The unbounded denominators conjecture for vector valued modular forms

Vesselin Dimitrov

University of Toronto

October 20, 2021

arXiv:2109.09040v1

Joint work with Frank Calegari and Yunqing Tang

Basic question

Problem

Is there an exact description of all the $\mathbb{Z}[[x]]$ formal solutions to linear ODEs $L(f) = 0$ where L ranges over all linear differential operators with coefficients in $\mathbb{C}(x)$ and no singularities outside of $0, 1/16$ and ∞?

A transcendental example is the complete elliptic integral

$$ {}_2F_1\left[\begin{matrix}1/2 \;\; 1/2\\ 1\end{matrix}; 16x\right] = \sum_{n=0}^{\infty} \binom{2n}{n}^2 x^n \in \mathbb{Z}[[x]] $$

$$ = \frac{1}{\sqrt{1-4x}} * \frac{1}{\sqrt{1-4x}} \quad \text{(Hadamard's product)} $$

On the arithmetic boundary in G-functions theory

A.3. Examinons le cas $(g, N) = (1, 1)$. L'espace de Hilbert H des états d'une théorie conforme de charge centrale c est une représentation d'un produit de deux algèbres de Virasoro, de la forme $H = \oplus_{i,j} V(h_i, c) \otimes \bar{V}(\bar{h}_j, c)$, où $V(h_i, c)$ est la représentation irréductible de plus haut poids $h_i \geq 0$ et de niveau c, et où les barres marquent les conjugués complexes. Les fonctions de corrélation de la théorie, et la fonction de partition en particulier $Z(t) = Tr q^{L_0 - c/24} \bar{q}^{\bar{L}_0 - c/24} = \sum_{i,j} N_{ij} \chi(h_i, c) \bar{\chi}(\bar{h}_j, c)$ sont des invariants modulaires sur le demi-plan de Poincaré ($q = e^{2\pi i \tau}$). Ce sont donc des fonctions du paramètre $\lambda = 16 q^{1/2} \prod_1^\infty (\frac{1+q^n}{1+q^{n-1/2}})^8$ de Legendre. Les théories rationnelles sont caractérisées par la finitude du rang de la matrice à coefficients entiers naturels (N_{ij}). G. Anderson et G. Moore ont montré que c et les plus hauts poids h_i de l'algèbre de Virasoro sont alors rationnels [1]. Ils s'appuient sur la remarque que $Z(t)$ s'écrit comme une somme finie f_k, où les $f_k = N_{kj} \chi(h_k, c)$ (resp. les g_k) engendrent des représentations de dimension finie du groupe modulaire ; ici $\bar{g}_k = \bar{\chi}(\bar{h}_k, c) + D_{jk} \bar{\chi}(\bar{h}_j, c)$, où les coefficients D_{jk} satisfont à $N_{ij} = \sum_{k=1}^{k=r} D_{jk} N_{ik}$ pour tout i et tout $j > r$. Il en découle que, vu comme fonction holomorphe multiforme en la variable λ, le vecteur $\vec{f}$ (resp.$\vec{g}$) de composantes les f_k (resp.g_k) vérifie un système différentiel linéaire sur $\mathbb{C} \setminus \{0, 1\}$.

A.4. En choisissant r assez grand, on peut supposer que les D_{jk} sont entiers. Il en découle que les développements de Puiseux en la variable q des fonctions f_k et g_k sont à coefficients entiers. Comme $q^{1/2}$ et $\lambda/16$ s'expriment mutuellement comme séries

From Yves André's Appendix A: Conjecture de Grothendieck et théorie des champs conformes in *Sur la conjecture des p-courbures de Grothendieck–Katz et un problème de Dwork*

On the arithmetic boundary in G-functions theory

(A) The Koebe map $\varphi(z) = z/(1+z)^2 : D(0,1) \to \mathbb{C} \setminus [1/4, \infty)$ is the conformally largest *univalent* map omitting the value $1/4$. Its *conformal size* is $|\varphi'(0)| = 1$.

If $f(x) \in \mathbb{Z}[1/N][[x]]$ has the global product $\prod_p R_p > 1$ of p-adic convergence radii, and has $f(\varphi(z)) \in \mathbb{C}[[z]]$ holomorphic (convergent) on $|z| < 1$ under some **univalent** *holomorphic map $\varphi : (D(0,1), 0) \hookrightarrow (\mathbb{C}, 0)$ with $|\varphi'(0)| \geq 1$, then $f(x) \in \mathbb{Q}(x)$ is rational. Sharp in view of the quadratic irrational example $f(x) = \sqrt{1-4x}$.*

This is the standard neat example illustrating the sharpness of the general Pólya-Bertrandias arithmetic rationality theorem.

But there are continuum many formal power series $f(x) \in \mathbb{Z}[[x]]$ such that $f(\varphi(z))$ converges. (*Raphaël Robinson*: An extension of Pólya's theorem on power series with integer coefficients, 1968)

André's algebraicity criterion: If we delete the word *univalent*, the conclusion holds with $f(x) \in \overline{\mathbb{Q}(x)}$ algebraic (instead of rational).

On the arithmetic boundary in G-functions theory

If we relax the univalent condition to "univalent above $\{0\}$," the constant $1/4$ improves to $1/16$, and we will see that this is sharp. The counterpart of the Koebe map here is a modular function:

(B) The modular lambda map

$$\lambda(q) = \frac{\left(\sum_{n\in\mathbb{Z}} q^{(n+1/2)^2}\right)^4}{\left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^4} = 16q\prod_{n=1}^{\infty}\left(\frac{1+q^{2n}}{1+q^{2n-1}}\right)^8$$

$$= 1 - \frac{\left(\sum_{n\in\mathbb{Z}}(-1)^n q^{n^2}\right)^4}{\left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^4} : D(0,1) \to \mathbb{C} \setminus \{1\}$$

is the conformally largest map omitting the value 1 *subject to the singleton fiber condition* $\lambda^{-1}(0) = \{0\}$. Its conformal size is $|\lambda'(0)| = 16$.

