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

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

\( \ell_n \): number of words of length \( n \)

---

<table>
<thead>
<tr>
<th>Formal languages</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>( L(x) )</td>
</tr>
</tbody>
</table>
Link between languages and combinatorics

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

<table>
<thead>
<tr>
<th>Formal languages</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>( L(x) )</td>
</tr>
<tr>
<td>Regular</td>
<td>rational ( L(x) = P(x)/Q(x) )</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
q_0(x) &= xq_0(x) + xq_1(x) \\
q_1(x) &= 1 + xq_1(x) + xq_0(x) \\
L(x) &= \frac{x}{1-2x}
\end{align*}
\]
Link between languages and combinatorics

\[ L(x_1, \ldots, x_r) = \sum_{w \in L} x_1^{w|a_1} \cdots x_r^{w|a_r} \quad \text{with} \quad \Sigma = \{a_1, \ldots, a_r\} \]

**Formal languages**  \quad **Generating series**

<table>
<thead>
<tr>
<th>Formal languages</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>( L(x) )</td>
</tr>
<tr>
<td>Regular</td>
<td>rational ( L(x) = P(x)/Q(x) )</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
q_0(x_a, x_b) &= x_a q_0(x_a, x_b) + x_b q_1(x_a, x_b) \\
q_1(x_a, x_b) &= 1 + x_b q_1(x_a, x_b) + x_a q_0(x_a, x_b) \\
L(x_a, x_b) &= \frac{x_b}{1 - (x_a + x_b)}
\end{align*}
\]
Link between languages and combinatorics

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

**Formal languages**

<table>
<thead>
<tr>
<th></th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>( L(x) )</td>
</tr>
<tr>
<td>Regular</td>
<td>rational ( L(x) = P(x)/Q(x) )</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>algebraic ( P(x, L(x)) = 0 )</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
S & \rightarrow aSB \mid \varepsilon \\
B & \rightarrow cB \mid bS
\end{align*}
\]

\[
x^2 S(x)^2 - (1 - x) S(x) + 1 - x = 0
\]
Link between languages and combinatorics

\[ L(x_1, \ldots, x_r) = \sum_{w \in L} x_1^{w|a_1} \cdots x_r^{w|a_r} \quad \sum = \{a_1, \ldots, a_r\} \]

<table>
<thead>
<tr>
<th>Formal languages</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>( L(x) )</td>
</tr>
<tr>
<td>Regular</td>
<td>rational ( L(x) = P(x)/Q(x) )</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>algebraic ( P(x, L(x)) = 0 )</td>
</tr>
</tbody>
</table>

\[
\left\{ 
\begin{align*}
S & \rightarrow aSB \mid \varepsilon \\
B & \rightarrow cB \mid bS
\end{align*}
\right.
\]

\[
\left\{ 
\begin{align*}
S(\vec{x}) &= x_a S(\vec{x}) B(\vec{x}) + 1 \\
B(\vec{x}) &= x_c B(\vec{x}) + x_b S(\vec{x})
\end{align*}
\right.
\]

\[
x_a x_b S(x_a, x_b, x_c)^2 - (1 - x_c) S(x_a, x_b, x_c) + 1 - x_c = 0
\]
Link between languages and combinatorics

\[ L(x_1, \ldots, x_r) = \sum_{w \in L} x_1^{w|a_1} \cdots x_r^{w|a_r} \quad \quad \Sigma = \{a_1, \ldots, a_r\} \]

<table>
<thead>
<tr>
<th>Formal languages</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>(L)</td>
<td>(L(x))</td>
</tr>
<tr>
<td>Regular</td>
<td>rational (L(x) = P(x)/Q(x))</td>
</tr>
<tr>
<td>ambiguous context-free</td>
<td>?</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
S & \rightarrow aSB \mid \varepsilon \\
B & \rightarrow cB \mid bS \\
\end{align*}
\]

\[
S(\vec{x}) = x_a S(\vec{x}) B(\vec{x}) + 1 \\
B(\vec{x}) = x_c B(\vec{x}) + x_b S(\vec{x})
\]

\[
x_a x_b S(x_a, x_b, x_c)^2 - (1 - x_c) S(x_a, x_b, x_c) + 1 - x_c = 0
\]
Link between languages and combinatorics

