Example

$(a+b)^* \to \ell_n = 2^n \to L(x) = \sum_n 2^n x^n = \frac{1}{1-2x}$

Example

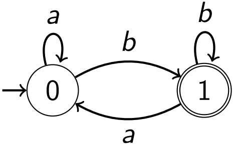
$b^*a(a+b)^* \to L(x) = \sum_{n\geq 0}(2^n - 1)x^n = \frac{x}{(1-x)(1-2x)}$

Example

Well bracketed words $\to L(x) = \sum_{n\geq 0} \frac{1}{n+1}\binom{2n}{n} x^{2n} = \frac{1-\sqrt{1-4x^2}}{2x^2}$

Link between automata and generating functions

regular languages $\subseteq$ unambiguous context-free languages



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

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

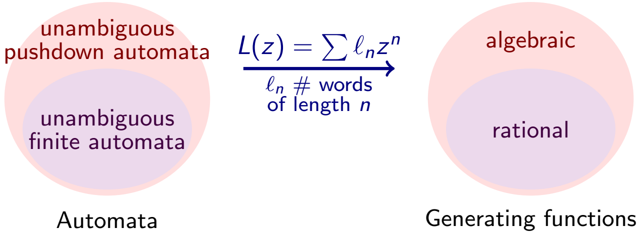
$$\begin{cases} S(x) = xS(x)B(x) + 1 \\ B(x) = xB(x) + xS(x) \end{cases}$$

$$q_0(x) = \frac{x}{1-2x}$$

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

rational series $L(x) = \dfrac{P(x)}{Q(x)}$ $\subseteq$ algebraic series $P(L(x), x) = 0$

Link between two hierarchies



Two remarkable applications :

- analytic proofs of inherent ambiguity [Flajolet 87]
- polynomial algorithm for the inclusion problem for unambiguous NFA's [Stearns & Hunt 85]

Analytic criteria for inherent ambiguity

Flajolet's idea: if the GF of a context-free language is not algebraic, then it is an inherently ambiguous context-free language.

Proposition [Useful criteria, Flajolet '87]:
Let $L(z) = \sum_{n \in \mathbb{N}} \ell_n z^n$ a series.

- If $L(z)$ has infinitely many singularities, then $L(z)$ is not algebraic.
- If ℓ_n does not satisfy a linear recurrence with polynomial coefficients in n, then $L(z)$ is not algebraic.
- If $\ell_n \sim_{n \to \infty} \gamma \beta^n n^r$, with $r \in \{-1, -2, -3, \ldots\}$ or $r \notin \mathbb{Q}$, or $\gamma \times \Gamma(r+1)$ transcendental, then $L(z)$ is not algebraic.

Analytic criteria for inherent ambiguity

Theorem [Flajolet '87]
$\Omega_3 = \{w \in \{a,b,c\}^* \,:\, |w|_a \neq |w|_b \text{ or } |w|_b \neq |w|_c\}$ is inherently ambiguous.

Analytic proof:

- Suppose that $\Omega_3(x)$ is algebraic
- Let $I = (a+b+c)^* \setminus \Omega_3$
- Then $I(x) = \frac{1}{1-3x} - \Omega_3(x)$ would be algebraic by closure properties
- But $I = \{w \in \{a,b,c\}^* \,:\, |w|_a = |w|_b = |w|_c\}$

$$[x^{3n}]I(x) = \binom{3n}{n,n,n} = \frac{(3n)!}{(n!)^3} \sim_{n\to\infty} 3^{3n}\frac{\sqrt{3}}{2\pi n}$$

If $\ell_n \sim_{n\to\infty} \gamma\beta^n n^r$, with $r \in \{-1,-2,-3,\ldots\}$ then $L(z)$ is not algebraic.

Flajolet's analytic method

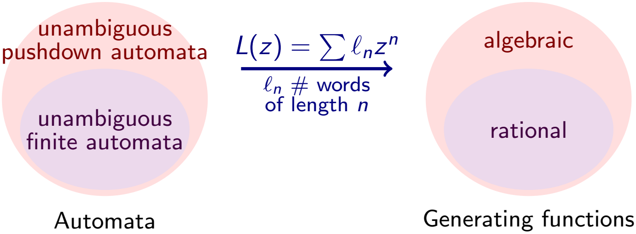
Advantages :

- is very powerful : P. Flajolet (re)proved the inherent ambiguity of 15 languages, some of which were conjectures, in only one article

Inconvenients :

- does not work on too simple languages, whose series are rational; for instance for $a^n b^m c^p$ with $n = m$ or $m = p$.
  $\rightarrow$ New methods needed [Makarov 21, Koechlin 21]

Link between two hierarchies



Two remarkable applications :

- analytic proofs of inherent ambiguity [Flajolet 87]
- polynomial algorithm for the inclusion problem for unambiguous NFA's [Stearns & Hunt 85]

Strict Inclusion problem for unambiguous automata

Problem: Given $\mathcal{A}$ and $\mathcal{B}$ two unambiguous NFA, with $L(\mathcal{A}) \subseteq L(\mathcal{B})$, is the inclusion strict?

Bad idea: compute an automaton recognizing $L_{\mathcal{C}}$, via determinizing $\mathcal{A}$ and $\mathcal{B}$

Strict Inclusion problem for unambiguous automata

Problem: Given $\mathcal{A}$ and $\mathcal{B}$ two unambiguous NFA, with $L(\mathcal{A}) \subseteq L(\mathcal{B})$, is the inclusion strict?

