Unambiguous Parikh automata and holonomic series

Florent Koechlin
LIPN Back-to-school day, Villetaneuse

October 10th, 2023

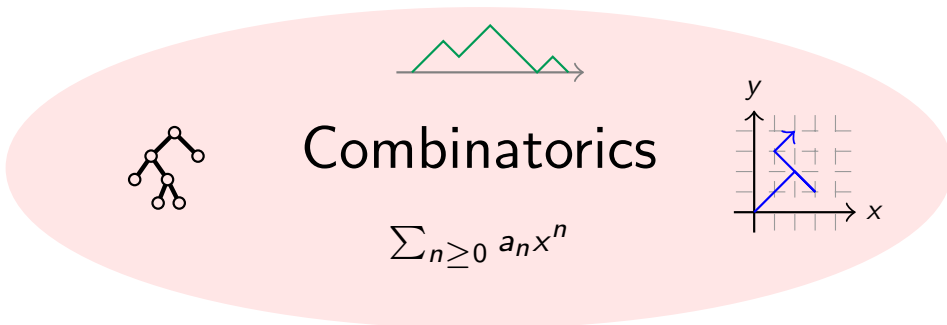
Curriculum

Arnaud Carayol
Cyril Nicaud
Pablo Rotondo

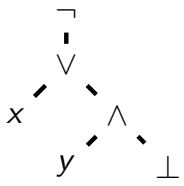
Alin Bostan (Inria Saclay)

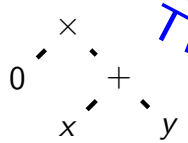
Mathilde Bouvel
Valentin Féray (IECL)
Xavier Goaoc

My research areas



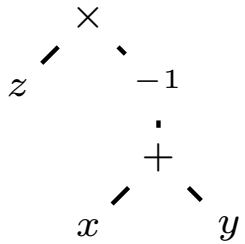
My research areas



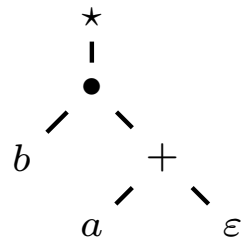


Random expression trees

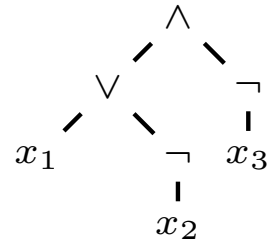
Thesis



$z \times (x+y)^{-1}$



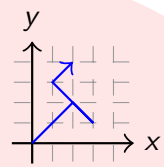
$(b \cdot (a+\varepsilon))^\star$



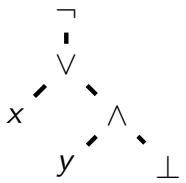
$(x_1 \vee \neg x_2) \wedge \neg x_3$

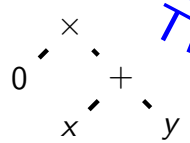
Combinatorics

$\sum_{n\geq 0} a_n x^n$



My research areas

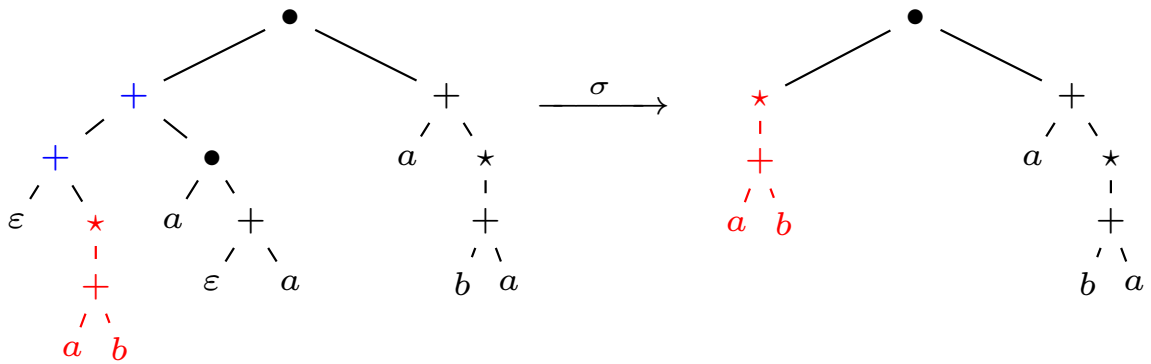




Random expression trees

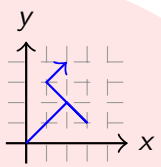
Thesis

Reduction by absorbing pattern:

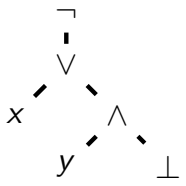


Combinatorics

$\sum_{n\geq 0} a_n x^n$

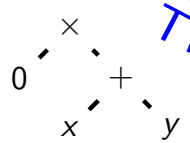


My research areas

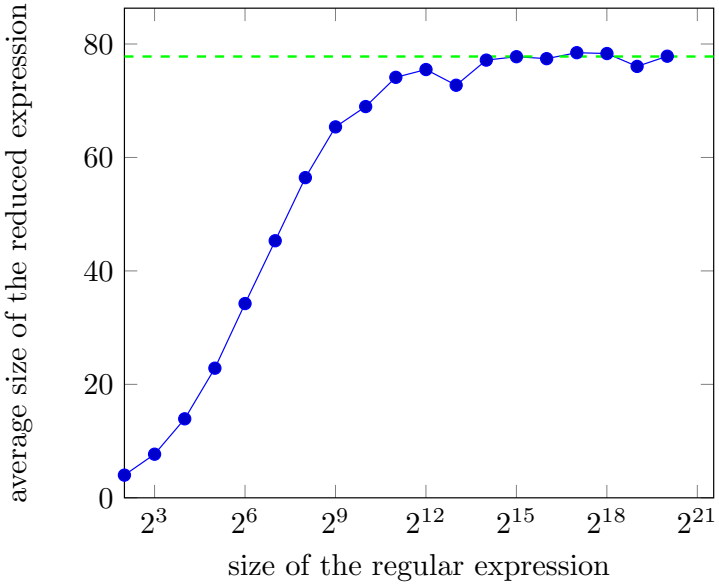


Uniform trees

Random expression trees

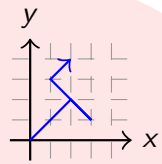


Thesis

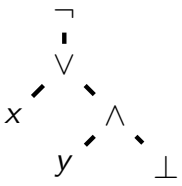


Combinatorics

$\sum_{n\geq 0} a_n x^n$

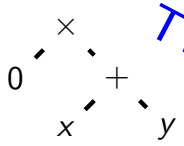


My research areas

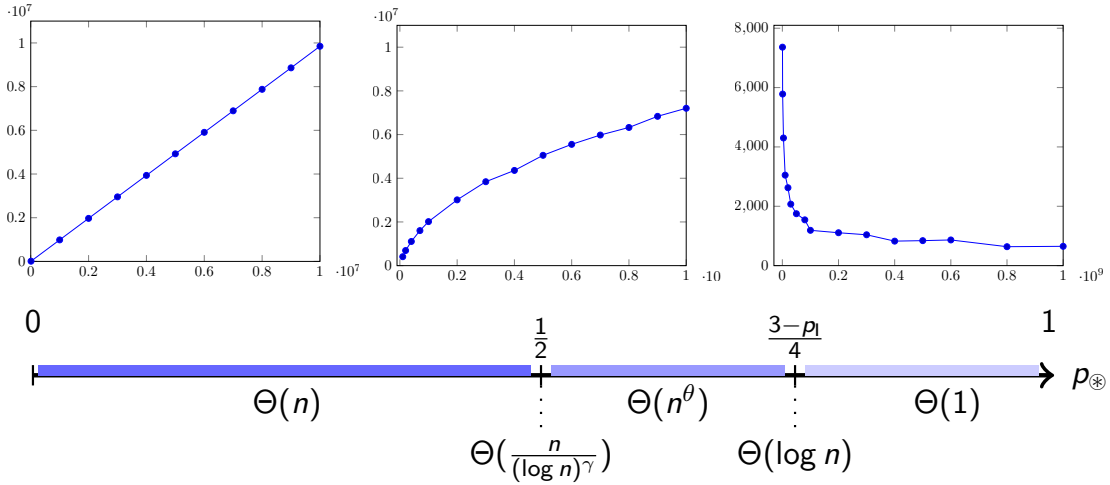


Uniform trees
BST distribution

Random expression trees

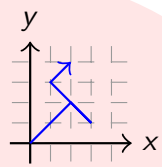


Thesis

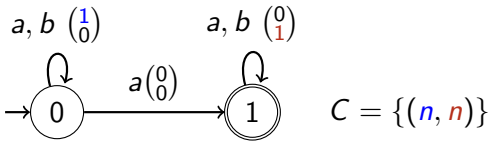


Combinatorics

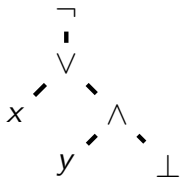
$\sum_{n\geq 0} a_n x^n$



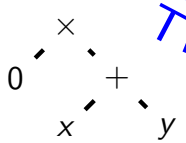
My research areas



Uniform trees
BST distribution



Random expression trees



Thesis

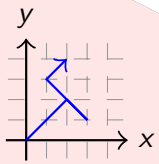
Unambiguous PA
Bounded CFL

Automata theory

Thesis

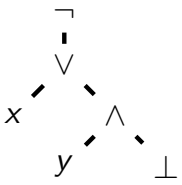
Combinatorics

$\sum_{n\geq 0} a_n x^n$



Computer Algebra

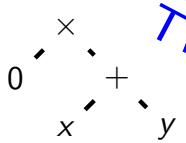
My research areas

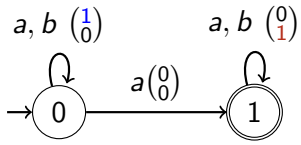


Uniform trees
BST distribution

Random expression trees

Thesis



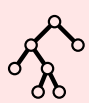


Unambiguous PA
Bounded CFL

Automata theory

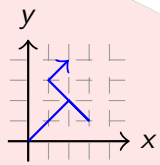
Thesis

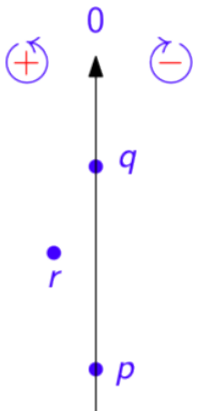
Computer Algebra



Combinatorics

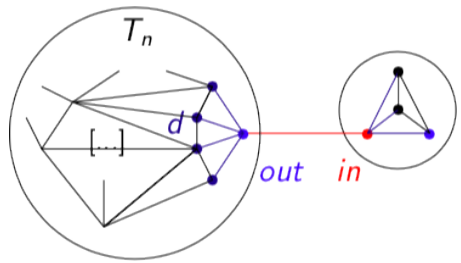
$\sum_{n\geq 0} a_n x^n$





Postdoc

Combinatorial geometry

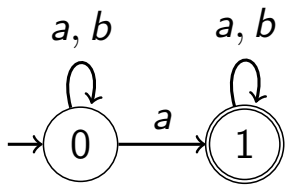


Automata theory

Language: abstract set of words built over a finite alphabet Σ.

Example

The set of words over $\{a, b\}$ containing at least on letter a :
a, ba, ab, aa, baa, aba, abb, ...



$\Sigma^* a \Sigma^*$

Automata theory
Describe

$$\begin{cases} S \to CS \mid B \\ C \to aCb \mid \varepsilon \\ B \to cB \mid \varepsilon \end{cases}$$

Link between describing and counting

Generating series

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

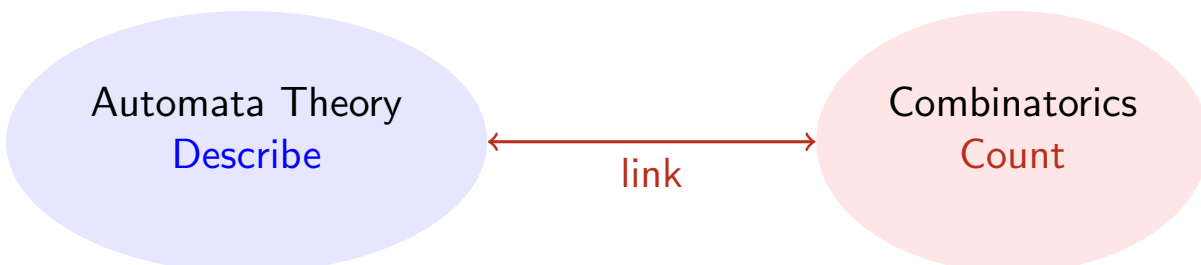
$$L(x) = \sum_{n=0}^{+\infty} \ell_n x^n$$

Example

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

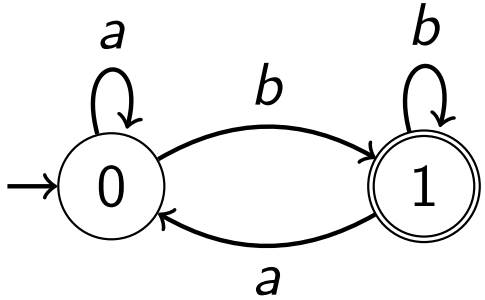
Example

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



Link between automata and generating series

regular languages $\subseteq$ unambiguous context-free languages



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

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

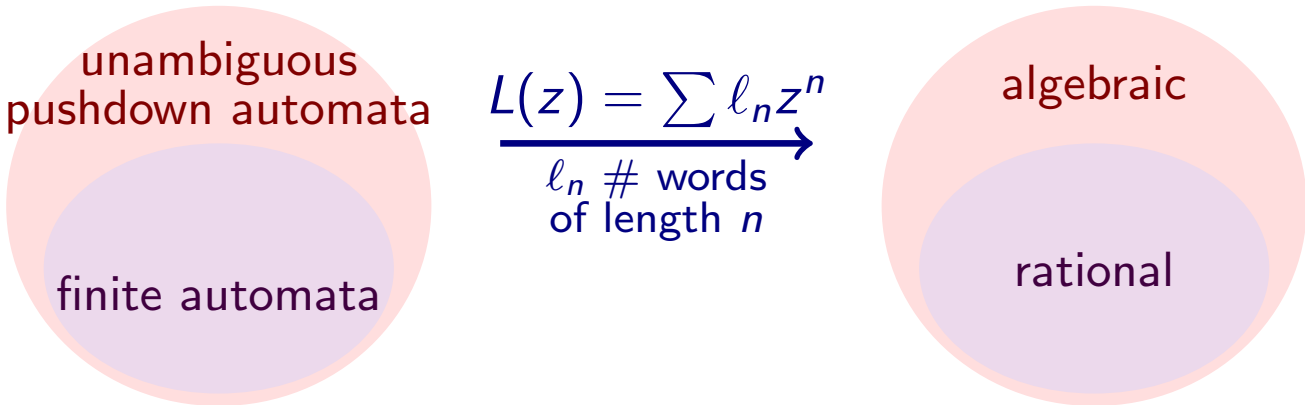
$$\begin{cases} S(x) = xS(x)B(x) + 1 \\ B(x) = xB(x) + xS(x) \end{cases}$$

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

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

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

Link between two hierarchies



Automata Generating series

Two remarkable applications :

- analytic proofs of inherent ambiguity [Flajolet 87]

Analytic criteria for inherent ambiguity

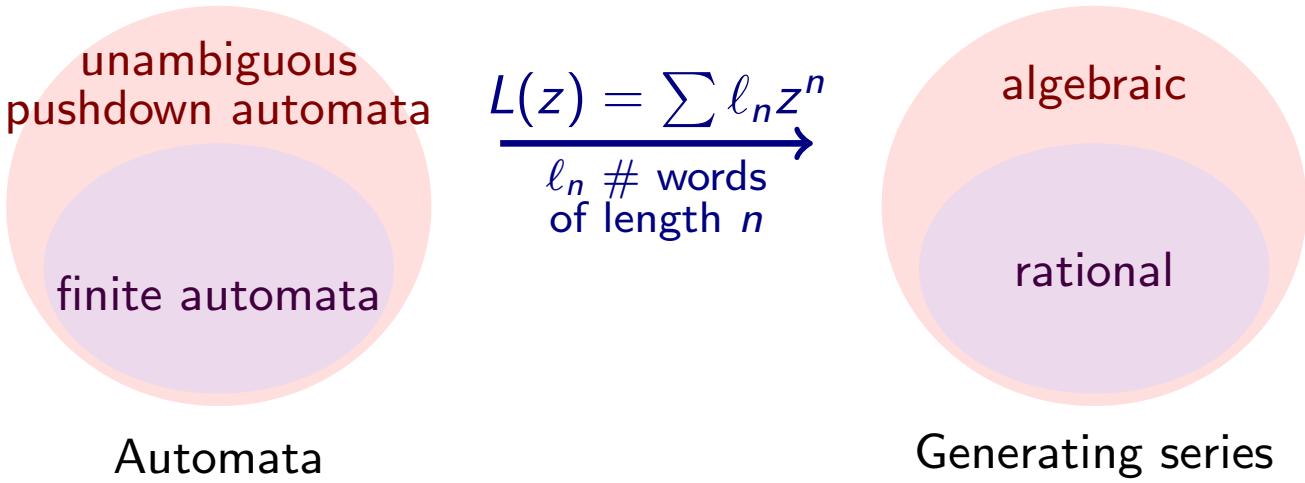
Theorem [Flajolet '87]
$\Omega_3 = \{w \in \{a,b,c\}^* : |w|_a \neq |w|_b \text{ or } |w|_b \neq |w|_c\}$ is inherently ambiguous.