On the arithmetic boundary in G-functions theory

Let $S, T \subset \mathbb{C}$, with $0 \in T$. Suppose that $f(x) \in \mathbb{Z}[1/N][[x]]$ has the global product $\prod_p R_p > 1$ of p-adic convergence radii, fulfills a linear ODE over $\mathbb{C}(x)$ with no singularities outside $S \cup T$, and that there exists a holomorphic map $\varphi : D(0,1) \to \mathbb{C} \smallsetminus S$ with $\varphi^{-1}(0) = \{0\}$, $\#\varphi^{-1}(t) \leq 1$ for all $t \in T$, and $|\varphi'(0)| \geq 1$. Then $f(x) \in \overline{\mathbb{Q}(x)}$ is an algebraic function.

Such as $f(x) = \sqrt{1-8x}$ with $\varphi(z) = \lambda(z)/8$, $T = \{0\}$, and S resp. $\{1/8, \infty\}$.

For then an easy monodromy calculation shows that $f(\varphi(z)) : D(0,1) \to \mathbb{C}$ is analytic on the full disc $|z| < 1$, even though its linear ODE has singularities at the singleton fibers $\varphi^{-1}(T)$, including at the origin $z = 0$.

Then André's criterion applies at once.

The singularity set $\{0, 1/16, \infty\}$ as a boundary case

The rescaled modular lambda map

$$\lambda(z)/16 : D(0,1) \to \mathbb{C} \setminus \{1/16\}$$

fulfills the preceding with $T = \{0\}$ and $S = \{1/16, \infty\}$.

It is now uniformized to a unit conformal size $\lambda'(0)/16 = 1$, and it is the conformally largest analytic map omitting $\{1/16, \infty\}$ and with $\varphi^{-1}(0) = \{0\}$.

Back to André's question at the arithmetic boundary of Grothendieck's p-curvature conjecture

à coefficients entiers l'un de l'autre, on en déduit, via la proposition 5.3.3, que les systèmes différentiels satisfaits par $\vec{f}$ et $\vec{g}$ sont à p-courbures nulles pour presque tout p. Que les f_k et g_k soient des fonctions algébriques de la variable λ découlerait donc de la conjecture de Grothendieck.

A.5. En fait, cette situation représente un cas très spécial d'application de la conjecture de Grothendieck : c'est le cas limite du critère 5.4.5 dans lequel *il existe une uniformisation* v*-adique simultanée de* $\vec{f}$, $\vec{g}$, *et* $x = \lambda$ *dans un disque* $D(0, R_v)$, *avec* $\prod R_v = 1$ (au lieu de $\prod R_v > 1$). Notons que ce critère ne s'étend pas sans restriction à ce cas limite, comme le montre l'exemple de la fonction hypergéométrique $F(\frac{1}{2}, \frac{1}{2}, 1; x)$ (uniformisation par les fonctions thêta, cf. [5]). Il faut donc tenir compte en outre de ce que les monodromies locales sont semi-simples dans la situation (du fait que les p-courbures sont presque toutes nulles). Cela suggère de rechercher des uniformisations de norme > 1 de (produits de) surfaces de Riemann compactes par des (poly)disques unité.

André posed this challenge question at the end of his appendix to his paper *Sur la conjecture des p-courbures de Grothendieck–Katz et un problème de Dwork*

Our main theorem, first form

Theorem (Calegari, D., Tang, 2021)

Suppose $f(x) \in \mathbb{Z}[[x]]$ satisfies a linear ODE $L(f) = 0$ where L has no singularities outside of $\{0, 1/16, \infty\}$, *and suppose additionally that L has a semisimple local monodromy at $x = 0$.* Then indeed $f(x) \in \overline{\mathbb{Q}(x)}$ (as prescribed by Christol's conjecture) and more precisely:

$\exists N \in \mathbb{N}$ such that the function $f\big(\lambda(q)/16\big)$, with $q = e^{\pi i \tau}$ on the upper half plane $\mathbf{H}$, is $\Gamma(N)$-automorphic:

$$\forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N, \quad f\left(\frac{a\tau + b}{c\tau + d}\right) = f(\tau).$$

Our main theorem, second form

Definition

A *vector-valued modular form* of weight k and dimension n for the group $\mathrm{SL}_2(\mathbb{Z})$ is a pair (F, ρ) comprised of:

- A holomorphic mapping $F = (F_1, \ldots, F_n) : \mathbf{H} \to \mathbb{C}^n$ on the upper half plane;
- A representation $\rho : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_n(\mathbb{C})$;
- Linked by $(c\tau + d)^{-k} F^{\mathrm{t}}\left(\frac{a\tau+b}{c\tau+d}\right) = \rho \begin{pmatrix} a & b \\ c & d \end{pmatrix} F^{\mathrm{t}}(\tau)$;
- Such that the matrix $\rho \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_n(\mathbb{C})$ is semisimple;
- All components $F_j : \mathbf{H} \to \mathbb{C}$ have *moderate growth in vertical strips*: for all $a < b$ and $C > 0$, there exist $A, B > 0$ such that

$$\forall \tau \in \mathbf{H}, \quad a \leq \mathrm{Re}\,\tau \leq b, \quad \mathrm{Im}\,\tau \geq C \quad \Longrightarrow \quad |F_j(\tau)| \leq A e^{B\,\mathrm{Im}\,\tau}.$$

Our main theorem, second form

Theorem

Let (F, ρ) be a vector-valued modular form for $\mathrm{SL}_2(\mathbb{Z})$ of dimension n and weight k. Suppose that some component function $F_j(\tau) : \mathbf{H} \to \mathbb{C}$ of $F = (F_1, \ldots, F_n) : \mathbf{H} \to \mathbb{C}^n$ has at $\tau = i\infty$ a formal Fourier expansion lying in $\mathbb{Z}[[q]] = \mathbb{Z}[[e^{2\pi i \tau}]]$. Then that component $F_j(\tau)$ is a classical modular form of weight k on a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$.

