Universal behavior of context-free grammars: complete characterization of the critical exponent

Hanane Tafat Bouzid

Laboratoire d'Informatique Paris Nord,
joint work with Cyril Banderier (Paris 13) and Michael Drmota (TU Wien).

March 26 2012
SLC, Ottrott

Plan

1. Introduction
2. Algebraic grammar
   - Definitions
   - Asymptotics
   - Critical exponent of strongly connected graph
   - Critical exponent of non-strongly connected graph
3. Conclusion

Motivations

- Grammars are a fundamental structure in computer science: information theory, language theory, compilation, bioinformatics, combinatorics (Schützenberger methodology)...
- Challenge: is it easy to test if a given generating function is $\mathbb{N}$-algebraic? (i.e. is it associated to a context-free grammar?)
- Can we find an easy criterion?

$1-(1-4z)^{1/3}+O((1-4z)^{2/3})$, $1-(1-4z)^{1/4}+O((1-4z)^{3/4})$.

Flajolet and Sedgewick, Analytic Combinatorics, p.493:
"It would at least be desirable to determine directly,
from a positive (but *reducible*) system,
the type of singular behaviour of the solution,
but the systematic research involved in such a programme
is yet to be carried out."

Plan

Combinatorial object	Specification	Grammar
Super-bicolored trees	$\mathcal{B} = 2z\mathcal{T}\,\mathrm{Seq}(\mathcal{B})$	$\begin{cases} \mathcal{B} \to a \cdot \mathcal{T} \cdot \mathcal{U} \mid b \cdot \mathcal{T} \cdot \mathcal{U} \\ \mathcal{U} \to \mathcal{B} \cdot \mathcal{U} \mid \epsilon \\ \mathcal{T} \to a \cdot \mathcal{T} \cdot \mathcal{A} \mid a \\ \mathcal{A} \to \mathcal{T} \cdot \mathcal{A} \mid \epsilon \end{cases}$

- Closure properties: union, concatenation, star, shuffle...
- Non closure properties: complement, intersection
  ($\Rightarrow$ no minus sign!)
- The generating function of an algebraic grammar is called $\mathbb{N}$-algebraic. (reminiscent of the set of $\mathbb{N}$-rational generating functions associated to automata.)

Algebraic branches

Combinatorial object	Algebraic grammar	Functional equation	Branches
Binary trees	$T \to a \cdot T \cdot T \mid a$	$T(z) = z + zT^2(z)$	$\begin{cases} \frac{1+\sqrt{1-4z^2}}{2z} \\ \frac{1-\sqrt{1-4z^2}}{2z} \end{cases}$

Algebraic grammar with m non-terminals (well-founded system):

1. Polynomial system of equations $\{Y_j = P_j(z, Y_1, \ldots, Y_m)\}_j$,
2. P_j have nonnegative coefficients,
3. $\overrightarrow{Y^{(0)}} = (0, \ldots, 0)$, $\overrightarrow{Y^{(t+1)}} = P_j(\overrightarrow{Y^{(t)}})$, $\overrightarrow{Y} = \lim_{t \to \infty} \overrightarrow{Y^{(t)}}$

$\Rightarrow$ unique branch analytic with nonnegative Taylor coefficients around 0.

Asymptotics of algebraic grammars

Theorem: asymptotics of the function via Puiseux expansion [Newton-Puiseux]

Let $f(z)$ the branch of an algebraic equation $P(z, f(z)) = 0$, near its singularities ρ, $f(z)$ admits a convergent fractional series expansion of the form: $f(z) = \sum_{k \geq k_0} c_k \,.\, (z-\rho)^{\frac{k}{K}}$ with $k_0 \in \mathbb{Z}$ and $K \geq 1$.

Theorem: asymptotics of coefficients via singularity analysis [Darboux, Flajolet–Odlyzko]

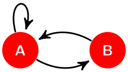
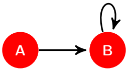
$f(z) \sim c_0 + C.(z-\rho)^{\alpha}$

$$\Rightarrow f_n \sim C \frac{1}{\Gamma(-\alpha)} \rho^{-n} n^{\gamma} \text{ with } \gamma = -\alpha - 1$$

What are the possible values of the "critical exponent" γ?

It is just known: $\gamma \in \mathbb{Q}/\{-\mathbb{N}\}$. (+problem also tackled by Schaeffer & Bousquet-Mélou)

A strongly connected graph is a directed graph that has a path from each vertex to every other vertex.