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

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

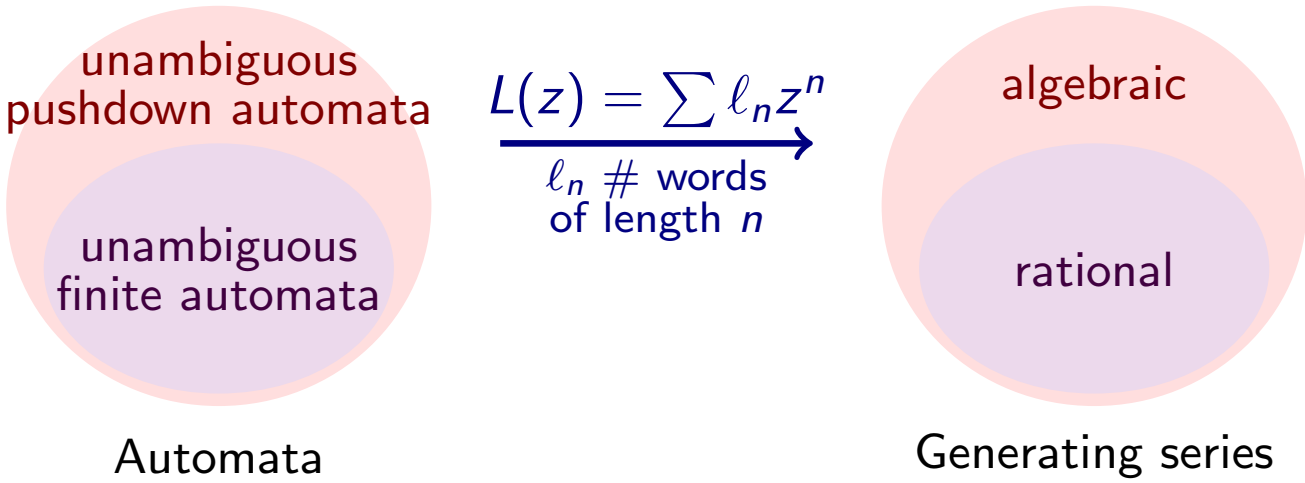
Link between two hierarchies



Two remarkable applications :

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

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]

Inclusion problem for unambiguous automata

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

Proposition: If $L(\mathcal{A}) \subsetneq L(\mathcal{B})$, there exists a small witness $w \in L(\mathcal{B}) \setminus L(\mathcal{A})$ of size at most $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

- $C(x) := \sum c_n x^n = B(x) - A(x)$ 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 $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

Inclusion problem for unambiguous automata

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

Proposition: If $L(\mathcal{A}) \subsetneq L(\mathcal{B})$, there exists a small witness $w \in L(\mathcal{B}) \setminus L(\mathcal{A})$ of size at most $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

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

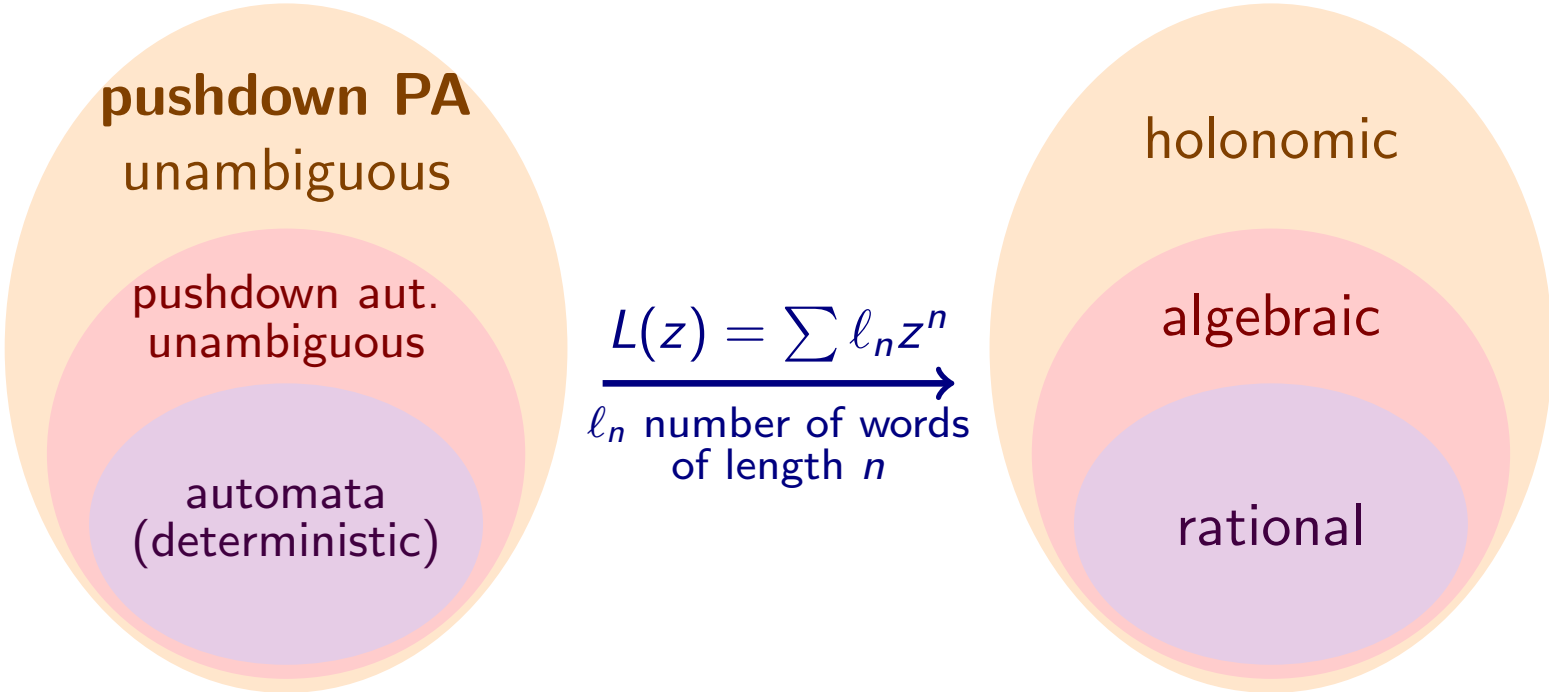
$$\forall n \geq r, c_n = \alpha_1 c_{n-1} + \cdots + \alpha_r c_{n-r}$$

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

Théorème [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.)

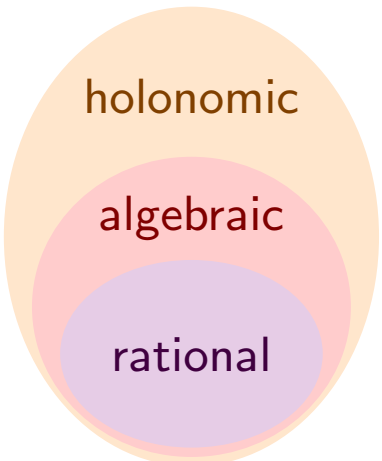
Extension of hierarchy



- During my phd: find a suitable model corresponding to holonomic series, that is relevant from a modeling / automata point of view

Holonomic series [Stanley 80]

Rational: $P(x)f(x) = Q(x)$ **Algebraic:** $P(x, f(x)) = 0$.



Generating series

Definition: A series $f(x) = \sum_n a_n x^n$ is holonomic (or D-finite) 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 $p_r(n)a_{n+r} + \ldots + p_0(n)a_n = 0$

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

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

Generalizable in several variables

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

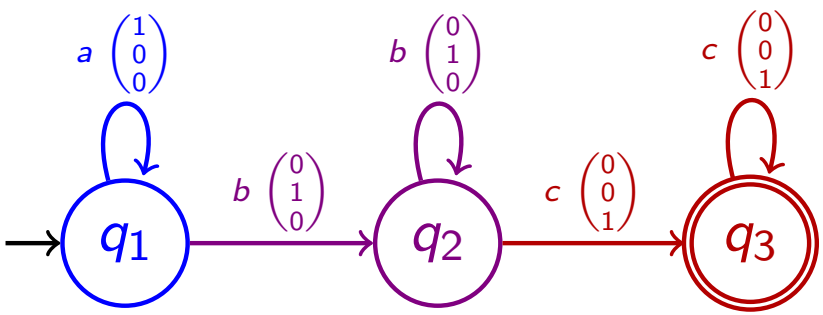
Parikh automata (PA) [Klaedtke, Rueß '03]

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

Idea: add constrained counters

- transitions labelled by vectors in $\mathbb{N}^d$
- test on the final value at the end of the run

Parikh automaton: finite automaton on $\Sigma \times \mathbb{N}^d$ with semilinear constraints

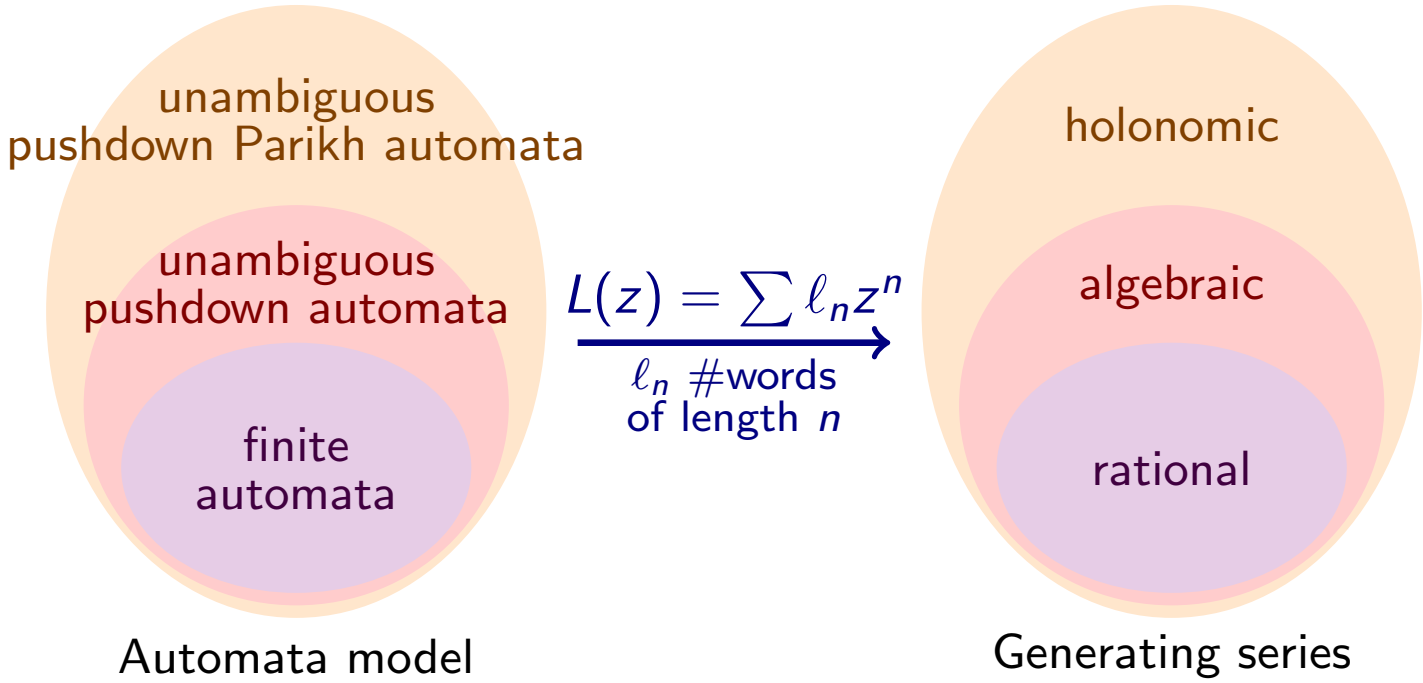


$$\mathcal{C} = \left\{ \begin{pmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{pmatrix} : \ell_1 = \ell_2 = \ell_3 \right\}$$

$$q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[b]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3 \xrightarrow[c]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3,\ w = aabbcc,\ v = \begin{pmatrix}2\\2\\2\end{pmatrix} \in \mathcal{C}$$

Can be extended with a stack.

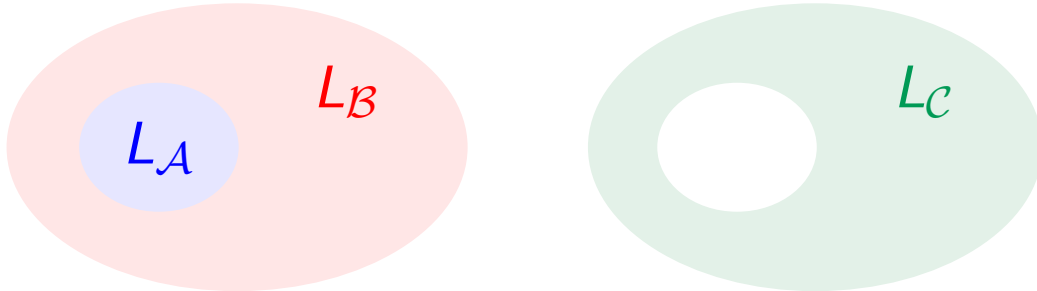
Extension



Two remarkable applications :

- analytic proofs of inherent ambiguity for PA
- doubly exponential bound for an algorithm for the inclusion problem for uPA

Algorithmic application: inclusion problem for uPA



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

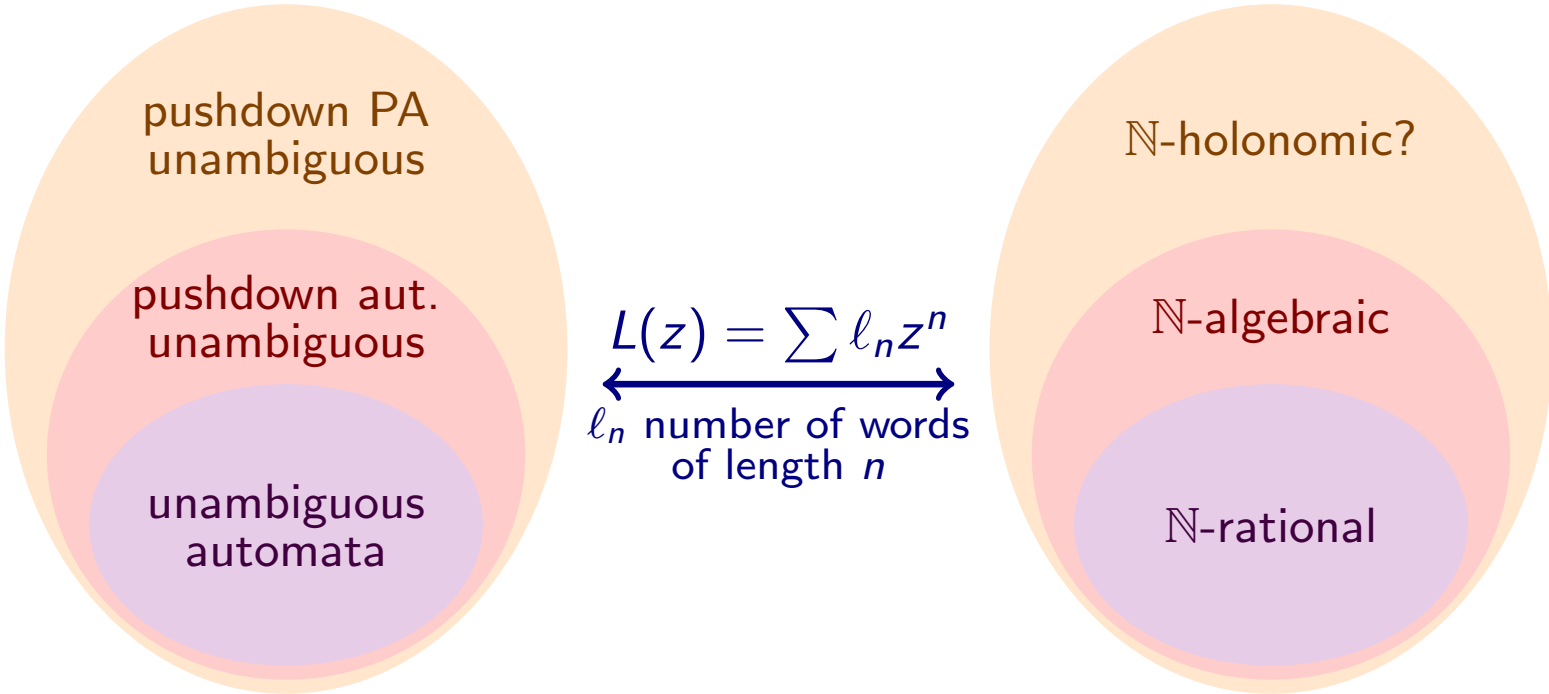
idea: replace $L_{\mathcal{C}} \stackrel{?}{=} \emptyset$ by $C(x) \stackrel{?}{=} 0$

- "Compute" $A(x)$ and $B(x)$ from $\mathcal{A}$ and $\mathcal{B}$
  $\rightarrow$ possible by 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 $(n - 100)c_n = 0$

Garrabrant & Pak's conjecture



Subclass of holonomic series:

- natural from combinatorics
- rich enough to have nice closure and algorithmic properties

Garrabrant & Pak's conjecture

If $f(x,y) = \sum_{n,m} a_{n,m} x^n y^m$, the diagonal of f is defined by

$$\Delta f(x) = \sum_{n \in \mathbb{N}} a_{n,n} x^n \,.$$

$$C(x) = \frac{1-\sqrt{1-4x}}{2x} = \Delta \frac{1-x/y}{1-x-y}$$

A series is the diagonal of a $\mathbb{N}$-rational series if and only if it the generating series of an unambiguous PA.