Corollary (Mason's conjecture)

If all components of a vector-valued modular form (F, ρ) for $\mathrm{SL}_2(\mathbb{Z})$ have q-expansions in $\mathbb{Z}[[q^{1/m}]]$, then the representation ρ has a finite image, and more precisely $\ker(\rho) \supseteq \Gamma(N)$ for some $N \in \mathbb{N}$.

Integral q-expansions arise from rational conformal field theories, encoding a graded (twisted) dimension

Let me just mention this simplest of all the illustrating examples:

$$j(\tau)^{1/3} = q^{-1/3} \frac{\Theta_{E_8}(q)}{\prod_{n=1}^{\infty}(1-q^n)^8}$$
$$= q^{-1/3}(1 + 248q + 4124q^2 + 34752q^3 + \cdots).$$

Kač and Peterson identified this (E_8 Moonshine style) with the graded dimension of the level one highest-weight representation of the affine Kac–Moody algebra $E_8^{(1)}$. One easily verifies that $j^{1/3}$ is in fact a Hauptmodul for the *congruence* group $\Gamma(3) \subset \mathrm{SL}_2(\mathbb{Z})$.

Contrast point: the automorphy group of $j(\tau)^{1/5} \in \mathbb{Z}[1/5][[q^{1/5}]] \setminus (\mathbb{Z}[[q^{1/5}]] \otimes \mathbb{C})$ is a noncongruence arithmetic group, because the denominators are 5-adically unbounded. In fact the complete list of n for which $j^{1/n}$ is congruence modular are the divisors of 24.

Atkin and Swinnerton-Dyer's conjecture: the core case

Theorem (Calegari, D., Tang, 2021)

Let $f(\tau) \in \overline{\mathbb{Z}}[[q^{1/N}]] \otimes \mathbb{C}$ be a holomorphic function on the upper half plane $\mathbf{H}$, expanded out in $q = e^{\pi i \tau}$. Suppose there exists a finite index subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ and an integer k such that f is a modular form of weight k and level Γ:

$$\forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, \quad f\left(\frac{a\tau + b}{c\tau + d}\right) = (a\tau + b)^k f(\tau).$$

(And $f(\tau)$ is meromorphic locally near every cusp of the compactification of $\mathbf{H}/\Gamma$.)
Then $f(\tau)$ is modular under a congruence *subgroup of $\mathrm{SL}_2(\mathbb{Z})$: there exists an $N \in \mathbb{N}$ such that*

$$\forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N, \quad f\left(\frac{a\tau + b}{c\tau + d}\right) = (a\tau + b)^k f(\tau).$$

The congruence property

$$\mathrm{SL}_2(\mathbb{Z}) = \mathrm{Mod}(\mathbb{T}^2) = \Gamma_{1,0} \cong \Gamma_{1,1}$$

- The vast majority of its sublattices are noncongruence (in contrast to the higher rank lattice $\mathrm{SL}_3(\mathbb{Z})$).
- We prove a characterization of the congruence sublattices in terms of integral Fourier expansions of their modular forms.
- The integrality of q-expansions of the congruence modular forms comes conceptually from Hecke theory.
- As Atkin conjectured and Serre and Berger proved, there is no Hecke theory in the noncongruence case. *No reason for integrality.*
- Their proof used that $\mathrm{SL}_2(\mathbb{Z}[1/p])$ has the congruence subgroup property (CSP), in contrast to $\mathrm{SL}_2(\mathbb{Z})$.
- We will use a Diophantine analysis (Gel'fond style) to extrapolate the CSP for $\mathrm{SL}_2(\mathbb{Z}[1/p])$ (using a whole range of primes p) $\rightsquigarrow$ our congruence characterization for $\mathrm{SL}_2(\mathbb{Z})$.

A reduction to weight 0 (modular functions)

The special case $k = 0$ of modular *functions* is no loss of generality, *viz.* multiplying by a power of Dedekind η, *and it will be assumed later on in our proof.*

To be more precise, consider the Ramanujan modular form of weight 12 on $\mathrm{SL}_2(\mathbb{Z})$:

$$\Delta(\tau/2) := q \prod_{n=1}^{\infty} (1 - q^n)^{24} \in \mathbb{Z}[[q]], \quad q = \exp(\pi i \tau).$$

WLOG, we may assume that $\Gamma \subset \Gamma(2)$. Now

$$f(\tau)^{12} j(\tau/2)^{-k} \Delta(q)^{-k} \in \mathbb{Z}[[q]]$$

is modular of weight 0 on Γ and has integer coefficients.

