How to conjecture and prove that the generating function of the Yang-Zagier numbers is algebraic[1]

CAP21 (IHES, France)

Sergey Yurkevich

Inria Saclay and University of Vienna

1st December, 2021

[1] Joint work with Alin Bostan and Jacques-Arthur Weil.

Two sequences

$$(a_n)_{n\geq 0} = (1\,,\; -48300\,,\; 7981725900\,,\; -1469166887370000, \ldots)$$
$$(b_n)_{n\geq 0} = (1\,,\; -144900\,,\; 88464128725\,,\; -62270073456990000\,,\; \ldots)$$

Origin of a_n and b_n

- In Arithmetic and Topology of Differential Equations, 2018 by Don Zagier:

$$c_{n-3} + 20\left(4500n^2 - 18900n + 19739\right)c_{n-2} + 80352000n(5n-1)(5n-2)(5n-4)c_n + 25\left(2592000n^4 - 16588800n^3 + 39118320n^2 - 39189168n + 14092603\right)c_{n-1} = 0,$$

with initial terms $c_0 = 1$, $c_1 = -161/(2^{10}\cdot 3^5)$ and $c_2 = 26605753/(2^{23}\cdot 3^{12}\cdot 5^2)$.

- Recursion comes from physics: integral over a moduli space ("topological ODE") [Bertola, et al, 2015].

Problem (Zagier, 2018)

Find $(u,v)\in\mathbb{Q}^*\times\mathbb{Q}^*$ such that $c_n\cdot(u)_n\cdot(v)_n\cdot w^n\in\mathbb{Z}$ for some $w\in\mathbb{Z}^*$.

$$(u)_n := u\cdot(u+1)\cdots(u+n-1).$$

- [Yang and Zagier]: $a_n = c_n\cdot(3/5)_n\cdot(4/5)_n\cdot(2^{10}\cdot 3^5\cdot 5^4)^n\in\mathbb{Z}$,
- [Dubrovin and Yang]: $b_n = c_n\cdot(2/5)_n\cdot(9/10)_n\cdot(2^{12}\cdot 3^5\cdot 5^4)^n\in\mathbb{Z}$.

Mystery of a_n and b_n

- "Yang and I found a formula showing that the numbers a_n are integers of exponential growth and hence can be expected to have a generating series that is a **period**, although we have not succeeded in finding it" – [Zagier, 2018]
- "Dubrovin and Yang found that the numbers b_n are *also* integral and that in this case the generating function is not only of Picard-Fuchs type, but is actually **algebraic**!" – [Zagier, 2018]
- "So this is a very mysterious example [...] of numbers defined by recursions with polynomial coefficients." – [Zagier, 2018]
- "My presumed arithmetic intuition [...] was entirely broken" – [Wadim Zudilin]

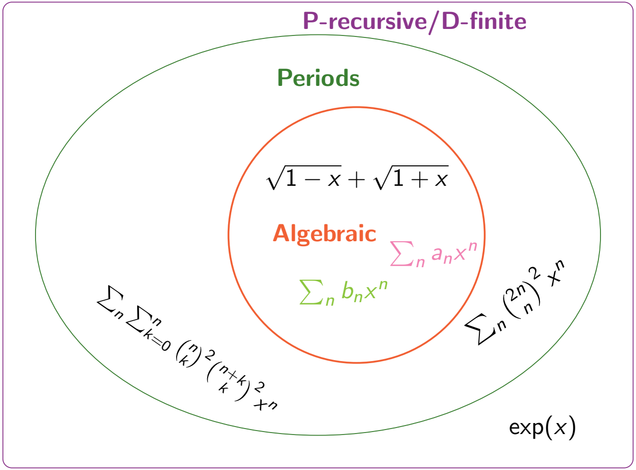
Problem

Investigate the nature of $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$ and similar sequences.

Theorem (Bostan, Weil, Y.)

The generating functions of both $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ are algebraic.