Conjecture [Garrabrant & Pak '14]:
The series of the Catalan numbers is not the diagonal of any $\mathbb{N}$-rational series.

Garrabrant & Pak's conjecture

If $f(x,y) = \sum_{n,m} a_{n,m} x^n y^m$, the diagonal of f is defined by

$$\Delta f(x) = \sum_{n \in \mathbb{N}} a_{n,n} x^n \,.$$

$$C(x) = \frac{1-\sqrt{1-4x}}{2x} = \Delta \frac{1-x/y}{1-x-y}$$

A series is the diagonal of a $\mathbb{N}$-rational series if and only if it the generating series of an unambiguous PA.

Conjecture [Garrabrant & Pak '14]:
The series of the Catalan numbers is not the diagonal of any $\mathbb{N}$-rational series.

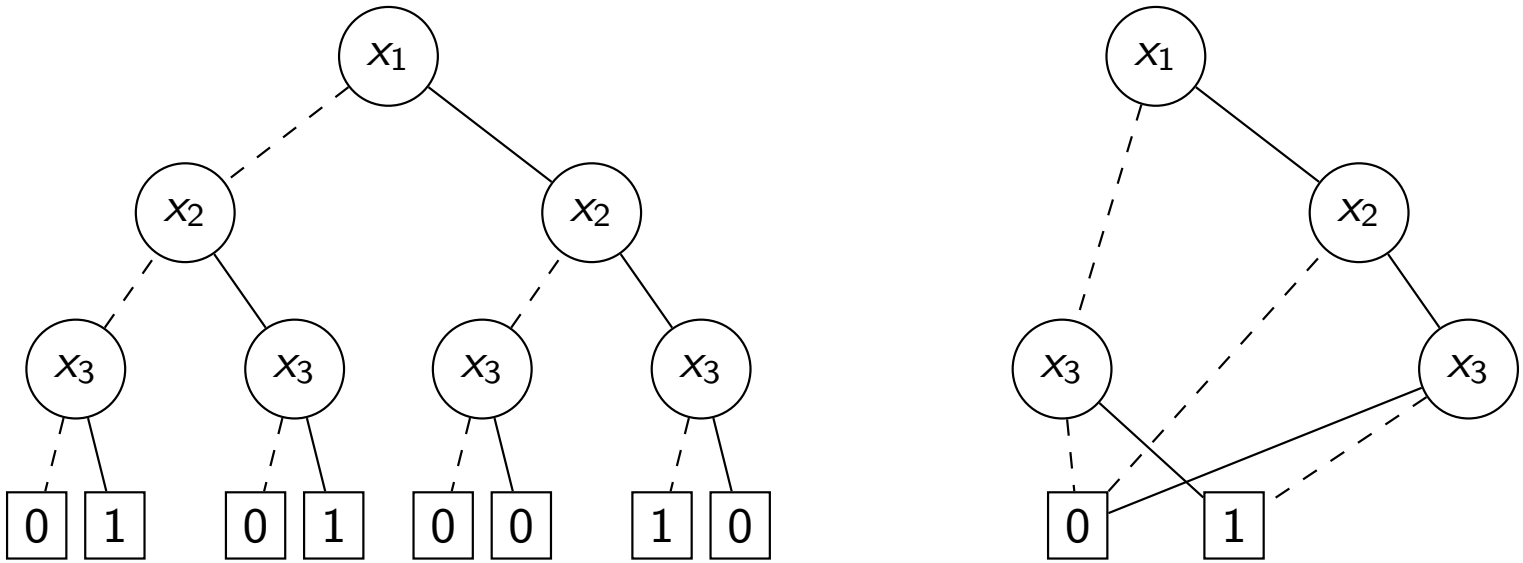
THANK YOU !

Sommaire annexes

Pistes de recherche sur les fonctions booléennes

Structures compactes: circuits et BDD

$$f = (x_1 \wedge x_2 \wedge \neg x_3) \vee (\neg x_1 \wedge x_3)$$



Pistes de recherche:

- les raisons qui font que l'élément absorbant / neutre n'agit plus aussi fortement que dans les structures d'arbres ?
- généraliser ces structures compactes à d'autres classes d'expressions?

Publications

retour

Conférences

- Arnaud Carayol, Philippe Duchon, Florent Koechlin and Cyril Nicaud. One Drop of Non-Determinism in a Random Deterministic Automaton. STACS'23.
- Florent Koechlin. New analytic techniques for proving the inherent ambiguity of context-free languages. FSTTCS'22
- Florent Koechlin and Pablo Rotondo. Analysis of an efficient reduction algorithm for random regular expressions based on universality detection. CSR'21
- Florent Koechlin and Pablo Rotondo. Absorbing patterns in BST-like expression-trees. STACS'21
- Alin Bostan, Arnaud Carayol, Florent Koechlin and Cyril Nicaud. Weakly-unambiguous Parikh automata and their link to holonomic series. ICALP'20
- Florent Koechlin, Cyril Nicaud and Pablo Rotondo. On the Degeneracy of Random Expressions Specified by Systems of Combinatorial Equations. DLT'20
- Florent Koechlin, Cyril Nicaud and Pablo Rotondo. Uniform Random Expressions Lack Expressivity. MFCS'19

Publications

retour

Versions longues

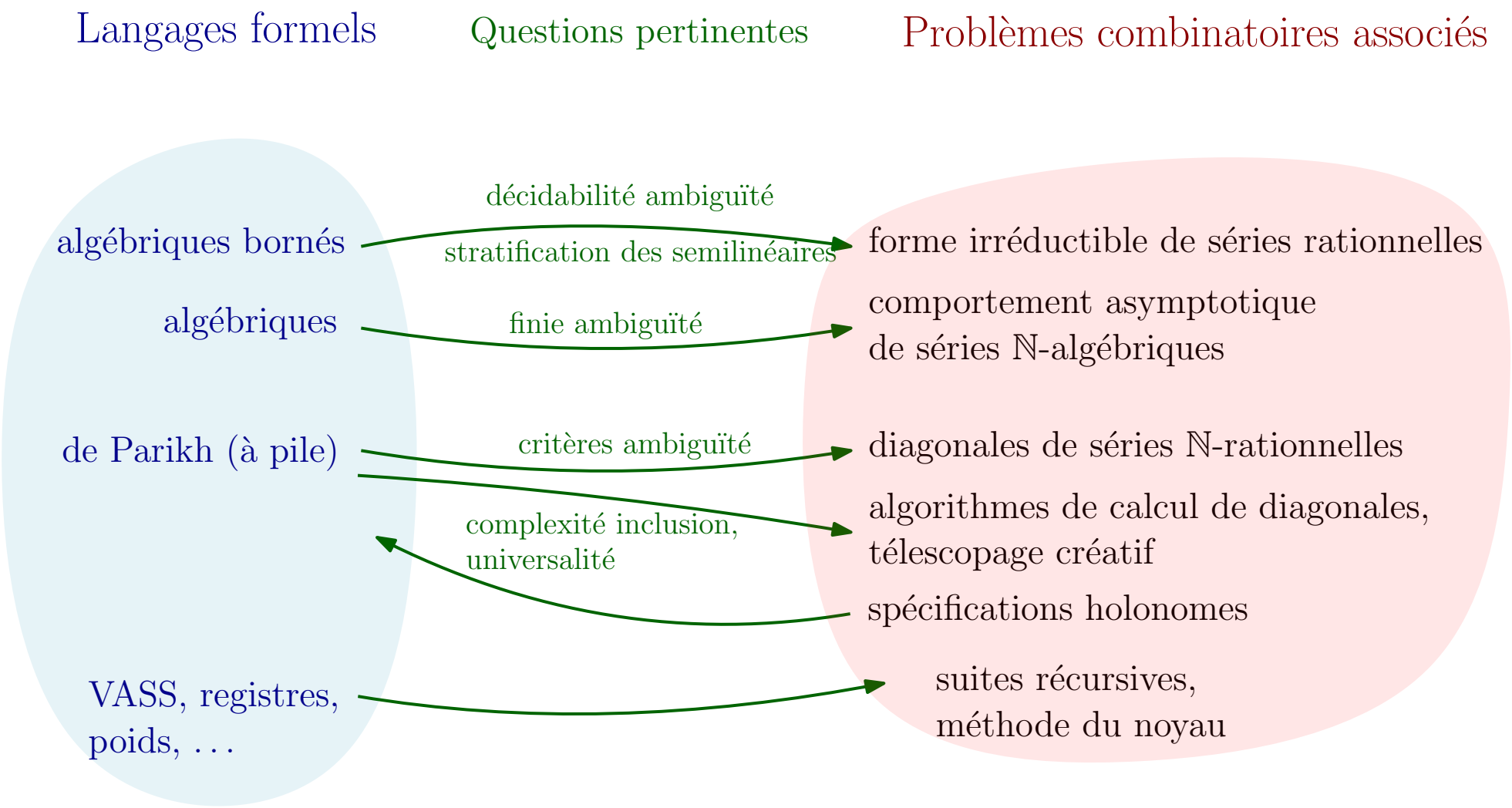
- Florent Koechlin, Cyril Nicaud and Pablo Rotondo.
  Simplifications of Uniform Expressions Specified by Systems.
  IJFCS'21, Volume No. 32, Issue No. 06, pp. 733 - 760.

Soumissions en cours

- Florent Koechlin et Pablo Rotondo. Analysis of an efficient reduction algorithm for random regular expressions based on universality detection.
  Soumis à Theory of Computing Systems (TCS), en tant que special issue de CSR'21.
- Florent Koechlin et Pablo Rotondo, The effects of semantic simplifications on random BST-like expression-trees,
  Soumis à Discrete Mathematics

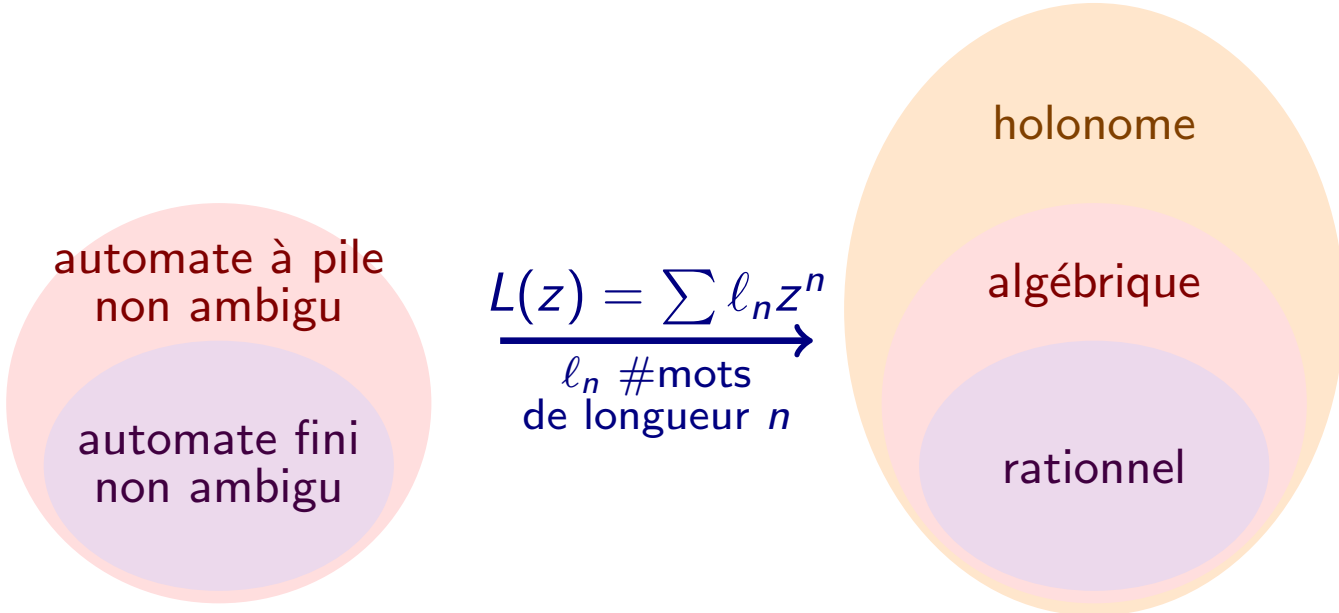
Projet de recherche : Interaction entre combinatoire et théorie des automates

retour



Focus: séries provenant de la théorie des langages

retour



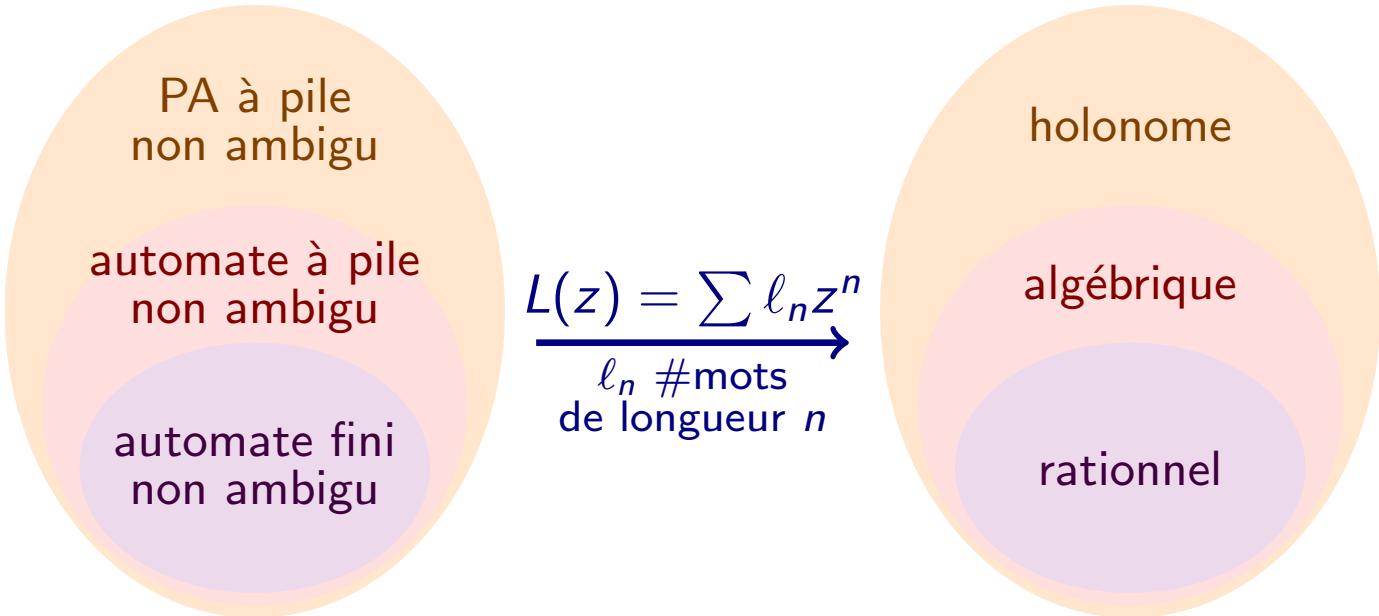
Modèle d'automates

Série génératrice

- Dans ma thèse: trouver un modèle correspondant aux séries holonomes, pertinent en modélisation et théorie des automates

Focus: séries provenant de la théorie des langages

retour

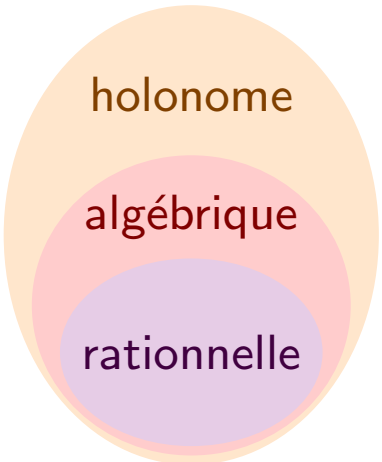


- Dans ma thèse: trouver un modèle correspondant aux séries holonomes, pertinent en modélisation et théorie des automates

Séries holonomes [Stanley 80]

retour

Rationnelle: $P(x)f(x) = Q(x)$ **Algébrique:** $P(x, f(x)) = 0$.



