Séminaire BOURBAKI Novembre 2022
74e année, 2022–2023, no 1197, p. 1 à 27
doi: 10.24033/ast.1206

THE UNBOUNDED DENOMINATORS CONJECTURE
[*after* F. Calegari, V. Dimitrov, and Y. Tang]

by **Javier Fresán**

Introduction

This written account of my talk at the Bourbaki seminar surveys some of the ideas in the beautiful proof by Calegari, Dimitrov, and Tang (2021) of the unbounded denominators conjecture, a long standing open problem in the theory of modular forms that gives a simple criterion to decide whether a modular form with algebraic Fourier coefficients at infinity is "invariant" under a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$ or not.

Throughout, we write $\overline{\mathbf{Q}}$ for the algebraic closure of $\mathbf{Q}$ in $\mathbf{C}$ and $\overline{\mathbf{Z}} \subset \overline{\mathbf{Q}}$ for the subring of algebraic integers. We let $\mathfrak{H} = \{\tau \in \mathbf{C} \mid \mathrm{Im}(\tau) > 0\}$ denote the upper half-plane and[(1)] $q = \exp(\pi i \tau)$. Recall that $\mathrm{SL}_2(\mathbf{Z})$ acts on $\mathfrak{H}^* = \mathfrak{H} \cup \mathbb{P}^1(\mathbf{Q})$ by Möbius transformations and that *congruence subgroups* of $\mathrm{SL}_2(\mathbf{Z})$ are those containing

$$\Gamma(M) = \ker\big(\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/M\mathbf{Z})\big) = \{A \in \mathrm{SL}_2(\mathbf{Z}) \mid A \equiv I \bmod M\}$$

for some integer $M \geqslant 1$. The *unbounded denominators conjecture*, originating from work of Atkin and Swinnerton-Dyer (1971), is now the following theorem:

Theorem 0.1 (Calegari–Dimitrov–Tang, 2021). *Let $f(\tau)$ be a holomorphic function on the upper-half plane $\mathfrak{H}$ such that*

(*a*) *there exists an integer k and a subgroup $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ of finite index such that*

$$f\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^k f(\tau) \tag{1}$$

holds for all matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in Γ;

(*b*) *f locally extends to a meromorphic function around each point of $\mathbb{P}^1(\mathbf{Q})$;*

(*c*) *f admits a Fourier expansion in $\overline{\mathbf{Z}}[\![q^{1/N}]\!]$ for some integer $N \geqslant 1$.*

Then the equality (1) *holds for all matrices in a congruence subgroup of* $\mathrm{SL}_2(\mathbf{Z})$.

[(1)] One reason for choosing this unusual convention for q will be explained in remark 0.2 below.

©Astérisque 446, SMF 2023

In what follows, we will refer to functions f satisfying assumptions (a) and (b) simply as *modular forms* of weight k, or *modular functions* if $k = 0$, for the group Γ. Note that every subgroup of finite index of $\mathrm{SL}_2(\mathbf{Z})$ contains a matrix of the shape $\left(\begin{smallmatrix} 1 & 2N \\ 0 & 1 \end{smallmatrix}\right)$ for some integer $N \geqslant 1$. More precisely, the *width* of each cusp $\zeta \in \mathbb{P}^1(\mathbf{Q})$ is defined as the smallest integer $m_\zeta \geqslant 1$ such that the stabiliser of ζ under the action of Γ on $\mathbb{P}^1(\mathbf{Q})$ contains, up to conjugation in $\mathrm{SL}_2(\mathbf{Z})$, one of the matrices $\pm \begin{pmatrix} 1 & m_\gamma \\ 0 & 1 \end{pmatrix}$. The assumption that f has a Fourier expansion in $\overline{\mathbf{Z}}[\![q^{1/N}]\!]$ implies that the width of the cusp at infinity divides $2N$. For $k = 0$, in the conclusion of the theorem we can take a congruence subgroup containing $\Gamma(L(\Gamma))$, where $L(\Gamma)$ stands for the lowest commun multiple of the widths of all cusps, a generalisation of the notion of level for non-congruence subgroups.