Language	Grammar	Dependency graph
$\mathcal{L} = \{(a^n.b^{2n})^*,\ n \geq 0\}$	$A \to aAB \mid \epsilon$ $B \to bbA \mid \epsilon$	 strongly connected graph.
$\mathcal{L} = \{a^n b^m a^m b^n,\ n \geq 0, m \geq 0\}$	$A \to aAb \mid B$ $B \to bBa \mid \epsilon$	 non-strongly connected graph.

Critical exponent of strongly connected graph

Theorem of [Drmota–Lalley–Woods, 97–93–97]

Let $\{Y_j = P_j(z, Y_1, \ldots, Y_m)\}_j$ a system satisfying:

1. all polynomials P_i have positive coefficients and $\frac{\partial^2 P_i}{\partial Y_i} \neq 0$;
2. the system admits a fixed point $\vec{Y}$;
3. the dependency graph of the system is strongly connected.

Then:

1. $Y(z) = C_0 - C.(1 - \frac{z}{\rho})^{\frac{1}{2}} + O(1 - \frac{z}{\rho})$ for $z \sim \rho$,
2. $[z^n]Y(z) \sim C.\frac{1}{2\sqrt{\pi}}.\rho^{-n}.n^{-\frac{3}{2}}$ with ρ, C_0, C algebraic numbers.

Examples

One single non-terminal $\Rightarrow$ universality of the $-3/2$ exponent (e.g., simple families of trees [Meir and Moon]).

$$\begin{cases} \mathcal{A} \to a \cdot \mathcal{A} | b \cdot \mathcal{B} \cdot \mathcal{B} | c \\ \mathcal{B} \to b \cdot \mathcal{A} | a \cdot \mathcal{B} \cdot \mathcal{B} | c \end{cases} \qquad \begin{cases} A(z) = zA(z) + zB(z)^2 + z \\ B(z) = zA(z) + zB(z)^2 + z \end{cases}$$

- Algebraic equation: $zA(z)^2 + (z-1)A(z) + z = 0$
- Branches of $A(z)$:

$$-\frac{z - 1 + \sqrt{-3z^2 - 2z + 1}}{2z}, \quad \frac{-z + 1 + \sqrt{-3z^2 - 2z + 1}}{2z}$$

- The singularities: $-1,\ 0, 1/3$
- Puiseux expansion: $A(z) = \frac{5}{2} - (3(1-3z))^{1/2} + O(1-3z)$

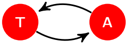
Sketch of proof [Drmota–Lalley–Woods Theorem, 97–93–97]

- Each component solution is an algebraic function with a positive radius of convergence
  (proof: combinatorial reason),
- each component has a unique branch with positive coefficients
  (proof: by construction, well-founded system),
- the Y_j's have the same dominant singularity
  (proof: if not, contradiction between $\lim_{x \to \rho} \partial_x^m Y_j(x) = \infty$),
- the Y_j's have a square root behavior
  (proof: via the implicit function theorem + Taylor expansion).

Critical exponent of non-strongly connected graph

general trees
$\mathcal{T} = z\,\mathrm{Seq}(\mathcal{T})$

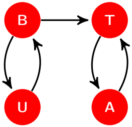
$$\begin{cases} \mathcal{T} \to a \cdot \mathcal{T} \cdot \mathcal{A} | a \\ \mathcal{A} \to \mathcal{T} \cdot \mathcal{A} | \epsilon \end{cases}$$



strongly connected.

super-bicolored trees:
$\mathcal{B} = \mathcal{T}\,[(z+z).\mathcal{T}]$
$\mathcal{B} = 2z\mathcal{T}\,\mathrm{Seq}(\mathcal{B})$

$$\begin{cases} \mathcal{B} \to a \cdot \mathcal{T} \cdot \mathcal{U} | b \cdot \mathcal{T} \cdot \mathcal{U} \\ \mathcal{U} \to \mathcal{B} \cdot \mathcal{U} | \epsilon \\ \mathcal{T} \to a \cdot \mathcal{T} \cdot \mathcal{A} | a \\ \mathcal{A} \to \mathcal{T} \cdot \mathcal{A} | \epsilon \end{cases}$$



non-strongly connected.

Branch: $B(z) = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4z + 4z\sqrt{1-4z}}$

Puiseux expansion: $B(z) = \frac{1}{2} - \frac{1}{2}(1-4z)^{1/4} + O((1-4z)^{1/2})$

Can be automatized via the Algolib Maple library [Flajolet–Salvy–Zimmermann].