Série génératrice

Holonome: Équation différentielle linéaire

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

Définition alternative: les coefficients a_n satisfont une récurrence linéaire $p_r(n)a_{n+r} + \ldots + p_0(n)a_n = 0$

Exemple: $F(x) = e^x$ holonome mais pas algébrique

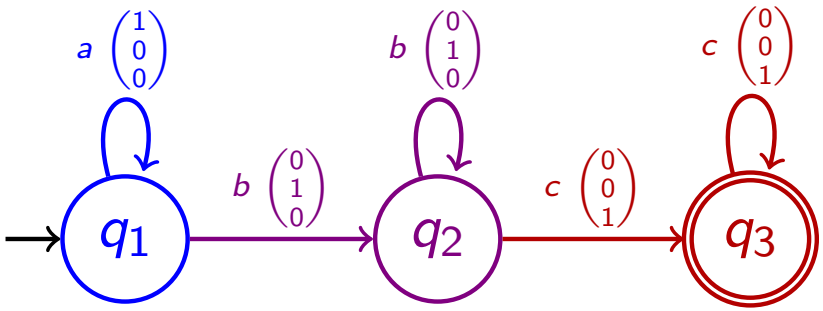
- équation différentielle: $F' - F = 0$
- récurrence: $(n+1)f_{n+1} - f_n = 0$

Généralisable à plusieurs variables: systèmes aux dérivées partielles

Automates de Parikh (PA) [Klaedtke, Rueß '03]

retour

Automate de Parikh: automate fini sur $\Sigma \times \mathbb{N}^d$ avec des contraintes semilinéaires sur des vecteurs d'entiers



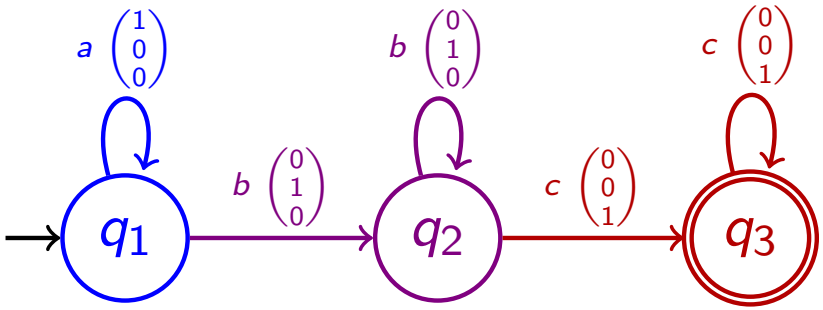
$$\mathcal{C} = \left\{ \begin{pmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{pmatrix} : \ell_1 = \ell_2 = \ell_3 \right\}$$

$$q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[c]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3 \xrightarrow[c]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3,\ w = aabbcc,\ v = \begin{pmatrix}2\\2\\2\end{pmatrix} \in \mathcal{C}$$

Automates de Parikh (PA) [Klaedtke, Rueß '03]

retour

Automate de Parikh: automate fini sur $\Sigma \times \mathbb{N}^d$ avec des contraintes semilinéaires sur des vecteurs d'entiers



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

$$q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[c]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3 \xrightarrow[c]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3,\ w = aabbcc,\ v = \begin{pmatrix}2\\2\\2\end{pmatrix} \in \mathcal{C}$$

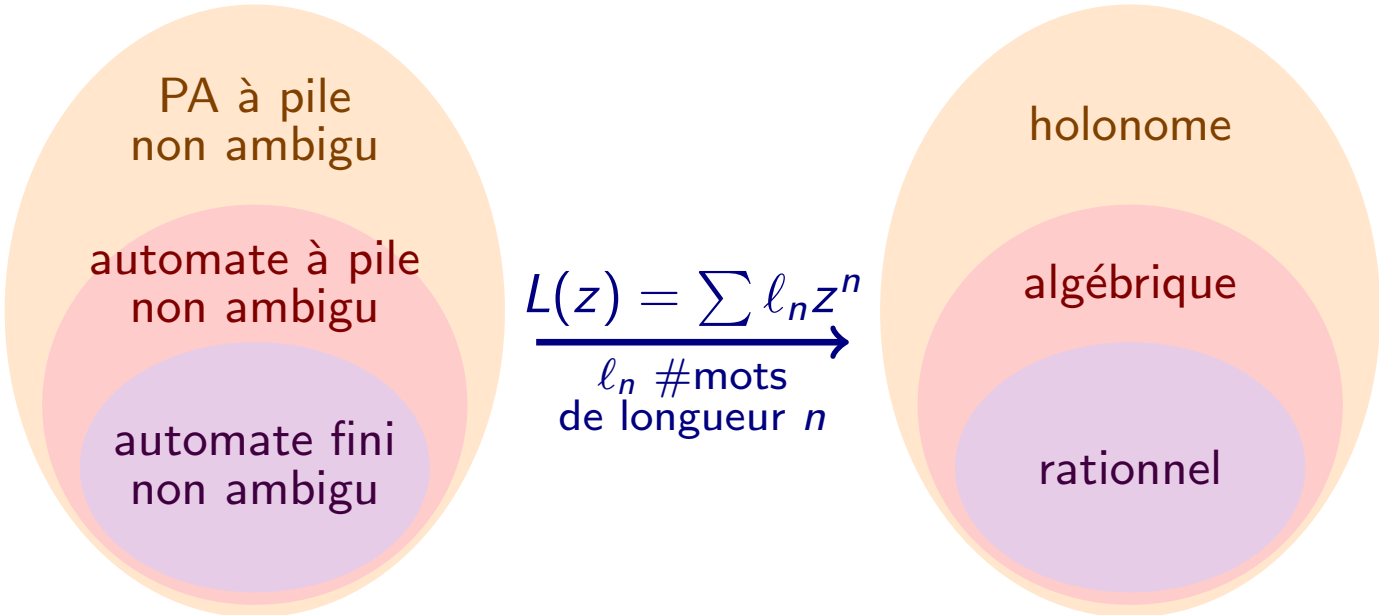
La série d'un langage de Parikh (à pile) non ambigu est holonome.

Systèmes associés holonomes:

$$\begin{cases} (a - 27\,a^2bc)\,\frac{\partial^2}{\partial a^2}F + (1 - 54\,abc)\,\frac{\partial}{\partial a}F + 6\,bc\,F = 0 \\ (b - 27\,b^2ca)\,\frac{\partial^2}{\partial b^2}F + (1 - 54\,bca)\,\frac{\partial}{\partial b}F + 6\,ca\,F = 0 \\ (c - 27\,c^2ab)\,\frac{\partial^2}{\partial c^2}F + (1 - 54\,cab)\,\frac{\partial}{\partial c}F + 6\,ab\,F = 0 \end{cases}$$

Focus: séries provenant de la théorie des langages

retour



Modèle d'automates

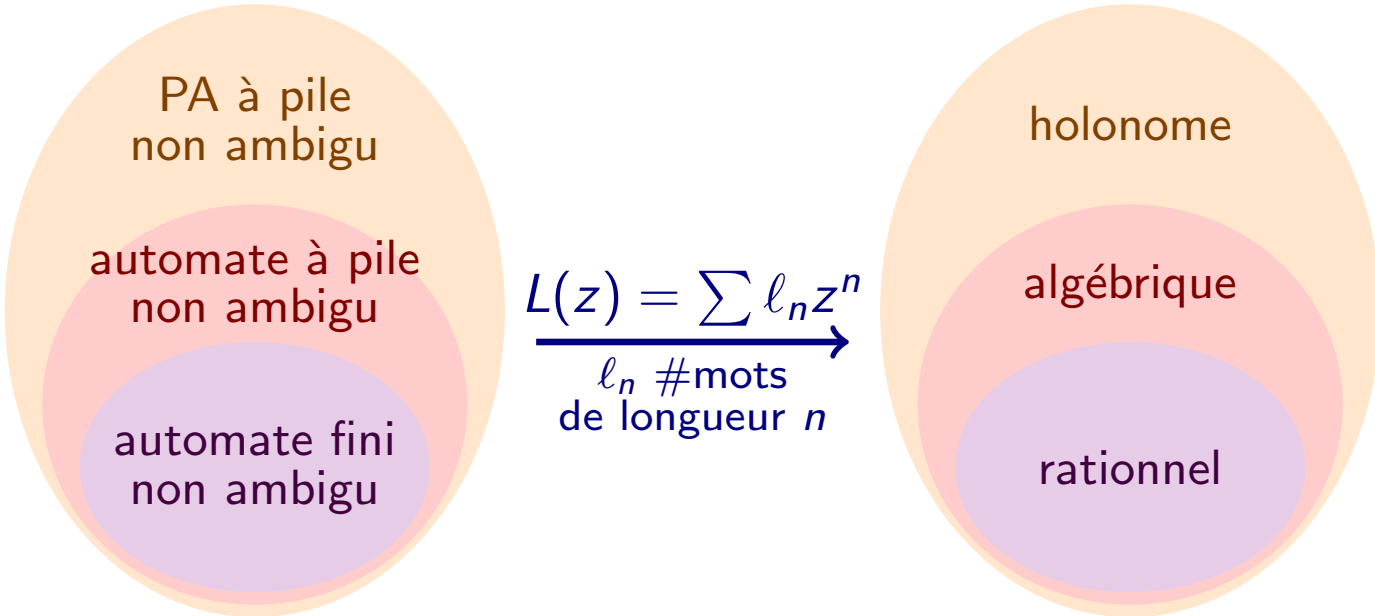
Série génératrice

Contributions [ICALP '20]:

- existence de langages intrinsèquement ambigus, même avec pile et compteurs + techniques pour en trouver
- l'inclusion des PA non ambigus est décidable en 2-EXPTIME
  → analyse fine des tailles des séries par des outils de calcul formel

Focus: séries provenant de la théorie des langages

retour



Modèle d'automates

Série génératrice

Systèmes associés holonomes:

$$\begin{cases} (a - 27\,a^2bc)\,\frac{\partial^2}{\partial a^2}F + (1 - 54\,abc)\,\frac{\partial}{\partial a}F + 6\,bc\,F = 0 \\ (b - 27\,b^2ca)\,\frac{\partial^2}{\partial b^2}F + (1 - 54\,bca)\,\frac{\partial}{\partial b}F + 6\,ca\,F = 0 \\ (c - 27\,c^2ab)\,\frac{\partial^2}{\partial c^2}F + (1 - 54\,cab)\,\frac{\partial}{\partial c}F + 6\,ab\,F = 0 \end{cases}$$

Limite des séries : Langages algébriques bornés

retour

Problème: Certains langages ambigus ont une série rationnelle.

$\rightarrow a^n b^m c^p$ avec $n = m$ ou $m = p$ et $(ab)^* c + a^*(bc)^*$

Séries génératrices associées sont *rationnelles*.

Exemple: $L = \{a^n b^m c^p$ avec $n \neq m$ ou $m \neq p\}$:

$S(a,b,c) = \frac{a+b+c-ab-ac-bc}{(1-a)(1-b)(1-c)(1-abc)}$ [Makarov 21, Koechlin 22]

Bornés: $L \subseteq w_1^* \dots w_d^*$

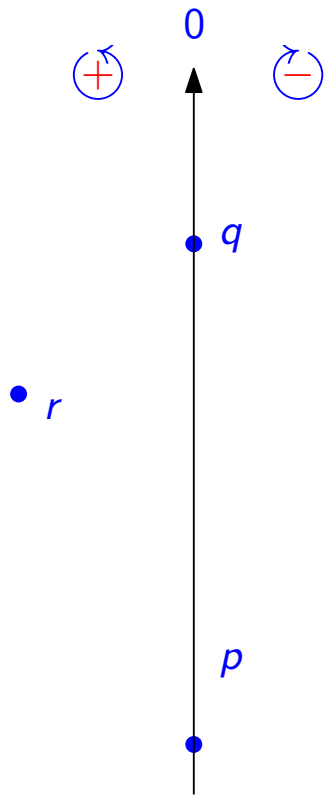
Conjecture [Ginsburg & Ullian '66] L'intrinsèque ambiguïté des langages algébriques bornés est décidable.

L'approche par les séries ouvre de nouvelles approches, qui s'avèrent plus efficaces que la plupart des critères existants.

Types d'ordre

retour

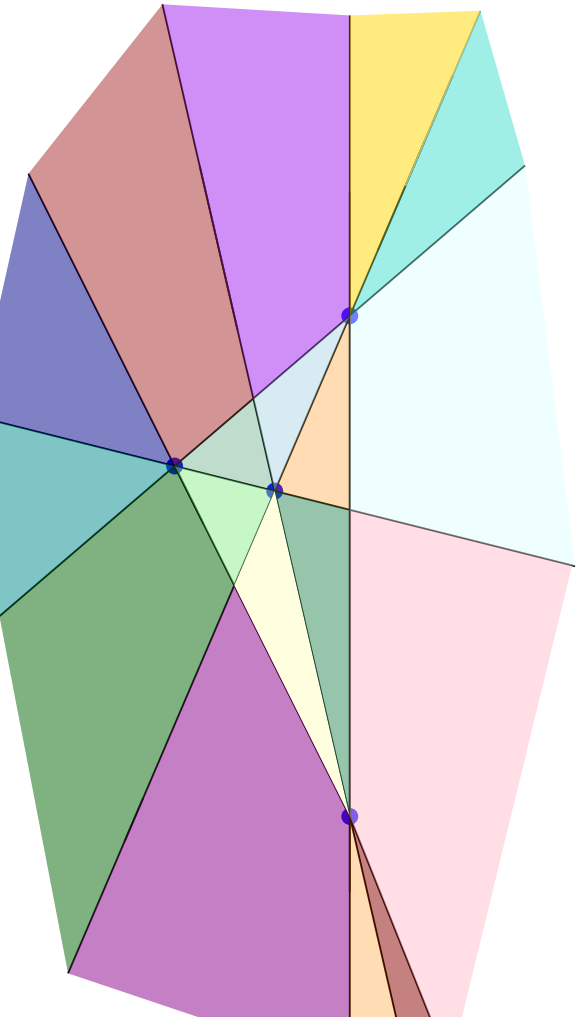
Orientation, type d'ordre



Orientation de trois points

- $\chi(p,q,r) = \begin{cases} +1 & \text{si } r \text{ est à gauche de } (pq) \\ 0 & \text{si } r \text{ est sur } (pq) \\ -1 & \text{si } r \text{ est à droite de } (pq) \end{cases}$
- $\chi(p,q,r) = sign \begin{vmatrix} x_p & x_q & x_r \\ y_p & y_q & y_r \\ 1 & 1 & 1 \end{vmatrix}$

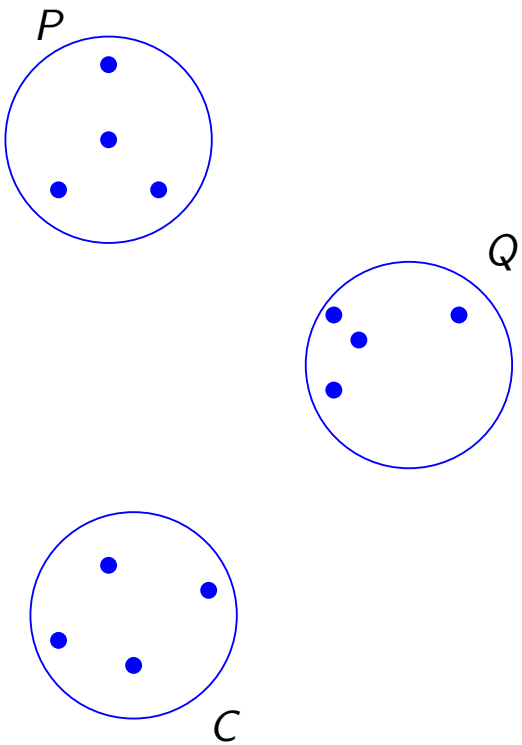
Orientation, type d'ordre



Orientation de trois points

- $\chi(p,q,r) = \begin{cases} +1 & \text{si } r \text{ est à gauche de } (pq) \\ 0 & \text{si } r \text{ est sur } (pq) \\ -1 & \text{si } r \text{ est à droite de } (pq) \end{cases}$
- $\chi(p,q,r) = sign \begin{vmatrix} x_p & x_q & x_r \\ y_p & y_q & y_r \\ 1 & 1 & 1 \end{vmatrix}$
- $f(z) = (\chi(p,q,z))_{p,q \in P}$ constant dans les cellules

Orientation, type d'ordre



Type d'ordre

- Deux ensembles P et Q de points de $\mathbb{R}^2$ *ont le même type d'ordre* s'il y a une bijection $f : P \to Q$ qui préserve les orientations :

$$\forall p, q, r \in P, \quad \chi(p, q, r) = \chi(f(p), f(q), f(r)).$$

- Type d'ordre = *classe d'équivalence* pour cette relation

Motivation

Dans la suite, les types d'ordre considérés sont *simples* : $\chi(p,q,r) = \pm 1$.