We pass from $\mathrm{SL}_2(\mathbb{Z}) = \Gamma(1)$ to the mod 2 level $\Gamma(2)$, since the latter acts ***freely*** *on* $\mathbf{H}$ *while the former does not.*

One of Jacobi's jewels

$$ {}_2F_1\left[\begin{matrix}1/2 & 1/2 \\ & 1\end{matrix}; \frac{\left(\sum_{n\in\mathbb{Z}} q^{(n+1/2)^2}\right)^4}{\left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^4}\right] = \left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^2 $$

(An equation of the form holonomic in λ = a modular form)

- Jacobi's *Thetanullwerte*:
$\left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^4 = \left(\sum_{n\in\mathbb{Z}} q^{(n+1/2)^2}\right)^4 + \left(\sum_{n\in\mathbb{Z}}(-1)^n q^{n^2}\right)^4$
- Gauss's *hypergeometric equation*
$(\lambda^2-\lambda)\frac{d^2f}{d\lambda^2} + ((a+b+1)\lambda - c)\frac{df}{d\lambda} + ab\,f = 0$, with the parameters $a = b = 1/2$, $c = 1$
- Thus we have pulled back the hypergeometric by $\lambda(q)$, and since $(\mathbf{H}, i) \to (\mathbb{C}\smallsetminus\{0,1\}, 1/2)$, $\tau \mapsto \lambda(q)$ is the (analytic) universal covering map, a weight 0 vvmf on $\Gamma(2)$ is nothing more nor less than a local system on $\mathbb{C}\smallsetminus\{0,1\}$ with a semisimple $x = 0$ local monodromy.

A formal passage of the integral expansion property

Equally important to our story, the integrality of a q-expansion can be equivalently interpreted as a $\mathbb{Z}[[x]]$ *holonomic* function under defining

$$x := \lambda(q)/16 = q - 8q^2 + \cdots \in q + q^2\mathbb{Z}[[q]] = x + x^2\mathbb{Z}[[x]]$$

and substituting formally

$$q = x + 8x^2 + 91x^3 + \cdots \in x + x^2\mathbb{Z}[[x]].$$

Thus we get, out of a $\mathbb{Z}[[q]]$ weight-k modular form on a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$:

- If $k = 0$, an algebraic function in $\mathbb{Z}[[x]]$ with branching only at $x = 0, 1/16, \infty$;
- In general, a holonomic function in $\mathbb{Z}[[x]]$ on a finite étale covering of $Y(2) = \mathbb{P}^1 \setminus \{0, 1/16, \infty\}$, of rank $k+1$ and monodromy commensurable with $\mathrm{Sym}^k\mathrm{SL}_2(\mathbb{Z}) \hookrightarrow \mathrm{SL}_{k+1}(\mathbb{Z})$.

Getting the linear ODE out of the modular form of weight k

If $f(\tau)$ is modular of weight k under some finite index subgroup $\Gamma < \Gamma(2)$, and $\mathbb{C}_\Gamma \subset \overline{\mathbb{C}(x)}$ is the field of Γ-automorphic functions, then the $\mathbb{C}_\Gamma$-linear span of the $k+1$ functions

$$f(\tau),\ \tau\, f(\tau),\ \tau^2\, f(\tau), \ldots, \tau^k\, f(\tau)$$

is closed under d/dx.

The elliptic nome $\tau = -i\, K'/K$ is the simplest instance

$$q = e^{\pi i \tau} = e^{-\pi K'/K},$$

where K, K' are the full elliptic integrals

$$K(\lambda) = 4\int_0^{\pi/2} \frac{d\theta}{\sqrt{1-\lambda\sin^2\theta}},$$
$$K'(\lambda) = K(1-\lambda) = -i\tau\, K(\lambda)$$

They satisfy the linear differential equation
$(\lambda^2-\lambda)\frac{d^2K}{d\lambda^2} + (2\lambda-1)\frac{dK}{d\lambda} + \frac{K}{4} = 0$ symmetric under $\lambda \leftrightarrow 1-\lambda$,
and they are the hypergeometric functions

$$K(\lambda) = (2\pi)\cdot {}_2F_1\Big(\frac{1}{2},\frac{1}{2};1;\lambda\Big), \quad K'(\lambda) = K(1-\lambda).$$

Fuchsian Uniformization

We need a notion of conformal size for a pointed (connected, open) Riemann surface (U, P). The Riemann–Fuchs–Koebe *uniformization theorem for Riemann surfaces* gives the correct such notion.

Let $D(0,1) = \{|z| < 1\} \subset \mathbb{C}$ be the open unit disc in the complex plane. *Then there exists an essentially* unique analytic universal covering map*

$$F : (D(0,1), 0) \to (U, P), \quad F(0) = P.$$

* *Up-to precomposing by a rotation* $z \mapsto e^{i\theta} z$

Definition

We define the **uniformization radius** of the pointed Riemann surface (U, P) to be $|F'(0)|$.

The analytic universal covering of $(\mathbb{C} \setminus \{\pm 1\}, 0)$

$$\tau \mapsto 2\lambda(\tau) - 1, \quad (\mathbf{H}, i) \to (\mathbb{C} \setminus \{\pm 1\}, 0)$$

Due to the accidental isomorphism $Y(2) \cong \mathbb{C} \setminus \mu_2$ mentioned before.

The Riemann uniformization radius is thus computed explicitly:

$$\frac{\Gamma(1/4)^2}{4\pi^2} \approx 4.376879\ldots.$$

Compare: If we view $\lambda(q) = 16q + \cdots : \{|q| < 1\} \to \mathbb{C} \setminus \{1\}$ instead as a holomorphic function with singleton fiber $\lambda^{-1}(0) = \{0\}$ on the unit q-disc, then the function $\sqrt{\lambda(q^2)} = 4q + \cdots : D(0,1) \to \mathbb{C} \setminus \{\pm 1\}$ is still holomorphic, and materializes the *strict* lower estimate > 4 on the above uniformization radius, by virtue of factorizing *properly* via the universal covering map: *overconvergence* $>$ *a priori radius*

The universal covering $F_N : (D(0,1),0) \to (\mathbb{C} \setminus \mu_N, 0)$

The last argument with the analyticity of $\sqrt[N]{\lambda(q^N)} = 16^{1/N} q + \cdots : D(0,1) \to \mathbb{C} \setminus \mu_N$ demonstrates immediately the *strict* (but asymptotically sharp "to zeroth order") lower bound $> 16^{1/N}$ on the uniformization radius $|F_N'(0)|$ at the origin of $\mathbb{C} \setminus \mu_N$.