Stearns and Hunt's idea:

- $C(x) := \sum c_n x^n = B(x) - A(x)$ is rational
- The coefficients of $C(x)$ satisfy a linear recurrence:

$$\forall n \geq r, c_n = \alpha_1 c_{n-1} + \cdots + \alpha_r c_{n-r}$$

- the order r is at most $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

If $L(\mathcal{A}) \subsetneq L(\mathcal{B})$, there exists a small witness $w \in L(\mathcal{B}) \setminus L(\mathcal{A})$ with

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

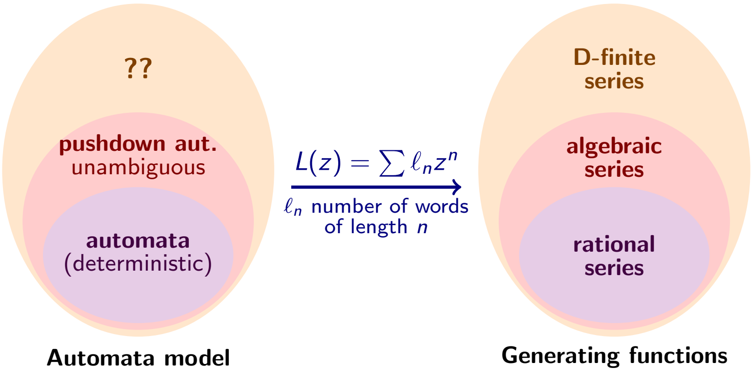
Inclusion problem for unambiguous automata

Problem: Given $\mathcal{A}$ and $\mathcal{B}$ two unambiguous NFA, $L(\mathcal{A}) \subseteq L(\mathcal{B})$?

Theorem [Stearns and Hunt 85]: The inclusion problem for unambiguous NFA is polynomial.

- $L(\mathcal{A}) \not\subseteq L(\mathcal{B}) \Leftrightarrow L(\mathcal{A}) \cap L(\mathcal{B}) \subsetneq L(\mathcal{A})$
- Compute coefficients up to $|Q_{\mathcal{A}}||Q_{\mathcal{B}}| + |Q_{\mathcal{A}}|$ (dynamic prog.)

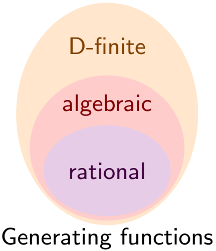
Extension to D-finite series



Goals of the talk

- suitable class of languages and unambiguous automata models
- proofs of inherent ambiguity, algorithmic consequences

D-finite series [Stanley 80]



Rational: $P(x)f(x) = Q(x)$
Algebraic: $P(x, f(x)) = 0$
D-finite: $P(x, \partial_x) \cdot f = 0.$

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

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

Alternative definition: the coefficients a_n satisfy a linear recurrence $p_r(n) a_{n+r} + \ldots + p_0(n) a_n = 0$

D-finite series [Stanley 80]

Example: $F(x) = e^x := \sum \frac{x^n}{n!}$ is D-finite but is not algebraic

- differential equation: $F' - F = 0$
- recurrence relation: $(n+1)f_{n+1} - f_n = 0$

D-finite series [Stanley 80]

Example: $F(x) = e^x := \sum \frac{x^n}{n!}$ is D-finite but is not algebraic

- differential equation: $F' - F = 0$
- recurrence relation: $(n+1)f_{n+1} - f_n = 0$

Example: $I = \{w \in \{a, b, c\}^* \, : \, |w|_a = |w|_b = |w|_c\}$

$$[x^{3n}]I(x) = \binom{3n}{n, n, n} = \frac{(3n)!}{(n!)^3} \sim_{n\to\infty} 3^{3n}\frac{\sqrt{3}}{2\pi n}$$

$I(x)$ is D-finite but is not algebraic.

Multivariate D-finite series [Lipshitz 88,89]

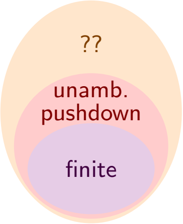
Multivariate D-finite series: satisfy a system of linear partial differential equations.

- Multivariate rational and algebraic series are D-finite.
  $\frac{1}{1-xy}$, $\frac{1-\sqrt{1-4xy}}{2xy}$
- Multivariate D-finite series are closed under :
  - arithmetic operations $+, \times, -$
    $\frac{1}{1-xy} - \frac{1-\sqrt{1-4xy}}{2xy}$
  - specialization to 1:
    $f(x_1, \ldots, x_n)$ D-finite $\Rightarrow f(x, 1, \ldots, 1)$ D-finite
  - Diagonals
  - Hadamard's product $\odot$ (component-wise product)
    $\frac{1}{1-xy} \odot \frac{1-\sqrt{1-4xy}}{2xy}$

$$f(x,y) = \sum_{(i_1,i_2)} a(i_1,i_2) x^{i_1} y^{i_2}, \quad g(x,y) = \sum_{(i_1,i_2)} b(i_1,i_2) x^{i_1} y^{i_2}$$

$$f \odot g(x,y) = \sum_{(i_1,i_2)} a(i_1,i_2) b(i_1,i_2) x^{i_1} y^{i_2},$$

Previous attempts on automata side



Automata model

Add linear constraints to the support of D-finite series :