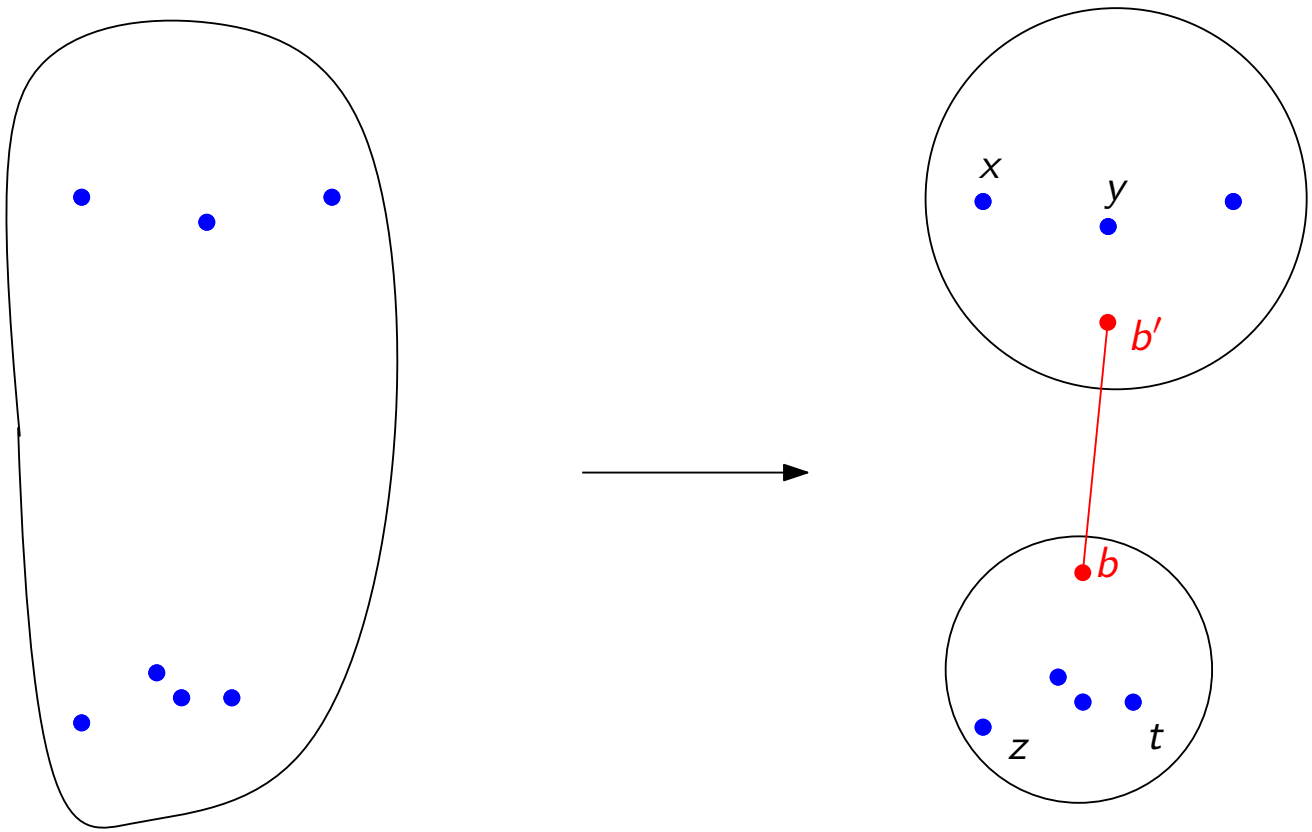
Motivation :

- Algorithmes exacts :
  - $[p,q] \in Conv(P) \Leftrightarrow \forall x \in P \setminus \{p,q\}, \chi(p,q,x) = cst$
  - $x \in \Delta(p,q,r) \Leftrightarrow \chi(p,q,x) = \chi(q,r,x) = \chi(r,p,x)$
  - (p,q) coupe $[a,b] \Leftrightarrow \chi(p,q,a) = -\chi(q,r,b)$
  - $[p,q]$ et $[a,b]$ se coupent $\Leftrightarrow \ldots$
- Benchmarks

Difficultés

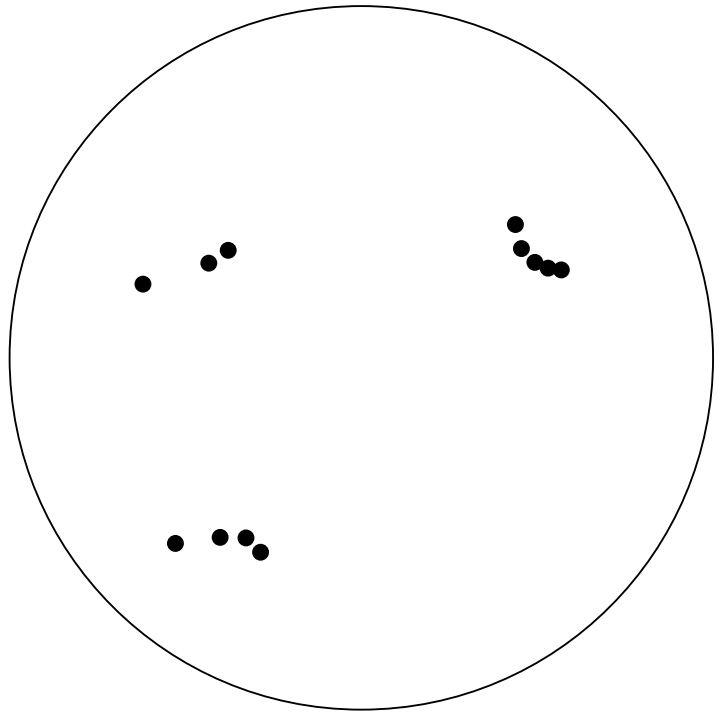
- Difficiles à décrire exhaustivement ($n = 11$ [Aichholzer et al.])
- Difficiles à compter $t_n = n^{3n+o(n)}$ [Alon, Warren]