Definitions and interactions

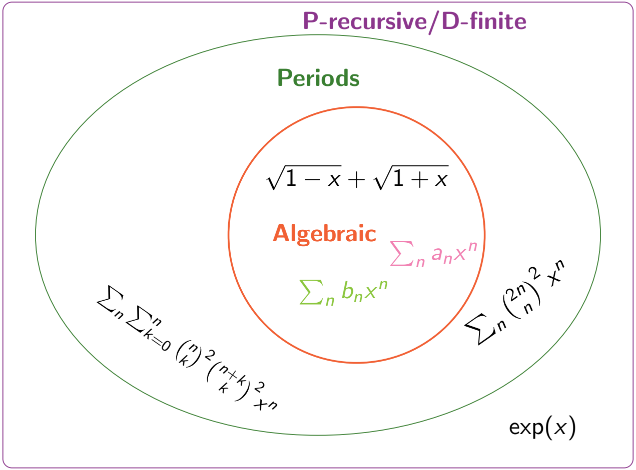


A sequence $(u_n)_{n\geq 0}$ is **P-recursive**, if it satisfies a linear recurrence with polynomial coefficients:

$$c_r(n)u_{n+r} + \cdots + c_0(n)u_n = 0.$$

$u_n = 1/n!$ satisfies $nu_n = u_{n-1}$.

Definitions and interactions



A power series $f(x) \in \mathbb{Q}[\![x]\!]$ is **D-finite** if it satisfies a linear differential equation with polynomial coefficients:

$$p_n(x) f^{(n)}(x) + \cdots + p_0(x) f(x) = 0.$$

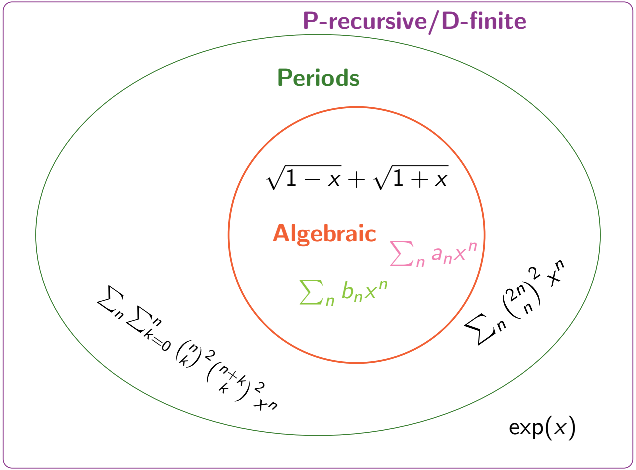
This equation can be rewritten: $L \cdot f = 0$,

$$L = p_n(x)\partial^n + \cdots + p_0(x) \in \mathbb{Q}(x)[\partial],$$

where $\partial := \frac{\mathrm{d}}{\mathrm{d}x}$.

$\exp(x)$ satisfies $\exp'(x) = \exp(x)$.
$L = \partial - 1$.

Definitions and interactions



A power series $f(x) \in \mathbb{Q}[\![x]\!]$ is called a **Period function** if it is an integral of a rational function in x and $t_1, \ldots, t_n$ over a semi-algebraic set.

$$p(e) = 4\int_0^1 \sqrt{\frac{1-e^2t^2}{1-t^2}}\mathrm{d}t$$

$$= 4 \oint\!\!\oint \frac{\mathrm{d}u\mathrm{d}v}{1-\frac{1-e^2u^2}{(1-u^2)v^2}} \text{ and}$$

$$((e-e^3)\partial^2 + (1-e^2)\partial + e)\cdot p = 0,$$

$$p(e) = 2\pi - \frac{\pi}{2}e^2 - \frac{3\pi}{32}e^4 - \cdots.$$

Back to a_n and b_n

- $(a_n)_n$ and $(b_n)_n$ are **P-recursive** sequences $\Rightarrow$ generating functions are **D-finite**.

$$L_a = 1800x(7x-62)\left(x^2+50x+20\right)\partial^2 + 720(42x^3+173x^2-14230x-620)\partial + 6048x^2 - 139453x - 249550 \in \mathbb{Q}(x)[\partial],$$

$$\begin{aligned} L_b = {} & 90000x^3(2911x+310)\left(x^2+50x+20\right)\partial^4 \\ & + 18000x^2\left(154283x^3+5185005x^2+1675710x+142600\right)\partial^3 \\ & + 50x\left(147290778x^3+2740219655x^2+566777510x+37497600\right)\partial^2 \\ & + 5\left(919899288x^3+5629046605x^2+1348939210x+10713600\right)\partial \\ & + 18\left(13937868x^2-1076845x+1247750\right) \in \mathbb{Q}(x)[\partial]. \end{aligned}$$

- The generating functions of $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ solve $L_a \cdot y = 0$ and $L_b \cdot y = 0$.

Main problem

Stanley's problem (1980)

Given a **D-finite** series how to prove or disprove that it is **algebraic**?