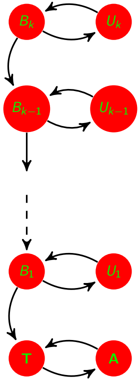
Explicit construction of critical behavior in $1/2^k$

- The critical behavior of bicolored trees $\mathcal{T}$ is $1/2$

- super-bicolored trees: $\mathcal{B} = \mathcal{T}[(z+z).\mathcal{T}]$
  - algebraic grammar,
  - schema of critical composition,
  - $\frac{1}{2^2}$ is the critical behavior.

- k-super-bicolored trees: $\mathcal{B}_k = \mathcal{T}[(z+z).\mathcal{B}_{k-1}]$
  - algebraic grammar,
  - schema of k nested critical compositions,
  - $\frac{1}{2^{k+1}}$ is the critical behavior.

Explicit construction of critical behavior in $1/2^k$ (proof)



- k-super-bicolored trees:
  $\mathcal{B}_k = \mathcal{T}\left[(z+z).\mathcal{B}_{k-1}\right]$
  $= ((z+z).\mathcal{B}_{k-1}.\,\mathrm{Seq}(\mathcal{B}_k)$
- Grammar of $\mathcal{B}_k$ is algebraic:
  $$\begin{cases} \mathcal{B}_k \to a \cdot \mathcal{B}_{k-1} \cdot \mathcal{S}_k | b \cdot \mathcal{B}_{k-1} \cdot \mathcal{S}_k \\ \mathcal{S}_k \to \mathcal{B}_k \cdot \mathcal{S}_k | \epsilon \end{cases}$$

The dependency graph of $\mathcal{B}_k$ is a non strongly connected graph.

- The composition is always a critical composition: (i.e. $h(z) = f(g(z))$ with $\rho_h = \rho_g$ and $\rho_f = g(\rho_g)$).

  $B_k(z) = T(2zB_{k-1}(z)) = \frac{1}{2}\sqrt{1 - 8zB_{k-1}(z)}.$

A first generalization of the Drmota–Lalley–Woods theorem

Let $\{Y_j = P_j(z, Y_1, \ldots, Y_m)\}_j$ a system satisfying:

1. all polynomials P_i have positive coefficients and $\frac{\partial^2 P_i}{\partial Y_i} \neq 0$;
2. the system admits a fixed point $\vec{Y}$;
3. ~~the dependency graph of the system is strongly connected~~.

Then:

1. $Y(z) = \sum_{i\geq 0} c_i \,.\, (1 - \frac{z}{\rho})^{\frac{i}{2^k}}$,
2. $[z^n]Y(z) \sim C.\frac{1}{2\sqrt{\pi}}.\rho^{-n}.n^{-\frac{1}{2^k}-1}$ with ρ, C_0, C algebraic numbers.

A second generalization of the Drmota–Lalley–Woods theorem

Let $\{Y_j = P_j(z, Y_1, \ldots, Y_m)\}_j$ a system satisfying:

1. all polynomials P_i have positive coefficients and ~~$\frac{\partial^2 P_i}{\partial Y_i} \neq 0$~~ $\frac{\partial P_j}{\partial Y_i} \neq 0$ and $\frac{\partial^2 P_j}{\partial Y_i \partial Y_h} \neq 0$ for all (i, j);
2. the system admits a fixed point $\vec{Y}$;
3. ~~the dependency graph of the system is strongly connected~~.

Then:

1. $Y(z) = \sum_{i \geq 0} c_i \,.\, (1 - \frac{z}{\rho})^{\frac{i}{2^k} - d}$,
2. $[z^n]Y(z) \sim C . \frac{1}{2\sqrt{\pi}} . \rho^{-n} . n^{-\frac{1}{2^k} - d - 1}$ with ρ, C algebraic numbers.

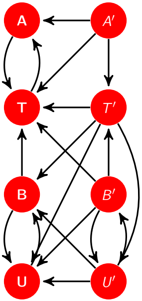
Proof: closure by sum, product and substitution.