Let us explain the name of the conjecture. If the coefficients of $f \in \overline{\mathbf{Q}}[\![q^{1/N}]\!]$ have *bounded denominators*, which amounts to saying that f lies in the subspace

$$\overline{\mathbf{Z}}[\![q^{1/N}]\!] \otimes_{\overline{\mathbf{Z}}} \overline{\mathbf{Q}} = \mathbf{Z}[\![q^{1/N}]\!] \otimes_{\mathbf{Z}} \overline{\mathbf{Q}} \subset \overline{\mathbf{Q}}[\![q^{1/N}]\!],$$

then we can apply theorem 0.1 to an integral multiple of f. The contrapositive statement then says the following:

> *Let $f(\tau)$ be a modular form for a subgroup of finite index of* $\mathrm{SL}_2(\mathbf{Z})$ *with a Fourier expansion in $\overline{\mathbf{Q}}[\![q^{1/N}]\!]$. If f is not modular for any congruence subgroup, then the Fourier coefficients of f have unbounded denominators.*

By contrast, all modular forms f for congruence subgroups have bounded denominators by the theory of Hecke operators (Shimura, 1971, Theorem 3.52). In a nutshell, after multiplying f by a large enough power of the modular discriminant to turn it into a holomorphic cusp form, we can write it as a linear combination of Hecke eigenforms, and the Fourier coefficients of those are algebraic integers since they are polynomial expressions with integer coefficients in the Hecke eigenvalues(2). Thus, the condition of having bounded denominators completely distinguishes congruence and non-congruence modular forms among all modular forms with algebraic Fourier coefficients at infinity.

By a theorem of Mennicke (1965) and Bass, Lazard, and Serre (1964), the group $\mathrm{SL}_n(\mathbf{Z})$ has the *congruence subgroup property* for each $n \geqslant 3$, meaning that all its subgroups of finite index contain a congruence subgroup. The same is true for other arithmetic groups such as $\mathrm{SL}_2(\mathbf{Z}[1/p])$ for each prime number p. Most subgroups of finite index of $\mathrm{SL}_2(\mathbf{Z})$, however, are *not* congruence. For example, given an integer $g \geqslant 0$, there is only a finite number of congruence subgroups Γ such that the

(2) This argument fails for non-congruence modular forms. Although there is still a way to define Hecke operators, their action is trivial on those forms that do not come from the smallest congruence subgroup containing Γ by results of Serre, presented in Thompson (1989), and Berger (1994).

curve $X(\Gamma) = \mathfrak{H}^*/\Gamma$ has genus g (Dennin, 1975), whereas there is an infinite number of non-congruence subgroups with the same property (Jones, 1979). Some explicit examples of non-congruence subgroups will be given in section 0.2 below.

One reason to care about modular forms for non-congruence subgroups is *Belyi's theorem*, according to which every smooth projective curve defined over $\overline{\mathbf{Q}}$ can be realised as a cover of the projective line $\mathbb{P}^1$ that is only ramified at $0, 1, \infty$ (such coverings are often called *Belyi maps*). Taking the isomorphism $\mathfrak{H}/\Gamma(2) \simeq \mathbb{P}^1 \setminus \{0, 1, \infty\}$ given by the modular lambda function into account, any such curve is hence isomorphic to $X(\Gamma)$ for a subgroup $\Gamma \subset \Gamma(2)$ of finite index. As we will see below, theorem 0.1 provides us with a criterion to decide whether Γ is a congruence subgroup or not in terms of the integrality properties of the associated Belyi map.

0.1. First reductions

It will be enough to prove the theorem under the assumption that f is a modular *function* with *integer* coefficients. We first explain the reduction to the case $k = 0$. For this, consider the q-series expansions

$$\begin{aligned} \frac{\lambda(\tau)}{16} &= q \prod_{n=1}^{\infty} \left(\frac{1 + q^{2n}}{1 + q^{2n-1}} \right)^8 = q - 8q^2 + 44q^3 - \cdots, \\ \eta\left(\tfrac{\tau}{2}\right)^2 &= q^{1/12} \prod_{n=1}^{\infty} (1 - q^n)^2 = q^{1/12} - 2q^{13/12} - q^{25/12} + \cdots, \end{aligned} \tag{2}$$