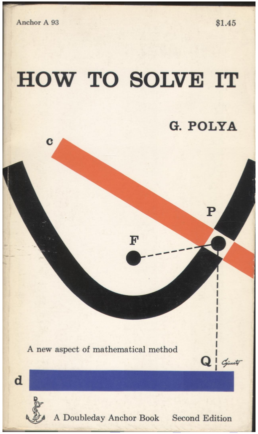
Useful (sub-)question

Given a **D-finite** series how to **conjecture** whether it is **algebraic**?

- Guess & Prove approach – **but** algebraicity degree can be arbitrarily high.
- Algorithm for **rational solutions** of linear ODE [Liouville, 1833], [Barkatou, 1998].
- Solved **in theory** [Singer, 1979, 2014] – **but** usually not applicable in practice.
- New practical algorithm for **disproving** **algebraicity** [Bostan, Rivoal, Salvy, 2021].
- Several tests for justifying **algebraicity** based on **conjectures** or **numerics**:
  work well in practice but do not provide proofs.
- Applied differential Galois theory sometimes efficient proving **algebraicity**.

The Guess-and-Prove approach

- Experimental mathematics and "Guess-and-Prove" propagated by G. Pólya.
- Extremely fruitful when using a computer.
- **P-recursive** sequences/**D-finite** functions: ideal data structure for guessing.
- Find new structure and simpler formulas.
- Very efficient and easy in practice (e.g. with Maple).
- For guessing: `gfun` [Salvy, Zimmermann 1992].
- For proving: Theory of $\mathbb{Q}(x)[\partial]$ and effective properties.



Maplesoft

Guess & Prove for P-recursive sequences/D-finite functions

- Given $u_0, u_1, \ldots, u_N \in \mathbb{Q}$, finding some $c_0(n), \ldots, c_r(n)$ for fixed r and of fixed maximal degree d such that

$$c_r(n)u_{n+r} + \cdots + c_0(n)u_n = 0, \quad \text{for } n = 0, \ldots, N - r$$

  amounts to solving a *linear system*.
- Showing equality of two P-recursive sequences/D-finite functions is decidable and easy/efficient in practice.
- $\to$ Maple (for $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$)

Guess & Prove for algebraicity: Toy example

- Let $f(x) = 2 - x^2/4 - 5x^4/64 + \cdots$ be a solution of
$$(4x^2 - 4) \cdot f''(x) + 4x \cdot f'(x) - f(x) = 0.$$
- How to prove that $f(x)$ is algebraic?
- Idea: *guess* a minimal polynomial $P(x,t)$ and then *prove* its correctness.
- Let $g(x) = 2 - x^2/4 - 5x^4/64$. Finding $(c_1, c_2, \ldots, c_9) \in \mathbb{Q}$ such that
$$P(x, f(x)) = (c_1 + c_2 x + c_3 x^2) + (c_4 + c_5 x + c_6 x^2) g^2 + (c_7 + c_8 x + c_9 x^2) g^4 = 0$$
results in a *linear system* which we can easily solve: $c_3 = 4, c_4 = -4, c_7 = 1$.
- Guess: $P(x,t) = 4x^2 - 4t^2 + t^4$.
- Effective version of Abel's Theorem: Any solution $h(x)$ of $P(x, h(x)) = 0$ satisfies:
$$(4x^2 - 4) \cdot h''(x) + 4x \cdot h'(x) - h(x) = 0.$$
- Proof: Conclude with uniqueness.
- In general algebraicity degree can be arbitrarily high: $N(1+x)f'(x) = f(x)$.

Grothendieck-Katz conjecture: "testing" **algebraicity**

- $L \cdot y = 0$ is equivalent to $Y' = A(x)Y$, where $A(x) \in M^{n \times n}(k)$ and $k = \mathbb{Q}(x)$.
- The p-curvature is the matrix $A_p(x) \in \mathbb{Q}(x)$, where

$$A_0(x) = \mathrm{Id}_n, \quad \text{and} \quad A_{\ell+1}(x) = A'_\ell(x) + A_\ell(x)A(x) \quad \text{for} \quad \ell \geq 0.$$

- It holds that $\partial^k Y = A_k Y$ for $k = 0, 1, \ldots$.

Conjecture (Grothendieck 1960's; Katz, 1972)

All solutions of $Y' = A(x)Y$ are algebraic if and only if $A_p = 0$ mod p for almost all primes p.