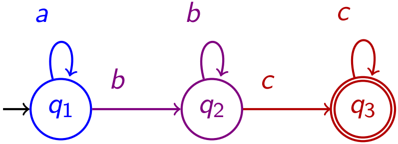
- Idea already hinted in [Lipshitz 88]
- Formalised by [Bertoni, Massazza, Sabadini '92], [Massazza '93], [Castiglione, Massazza '17]
- Family of languages : RCM et LCL, built purposely to have D-finite series

$$\{a^n b^n c^n\} = a^* b^* c^* \cap [\#a = \#b \wedge \#b = \#c]$$

- No associated automata model
  $\rightarrow$ conjectured link between RCM and deterministic counter machines (RBCM)[Castiglione, Massazza '17]

Parikh automata (PA) [Klaedtke, Rueß '03]

Motivation: $\{a^n b^n c^n\}$ is simple but not context-free

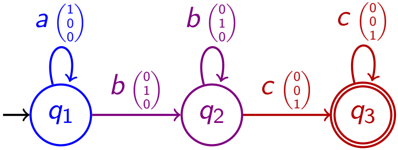


$$\mathcal{L}(\mathcal{A}) = a^* b^+ c^+$$

- Run is labeled by a word
- Word is recognised if the run ends in final state q_3

Parikh automata (PA) [Klaedtke, Rueß '03]

Motivation: $\{a^n b^n c^n\}$ is simple but not context-free

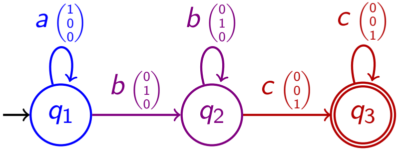


$$\mathcal{L}(\mathcal{A}) = a^* b^+ c^+$$

- Run is labeled by a word and a vector
- Word is recognised if the run ends in final state q_3

Parikh automata (PA) [Klaedtke, Rueß '03]

Motivation: $\{a^n b^n c^n\}$ is simple but not context-free



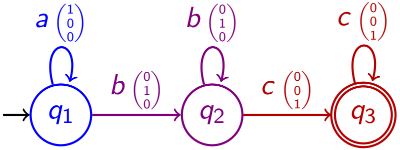
$$\mathcal{C} = \left\{ \begin{pmatrix}\ell_1\\ \ell_2\\ \ell_3\end{pmatrix} : \ell_1 = \ell_2 = \ell_3 \right\}$$

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

- Run is labeled by a word and a vector
- Word is recognised if the run ends in final state q_3 and if its vector is in $\mathcal{C}$

Parikh automata (PA) [Klaedtke, Rueß '03]

Motivation: $\{a^n b^n c^n\}$ is simple but not context-free



$$\mathcal{C} = \left\{ \begin{pmatrix}\ell_1\\\ell_2\\\ell_3\end{pmatrix} : \ell_1 = \ell_2 = \ell_3 \right\}$$

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

- Run is labeled by a word and a vector
- Word is recognised if the run ends in final state q_3 and if its vector is in $\mathcal{C}$
- Can be extended with a stack.

Semilinear sets in $\mathbb{N}^d$

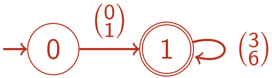
The accepting set of vectors $\mathcal{C}$ is a semilinear set

$\bigwedge \bigvee$ of linear inequalities or equalities modulo constants

$$\{(3n, 6n+1) : n \in \mathbb{N}\} = \{(x_1, x_2) : x_1 \equiv 0[3] \wedge x_2 = 2x_1 + 1\}$$

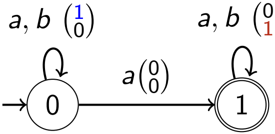
Equivalent definitions (all very useful!)

- Finite union of linear sets $\vec{c} + P^*$ where $P = \{p_1, \ldots, p_r\}$
  $(0,1) + \{(3,6)\}^*$
- Presburger arithmetic [Ginsburg and Spanier, 66]
  $\Phi(x_1, x_2) := \exists x, x_1 - 3x = 0 \ \wedge \ 1 + 2x_1 - x_2 = 0$
- (Unambiguous) rational subsets of $(\mathbb{N}^d, +)$ [Eilenberg and Schützenberger, Ito, 69]



(Weakly) unambiguous Parikh automata

Weak Unambiguity: at most one accepting run (final state + semilinear constraint)



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

$L(\mathcal{A}) = \{$words with an a in the middle$\} = \{\ldots, abbabab, \ldots\}$

$\neq$ **unambigous Parikh automata from [Cadilhac, Finkel, McKenzie 13]**

Relevant automata model

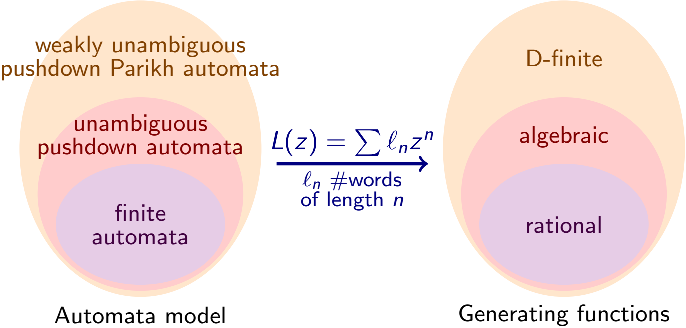
Theorem [Bostan, Carayol, K., Nicaud '20] : The class of weakly unambiguous Parikh languages coincide with :