which define a modular function for the group $\Gamma(2)$ and a modular form of weight 1 for $\Gamma(12)$ respectively. The first one induces an isomorphism[3]

$$\mathfrak{H}/\Gamma(2) \xrightarrow{\sim} \mathbb{P}^1 \setminus \{0, 1/16, \infty\}$$

that extends to a map sending the cusp at infinity to 0. The second one, a 12th root of the modular discriminant $\Delta(\tau/2)$, does not vanish on the upper half-plane and has the property that its inverse has integer Fourier coefficients at infinity. Therefore,

$$F(\tau) = \left(\frac{\lambda(\tau)}{16} \right)^k \frac{f(\tau)}{\eta\left(\frac{\tau}{2}\right)^{2k}} \in \overline{\mathbf{Z}}[\![q^{1/N}]\!]$$

satisfies the assumptions of theorem 0.1 with the weight $k = 0$ and the subgroup $\Gamma \cap \Gamma(12)$ of $\mathrm{SL}_2(\mathbf{Z})$. If F is a modular function for a congruence subgroup, then f is a modular form for a congruence subgroup. Note that the first factor is there to kill the pole at $q = 0$ introduced by η, thus keeping the condition that $f(\tau)$ is holomorphic at infinity. This operation could, however, create new poles at other cusps; this explains the lack of symmetry between infinity and the other cusps in the statement of the theorem.

[3] One says that λ is a *Hauptmodul* for $\Gamma(2)$.

Remark 0.2. One explanation for the normalisations $x = \lambda/16$ and $q = \exp(\pi i \tau)$ is that they allow for the identity $\mathbf{Z}[\![q]\!] = \mathbf{Z}[\![x]\!]$ coming from the expressions

$$x = q - 8q^2 + 44q^3 + \cdots, \quad q = x + 8x^2 + 84x^3 + \cdots$$

of x and q as power series with *integer* coefficients in q and x respectively.

Let us now explain how to reduce to the case $f \in \mathbf{Z}[\![q^{1/N}]\!]$ following a suggestion of John Voight (Calegari, Dimitrov, and Tang, 2021, Remark 6.3.2). Let Γ be a finite index subgroup of $\mathrm{SL}_2(\mathbf{Z})$. By Belyi's theorem, the curve $X(\Gamma)$, its cusp at infinity, the uniformiser $q^{1/N}$, and the covering $X(\Gamma) \to \mathbb{P}^1$ are defined over some number field K. Moreover, the algebro-geometric interpretation of modular functions as rational functions on the curve $X(\Gamma)$ shows that they carry a natural structure of a K-vector space, corresponding to those modular functions whose q-expansion at infinity has coefficients in K. After enlarging K to its Galois closure if necessary, an element σ of the Galois group $\mathrm{Gal}(K/\mathbf{Q})$ transforms the covering $X(\Gamma) \to \mathbb{P}^1$ into a covering $X(\Gamma_\sigma) \to \mathbb{P}^1$ for possibly another subgroup Γ_σ of finite index, that we may conjugate so that the cusp at infinity maps again to ∞. Since the Galois action on q-expansions is given by applying σ to the coefficients, the conjugate of a modular function will still be modular for a subgroup of finite index. Now, the modularity assumption on $f \in \overline{\mathbf{Z}}[\![q^{1/N}]\!]$ implies that there exists a number field L, with ring of integers $\mathscr{O}_L$, such that f lies in $\mathscr{O}_L[\![q^{1/N}]\!]$. If $\alpha_1, \ldots, \alpha_d$ is a $\mathbf{Z}$-basis of $\mathscr{O}_L$, then $f_i(\tau) = \mathrm{Tr}_{L/\mathbf{Q}}(\alpha_i f(\tau))$ lies in $\mathbf{Z}[\![q^{1/N}]\!]$ and is still modular for a finite index subgroup of $\mathrm{SL}_2(\mathbf{Z})$ by the above. By the special case of theorem 0.1 in which the function is assumed to have integer coefficients, each of these functions is modular for a congruence subgroup Γ_i, so f is modular for $\Gamma_1 \cap \cdots \cap \Gamma_d$.