- If $A(x)$ is given and we find that $A_{p_i} = 0$ mod p_i for all primes $p_1, p_2, \ldots, p_N$, we can conjecture that all solutions of $Y' = A(x)Y$ are algebraic.
- A_p mod p can be efficiently computed [Bostan, Caruso, Schost, 2015].

"Testing" **algebraicity** for L_a and L_b

- It holds that $\partial^k Y = A_k Y$ for $k = 0, 1, \dots$.
- The right Euclidean division of ∂^k by L in $\mathbb{Q}(x)[\partial]$ reads:

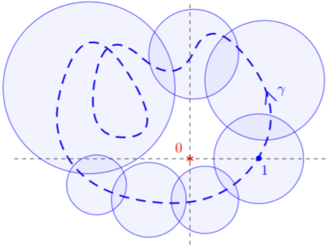
$$\partial^k = (\dots) \cdot L + d_{n-1}(x)\partial^{n-1} + \cdots + d_0(x),$$

  for $d_0(x), \dots, d_{n-1}(x) \in \mathbb{Q}(x)$ depending on k.
- Hence, the first row of A_k is $d_0(x), \dots, d_{n-1}(x)$.
- $\to$ Maple (for L_a and L_b)

Monodromy group: quantifying algebraicity

- $L \cdot y = 0$ for $L \in \mathbb{Q}(x)[\partial]$ has $n = \mathrm{ord}(L)$ linearly independent solutions.
- Assume $f_1, \ldots, f_n$ are linearly independent solutions at 0. If we analytically continue them along a closed loop in $\mathbb{C}$, we find $\widetilde{f_1}, \ldots, \widetilde{f_n}$ possibly different.
- There exists $M_{\underline{f}} \in \mathrm{GL}(n, \mathbb{C})$ such that

$$\begin{pmatrix} \widetilde{f_1} \\ \vdots \\ \widetilde{f_n} \end{pmatrix} = M_{\underline{f}} \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}.$$



- The matrices $M_{\underline{f}}$ define the so-called monodromy group M.

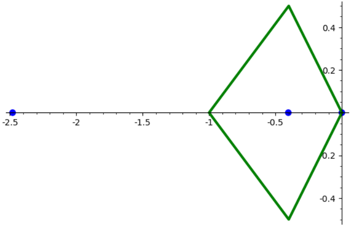
Theorem (Singer, Ulmer, 1993)

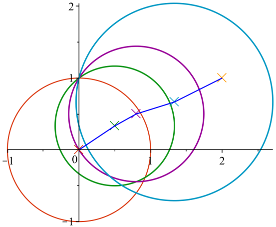
Let f be a solution of $L \cdot y = 0$. The algebraicity degree of f is equal to the cardinality of the orbit of f under the action of M.

- Analytic continuation of **D-finite** functions can be efficiently computed numerically[Chudnovsky[2], 1987],[van der Hoeven, 1999, 2001],[Mezzarobba, 2010].

Quantifying **algebraicity** for L_a and L_b

- Very efficient analytic continuation implemented by Mezzarobba in SageMath.
- $\rightarrow$ SageMath.
- Numerical computations suggest: solutions of L_a and L_b have alg. degree 120.





Differential Galois theory: proving **algebraicity**

- $L \cdot y = 0$ is equivalent to $Y' = A(x)Y$, where $A(x) \in M^{n \times n}(k)$ and $k = \overline{\mathbb{Q}}(x)$.
- Picard-Vessiot extension: $K = k(U)$, where U is a fundamental solution matrix.
- The differential Galois group G is the group of field automorphisms of K which commute with the derivation and leave all elements of k invariant:

$$G := \mathrm{Aut}_\partial(K/k) = \{\sigma \in \mathrm{Aut}(K) \colon \sigma|_k \equiv \mathrm{id}_k \text{ and } \sigma \circ \partial \equiv \partial \circ \sigma\}.$$

- G is a linear algebraic subgroup of $\mathrm{GL}_n(\overline{\mathbb{Q}})$.
- G stabilizes the ideal of differential relations between solutions. Moreover:

Theorem (Kolchin, 1948)

$L \cdot y = 0$ has a basis of algebraic solutions if and only if G is finite.

- In practice G is difficult to compute [Hrushovski, 2002], [Feng, 2015], [van der Hoeven, 2007], [Amzallag, Minchenko, Pogudin, 2018], [Sun, 2019].