Example 1

- The generating function of super-bicolored trees:
  $B(z) = \frac{1}{2} - \frac{1}{2}(1-4z)^{1/4} + O((1-4z)^{3/4})$

  $\frac{\partial}{\partial z} B(z) = \frac{1}{2}(1-4z)^{-3/4} + O((1-4z)^{-1/4})$

$$
\left\{
\begin{array}{c}
\mathcal{B}' \to \mathcal{T} \cdot \mathcal{U} | a \cdot \mathcal{T}' \cdot \mathcal{U} | a \cdot \mathcal{T} \cdot \mathcal{U}' | b \cdot \mathcal{T}' \cdot \mathcal{U} | b \cdot \mathcal{T} \cdot \mathcal{U}' \\
\mathcal{B} \to a \cdot \mathcal{T} \cdot \mathcal{U} | b \cdot \mathcal{T} \cdot \mathcal{U} \\
\mathcal{T}' \to \mathcal{T}\mathcal{A} | a \cdot \mathcal{T}' \cdot \mathcal{A} | a \cdot \mathcal{T} \cdot \mathcal{A}' | \epsilon \\
\mathcal{T} \to a \cdot \mathcal{T} \cdot \mathcal{A} | a \\
\mathcal{U}' \to \mathcal{B}' \cdot \mathcal{U} | \mathcal{B} \cdot \mathcal{U}' | \epsilon \\
\mathcal{U} \to \mathcal{B} \cdot \mathcal{U} | \epsilon \\
\mathcal{A}' \to \mathcal{T}' \cdot \mathcal{A} | \mathcal{T} \cdot \mathcal{A}' | \epsilon \\
\mathcal{A} \to \mathcal{T} \cdot \mathcal{A} | \epsilon
\end{array}
\right.
$$

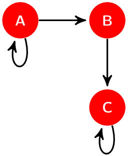


non-strongly connected.

The dependency graph is non-strongly connected.

Example 2

$$\begin{cases} \mathcal{A} \to a \cdot \mathcal{A} \cdot \mathcal{A} \cdot \mathcal{A} | b \cdot \mathcal{A} \\ \mathcal{B} \to a | a \cdot \mathcal{B}\mathcal{C} | b \cdot \mathcal{B}\mathcal{C} \\ \mathcal{C} \to a | b \cdot \mathcal{C}\mathcal{C} \end{cases}$$



$$\begin{cases} A(z) = \frac{z^4(1-4z^2)^{-3/2}}{1-z} \\ B(z) = z(1-4z^2)^{-1/2} \\ C(z) = \frac{1-(1-4z^2)^{1/2}}{2z} \end{cases}$$

$$\begin{cases} A(z) = \frac{\sqrt{2}}{32}(1-2z)^{-3/2} + \dots \\ B(z) = \frac{\sqrt{2}}{4}(1-2z)^{-1/2} + \dots \\ C(z) = 1 - \sqrt{2}(1-2z)^{1/2} + \dots \end{cases}$$

$$[z^n]A(z) \sim \frac{\sqrt{2}}{56\sqrt{\pi}} 2^n n^{1/2}$$

Counter-example

- walks in the quarter-plane: Kreweras (1965), Gessel (2001) [Bostan & Kauers, 2010]: algebraic GF, asymptotics compatible with $\mathbb{N}$-algebraicity . . . but non $\mathbb{N}$-algebraic (proof via Ogden's pumping lemma).

- Some families of maps: algebraicity proven via the kernel method (Tutte, Brown, Bousquet-Mélou, . . .) but non $\mathbb{N}$-algebraic because their critical exponents are not in our set of dyadic possible exponents.

Plan

Conclusion

- We gave a characterization of the critical exponents for $\mathbb{N}$-algebraic functions (generating function associated to an algebraic grammar).
- Other work done: Perron-Frobenius "multivariate" generalisation= the possible limit laws in non strongly connected context-free linear grammars (=automata) ? Answer = "we can asymptotically get <u>any</u> limit law" [Banderier–Bodini–Ponty–Tafat, 2011] + applications to bioinformatics, Boltzmann random generation.
- Work in progress: the possible limit laws in non strongly connected context free-grammars? (Gaussian case: [Bender, Drmota, Soria, Flajolet, Hwang], what else?)
- A description of the ring of the algebraic constants C, ρ, ... in $f_n \sim C\rho^{-n}/\Gamma(\gamma+1)n^{\gamma}$.

Conclusion (implementations)

- an effective Soittola-like theorem for $\mathbb{N}$-algebraic functions:
  - Input: algebraic equation
  - Output: context-free grammar.

  (decidable in the case of 2 non-terminals = genus 0 [Abhyankar])
  (decidable for $\mathbb{N}$-rational functions [Soittola, implementation: Koutschan & Strehl])

- Future implementation (in SageMath):
  - Input: language, pattern,
  - Output: asymptotics, limit law.