Split selon un module

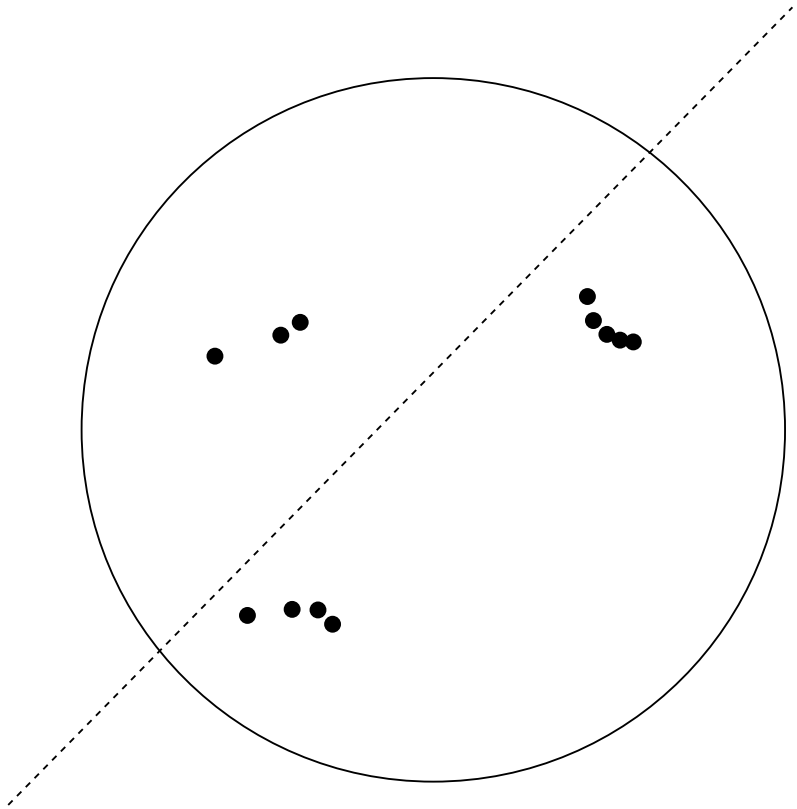


Règle : $\chi(x,y,z) = \chi(x,y,b')$ et $\chi(x,z,t) = \chi(b,z,t)$

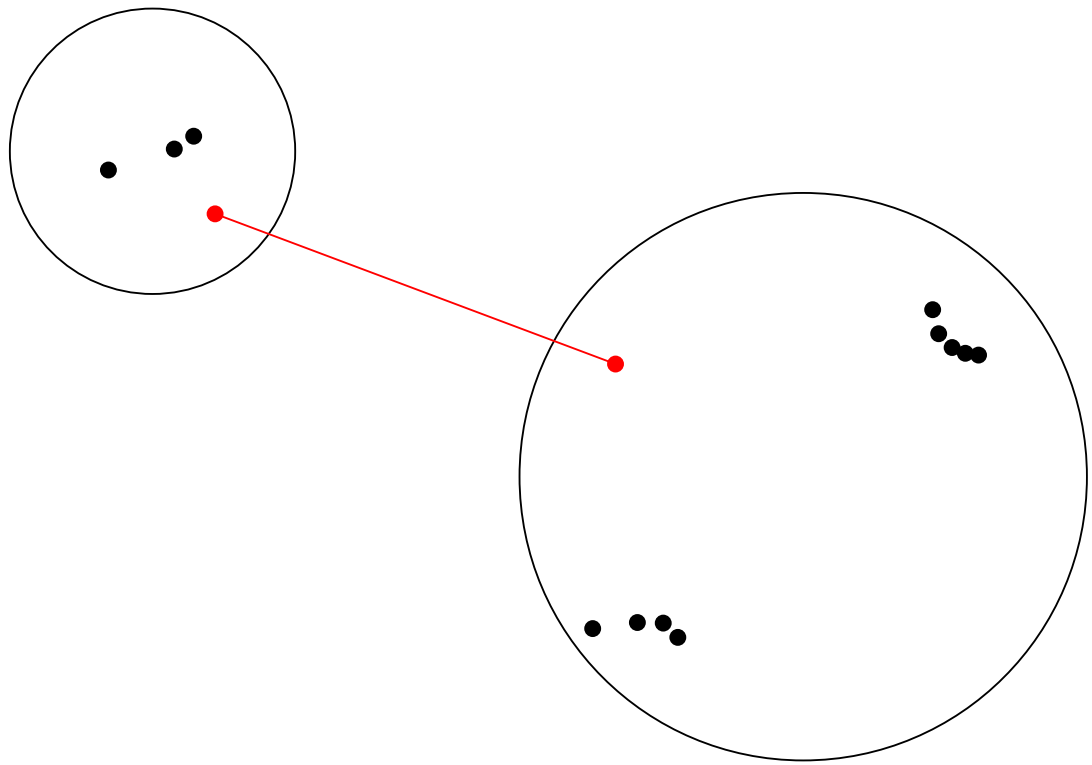
Décomposition modulaire



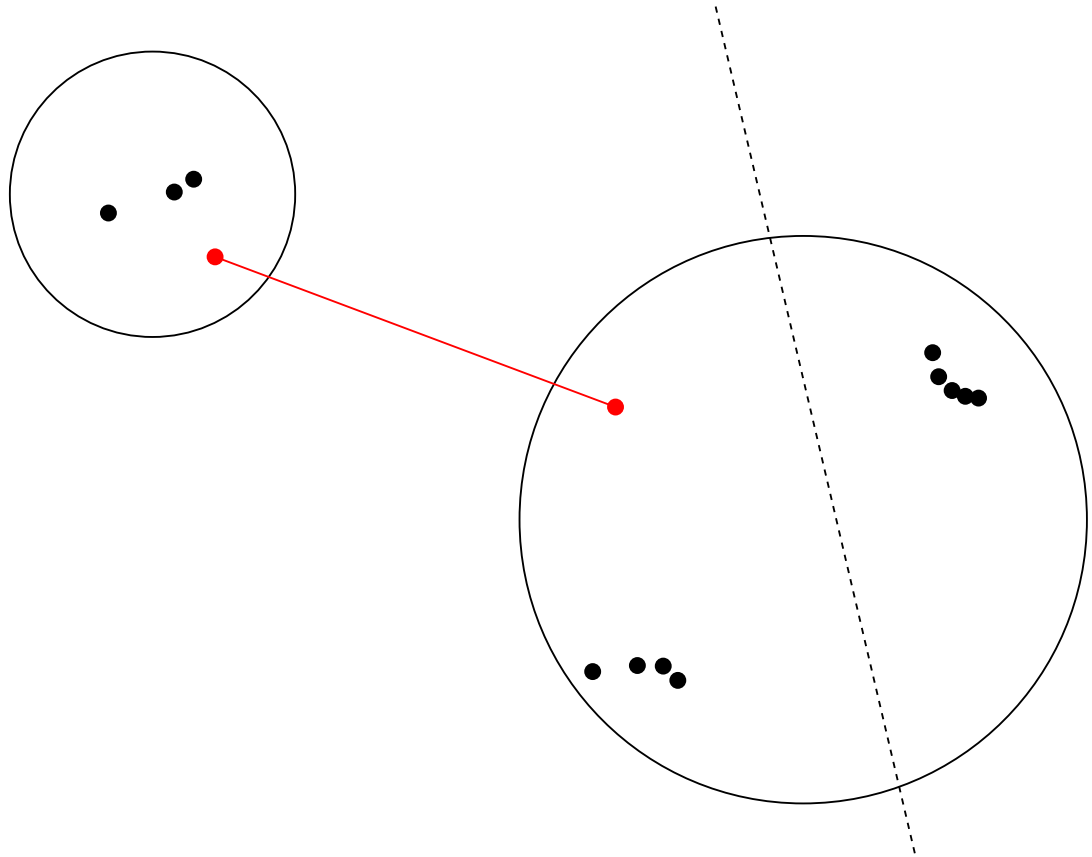
Décomposition modulaire



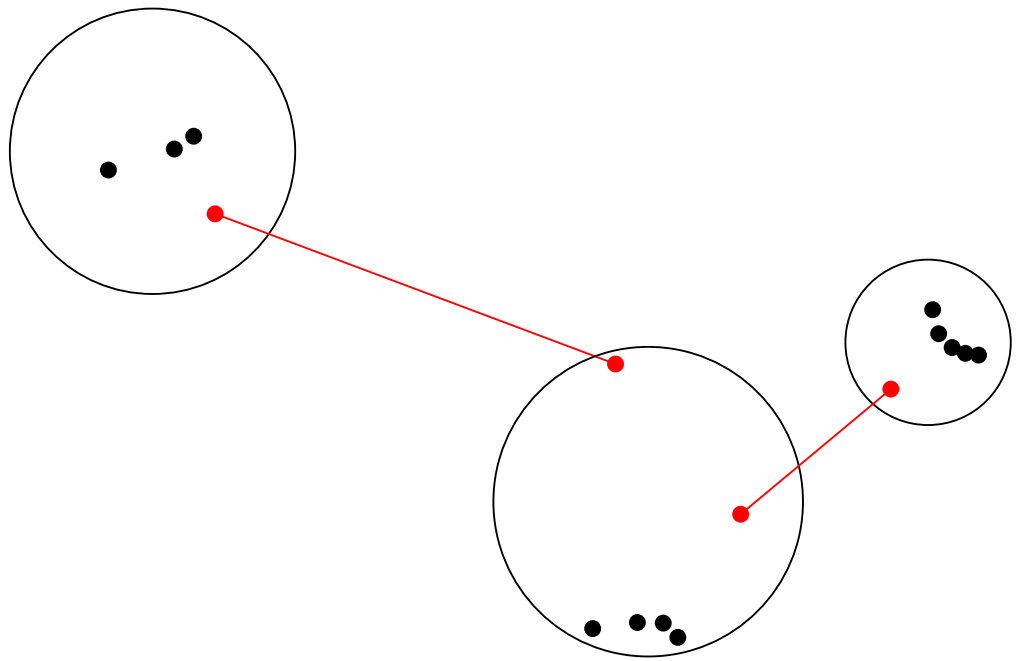
Décomposition modulaire



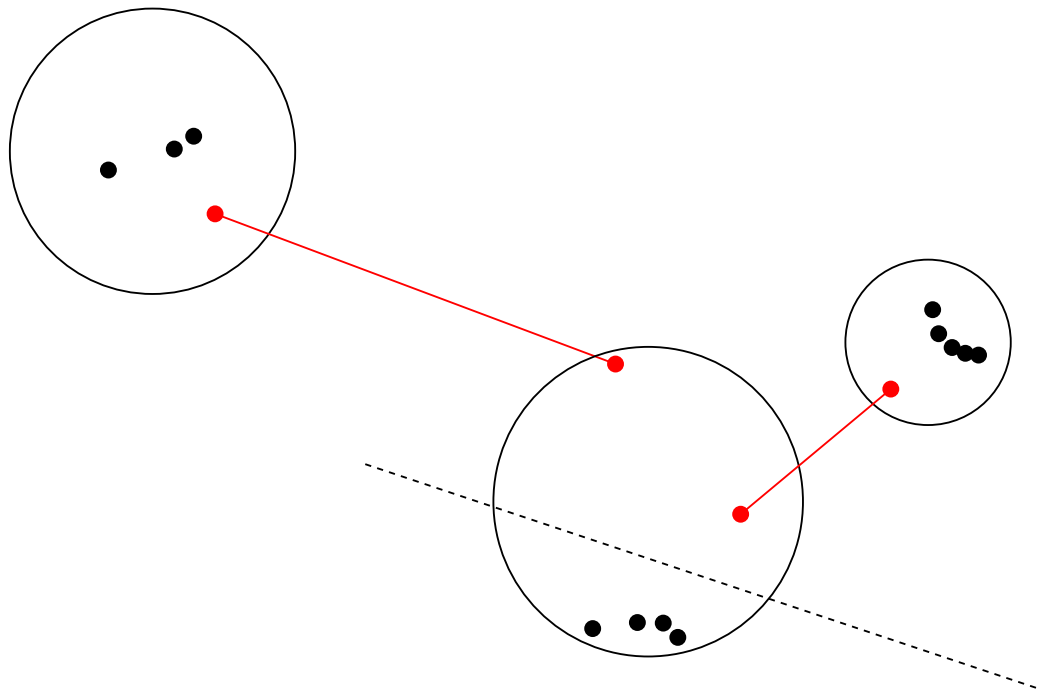
Décomposition modulaire



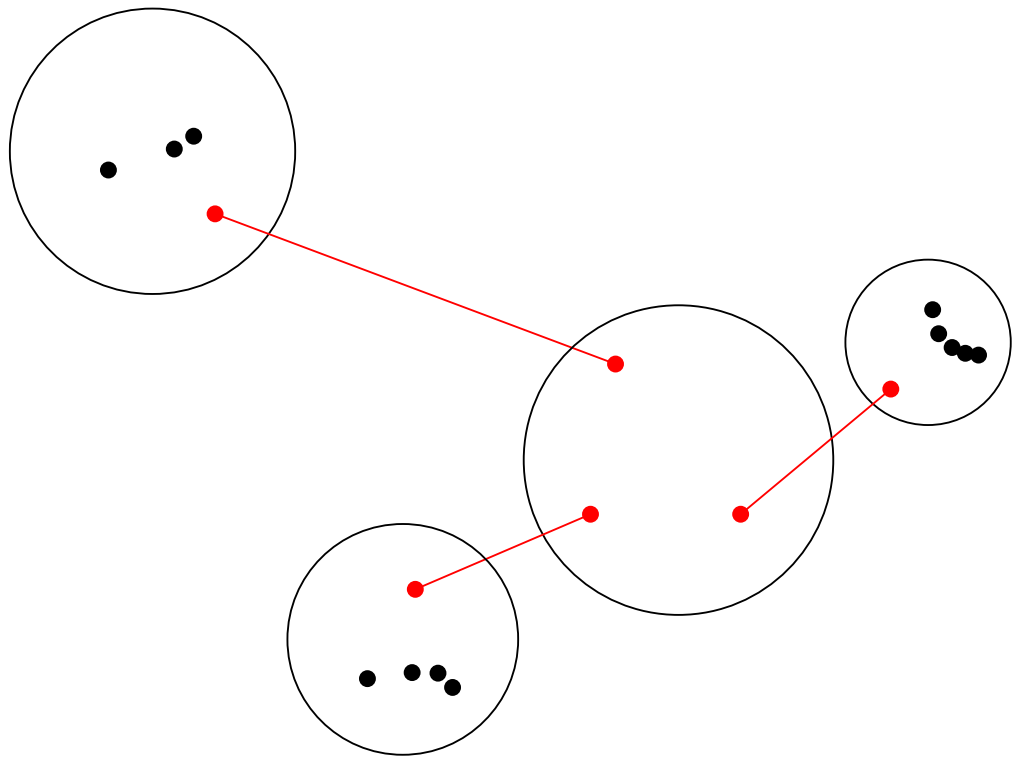
Décomposition modulaire



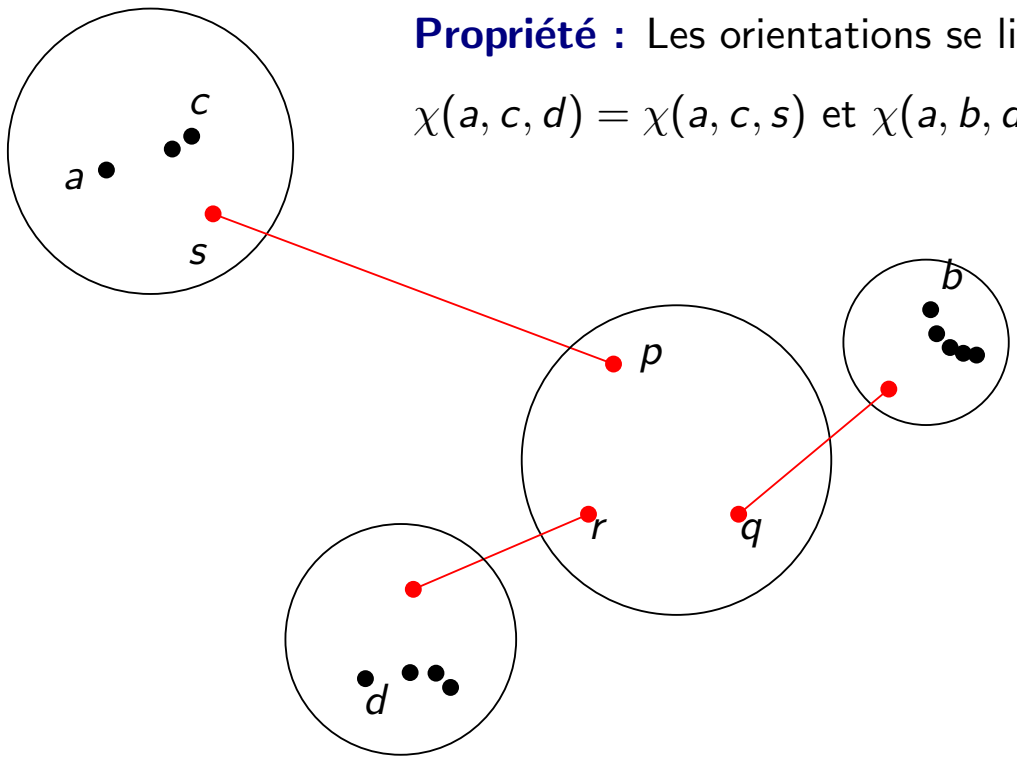
Décomposition modulaire



Décomposition modulaire



Décomposition modulaire



Propriété : Les orientations se lisent en suivant les proxies !

$\chi(a, c, d) = \chi(a, c, s)$ et $\chi(a, b, d) = \chi(p, q, r)$

Décomposition modulaire

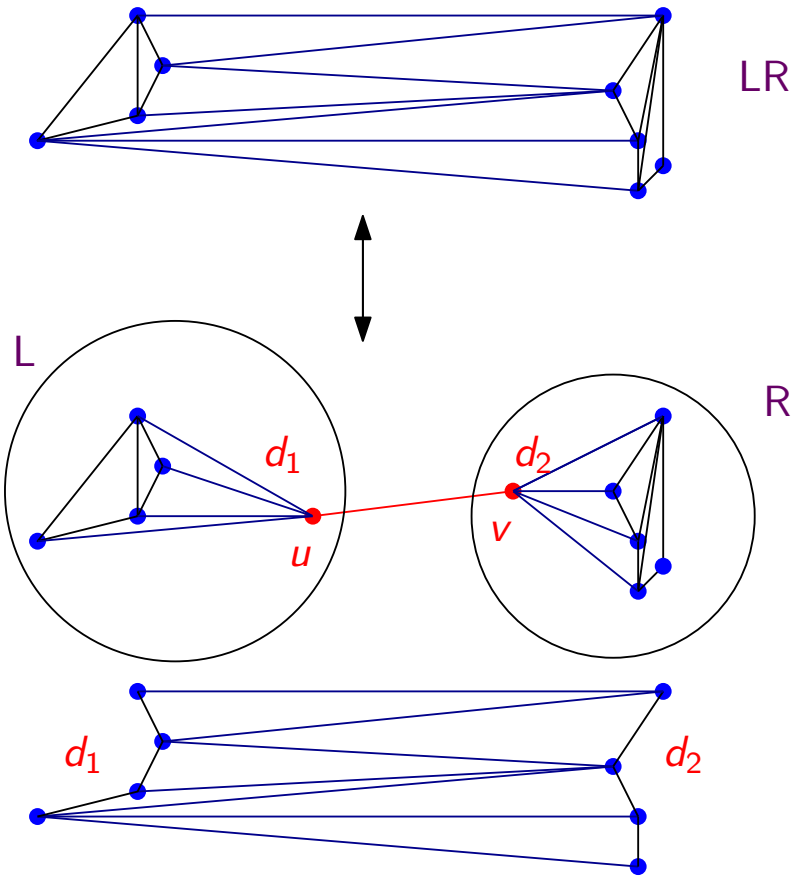
Décomposition modulaire

- On part d'un noeud seul contenant un type d'ordre représenté par un ensemble de points
- On applique $\Rightarrow := \rightarrow_{split\ \ non\ \ convexe} \cup \rightarrow_{fusion\ \ convexe}$

Résultat

$\Rightarrow$ est *confluente* : à isomorphisme de graphe prêt, un type d'ordre a une *unique décomposition modulaire normale*

Calcul de triangulations



Notations

- $\tau(LR)$: nombre de triangulations de LR.
- $\tau_{d_1}(L)$: nombre de triangulations de L telles que u est d'arité d_1.
- $\tau_{d_2}(R)$: nombre de triangulations de R telles que v est d'arité d_2.
- $\left(\binom{n}{k}\right) := \binom{k+n-1}{k}$

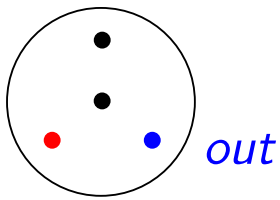
Bijection :

$$\tau(LR) = \sum_{d_1,d_2} \left(\binom{d_2}{d_1-1}\right) \tau_{d_1}(L)\,\tau_{d_2}(R)$$

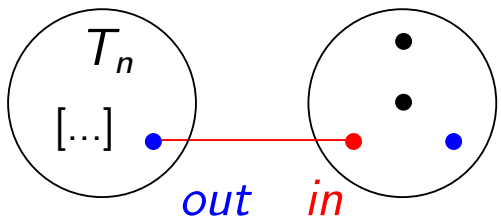
Un exemple analysable : une chaîne

Relation de récurrence

$T_1 =$



$T_{n+1} =$



Bijection :

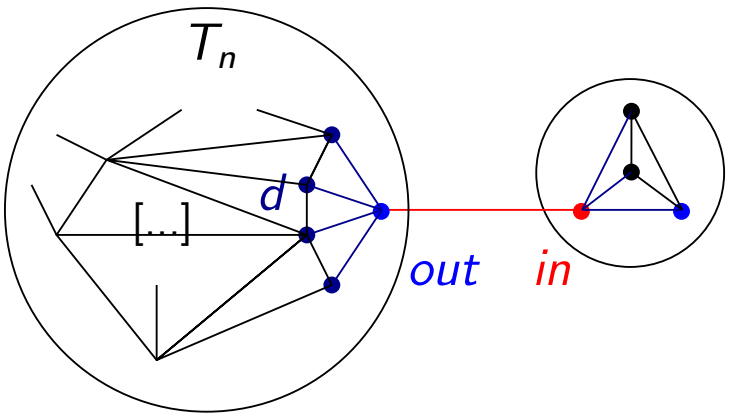
$$\tau(T_{n+1}) = \sum_{d_1,d_2} \left(\!\!\binom{d_2}{d_1-1}\!\!\right) \tau_{d_1}(T_n)\,\tau_{d_2}(T_1)$$

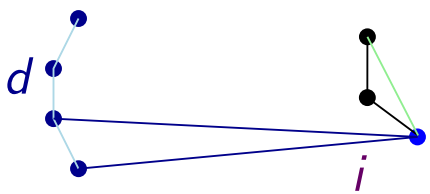
Polynôme des triangulations

- $F_n(u) = \sum_{\Delta \in \mathcal{T}_n} u^{\deg_\Delta(out)}$
- $F_n(1) = \tau(T_n)$
- $F_1(u) = u^3$

Un exemple analysable : une chaîne

Relation de récurrence





Polynôme des triangulations

- $F_n(u) = \sum_{\Delta \in \mathcal{T}_n} u^{\deg_\Delta(out)}$
- $F_n(1) = \tau(T_n)$
- $F_1(u) = u^3$

Récurrence

$$F_{n+1}(u) = u^3 \sum_{d \geq 0} [u^d] F_n(u) \sum_{i=0}^{d-1} u^i (d-i)$$

Méthode du noyau

Récurrence : $F_{n+1}(u) = u^3 \sum_{d\geq 0} [u^d] F_n(u) \sum_{i=0}^{d-1} u^i (d-i)$

Série multivariée

- $F(z,u) = \sum_{n\geq 0} F_n(u) z^n$
- $F(z,1) = \sum_{n\geq 0} \tau(T_n) z^n$

Relation vérifiée par F :

$$F(z,u)\left(1 - \frac{u^4 z}{(u-1)^2}\right) = u^3 z \left(1 - \frac{\partial_u F(z,1)}{u-1} - \frac{F(z,1)}{(u-1)^2}\right)$$

Méthode du noyau

- Le terme de gauche est *linéaire* en $\partial_u F(z,1), F(z,1)$
- Chaque racine $u(z)$ de $1 - \frac{u^4 z}{(u-1)^2}$ fournit une équation linéaire. Si on en trouve assez on peut résoudre et trouver $F(z,1)$

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

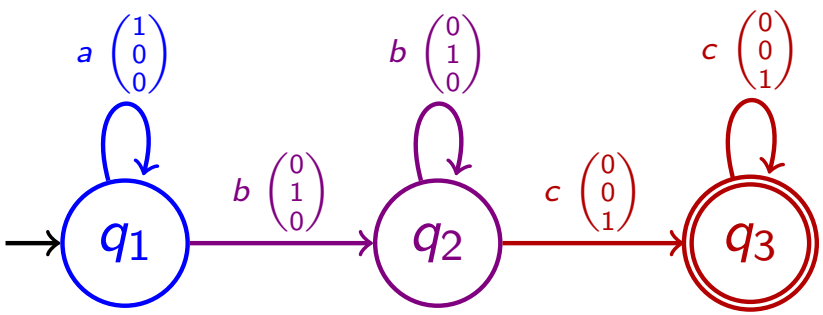
retour

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

Idea: add constrained counters

- transitions labelled by vectors in $\mathbb{N}^d$
- test on the final value at the end of the run

Parikh automaton: finite automaton on $\Sigma \times \mathbb{N}^d$ with semilinear constraints



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

$$q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[a]{\begin{pmatrix}1\\0\\0\end{pmatrix}} q_1 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[b]{\begin{pmatrix}0\\1\\0\end{pmatrix}} q_2 \xrightarrow[b]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3 \xrightarrow[c]{\begin{pmatrix}0\\0\\1\end{pmatrix}} q_3,\ w = aabbcc,\ v = \begin{pmatrix}2\\2\\2\end{pmatrix} \in \mathcal{C}$$

Can be extended with a stack.