Differential Galois theory: proving algebraicity

- $L \cdot y = 0$ is equivalent to $Y' = A(x)Y$, where $A(x) \in M^{n\times n}(k)$ and $k = \overline{\mathbb{Q}}(x)$.
- G is the differential Galois group.
- Galois-Lie algebra $\mathfrak{g} := \mathrm{Lie}(G)$: Lie algebra of G, i.e. the tangent space of G at id.
- $\mathfrak{g}$ measures the transcendence of K over k:

Theorem (Kolchin, 1948)

If K is the Picard-Vessiot extension of $Y' = A(x)Y$ and $\mathfrak{g} = \mathrm{Lie}(G)$, then

$$\dim_{\mathbb{C}}(G) = \dim_{\mathbb{C}}(\mathfrak{g}) = \mathrm{trdeg}(K/k).$$

- Theory and algorithm for computing $\mathfrak{g}$ [Barkatou, Cluzeau, Di Vizio, Weil, 2020].
- Idea: Compute symmetric powers of L and find **rational solutions** of them. These solutions yield information for $\mathfrak{g}$ via solving a **linear** system.

Toy example

- The operator $L = (4x^2 - 4)\partial^2 + 4x\partial - 1$ has a basis of algebraic solutions:

$$\sqrt{1+x} + \sqrt{1-x} \text{ and } \sqrt{1+x} - \sqrt{1-x}.$$

- $L \cdot y = 0$ is equivalent to $Y' = A(x)Y$ where $A(x) = \begin{pmatrix} 0 & 1 \\ \frac{1}{4x^2-4} & \frac{-4x}{4x^2-4} \end{pmatrix}.$

Toy example

- The operator $L = (4x^2-4)\partial^2 + 4x\partial - 1$
- $L \cdot y = 0$ is equivalent to $Y' = A(x)Y$ where $A(x) = \begin{pmatrix} 0 & 1 \\ \frac{1}{4x^2-4} & \frac{-4x}{4x^2-4} \end{pmatrix}$.
- If $Y = (y_1, y_2)^t$ is a solution to $Y' = A(x)Y$ then $Y = (y_1^2, 2y_1y_2, y_2^2)^t$ is a solution to the symmetric square system $Y' = A^{(2)}(x)Y$, where now

$$A^{(2)}(x) = \frac{1}{4(x^2-1)} \begin{pmatrix} 0 & 4x^2-4 & 0 \\ 2 & -4x & 8x^2-8 \\ 0 & 1 & -8x \end{pmatrix}.$$

- It has rational solutions! $F_1 = (4x, 4, x/(x^2-1))^t$, $F_2 = (-4, 0, 1/(x^2-1))^t$.
- If $M \in \mathfrak{g}^{(2)}$ then $MF = 0$ and M comes from a symmetric square. I.e. M satisfies

$$\begin{pmatrix} 2m_{1,1} & m_{1,2} & 0 \\ 2m_{2,1} & m_{1,1}+m_{2,2} & 2m_{1,2} \\ 0 & m_{2,1} & 2m_{2,2} \end{pmatrix} \cdot F_\ell = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad m_{i,j} \in \mathbb{Q}(x), \ell = 1, 2.$$

- The only solution is $m_{i,j} = 0$. Hence $\mathfrak{g}^{(2)} = \mathfrak{g} = 0$. All solutions of L are algebraic.

The generating sequence of $(b_n)_n$ is algebraic (known to Dubrovin & Yang)

- For L_b same method as in the toy example works.
- $L_b \cdot y = 0$ equivalent to $Y' = A(x)Y$ for $A(x) \in M^{4\times 4}(\mathbb{Q}(x))$.
- The fifth symmetric power $Y' = A^{(5)}(x)Y$ has rational solutions.
- $A^{(5)}(x) \in M^{N\times N}(\mathbb{Q}(x))$, where $N = \binom{4+5-1}{4-1} = 56$.
- Finding the rational solutions takes ≈2 min on a regular PC.
- The corresponding system in $m_{i,j}$ has no non-zero solutions in $\mathbb{Q}(x)$ (≈15 sec).
- $\Rightarrow \mathfrak{g}_b = 0$, therefore L_b has only algebraic solutions.

The generating sequence of $(a_n)_n$ is algebraic (new)

- For the generating function of $(a_n)_{n\geq 0}$ same method as for $(b_n)_{n\geq 0}$ works.
- The 20th symmetric power has rational solutions (≈ 1 sec).
- $A^{(20)} \in M^{N\times N}(\mathbb{Q}(x))$, where $N = \binom{2+20-1}{2-1} = 21$.
- The corresponding system in $m_{i,j}$ has no non-zero solutions in $\mathbb{Q}(x)$ ($\approx$4 sec).
- $\Rightarrow \mathfrak{g}_a = 0$, therefore L_a has only algebraic solutions.