- RCM of [Castiglione, Massazza '17]
- unambiguous two-way RBCM [Ibarra '78]
  $\Rightarrow$ stronger version of [Castiglione, Massazza '17]'s conjecture

Theorem [Bostan, Carayol, K., Nicaud '20] : The class of weakly unambiguous pushdown Parikh languages coincide with :

- LCL adapted from [Massazza '93]
- unambiguous one-way RBCM with a stack [Ibarra '78]

Extension



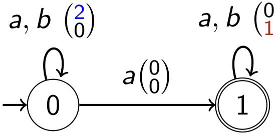
Theorem **[Bostan, Carayol, K., Nicaud '20] :** The generating function of a language recognized by a weakly unambiguous pushdown Parikh automaton is D-finite.

D-finiteness of the generating function

Theorem: the series of a weakly-unambiguous PA is D-finite

Counting the number of runs by vectors

$$\overline{A}(x, y_1, \ldots, y_d) = \sum_{n, v_1, \ldots, v_d} a_{n, v_1, \ldots, v_d} x^n y_1^{v_1} \cdots y_d^{v_d} \quad \text{is rational}$$



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

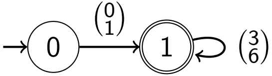
D-finiteness of the generating function

Theorem: the series of a weakly-unambiguous PA is D-finite

Support series of the semilinear constraint C

$$C(y_1, \ldots, y_d) := \sum_{(v_1,\ldots,v_d)\in C} y_1^{v_1} \ldots y_d^{v_d} \quad \text{is rational}$$

$\rightarrow$ (Unambiguous) rational subsets of $(\mathbb{N}^d, +)$



$$C(y_1, y_2) = \frac{y_2}{1 - y_1^3 y_2^6}$$

D-finiteness of the generating function

Theorem: the series of a weakly-unambiguous PA is D-finite

Counting the number of runs by vectors

$$\overline{A}(x, y_1, \ldots, y_d) = \sum_{n, v_1, \ldots, v_d} a_{n, v_1, \ldots, v_d} x^n y_1^{v_1} \cdots y_d^{v_d} \quad \text{is rational}$$

Support series of the semilinear constraint C

$$C(y_1, \ldots, y_d) := \sum_{(v_1, \ldots, v_d) \in C} y_1^{v_1} \ldots y_d^{v_d} \quad \text{is rational}$$

Filtering out non-accepting runs

$$\overline{A}(x, y_1, \ldots, y_d) \odot \frac{C(y_1, \ldots, y_d)}{1 - x} \; [y_1 \to 1, \ldots, y_d \to 1]$$

D-finiteness of the generating function

Theorem: the series of a weakly-unambiguous PA is D-finite

Counting the number of runs by vectors

$$\overline{A}(x, y_1, \ldots, y_d) = \sum_{n, v_1, \ldots, v_d} a_{n, v_1, \ldots, v_d} x^n y_1^{v_1} \cdots y_d^{v_d} \quad \text{is rational}$$

Support series of the semilinear constraint C

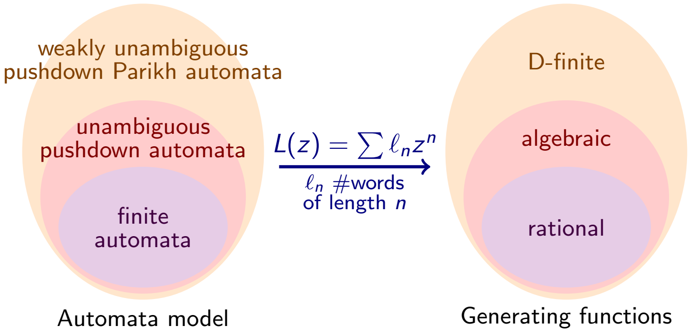
$$C(y_1, \ldots, y_d) := \sum_{(v_1, \ldots, v_d) \in C} y_1^{v_1} \ldots y_d^{v_d} \quad \text{is rational}$$

Filtering out non-accepting runs

$$\overline{A}(x, y_1, \ldots, y_d) \odot \frac{C(y_1, \ldots, y_d)}{1 - x} \, [y_1 \to 1, \ldots, y_d \to 1]$$

weak-unambiguity $\Rightarrow$ counting accepting runs = counting words $\square$

Extension



Two remarkable applications :

- analytic proofs of inherent ambiguity for PA
- complexities bound for an algorithm for the inclusion problem for wuPA

Inherent ambiguity using non-holonomy

If the series of a language accepted by a PA is not D-finite, then it is inherently weakly-ambiguous for PA.

Theorem [Stanley 1980]: Let $f(x) = \sum a_n x^n$:

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

Inherent ambiguity using non-holonomy

If the series of a language accepted by a PA is not D-finite, then it is inherently weakly-ambiguous for PA.