We need to be much more precise: the above classical theory generalizes, with Poincaré's ODE approach to the uniformization of Riemann surfaces, to describe the multivalued inverse "fairly explicitly" in terms of hypergeometric functions, and derive an exact formula for the uniformization radius (Kraus and Roth, 2016):

$$|F_N'(0)| = \frac{\Gamma\left(\frac{N-1}{2N}\right)^2 \Gamma\left(1+\frac{1}{N}\right)}{\Gamma\left(\frac{N+1}{2N}\right)^2 \Gamma\left(1-\frac{1}{N}\right)}$$

$$= 16^{1/N}\left(1 + \frac{\zeta(3)}{2N^3} + \frac{3\zeta(5)}{8N^5} + \dots\right) > 16^{1/N}\left(1+\zeta(3)/(2N^3)\right).$$

The universal covering of $\mathbb{C} \setminus \mu_N$ is our path to resolving a $\mathbb{Z}/N$ local monodromy at $x = 0$

- Firstly, we have seen that the unbounded denominators conjecture is secretly an arithmetic property about local systems on $Y(2) \cong \mathbb{P}^1 \setminus \{0, 1, \infty\}$.
- Secondly, if our local system (as is the case in our $k = 0$ situation with UBD, but not in the hypergeometric equation discussed above) has a finite $\mathbb{Z}/N$ local monodromy at the point $\lambda = 0$, then the N-isogeny $\lambda \mapsto \lambda^N$ trades our $\mathbb{C} \setminus \{0, 1\}$ local system into a $\mathbb{C} \setminus \mu_N$ local system.
- We want to exploit an "overconvergence boost" from the fact that the $F_N : D(0, 1) \to \mathbb{C} \setminus \mu_N$ pullback of the latter local system is a trivial local system (*no singularities throughout!*) on the disc $D(0, 1)$.

"Overconvergence," since that — with or without a finite $\lambda = 0$ local monodromy — were true by fiat *for the pullback under the holomorphic map* $\sqrt[N]{\lambda(q^N)} : D(0, 1) \to \mathbb{C} \setminus \mu_N$.

A comparison: a Γ-automorphic function $g(\lambda(q)/16)$ versus the hypergeometric example $G(x) = \sum_{n=0}^{\infty} \binom{2n}{n}^2 x^n$

The former has a $\mathbb{Z}/N$ local monodromy (Puiseux branching) for some N. Then $g(F_N^N(z)/16)$ converges on the full unit disc $|z| < 1$ by Cauchy's analyticity theorem: for it satisfies a linear ODE with analytic coefficients and no singularities on that complex disc. *Overconvergence comes by resolving the $x = 0$ singularity via a suitable $x \mapsto x^N$ isogeny.*

The latter has an infinite (and unipotent) local monodromy at $x = 0$. Now $G(F_N^N(z)/16)$ converges only up to the "first" nonzero fiber point $F_N^{-1}\{0\} \setminus \{0\}$, giving a certain radius rather smaller than 1.

The holonomy theorem

Following our $x \mapsto x^N$ isogeny change, we switch our ongoing convention to $x = x(q) := \sqrt[N]{\lambda(q^N)/16} : D(0,1) \to U := \mathbb{C} \setminus 16^{-1/N}\mu_N$.

Theorem

Let $x(t) = t + \cdots \in \mathbb{Q}[[t]]$ be such that $x(t)^N \in \mathbb{Z}[[t]]$. Let $\sigma \in \mathbb{N} \cup \{0\}$ (in our application, $\sigma = 0$). Fix the holomorphic mapping $\varphi : \overline{D(0,1)} \to U$ with $\varphi(0) = 0$ and $|\varphi'(0)| > e^{\sigma}$. Then, the totality of formal functions $f(x) \in \mathbb{Q}[[x]]$ that

- fulfill a linear ODE over $\mathbb{Q}(x)$ without singularities on U, and
- have t-expansions $f(x(t)) = \sum \frac{a_n}{[1,\ldots,n]^{\sigma}} t^n$ with all $a_n \in \mathbb{Z}$,

span over $\mathbb{Q}(x^N)$ a finite-dimensional vector space of dimension at most

$$e \cdot \frac{\int_{|z|=1} \log^{+} |\varphi^N| \, \mu_{\mathrm{Haar}}}{\log |\varphi'(0)| - \sigma}.$$

($e = 2.71\ldots$ is Euler's constant)

The holonomy theorem: proof idea

It follows a method of André, itself going back to D. & G. Chudnovsky in their Diophantine approximations proof of the Faltings isogeny theorem for elliptic curves over $\mathbb{Q}$. A crucial new twist (obviously inspired by Thue–Siegel–Schneider–Roth) is to let the number of auxiliary variables $\mathbf{x} := (x_1, \ldots, x_d)$ to $d \to \infty$.