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

\[ \ell_n : \text{number of words of length } n \]

<table>
<thead>
<tr>
<th>Formal languages</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>( L(x) )</td>
</tr>
<tr>
<td>Regular</td>
<td>rational ( L(x) = P(x)/Q(x) )</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>algebraic ( P(x, L(x)) = 0 )</td>
</tr>
</tbody>
</table>

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

\[
G(x) = 1 + 2x + 11x^2 + \ldots \quad [\text{Bostan & Kauers 10, Drmota & Banderier 13}]
\]
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.
Example (Flajolet 87)

\[ D = \{ a^{n_1} b \ a^{n_2} b \ldots \ a^{n_k} b : k \in \mathbb{N}^*, \ n_1 = 1 \ \text{and} \ \exists j < k, n_{j+1} \neq 2n_j \} \]

is inherently ambiguous.

- \( aab \notin D \)
- \( abaabaaab \in D \)
- \( abaabaaab \notin D \)
- \( ab \ a^2 b \ a^4 b \ldots a^{2^{k-1}} b \notin D \)
Example (Flajolet 87)

\[ D = \{ a^{n_1}b \ a^{n_2}b \ldots a^{n_k}b : k \in \mathbb{N}^*, \ n_1 = 1 \text{ and } \exists j < k, n_{j+1} \neq 2n_j \} \]

is inherently ambiguous.

- **By contradiction**, suppose \( D \) is **unambiguous**. Then \( D(x) \) is algebraic.

- **Aim**: build from \( D(x) \) a series that is not algebraic and use closure properties.
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.

$B = ab(ab^*)^* \setminus \mathcal{D} = \{ ab \, a^2 b \, a^4 b \ldots a^{2^k-1} b : k \in \mathbb{N}^* \}$

$B(x) = \frac{x^2}{1-x^2} - D(x) = \text{algebraic}$

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.
  → language of primitive words \( \mathcal{L}_P \)

\[ aabb \in \mathcal{L}_P, \quad abab \notin \mathcal{L}_P \]

CFL: open not unambiguous CFL: [Peterson 96]
### Hierarchy of languages and series

<table>
<thead>
<tr>
<th>Language</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$L(x)$</td>
</tr>
<tr>
<td>Regular</td>
<td>$\rightarrow$ rational $Q(x)L(x) = P(x)$</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>$\rightarrow$ algebraic $P(x, L(x)) = 0$</td>
</tr>
<tr>
<td>?</td>
<td>$\rightarrow$ holonomic $P(x, \partial_x) \cdot L(x) = 0$</td>
</tr>
</tbody>
</table>
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:

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

Equivalently $a_n$ satisfies a linear recurrence of the form

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

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)
  $F(x) = \sqrt{1 - x} := \sum \frac{4^{-n}(2^n)x^n}{1-2n} satisifies F^2 - 1 - x = 0$
  $2(1 - x)F' - F = 0$
  $2(n + 1)u_{n+1} - (2n + 1)u_n = 0$

- $F(x) = e^x := \sum x^n/n!$ is holonomic but is **not algebraic**
  $F' - F = 0$
  $(n + 1)u_{n+1} - u_n = 0$
A series \( f(x_1, \ldots, x_n) \) is \textit{holonomic} (or \textit{D-finite}) if it satisfies a system of partial derivative equations of the form:

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

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 $f(x_1, \ldots, x_n)$ is holonomic, then $f(x, 1, \ldots, 1)$ is holonomic too 
3. Hadamard’s product $\odot$

\[
\begin{align*}
  f(x_1, \ldots, x_n) &= \sum_{i \in \mathbb{N}^n} a(i_1, \ldots, i_n)x_1^{i_1} \ldots x_n^{i_n} \\
  g(x_1, \ldots, x_n) &= \sum_{i \in \mathbb{N}^n} b(i_1, \ldots, i_n)x_1^{i_1} \ldots x_n^{i_n} \\
  f \odot g(x_1, \ldots, x_n) &= \sum_{i \in \mathbb{N}^n} a(i_1, \ldots, i_n)b(i_1, \ldots, i_n)x_1^{i_1} \ldots x_n^{i_n}
\end{align*}
\]
Let $S \subseteq \mathbb{N}^n$. The support series of $S$ is