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

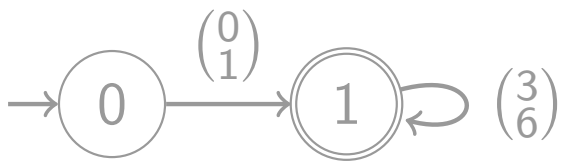
retour

$\bigwedge \bigvee$ 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

- Finite union of linear sets $\vec{c} + P^*$ where $P = \{p_1, \ldots, p_r\}$
  $(0,1) + \{(3,6)\}^*$
- Presburger arithmetic
  $\Phi(x_1, x_2) := \exists x, x_1 - 3x = 0 \ \wedge \ 1 + 2x_1 - x_2 = 0$
- (Unambiguous) rationnel subsets of $(\mathbb{N}^d, +)$

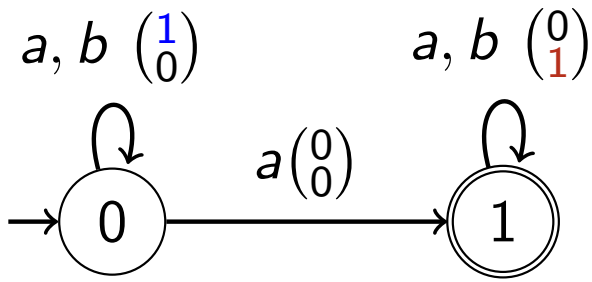


Weakly unambiguous Parikh automata

retour

Universality: undecidable for non deterministic Parikh automata

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



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

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

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

Relevant automata model

retour

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

Théorème [ICALP '20] : La classe classe des automates de Parikh à pile non ambigus of weakly unambiguous pushdown Parikh languages coincide with :

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

Description of expression trees

retour

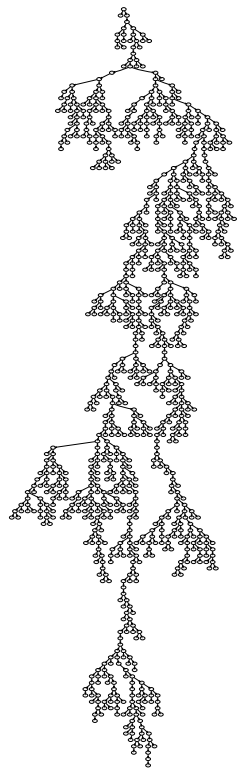
1 equation $E = a + b + \varepsilon + \overset{\star}{|}_{E} + \overset{+}{/\backslash}_{E\ E} + \overset{\bullet}{/\backslash}_{E\ E}$

Systems of equations

$$\begin{cases} \mathcal{L}_R = \overset{\star}{|}_{\mathcal{S}} + \mathcal{S}, \\ \mathcal{S} = a + b + \varepsilon + \overset{+}{/\backslash}_{\mathcal{L}_R\ \mathcal{L}_R} + \overset{\bullet}{/\backslash}_{\mathcal{L}_R\ \mathcal{L}_R}. \end{cases}$$

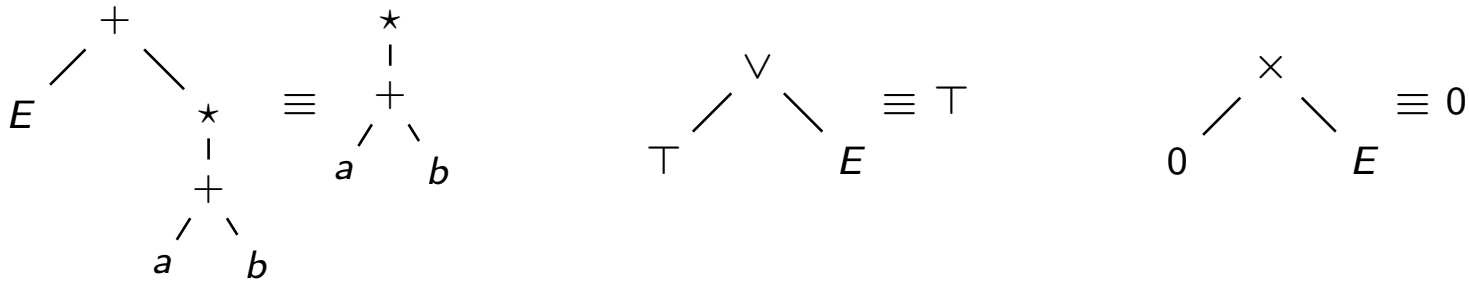
useful for

- average case analysis (uniform distribution)
- counting of syntactic trees, bounds on represented objects [Lee, Shallit '05]



Semantic rule: absorbing pattern

retour



Absorbing operator: $\circledast$ operator of arity $a \geq 2$

Absorbing pattern: constant tree $\mathcal{P}$ os size p

Semantic simplification: $\overset{\circledast}{C_1 \cdots C_a} \equiv \mathcal{P}$, whenever $C_i = \mathcal{P}$ of one $i \in \{1, \ldots, a\}$

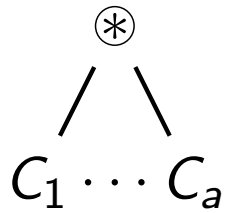
Example

- $\mathcal{P} = (a + b)^*$ is absorbing for $+$
- 0 is absorbing for $\times$
- $\top$ is absorbing for $\vee$, $\bot$ is absorbing for $\wedge$

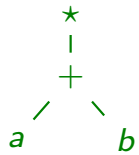
Simplification with absorbing pattern

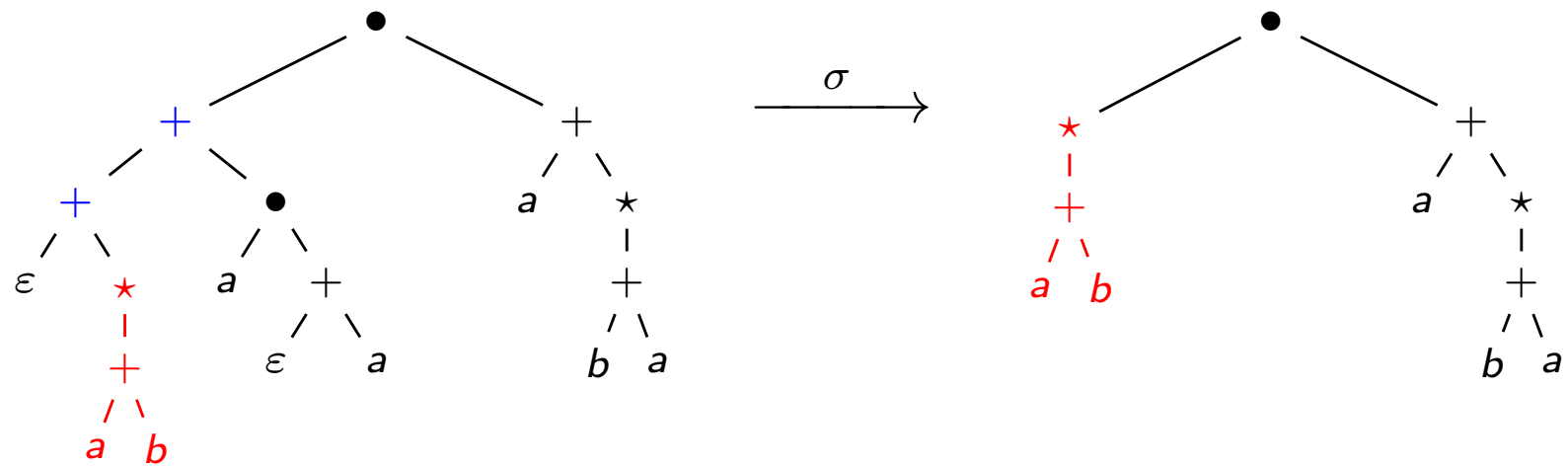
retour

Bottom-up simplification:



$\rightsquigarrow \mathcal{P}$, if $C_i = \mathcal{P}$ for some $i \in \{1, \ldots, a\}$.

Example ($\mathcal{P} =$  and $\circledast = +$)



Question : What is the average size of $\sigma(T)$?

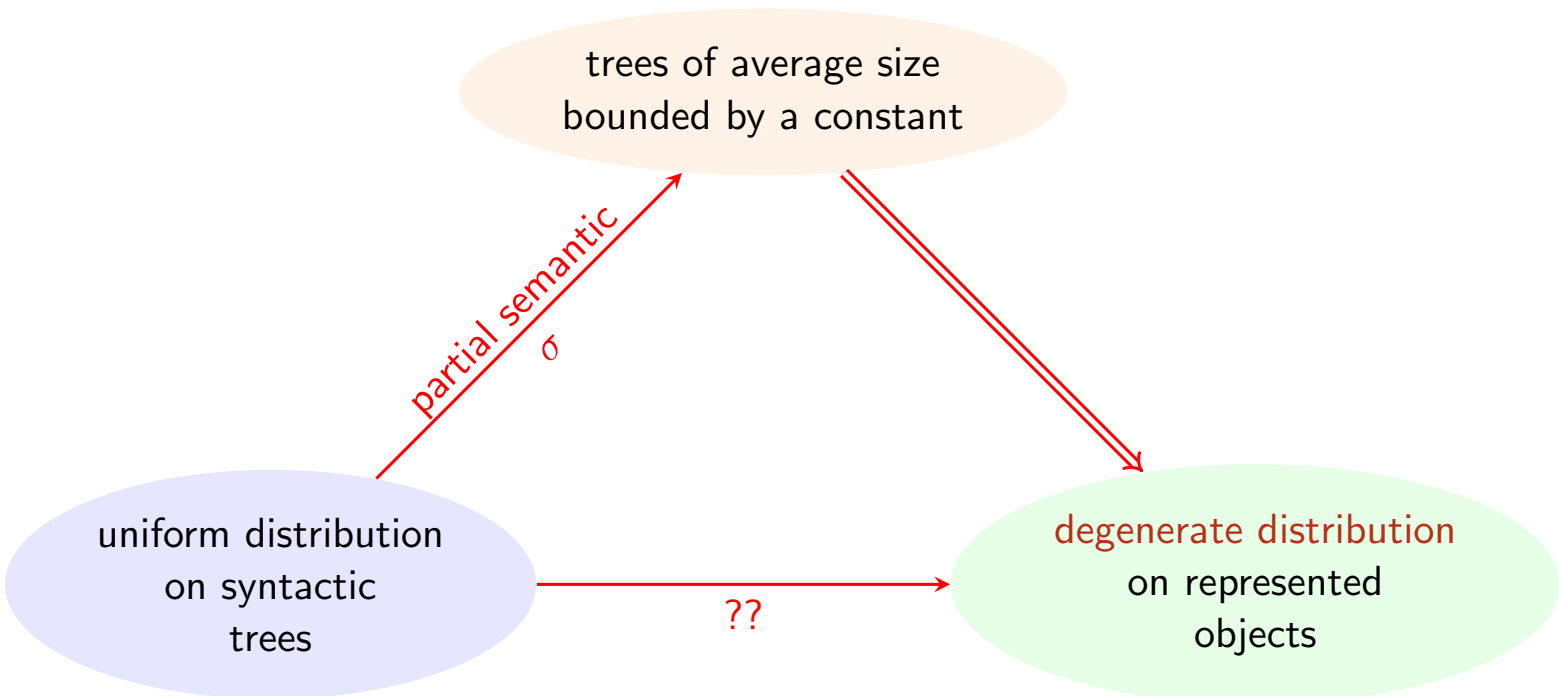
Uniform distribution is often degenerate

retour

Theorem : (K., Nicaud, Rotondo 20])

Let $\mathcal{F}$ a set of expressions described by a well founded system, with an absorbing pattern.

Then, uniform expressions have, after reduction, an average size bounded by a constant C independent of n.



Commentaires

retour

- Systèmes naturels pour décrire des expressions

$$\begin{cases} \mathcal{B} = \overset{\neg}{|}_{\mathcal{S}} + \mathcal{S}, \\ \mathcal{S} = \bot + \top + x + \overset{\wedge}{\underset{\mathcal{B}\ \mathcal{B}}{\wedge}} + \overset{\vee}{\underset{\mathcal{B}\ \mathcal{B}}{\wedge}} \end{cases}$$

$$\begin{cases} \mathcal{L} & = \mathcal{L}_+ + \mathcal{L}_\bullet + \mathcal{L}_2, \\ \mathcal{L}_+ & = \mathcal{L} \overset{+}{\diagup\diagdown} \mathcal{L}_\bullet + \mathcal{L}_2\,, \\ \mathcal{L}_\bullet & = \mathcal{L} \overset{\bullet}{\diagup\diagdown} \mathcal{L}_+ + \mathcal{L}_2\,, \\ \mathcal{L}_2 & = a + b + \varepsilon + \overset{\star}{|}_{\mathcal{L}} \end{cases}$$

[Lee, Shallit '05]

- Éléments absorbants difficiles à éviter : $((a+b)^*)^*$, $(a^*b^*)^* \ldots$

[Broda, Machiavelo, Moreira, Reis'21]

$$\begin{cases} \alpha = \varepsilon + a + b + \overset{\bullet}{\underset{\alpha\ \alpha}{\wedge}} + \overset{\star}{|}_{\alpha} + \overset{+}{\underset{\alpha_P\ \alpha_P}{\wedge}} \\ \alpha_P = \varepsilon + a + b + \overset{\bullet}{\underset{\alpha\ \alpha}{\wedge}} + \overset{\star}{|}_{\alpha_\Sigma} + \overset{+}{\underset{\alpha_P\ \alpha_P}{\wedge}} \\ \alpha_\Sigma = \varepsilon + a + b + \overset{\bullet}{\underset{\alpha\ \alpha}{\wedge}} + \overset{\star}{|}_{\alpha} + \gamma \\ \gamma = \overset{+}{\underset{\alpha_{ab}\ \alpha_{ab}}{\wedge}} + \overset{+}{\underset{\alpha_{ab}\ a}{\wedge}} + \overset{+}{\underset{\alpha_{ab}\ b}{\wedge}} + \overset{+}{\underset{a\ \alpha_{ab}}{\wedge}} + \overset{+}{\underset{b\ \alpha_{ab}}{\wedge}} + \overset{+}{\underset{a\ a}{\wedge}} + \overset{+}{\underset{b\ b}{\wedge}} \\ \alpha_{ab} = \varepsilon + \overset{\bullet}{\underset{\alpha\ \alpha}{\wedge}} + \overset{\star}{|}_{\alpha_\Sigma} + \overset{+}{\underset{\alpha_P\ \alpha_P}{\wedge}} \end{cases}$$

$\rightarrow$ évite $(a+b)^*$ mais pas $((a+b)^*)^*$!

Natural conditions ?

retour

- Natural conditions ensuring the system is describing expression trees correctly
- The system must not prevent the simplification from happening
- Some technical restrictions to keep the system manageable without having to know anything else about the semantics of the objects represented by the trees

Natural conditions ?

retour

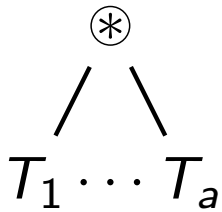
(H_1) The graph of unit rules is acyclic ($\mathcal{L}_R = \mathcal{T} + \ldots$)

(H_2) The system is *non-ambiguous*: each complete expression can be built in at most one way.