Suppose effectively there are m such functions $f_1(x), \ldots, f_m(x) \in \mathbb{Q}[[x]]$ linearly independent over $\mathbb{Q}(x^N)$. We use the m^d split variables univariate products $\prod_{s=1}^{d} f_{i_s}(x_s)$ and Siegel's lemma to create an auxiliary function of the form:

$$F(\mathbf{x}) = \sum_{\substack{\mathbf{i} \in \{1,\ldots,m\}^d \\ \mathbf{k} \in \{0,\ldots,D-1\}^d}} a_{\mathbf{i},\mathbf{k}} \, \mathbf{x}^{N\mathbf{k}} \prod_{s=1}^{d} f_{i_s}(x_s) \in (\mathbf{x})^{\alpha} \, \mathbb{Q}[[\mathbf{x}]] \setminus \{0\},$$

with sub-exponentially small coefficients $a_{\mathbf{i},\mathbf{k}} = \exp(o(\alpha))$ as firstly $\alpha \to \infty$ and secondly $d \to \infty$. *With a degree D as low as possible.*

Siegel's lemma: the parameter count

$$F(\mathbf{x}) = \sum_{\substack{\mathbf{i}\in\{1,\ldots,m\}^d \\ \mathbf{k}\in\{0,\ldots,D-1\}^d}} a_{\mathbf{i},\mathbf{k}} \, \mathbf{x}^{N\mathbf{k}} \prod_{s=1}^{d} f_{i_s}(x_s) \in (\mathbf{x})^{\alpha}\,\mathbb{Q}[[\mathbf{x}]] \setminus \{0\},$$

- $(mD)^d$ free parameters $a_{\mathbf{i},\mathbf{k}}$
- $\binom{\alpha+d}{d} \sim \alpha^d/d! \approx (e\alpha/d)^d$ equations to solve
- $\#\texttt{parameters} > \#\texttt{equations}$ if $dD > e(1+o(1))\frac{\alpha}{m}$ asymptotically
- by letting also $d \to \infty$, we can also make sure the Dirichlet exponent $\to 0$, and the coefficients a are $\exp(o(\alpha))$
- then we can asymptotically take the degree parameter $dD \sim e\,\frac{\alpha}{m}$.

The extrapolation, assuming (our case at hand) $\sigma = 0$

The idea is that the function $G(\mathbf{z}) := F(\varphi(\mathbf{z})) \in \mathbb{C}[[\mathbf{z}]]$ is analytic on $\overline{D(0,1)}$ (by Cauchy's theorem), and yet since $\varphi(z) = \varphi'(0)z + \cdots$, it also inherits from $f_i(x(t)) \in \mathbb{Z}[[t]]$ an integrality property of its *lexicographically lowest* term $c\,\mathbf{z}^{\boldsymbol{\beta}}$:

- $c \in \varphi'(0)^{|\boldsymbol{\beta}|}\,\mathbb{Z} \setminus \{0\}$, with total degree $|\boldsymbol{\beta}| \geq \alpha$
- hence the Liouville lower bound for that coefficient: $\log|c| \geq \alpha \log|\varphi'(0)|$
- (A simplification step pointed out to us by André) We can use the plurisubharmonic property of $\log|\mathrm{holomorphic\ function}|$ together with an easy induction scheme on d to prove that, *for our lexicographically lowest monomial* $c\,\mathbf{z}^{\boldsymbol{\beta}}$, we have a bound in the other direction:

$$\log|c| \leq \int_{\mathbf{T}^d} \log|F|\,\mu_{\mathrm{Haar}}.$$

The base case $d = 1$ is simply the subharmonic property of $\log|z^{-\beta}F(z)|$.

The holonomy rank bound: proof completion

- $\alpha \log|\varphi'(0)| \leq \int_{\mathbf{T}^d} \log|F|\, \mu_{\mathrm{Haar}}$
- the RHS is upper estimated by our arithmetic information from the shape of F and the asymptotically subexponential coefficients bound in Siegel's lemma:

$$\alpha \log|\varphi'(0)| \leq \int_{\mathbf{T}^d} \log|F|\, \mu_{\mathrm{Haar}} \leq dD \int_{\mathbf{T}} \log^+|\varphi^N|\, \mu_{\mathrm{Haar}} + o(\alpha)$$

- With the degree parameter asymptotic estimate $dD \sim e\alpha/m$, the last inequality amounts in the $\alpha \to \infty$, $d \to \infty$ limit to

$$m \leq e \frac{\int_{\mathbf{T}} \log^+|\varphi^N|\, \mu_{\mathrm{Haar}}}{\log|\varphi'(0)|},$$

that is precisely what we aimed to prove.

The effectivization problem

A general purpose procedure to find the full solution space (all functions) in this holonomy theorem?

We only gave an upper bound on the total number of independent functions. This situation is not unlike the notorious non-effectivity issue with Roth's and Faltings's theorems. It would be of a great value for irrationality or linear independence proofs involving certain L-values (both real and p-adic) to be able to compute the solution space effectively.

The UBD turns out one of the few cases where we are able to (ultimately at the end of this proof) determine the entire solution space exactly: the congruence modular functions.

Another simple case of an effective determination, with $k = 1$: the formal power series $\sum a_n x^n / [1, \ldots, n]$ with all $a_n \in \mathbb{Z}$ and satisfying a linear ODE on $\mathbb{P}^1 \setminus \{0, 1, \infty\}$ span the two-dimensional $\mathbb{Q}(x)$-linear space with basis 1 and $\log(1 - x)$.

What have we got so far

We use the preceding with the choices

$$t := q^{1/N} = e^{\pi i \tau/N}, \quad x(t) := \sqrt[N]{\lambda(t^N)/16}, \quad U := \mathbb{C} \setminus 16^{-1/N}\mu_N,$$
$$\varphi(z) := 16^{-1/N} F_N(rz) \quad : \quad \overline{D(0,1)} \to U$$

for $r := 1 - 1/(2N^3)$.