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

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

$$(f \odot g)(x_1, \ldots, x_n) = \sum_{(i_1, \ldots, i_n) \in S} a(i_1, \ldots, i_n)x_1^{i_1} \ldots x_n^{i_n}$$
Example of Hadamard’s product

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

- \( abbca \in \Omega_3, \quad abbcca \notin \Omega_3 \).
- \( \Omega_3 \) is context-free, inherently ambiguous as a CFL.

\[
\Omega_3(x_a, x_b, x_c) = \frac{1}{1 - (x_a + x_b + x_c)} \odot \left( \frac{1}{(1-x_a)(1-x_b)(1-x_c)} - \frac{1}{1-x_a x_b x_c} \right)
\]

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

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

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

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: $a^n b^m a^n b^m$

→ not fully satisfactory from an automaton point of view. Conjectured a link with deterministic Reversal Bounded Counter Machines.
## Hierarchy of languages and series

<table>
<thead>
<tr>
<th>Language</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$L(x)$</td>
</tr>
<tr>
<td>Regular</td>
<td>$\rightarrow$ rational $Q(x)L(x) = P(x)$</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>$\subseteq$ algebraic $P(z, L(x)) = 0$</td>
</tr>
<tr>
<td>Weakly-unambiguous Pushdown PA</td>
<td>$\subseteq$ holonomic $P(x, \partial_x) \cdot L(x) = 0$</td>
</tr>
</tbody>
</table>

For the presentation we work with PA and not Pushdown PA.
## Hierarchy of languages and series

<table>
<thead>
<tr>
<th>Language</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$L(x)$</td>
</tr>
<tr>
<td>Regular</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>$\subseteq$</td>
</tr>
<tr>
<td>Weakly-unambiguous</td>
<td>$\subseteq$</td>
</tr>
<tr>
<td>Pushdown PA</td>
<td></td>
</tr>
</tbody>
</table>
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbbccc \rightarrow (0_0) \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = a a a b b b c c c \rightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aabbccc \rightarrow \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbbccc \rightarrow (3, 0, 0) \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaaa bbbccc \rightarrow \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbbccc \rightarrow (3 \ 2 \ 0) \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbbcccc \rightarrow (3, 3, 0) \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbbcccc \rightarrow (3, 3, 1) \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbbbc cc \rightarrow \binom{3}{3} \binom{3}{2} \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbbccc \rightarrow (3, 3, 3) \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ w = aaabbccc \rightarrow \left\{ \begin{array}{c} 3 \\ 3 \\ 3 \end{array} \right\} \in C \]

\[ w \in L(A) \]
Parikh automata [Klaedtke and Rueß 03]

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

\[ \ell = \{ (a^n b^m c^p, (m \choose p)) : n, m, p \in \mathbb{N}^* \} \]

\[ L(A) = \{ a^n b^n c^n : n \in \mathbb{N}^* \} \]
Semilinear sets of $\mathbb{N}^d$ (Parikh 1966)

- Intuitively: boolean combination of linear (affine) inequalities defining subsets of $\mathbb{N}^d$
  $$x_1 - x_2 = 0 \land x_2 - x_3 = 0 \rightarrow C = \{(n, n, n) : n \in \mathbb{N}\}$$

- More generally, subsets defined by the Presburger arithmetic
  [Ginsburg and Spanier 66]
  $$\Phi(x_1, x_2) := \exists x, x_1 - 3x = 0 \land 1 + 2x_1 - x_2 = 0 \rightarrow \{(3n, 6n + 1) : n \in \mathbb{N}\}$$
Semilinear = Finite union of linear sets $\vec{c} + P^*$ where $P = \{p_1, \ldots, p_r \}$ and $P^* = \{\lambda_1 p_1 + \ldots + \lambda_r p_r : \lambda_i \in \mathbb{N}\}$
Theorem (Eilenberg and Schützenberger 69, Ito 69)

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

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

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

Remark

In the sequel we will deal with holonomic series of the form $f \circ C$, where $C$ is the support series of a semilinear set.
Weakly-unambiguous Parikh automaton

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

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

\[ L(A) = \{ w_1 aw_2 : |w_1| = |w_2|, w_1, w_2 \in \Sigma^* \} \text{ with } \Sigma = \{ a, b \}. \]

≠ [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.
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.