- $\to$ Maple

Experimental mathematics: more similar examples

Problem

Find $(u, v) \in \mathbb{Q}^* \times \mathbb{Q}^*$ such that $c_n \cdot (u)_n \cdot (v)_n \cdot w^n \in \mathbb{Z}$ for some $w \in \mathbb{Z}^*$.

$$(u)_n := u \cdot (u+1) \cdots (u+n-1).$$

#	u	v	ODE order	degree	#	u	v	ODE order	degree
a_n	3/5	4/5	2	120	f_n	19/60	49/60	4	155520
b_n	2/5	9/10	4	120	g_n	19/60	59/60	4	46080
c_n	1/5	4/5	2	120	h_n	29/60	49/60	4	46080
d_n	7/30	9/10	4	155520	i_n	29/60	59/60	4	155520
e_n	9/10	17/30	4	155520					

- "Test": 0 p-curvatures for primes $< 100 \to$ expect **algebraic** generating functions.
- Quantify: Guesses for degrees based on numerics.
- Proof: Done: a_n, b_n, c_n. In progress: $d_n, e_n, f_n, g_n, h_n, i_n$.

Summary

- Both sequences $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ have algebraic generating functions, hence they are particular **periods**.
- Guess & Prove approach often provides useful insight but is sometimes infeasible.
- The Grothendieck-Katz conjecture allows efficient "testing" whether a **D-finite** series is **algebraic**.
- Numerical monodromy group calculations allow efficient quantifying **algebraicity** of **D-finite** series.
- Differential Galois theory allows efficient proving that **D-finite** series is **algebraic**.

Bonus: explicit solution for $\sum_{n\geq 0} a_n x^n$

We saw that $\sum_{n\geq 0} a_n x^n$ is a solution of

$$q_2(x)y''(x) + q_1(x)y'(x) + q_0(x)y(x) = 0, \quad \text{where} \tag{1}$$

$$q_2(x) = 5x(302400x - 31)(373248000x^2 + 216000x + 1),$$

$$q_1(x) = 135444234240000x^3 + 64571904000x^2 - 61473600x - 31,$$

$$q_0(x) = 300(902961561600x^2 - 240974784x - 4991).$$

Maple's `dsolve(deq)` shows that every solution of (1) is a linear combination of

$$u_1(x) \cdot {}_2F_1\left[\begin{matrix} -1/60 \ \ 11/60 \\ 2/3 \end{matrix}; \frac{p_1(x)}{p_2(x)}\right] \quad \text{and} \quad u_2(x) \cdot {}_2F_1\left[\begin{matrix} 19/60 \ \ 31/60 \\ 4/3 \end{matrix}; \frac{p_1(x)}{p_2(x)}\right],$$

where ${}_2F_1\left[\begin{matrix} a \ b \\ c \end{matrix}; x\right]$ is the Gaussian hypergeometric function

$${}_2F_1\left[\begin{matrix} a \ \ b \\ c \end{matrix}; x\right] := \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!}, \quad (u)_j := u(u+1)\cdots(u+j-1).$$

Bonus: Origin of $(c_n)_{n\geq 0}$

- For a simple Lie-algebra $(\mathfrak{g}, [\cdot,\cdot])$ [Bertola, Dubrovin, Yang, 2015] define the co-called *topological ordinary differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} M = [M, \Lambda],$$

where $M = M(\lambda)$ and $\Lambda = I_+ + \lambda E_{-\theta}$, for a principal nilpotent element $I_+ = \sum_{i=1}^{n} E_i$ and (normalized) $E_{-\theta} \in \mathfrak{g}_{-\theta}$.

- For $\mathfrak{g} = \mathrm{sl}_{n+1}(\mathbb{C})$ one finds

$$\Lambda = \begin{pmatrix} 0 & I_n \\ \lambda & 0 \end{pmatrix}, \quad I_n \text{ is the } n \times n \text{ identity matrix.}$$

and the (normalized) (dominant) ODE reads

$$64800000x^3(x+155)y^{(iv)}(x) + (x^2 - 1220x + 623875)y(x) + 7200(10x^2 + 3209x + 133920)y'(x) + 18000x(5x^2 + 6091x + 1874880)y''(x) + 12960000x^2(18x + 3565)y'''(x) = 0$$

- Then $\sum_{n\geq 0} c_n x^n$ is the unique power series solution.