Conclusion: the modular functions (weight $k = 0$, both congruence and noncongruence) that have

- $\mathbb{Z}[[q^{1/N}]]$ *Fourier expansions at the one cusp $i\infty$, and*
- *cusp widths dividing N at all the cusps,*

span over $\mathbb{C}(\lambda) = \mathbb{C}(x^N)$ a vector space of dimension at most

$$e \cdot \frac{\int_{|z|=1-1/(2N^3)} \log^+ |F_N^N| \, \mu_{\mathrm{Haar}}}{\log |16^{-1/N} F_N'(0)| + \log r} \ll N^3 \int_{|z|=1-1/(2N^3)} \log^+ |F_N^N| \, \mu_{\mathrm{Haar}}.$$

The $\mathbb{Z}[[q^{1/N}]]$ modular functions with Wohlfahrt level N

By Hecke theory, these include all the $\Gamma(N)$-automorphic functions. Their $\mathbb{C}(\lambda)$-linear span dimension (say N is even) equates to the index formula

$$\frac{1}{2}[\Gamma(2) : \Gamma(N)] = \frac{N^3}{2[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(2)]} \prod_{p|N} \left(1 - \frac{1}{p^2}\right) > \frac{N^3}{12\zeta(2)}.$$

Thus we have these true solutions as an $\gg N^3$ lower bound against the

$$\ll N^3 \int_{|z|=1-1/(2N^3)} \log^+ |F_N^N| \, \mu_{\mathrm{Haar}}$$

upper bound that we just proved.

Extrapolation from a single counterexample $f(q)$

But suppose we have even a single noncongruence counterexample $f(q) \in \mathbb{Z}[[q]]$, of Wohlfahrt (LCM of cusp widths) level N. *Then $f(q^p) \in \mathbb{Z}[[q]]$ is another counterexample at a Wohlfahrt level Np.*

An idea going back to Serre from his proof of the triviality of the Hecke operators over noncongruence subgroups — based on an amalgamated sum presentation of $\mathrm{SL}_2(\mathbb{Z}[1/p])$, and on the congruence subgroup property of that S-arithmetic group — proves that this construction is independent over the *congruence* modular forms. And thus by this construction out of a single counterexample at Wohlfahrt level N we reach as many as $2^{\pi(X)}$ *independent* counterexamples at Wohlfahrt level $N \prod_{p<X} p$.

Matching up

Hence, at the Wohlfahrt level $M := N\prod_{p<X} p \asymp_N e^{X+o(X)}$, *we have*

$$\gg M^3 2^{\pi(X)} \gg M^3 2^{X/(2\log X)}$$

examples against our upper bound of

$$\ll M^3 \int_{|z|=1-1/(2M^3)} \log^+ |F_M^M| \, \mu_{\mathrm{Haar}} \ll M^3 X$$

Now $X \sim \log M$ as we let the parameter $X \to \infty$, and so to get the desired constradiction out of a single counterexample $f(q)$, it remains to prove that the integral is sub-exponentially small in $\log M / \log\log M$.

In fact we prove that the integral is $O(\log M)$. *(But an* $\ll_\epsilon M^\epsilon$ *bound would not have sufficed.)*

The mean growth of the universal covering map F_N

And this was how in our *arithmetico-analytic continuation* (Gel'fond) argument with Diophantine approximations we came to require a precise — doubly uniform in both $N \geq 2$ and $r < 1$ — upper estimate on the *Nevanlinna mean proximity function at* ∞:

$$m(r, F_N) := \int_{|z|=r} \log^+ |F_N| \, \mu_{\mathrm{Haar}} \ll \frac{1}{N-1} \log \frac{N}{1-r}.$$

The general fact of the matter is: for *any* universal covering map $F : D(0,1) \to \mathbb{C} \setminus \{a_1, \ldots, a_N\}$, we have (Tsuji, 1952) the precise asymptotics as $r \to 1^-$:

$$m(r, F) = \frac{1}{N-1} \log \frac{1}{1-r} + O_{a_1, \ldots, a_N; F(0)}(1).$$

To be contrasted with the crude supremum growth asymptotic formula:

$$\sup_{|z|=r} \log |F| \asymp \frac{1}{1-r}.$$

$m(r, F_N^N) := \int_{|z|=r} \log^+ |F_N^N| \, \mu_{\mathrm{Haar}} \ll \log \frac{N}{1-r}$

- Heuristically this is plausible upon comparing to the renormalized function $F_N(q^{1/N})^N \to \lambda(q)$, the convergence taking place as q-expansions under $N \to \infty$ on any fixed disc of radius $r < 1$. But this convergence is not uniform as $r \to 1$, whereas we will need to take $r = 1 - 1/(2N^3)$.
- The growth of the map F_N is governed by the growth of the cusps of a bounded height of the (N, ∞, ∞) triangle (Fuchsian) group. The above convergence tempts us to compare these cusps (which are to some extent explicit, but of course they vary with N) to the cusps of the limit (∞, ∞, ∞) triangle group $\Gamma(2)$.
- Turns out quite hard! Although in this way we could compute a precise upper estimate on the conformal radii of the sup-level sets $|F_N| < e^M$, that turns out not sufficiently precise for our arithmetico-analytic continuation argument. The requisite mean (integrated) bound above can be translated in terms of a uniform cusp count, but the latter goes beneath what we could directly prove.

Enter Nevanlinna's lemma on the logarithmic derivative

Instead, we were surprised to find that inner workings of the (general, abstract) *second main theorem* of Nevanlinna's value distribution theory exactly sufficed to give the double uniformity that we needed:

Theorem

$m(r, F_N^N) := \int_{|z|=r} \log^+ |F_N^N| \, \mu_{\mathrm{Haar}} \ll \log \frac{N}{1-r}$ *(an absolute and computable explicit coefficient)*

Idea: By construction, the function $f := 1 - F_N^N$ is a *functional unit*, and we have a factorization in terms of logarithmic derivatives:

$$\frac{F_N^N}{F_N^N - 1} = \frac{F_N}{NF_N'} \frac{f'}{f} = N^{-1} \frac{1 - F_N}{(1 - F_N)'} \frac{f'}{f}$$

This formula is what gets specially used about the target set $\mu_N \cup \{\infty\}$ *of omitted values.*

The LHS has mean proximity function $= N\, m(r, F_N) + O(1)$, so it is what we want to bound. The RHS is $m(r, F_N) + \texttt{small}$, essentially because logarithmic derivatives are small on average over circles. (Lemma on the logarithmic derivative.)