Theorem [Stanley 1980]: Let $f(x) = \sum a_n x^n$:

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

Example: the following language is recognized by a PA
$\mathcal{D} = \{a^{n_1} b\ a^{n_2} b \dots a^{n_k} b : k \in \mathbb{N}^*,\ n_1 = 1 \text{ and } \exists j < k, n_{j+1} \neq 2n_j\}$

Observe that:

- $w = aba^2ba^4ba^8b$ is a typical word that is not in $\mathcal{D}$
- $D(x) = \frac{x^2}{1-\frac{x}{1-x}} - \sum_{k\geq 1} x^{2^k-1+k}$
- $D(x)$ has infinitely many singularities: it is not D-finite

Hence $\mathcal{D}$ is inherently weakly-ambiguous for PA.

Limit of the method

Proposition: the following language is recognized by a PA

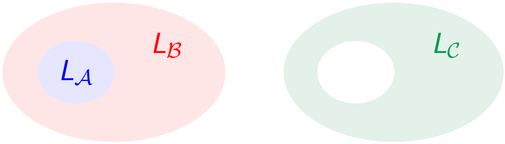
$$\mathcal{L}_{even} = \{a^{n_1}b \dots a^{n_{2k}}b \,:\, k \in \mathbb{N}^*, \exists i \in [1,k], n_{2i-1} = n_{2i}\}$$

- actually it is an unambiguous context-free language
- its generating function is even rational!
- it is inherently weakly-ambiguous [Bostan, Carayol, K., Nicaud '20]

Non-D-finiteness is only a sufficient condition for inherent weak-unambiguity for PA

Proposition: inherent weak-unambiguity is undecidable.

Algorithmic application: strict inclusion problem for wuPA



Pose $L_{\mathcal{C}} := L_{\mathcal{B}} \setminus L_{\mathcal{A}}$

idea: replace $L_{\mathcal{C}} \stackrel{?}{=} \emptyset$ by $C(x) \stackrel{?}{=} 0$

- "Compute" $A(x)$ and $B(x)$ from $\mathcal{A}$ and $\mathcal{B}$
  $\rightarrow$ possible by weak-unambiguity
- Differential equation satisfied by $C(x) = B(x) - A(x)$
- Linear recurrence satisfied by $c_n = b_n - a_n$

$$-p_r(n)c_{n+r} = p_{r-1}(n)c_{n+r-1} + \ldots + p_0(n)c_n$$

- Bound B such that $c_n = 0$ for $n \leq B$ implies $\forall n, c_n = 0$
  $C(x) = x^{100}$ satisfies $xC'(x) - 100C(x) = 0$ and $(n-100)c_n = 0$

Algorithmic application: inclusion problem for wuPA

Proposition [Bostan, Carayol, K., Nicaud '20] :
If $L(\mathcal{A}) \subsetneq L(\mathcal{B})$, then there exists a word $w \in L(\mathcal{B}) \setminus L(\mathcal{A})$ such that

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

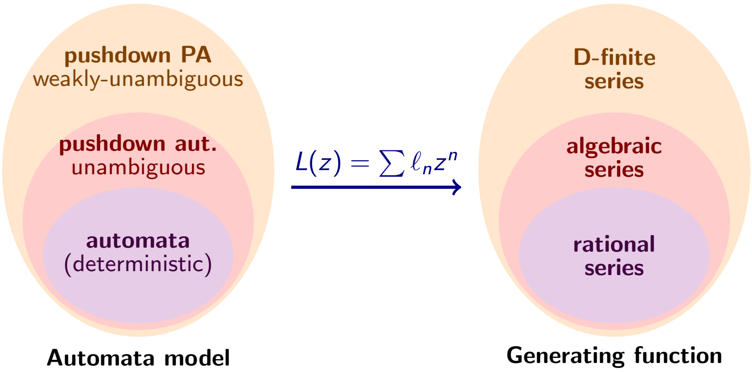
with $d = d_{\mathcal{A}} + d_{\mathcal{B}}$, M depends on $\mathcal{A}$ and $\mathcal{B}$.

Theorem [Bostan, Carayol, K., Nicaud '20] :
The inclusion problem

$$L(\mathcal{A}) \overset{?}{\subseteq} L(\mathcal{B})$$

is in $2-\mathrm{EXPTIME}$ for weakly-unambiguous Parikh automata.

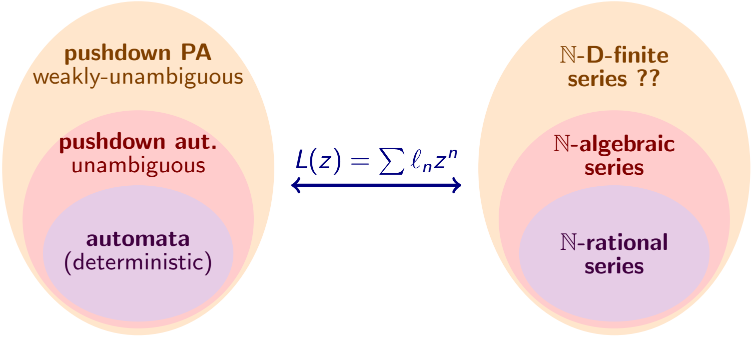
Let us be more precise



$$\frac{1-2x+225x^2}{(1-25x)(625x^2+14x+1)} = 1+9x+49x^2+\dots$$ [Salomaa&Soittola 78, Bousquet-Mélou 08]