(H_3) The system is *aperiodic*.

(H_4) The dependency graph of the system is strongly connected.

(H_5) [Reduction is non-trivial] There is a rule of the form



with $\mathcal{P} \in \sigma(\mathcal{L}(T_i))$ and $a(T_j) \geq 1$, $i \neq j$.

(H_6) The system is *not linear*: there is a rule of arity at least 2.

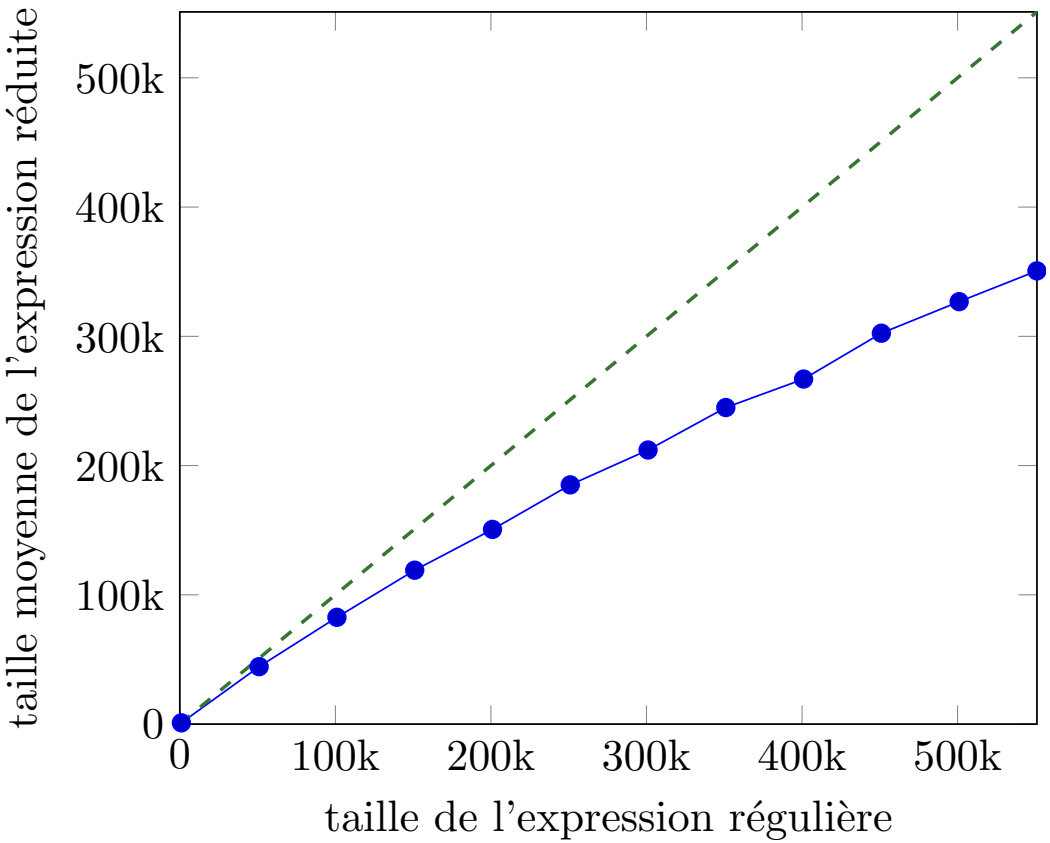
Theorem (K., Nicaud, Rotondo, 2020)

Under the hypotheses $H_\star$, all moments associated to the size after reduction of a uniform tree of size n are bounded by a constant.

Focus sur les expressions régulières

retour

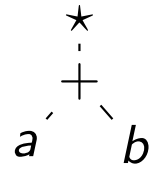
$$\mathcal{L}_R = a + b + \varepsilon + \overset{\star}{|}_{\mathcal{L}_R} + \overset{\bullet}{/\backslash}_{\mathcal{L}_R\ \mathcal{L}_R} + \overset{+}{/\backslash}_{\mathcal{L}_R\ \mathcal{L}_R} \qquad \text{avec } \mathcal{P} = \overset{\star}{|}_{+}\ {}_{a\ \ b}$$

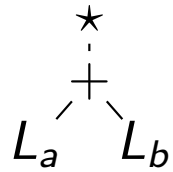


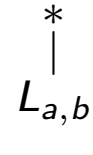
Proposition (K, Nicaud, Rotondo '19):
La taille réduite moyenne tend vers 3 624 217.

Focus sur $\mathcal{L}_R = a + b + \varepsilon + \overset{\star}{|}_{\mathcal{L}_R} + \underset{\mathcal{L}_R\ \mathcal{L}_R}{\overset{\bullet}{/\backslash}} + \underset{\mathcal{L}_R\ \mathcal{L}_R}{\overset{+}{/\backslash}}$

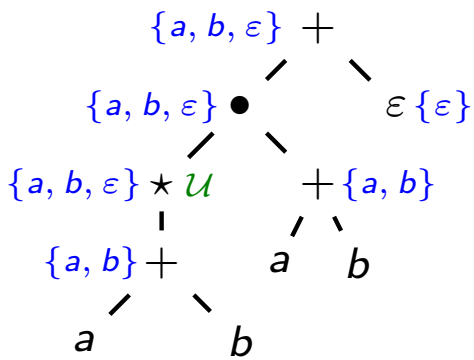
retour

- $\mathcal{P} =$ 

, mais marche aussi avec tout 

ou 

- Savoir si E reconnaît a ou b est facile.
- On a aussi besoin d'ε pour la concaténation



$$(a + b)^\star \cdot (a + b) \ + \ \varepsilon$$

Règles:

$$\overset{\star}{|}_{E} \equiv \mathcal{U} \text{ si } \{a, b\} \subseteq \mathcal{L}(E)$$

$$\underset{\mathcal{U}\ E}{\overset{+}{/\backslash}} \equiv \mathcal{U}$$

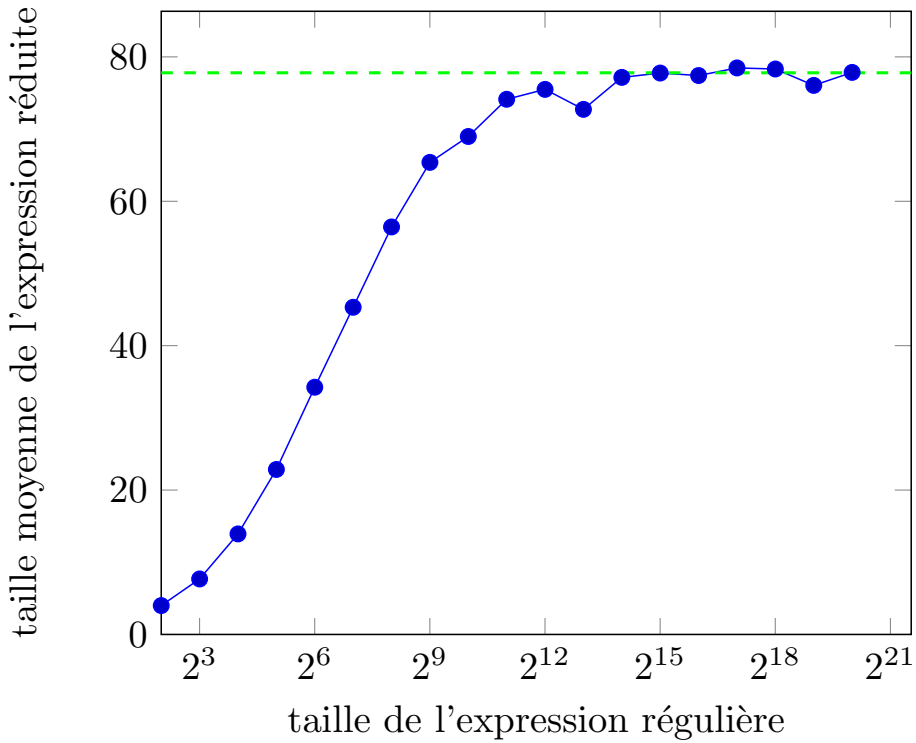
$$\underset{\mathcal{U}\ E}{\overset{\bullet}{/\backslash}} \equiv \mathcal{U} \text{ si } \varepsilon \in \mathcal{L}(E)$$

Focus sur $\mathcal{L}_R = a + b + \varepsilon + \mathcal{L}_R^{\star} + \mathcal{L}_R\,\mathcal{L}_R + \mathcal{L}_R\,\mathcal{L}_R$

retour

- $\mathcal{P} = a + b$, mais marche aussi avec tout $L_a + L_b$ ou $L_{a,b}^{*}$
- Savoir si E reconnaît a ou b est facile.
- On a aussi besoin d'ε pour la concaténation

Théorème (K., Rotondo '21):
La taille moyenne tend vers une constante $C \approx 77,8$.



Proportion d'expressions universelles

retour

$$\mathcal{L}_R = a + b + \varepsilon + \overset{\star}{|}_{\mathcal{L}_R} + \overset{\bullet}{/\!\backslash}_{\mathcal{L}_R\ \mathcal{L}_R} + \overset{+}{/\!\backslash}_{\mathcal{L}_R\ \mathcal{L}_R}$$

L'analyse introduit deux classes intéressantes :

$\mathcal{U}$: les entièrement réductibles et $\mathcal{T}_{\Sigma,\varepsilon}$: reconnaissent Σ et ε.

$$\mathcal{U} \subsetneq \texttt{expressions universelles} \subsetneq \mathcal{T}_{\Sigma,\varepsilon}$$

Théorème (K., Rotondo '21):

La proportion d'expressions universelles sur un alphabet à deux lettres est asymptotiquement comprise entre 31% et 46%.

Génération aléatoire utilisée pour les benchmarks

retour

```
function RandomFormula(n):
if n = 1 then
    p := random symbol in AP ∪ {⊤, ⊥};
    return p;
else if n = 2 then
    op := random operator in {¬, X, □, ◊};
    f := RandomFormula(1);
    return op f;
else
    op := random operator in {¬, X, □, ◊, ∧, ∨, →, ↔, U, R};
    if op in {¬, X, □, ◊} then
        f := RandomFormula(n − 1);
        return op f;
    else
        x := uniform integer in [1, n − 2];
        f1 := RandomFormula(x);
        f2 := RandomFormula(n − x − 1);
        return (f1 op f2);
```

Algorithme 1: Pseudo-code utilisé `lbtt` [Tauriainen '00]

Génération aléatoire utilisée pour les benchmarks

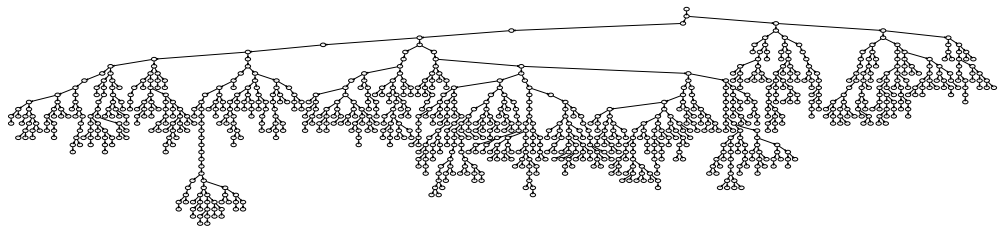
retour

Distribution ABR: différente de l'uniforme

uniforme

$h \sim c\sqrt{n}$

[Flajolet, Odlyzko '82]



ABR

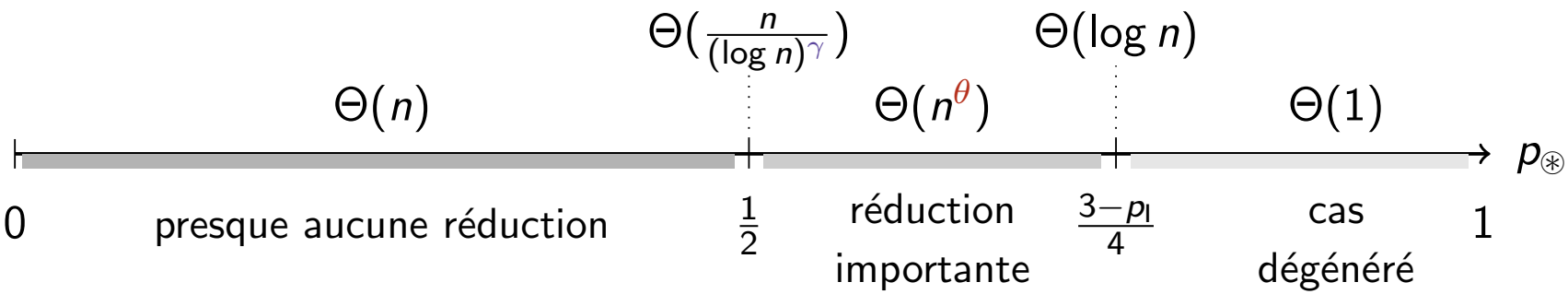
$h \sim d \log n$

[Devroye '86]

La distribution ABR est plus complexe

retour

Théorème [K. Rotondo '21]: cinq régimes pour la taille moyenne réduite, dépendant de la probabilité $p_{\circledast}$ de l'opérateur absorbant



avec $\gamma = \frac{2}{1-p_{\mathrm{I}}}$ et $\theta = 1 - \frac{4p_{\circledast}-2}{1-p_{\mathrm{I}}}$.

Idée des techniques

retour

- Récurrences sur e_n l'espérance de la taille réduite et sur γ_n la probabilité des entièrement réductibles

- Équations différentielles
  sur les séries génératrices $E(z)$ et $A(z) = \sum_n \gamma_n z^n$
  $E(z)$: équation linéaire non homogène du premier ordre dépendant de $A(z)$
  $A(z)$: équation de Riccati $A'(z) = a(z) + b(z)A(z) - p_{\circledast} A(z)^2$

- Principe de la combinatoire analytique : étude de A et E vues comme des fonctions analytiques $\mathbb{C} \to \mathbb{C}$, autour de leur singularité $z = 1$

- Différentes techniques : intégrations singulières pour $E(z)$, linéarisation et méthode de Frobenius pour $A(z)$

Inclusion problem for unambiguous automata

retour

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

Proposition: If $L(\mathcal{A}) \subsetneq L(\mathcal{B})$, there exists a small witness $w \in L(\mathcal{B}) \setminus L(\mathcal{A})$ of size at most $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

- $C(x) := \sum c_n x^n = B(x) - A(x)$ 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 $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

Inclusion problem for unambiguous automata

retour

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

Proposition: If $L(\mathcal{A}) \subsetneq L(\mathcal{B})$, there exists a small witness $w \in L(\mathcal{B}) \setminus L(\mathcal{A})$ of size at most $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

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

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

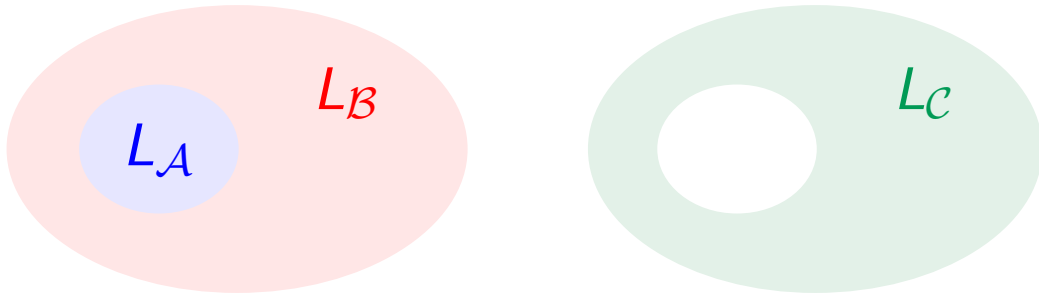
- the order r is at most $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$

Théorème [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.)

Algorithmic application: inclusion problem for wuPA

retour



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 $(n - 100)c_n = 0$

Bounds for the series of a weakly-unambiguous PA

retour

$\mathcal{A}$ weakly-unambiguous PA with a semilinear constraint C of dimension d (of the vectors)

$$A(x) = \underbrace{\bar{A}(x, y_1, \ldots, y_d)}_{\substack{\text{rat. series} \\ \text{obtained from } \mathcal{A}}} \odot \underbrace{C(x, y_1, \ldots, y_d)}_{\substack{\text{rat. series} \\ \text{obtained from } C}} \quad [y_1 \to 1, \ldots, y_d \to 1]$$

At each step, we bound the size of the system of differential equation satisfied by the series (order + coefficient)

- Main difficulty: the Hadamard product $\odot$
- Detailed analysis of the proof of closure from [Lipshitz 89]