- Languages recognized by weakly-unambiguous PA have holonomic generating series
Definition (Generating series of the runs of a PA)

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

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

The generating series of these runs are classically rational.
Example

\[ q_0(x, y_1, y_2, y_3) = xy_1 y_2^3 y_3 q_1(x, y_1, y_2, y_3) \]
\[ q_1(x, y_1, y_2, y_3) = xy_1^3 q_1(x, y_1, y_2, y_3) + xy_2 q_2(x, y_1, y_2, y_3) \]
\[ q_2(x, y_1, y_2, y_3) = xy_2 q_2(x, y_1, y_2, y_3) + xy_3^2 q_3(x, y_1, y_2, y_3) \]
\[ q_3(x, y_1, y_2, y_3) = xy_1^2 y_3 q_3(x, y_1, y_2, y_3) + 1 \]
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.
- $C(y_1, \ldots, y_d) = \sum_{(i_1, \ldots, i_d) \in \mathcal{C}} y_1^{i_1} \ldots y_d^{i_d}$ supports series of the semilinear set $\mathcal{C}$, 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.
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

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

\[ \{a, b\}, (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}) \quad a, (\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}) \quad a, (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}) \quad b, (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}) \quad a, (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}) \quad b, (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}) \quad a, (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}) \]

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

\[
2^{k+1} - 1 + k + 1 - (2^k - 1 + k) \rightarrow \infty \text{ incompatible with any } \\
pr(n)a_{n+r} + \ldots + p_0(n)a_n = 0 \quad \text{with } p_i(n) \in \mathbb{Q}[n]
\]
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**

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

- \textit{aaabaaab aabaab abaab} \in \mathcal{L}_{\text{even}}
- It is deterministic context-free \Rightarrow \textit{algebraic} generating series
- The proof uses Ramsey’s theorem, and is very \textit{specific} to this language. It shows inherent ambiguity for a \textit{wider} family of automata.
An algorithmic consequence of holonomy

Holonomy of the generating series has **algorithmic consequences**

→ 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.
Proposition (Stearns and Hunt 85)

Given two unambiguous finite automata $A$ and $B$ such that

$$L(B) \subsetneq L(A)$$

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

$$|w| < |Q_A| + |Q_B|$$
Sketch of the proof

- \( L_A(x) = \sum_n a_n x^n \) generating series of \( L(A) \)
- \( L_B(x) = \sum_n b_n x^n \) generating series of \( L(B) \).
- \( G(x) = L_B(x) - L_A(x) \) rational, degrees at most \( r \leq |Q_A| + |Q_B| \)
- Then \( g_n = b_n - a_n \) satisfies:

\[
\forall n \geq r, \quad c_r g_n = c_{r-1} g_{n-1} + \cdots + c_0 g_{n-r}
\]
Sketch of the proof

- \( L_A(x) = \sum_n a_n x^n \) generating series of \( L(A) \)
- \( L_B(x) = \sum_n b_n x^n \) generating series of \( L(B) \).
- \( G(x) = L_B(x) - L_A(x) \) rational, degrees at most \( r \leq |Q_A| + |Q_B| \)
- 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 \( L(A) \subsetneq L(B) \), there exists \( N < r \) such that \( a_N < b_N \).

\( \rightarrow \) There is a small witness word of length < \( |Q_A| + |Q_B| \) in \( L(B) \setminus L(A) \).
Inclusion problem

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

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

\[ \rightarrow \text{Our contribution is to give explicit bounds in the weakly-unambiguous case} \]
Inclusion separation for weakly-unambiguous automata?