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

Let us be more precise

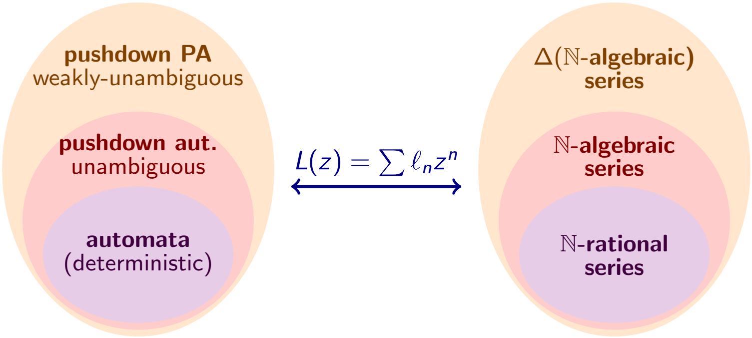


Automata model **Generating function**

$\frac{1-2x+225x^2}{(1-25x)(625x^2+14x+1)} = 1+9x+49x^2+\dots$ [Salomaa&Soittola 78, Bousquet-Mélou 08]

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

Let us be more precise



Automata model **Generating function**

Theorem [Koechlin '21]: A series $f(x)$ is the generating series of a weakly-unambiguous (pushdown) Parikh automaton if and only if it is the diagonal of an $\mathbb{N}$-rational ($\mathbb{N}$-algebraic) series

Diagonals of $\mathbb{N}$-rational functions

- A multivariate series

$$A(x_1, \ldots, x_k) = \sum_{i_1, \ldots, i_k} \ell_{i_1, \ldots, i_k} x_1^{i_1} \ldots x_k^{i_k}$$

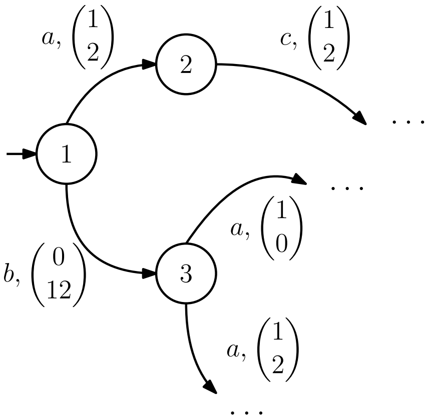
  is $\mathbb{N}$-rational if it counts the number of accepting runs of a finite automaton over the alphabet $\Sigma = \{a_1, \ldots, a_k\}$, where each variable counts the number of occurrences of each letter in the run.

- The diagonal of $A(x_1, \ldots, x_k)$ is the univariate function defined by

$$\Delta A(x) = \sum_n \ell_{n, \ldots, n} x^n$$

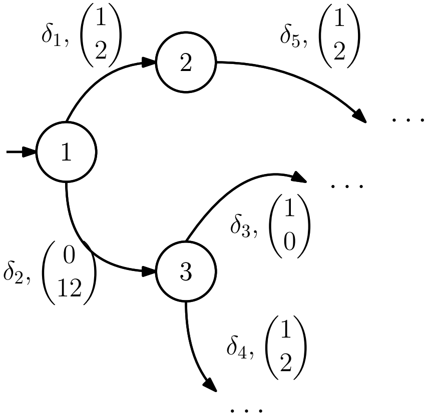
$\rightarrow$ Diagonals of $\mathbb{N}$-rational series $\simeq$ generating functions of regular languages with constraint "same number of occurrences of each letter"

wuPA implies diagonal of ℕ-rational functions



$$\phi(x_1, x_2) = (x_1 = 2x_2)$$

wuPA implies diagonal of $\mathbb{N}$-rational functions



$$\phi(x_1, x_2) = (x_1 = 2x_2)$$

wuPA implies diagonal of $\mathbb{N}$-rational functions

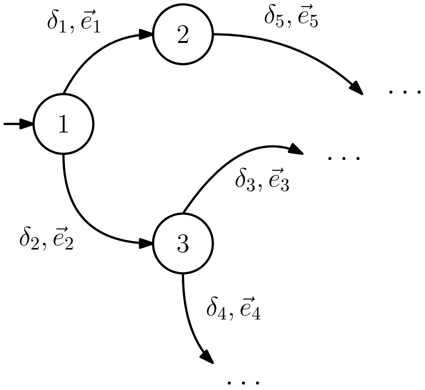
$$\delta_i = \quad (n) \xrightarrow{\delta_i, \binom{1}{3}} (m) \qquad \longrightarrow \qquad (n) \xrightarrow{\delta_i, \vec{e}_i} (m)$$

$$\vec{e}_i = (0, \cdots, 0, \underset{i}{1}, 0, \cdots, 0)$$

$$\phi(x_1, x_2) = (x_1 = 2x_2) \qquad \longrightarrow \qquad \begin{gathered} \phi'(y_1, \ldots, y_r) = \\ \sum_i v_{i,1} y_i = 2 \sum_i v_{i,2} y_i \end{gathered}$$