A simplified, explicit case of the lemma of the logarithmic derivative suffices for our purposes

Apply the lemma on the logarithmetic derivative to the two functional units $g := 1 - F_N^N$ and $g := 1 - F_N$, followed by a standard transformation sequence from the lemma on the logarithmic derivative $\to$ Nevanlinna's second main theorem, and a trivial supremum estimation of the emerging *double* log growth term in this bound:

Lemma

Let $g : \overline{D(0,R)} \to \mathbb{C}^\times$ *be a nowhere vanishing holomorphic function on some open neighborhood of the closed disc* $|z| \leq R$*. Assume that* $g(0) = 1$*. Then, for all* $0 < r < R$*,*

$$m\left(r, \frac{g'}{g}\right) < \log^+ \left\{ \frac{m(R,g)}{r} \frac{R}{R-r} \right\} + \log 2 + 1/e.$$

Quintessence of the proof of this form of the lemma of the logarithmic derivative

It is based on Poisson's kernel, which by a simple differentiation in term yields the reproducing formula for our logarithmic derivative:

$$\frac{g'(z)}{g(z)} = \int_{|w|=R} \frac{2w}{(w-z)^2} \log|g(w)|\, \mu_{\mathrm{Haar}}(w), \quad \forall z \in D(0,R).$$

Follow it by easy estimations based on Jensen's inequality and the concavity of $\log^+|x|$ on $[1,\infty)$.

To complete the general picture: For an arbitrary meromorphic mapping $g : D(0,1) \to \mathbb{P}^1$ with $g(0) = 1$ (now possibly having a nonempty divisor of zeros and poles), if we use the full *Nevanlinna characteristic* $T(r,g) = m(r,g) + N(r,g)$, Gol'dberg and Grinshtein proved that the same type of bound persists:

$$m\Big(r, \frac{g'}{g}\Big) < \log^+ \Big\{ \frac{T(R,g)}{r} \frac{R}{R-r} \Big\} + 5.8501.$$

Essentially best-posible in form, and comparable to Lang's conjecture on the error term in Roth's theorem under the Osgood-Vojta dictionary to Diophantine approximation.

Conclusion of the doubly uniform mean growth bound on F_N

By the factorization identity and the lemma on the logarithmic derivative we got, uniformly $\forall 0 < r < R$ and $\forall N \geq 2$, to

$$N\, m(r, F_N) \leq m(r, F_N) + O\Big(\log^+ \Big\{ N \frac{m(R, F_N)}{r} \frac{R}{R-r} \Big\} \Big).$$

Choose $R = (1+r)/2$.
Since the $\log^+ m(R, F_N)$ erro term is *logarithmic*, it is alright to estimate it crudely by the supremum

$$\log^+ \sup_{|z|=R} \log |F_N| \ll \log \frac{N}{1-R},$$

by a simple geometric estimate based on Shimizu's lemma.

A (naive) question of the Lehmer variety

Let

$$f(q) = \sum_{r\in\mathbb{Q}^{\geq 0}} \frac{a(r)}{b(r)} q^r \in \mathbb{Q}[[q^{1/N}]], \quad \gcd\big(a(n), b(n)\big) = 1$$

be the q-expansion of a noncongruence modular form. Having proved the unboundedness $\limsup_{r\to\infty} b(r) \to \infty$, it becomes a natural question how slowly may these denominators grow in terms of the rate

$$\delta(f) := \limsup_{r\to\infty} \frac{1}{r} \log |b(n)|.$$

A (naive) question of the Lehmer variety

$$\delta(f) := \limsup_{r\to\infty} \frac{1}{r} \log |b(n)|.$$

In our discussion, it is natural to measure the growth with respect to the parameter $N = N(f) :=$ the LCM of the cusp widths of f in the $\Gamma(2)$-orbit of the infinite cusp $i\infty$.

Observe that the product $N(f) \cdot \delta(f)$ remains invariant under changing $q \mapsto q^m$, $\forall m \in \mathbb{N}$.

Hence the theoretically best-possible lower growth lower boun would look like $\delta(f) \gg N(f)^{-1}$ (unless f is congruence). *Is that bound possibly true?*

To compare: tracking our $f(q) \mapsto f(q^p)$ extrapolation argument, one can easily make everything quantitative (how large the parameter $X \to \infty$ really needed to be in terms of the initial N, etc.), and prove the following weaker lower bound on denominators growth of noncongruence modular forms: $\delta(f) \gg N(f)^{-3-o(1)}$.

Back to our initial question: can we still say something if the ODE has a $\log x$ branching?

Problem

Is there an exact description of all the $\mathbb{Z}[[x]]$ formal solutions to linear ODEs $L(f) = 0$ where L ranges over all linear differential operators with coefficients in $\mathbb{C}(x)$ and no singularities outside of $0, 1/16$ and ∞?

Partial answer:

- Yes, if the $x = 0$ local monodromy is semisimple: the solutions then correspond precisely to the congruence modular functions, via $\tau \mapsto f\big(\lambda(\tau)/16\big)$.
- More generally, if A is the local monodromy operator at $x = 1/16$, the answer is the same under the (generally) weaker condition that f and $A(f)$ are simultaneously holomorphic at $x = 0$.
- In general, the examples include at least all the congruence modular forms of *any* weight, expanded out formally in terms of $x = \lambda/16$. Are these all the examples?

Final slide!

Thank you for your invitation and attention, everyone stay safe and well!