Essentially same ideas as regular case, however:

- $L_A(x) = A(x, 1, \ldots, 1)$ where:
  
  $A(x, y_1, \ldots, y_d) := q_l(x, y_1, \ldots, y_d) \odot C(x, y_1, \ldots, y_d)$
  
  → Same problem with $L_B(x)$

- Then $g_n = v_n - u_n$ satisfies a linear recurrence of the form
  
  $\forall n \geq r, \ c_r(n)g_n = c_{r-1}(n)g_{n-1} + \ldots + c_0(n)g_{n-r}$

  $G(x) = x^{1000} \rightarrow (1000 - n)g_n = 0$

  → 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 $A$ and $B$.
- 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 $L(A) \not\subseteq L(B)$, there exists a word $w \in L(B) \setminus L(A)$ such that

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

where $d = d_A + d_B$, $M = |A| |B| \|A\|_{\infty} \|B\|_{\infty}$. 
Consequence: inclusion problem

Input: two weakly-unambiguous Parikh automata $A, B$

Question: $L(A) \subseteq L(B)$?

**Proposition**

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

$\rightarrow$ dynamic programming approach to avoid another exponential when enumerating every word of length less than the witness!
<table>
<thead>
<tr>
<th>Language</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$L(x)$</td>
</tr>
<tr>
<td>Regular</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>$\Rightarrow$</td>
<td>rational $Q(x)L(x) = P(x)$</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td>$\Rightarrow$</td>
<td>algebraic $P(x, L(x)) = 0$</td>
</tr>
<tr>
<td>Weakly-unambiguous Pushdown PA</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td>$\Rightarrow$</td>
<td>holonomic $P(x, \partial_x) \cdot L(x) = 0$</td>
</tr>
</tbody>
</table>
## Conclusion

<table>
<thead>
<tr>
<th>Language</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$L(x)$</td>
</tr>
<tr>
<td>Regular</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>$\subseteq$</td>
<td>rational $Q(x)L(x) = P(x)$</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>$\subseteq$</td>
</tr>
<tr>
<td>$\subseteq$</td>
<td>algebraic $P(x, L(x)) = 0$</td>
</tr>
<tr>
<td>Weakly-unambiguous Pushdown PA</td>
<td>$\subseteq$</td>
</tr>
<tr>
<td>$\subseteq$</td>
<td>holonomic $P(x, \partial_x) \cdot L(x) = 0$</td>
</tr>
</tbody>
</table>

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

→ 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 → need for new ideas to find other links between holonomic series and formal languages.

Thank you!
References I


References II


<table>
<thead>
<tr>
<th>Language</th>
<th>Generating series</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$L(x)$</td>
</tr>
<tr>
<td>Regular</td>
<td>$\rightarrow$ rational $Q(x)L(x) = P(x)$</td>
</tr>
<tr>
<td>Unambiguous context-free</td>
<td>$\subseteq$ algebraic $P(x, L(x)) = 0$</td>
</tr>
<tr>
<td>Weakly-unambiguous Pushdown PA</td>
<td>$\subseteq$ holonomic $P(x, \partial_x) \cdot L(x) = 0$</td>
</tr>
</tbody>
</table>
Remark

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

Example

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

**Consequence**

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

\[ (c^*)^{-1} \mathcal{D}_2 \cap (a + b)^* = \mathcal{D} \]
\{a^n b^m c^p : n = m \text{ or } m = p\} \text{ is}

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

\mathcal{L}_{\text{even}} = \{a^{n_1} b \ldots a^{n_2k} b : k \in \mathbb{N}^*, \exists i \in [1, k], n_{2i-1} = n_{2i}\} \text{ is}

- deterministic as a CF language
- inherently ambiguous as a PA language
General method [Greibach 68], by reducing the universality problem

\[ L_1 = \Sigma_1^*? \]

\[ L = L_1\#\Sigma^* \cup \Sigma_1^*\#D \]. Then:

\[ L_1 = \Sigma_1^* \iff L \text{ is weakly-unambiguous} \]

⇒ If \( L_1 = \Sigma_1^* \), \( L = \Sigma_1^*\#\Sigma^* \) is regular.

⇐ By contraposition, let \( y \notin L_1 \). As \((y\#)^{-1}L = D\) is not weakly-unambiguous, neither is \( L \).
A given under the form \((\Sigma, Q, q_I, F, C, \Delta)\).

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

\(C\) given under a unambiguous form \(\bigcup_{i=1}^{p} c_i + P_i^*\).

\(\|A\|_\infty\) maximum coordinate of the vectors in the description of \(\Delta\) and \(C\)

\(|A| = |Q| + |\Delta| + p + \sum |P_i|\)