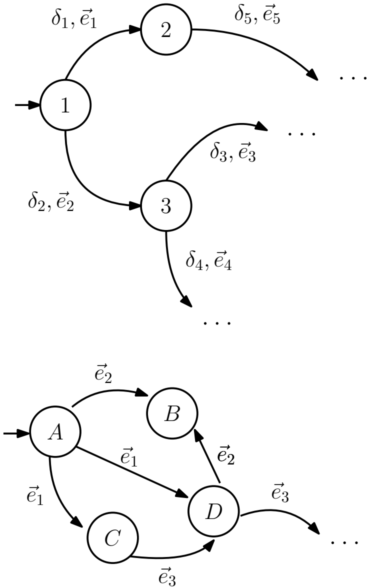
"Replace x_1 by $\ldots + y_i + \ldots$
and x_2 by $\ldots + 3y_i + \ldots$"

wuPA implies diagonal of $\mathbb{N}$-rational functions

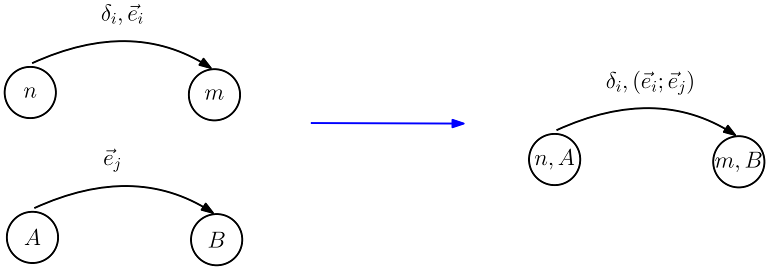


$$\phi'(y_1, \dots, y_r)$$

wuPA implies diagonal of ℕ-rational functions

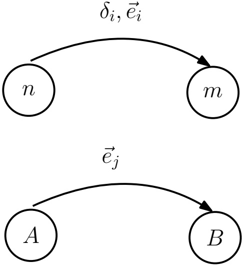


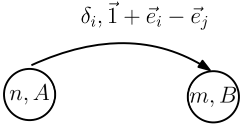
wuPA implies diagonal of $\mathbb{N}$-rational functions



$$\phi'(y_1, \ldots, y_r) \longrightarrow \Phi(y_1, \ldots, y_r; z_1, \ldots, z_r) = (y_1 = z_1 \wedge \ldots \wedge y_r = z_r)$$

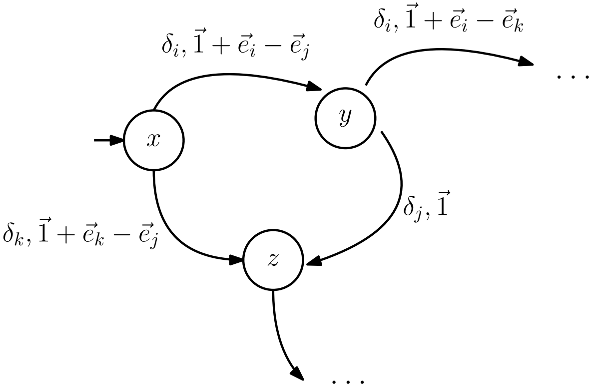
wuPA implies diagonal of $\mathbb{N}$-rational functions





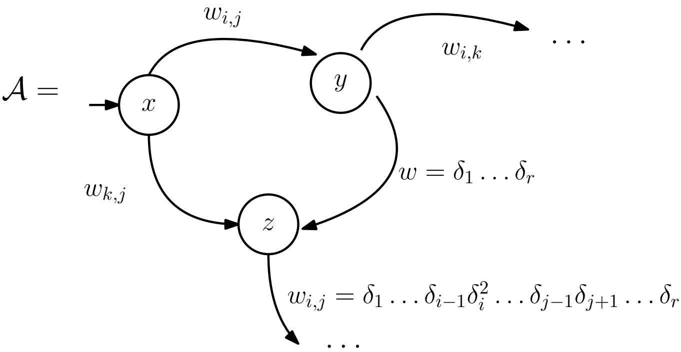
$\phi'(y_1, \ldots, y_r) \longrightarrow \Phi(y_1, \ldots, y_r) = (y_1 = \ldots = y_r)$

wuPA implies diagonal of $\mathbb{N}$-rational functions



$$\Phi(y_1, \ldots, y_r) = (y_1 = y_2 = \ldots = y_{r-1} = y_r)$$

wuPA implies diagonal of ℕ-rational functions

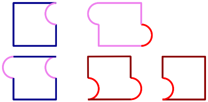


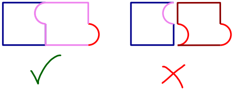
$$L(x) = \Delta A(y_1, \dots, y_r)$$

Diagonals of $\mathbb{N}$-rational functions

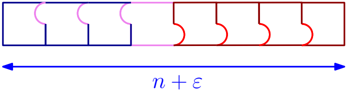
Irrational tiles [Garrabrant and Pak]

- set of tiles with height 1, with possible irrational length

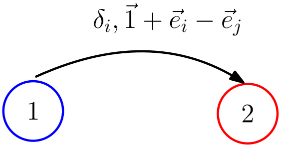


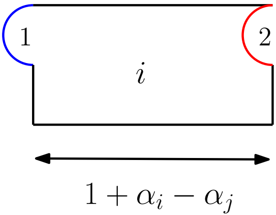
- overlapping compatibility 

[Garrabrant and Pak] $L(x) = \sum_n \ell_n x^n$ is the diagonal of an $\mathbb{N}$-rational function if and only if there is a set of tiles and some $\varepsilon > 0$ such that ℓ_n = number of tilings of the rectangle of length $n + \varepsilon$.



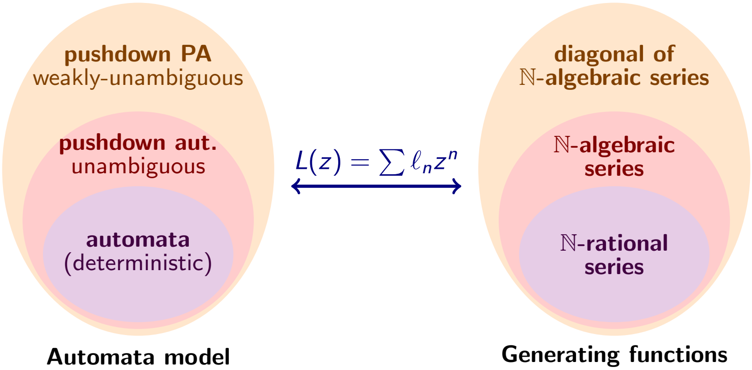
Diagonals of $\mathbb{N}$-rational functions





where $1, \alpha_1, \ldots, \alpha_r$ independent over $\mathbb{Q}$

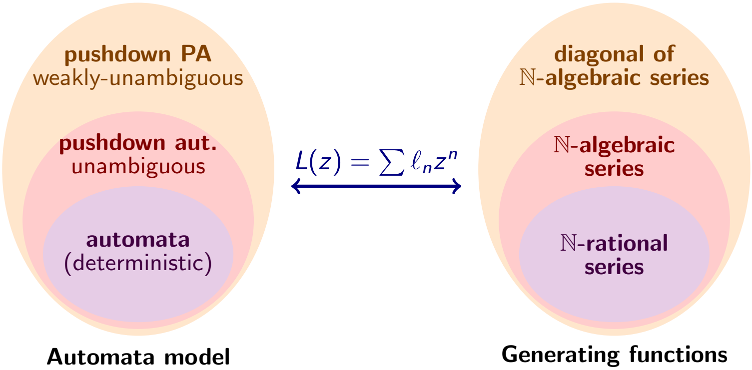
Conclusion



- more precise but less useful? Not enough closure properties
- diagonals of $\mathbb{N}$-rational series are not well understood
  Conjecture: Catalan's numbers are not diagonal of $\mathbb{N}$-rational series [Garrabrant and Pak '14]

Conclusion

THANK YOU !



- more precise but less useful? Not enough closure properties
- diagonals of $\mathbb{N}$-rational series are not well understood
  Conjecture: Catalan's numbers are not diagonal of $\mathbb{N}$-rational series [Garrabrant and Pak '14]

Bonus Analytic criteria for inherent ambiguity V2

Proposition

$\Omega_4 = \{w \in \{a,b,c,d\}^* : \neg(|w|_a = |w|_b = |w|_c = |w|_d)\}$ is inherently ambiguous.

Analytic proof:

- Suppose that $\Omega_4(x)$ is algebraic
- Let $I = (a+b+c+d)^* \setminus \Omega_4$
- But $I = \{w \in \{a,b,c,d\}^* : |w|_a = |w|_b = |w|_c = |w|_d\}$

$$[x^{4n}]I(x) = \binom{4n}{n,n,n,n} = \frac{(4n)!}{(n!)^4} \sim_{n\to\infty} \frac{\sqrt{2}}{2\pi^{3/2}} \frac{256^n}{n^{3/2}}$$

But $\Gamma(-3/2+1)\frac{\sqrt{2}}{2\pi^{3/2}} = -\frac{\sqrt{2}}{\pi}$ is transcendent!

Bonus Analytic criteria for inherent ambiguity V2

Proposition

$\Omega_4 = \{w \in \{a,b,c,d\}^* : \neg(|w|_a = |w|_b = |w|_c = |w|_d)\}$ is inherently ambiguous.

Analytic proof:

- Suppose that $\Omega_4(x)$ is algebraic
- Let $I = (a+b+c+d)^* \setminus \Omega_4$
- But $I = \{w \in \{a,b,c,d\}^* : |w|_a = |w|_b = |w|_c = |w|_d\}$

$$[x^{4n}]I(x) = \binom{4n}{n,n,n,n} = \frac{(4n)!}{(n!)^4} \sim_{n\to\infty} \frac{\sqrt{2}}{2\pi^{3/2}} \frac{256^n}{n^{3/2}}$$

But $\Gamma(-3/2+1)\frac{\sqrt{2}}{2\pi^{3/2}} = -\frac{\sqrt{2}}{\pi}$ is transcendent!

- Lazy proof: $\Omega_4 \cap (a+b+(cd))^* \simeq \Omega_3$.

More specific criteria on series (2)

Motivation: there exist many inherently ambiguous languages with very simple (rational !) series.

The language $\{a^n b^m c^p \text{ with } n \neq m \text{ or } m \neq p\}$ is inherently ambiguous [Makarov 21, Koechlin 21]

$$\frac{1}{(1-a)(1-b)(1-c)} - \frac{1}{1-abc} = \frac{a+b+c-ab-ac-bc}{(1-a)(1-b)(1-c)(1-abc)}$$

The 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 inherently weakly-ambiguous for PA.
$\rightarrow$ proof based on generating functions ?