To summarise, we are reduced to proving the following statement:

Theorem 0.3. *Let $N \geqslant 1$ be an integer and let $f(\tau) \in \mathbf{Z}[\![q^{1/N}]\!]$ be a holomorphic function on $\mathfrak{H}$ that locally extends to a meromorphic function around each point of $\mathbb{P}^1(\mathbf{Q})$ and is invariant under the action of a subgroup $\Gamma \subset \Gamma(2)$ of finite index. Then f is a modular function for a congruence subgroup.*

0.2. An interpretation in terms of Belyi maps

In the notation of theorem 0.3, let $Y(\Gamma) = \mathfrak{H}/\Gamma$ and consider the diagram

$$\begin{tikzcd} Y(\Gamma) \arrow[r, "f"] \arrow[d, "\pi"'] & \mathbf{C} \\ Y(2) \simeq \mathbb{P}^1 \setminus \{0, 1/16, \infty\} \arrow[ur, dotted] & \end{tikzcd}$$

where π is an étale cover and $Y(2)$ and $\mathbb{P}^1 \setminus \{0, 1/16, \infty\}$ are identified through the isomorphism $\lambda/16$. We can then think of f as a multivalued *algebraic* function of the variable

$\lambda/16$ ramified at the points $0, 1/16, \infty$, as indicated by the dotted arrow. By expanding it as a Puiseux series at a branch above 0, the theorem can be rephrased as saying that

f lies in $\mathbf{Z}[\![\frac{\lambda(\tau/m)}{16}]\!] \otimes \mathbf{C}$ for some $m \geqslant 1$ if and only if Γ is a congruence subgroup.

Example 0.4 (Fermat curves). Let $n \geqslant 1$ be an integer and consider the Fermat curve X_n with affine equation $x^n + y^n = 1$. Since the modular lambda function λ does not take the values 0 and 1, there exist holomorphic functions $x, y\colon \mathfrak{H} \to \mathbf{C}$ satisfying

$$x(\tau)^n = \lambda(\tau) \quad \text{and} \quad y(\tau)^n = 1 - \lambda(\tau).$$

The diagonal arrow in the diagram

$$\begin{tikzcd}
\mathfrak{H} \arrow[d, "\lambda"'] \arrow[dr, "\tau\mapsto(x(\tau),y(\tau))"] & \\
\mathbb{P}^1 \setminus \{0,1,\infty\} & X_n \arrow[l, "(x,y)\mapsto x"]
\end{tikzcd}$$

factors through an isomorphism $\mathfrak{H}/\Phi(n) \simeq X_n$, where the *Fermat group* $\Phi(n)$ is defined as the kernel of the composition $\Gamma(2) \to \Gamma(2)^{\mathrm{ab}} \to \Gamma(2)^{\mathrm{ab}}/n$. Explicitly, $\Phi(n)$ is generated by the n-th powers of the matrices $A = \left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right)$, and by the commutator $[\Delta, \Delta]$ of the subgroup $\Delta = \langle A, B\rangle = \Gamma(2)/\{\pm I\}$ that they generate. It is a classical result of Klein and Fricke (2017, page 534) that $\Phi(n)$ is a congruence subgroup if and only if $n \in \{1, 2, 4, 8\}$. This property is reflected by the fact that the modular functions $x(\tau)$ and $y(\tau)$ have unbounded denominators unless n takes one of those values, or yet by the fact that the coefficients of the power series[4]

$$\sqrt[n]{1-x} = \sum_{m=0}^{\infty} \frac{16^m \left(\frac{-1}{n}\right)_m}{m!} \left(\frac{x}{16}\right)^m \in \mathbf{Q}\left[\!\!\left[\frac{x}{16}\right]\!\!\right]$$

have bounded denominators if and only if $n \in \{1, 2, 4, 8\}$, in which case they are all integers. Indeed, writing the m-th coefficient as

$$a_m = (-16)^m \frac{(n-1)(2n-1)\cdots[(m-1)n+1]}{n^m m!},$$

we see that, for each odd prime number p dividing n, the p-adic valuation $v_p(a_m)$ is smaller than $-v_p(m!)$, which tends to $-\infty$ as $m \to +\infty$. If 2 divides n, then $v_2(a_m)$ is equal to $4m - mv_2(n) - v_2(m!)$, so that again it tends to $-\infty$ as soon as $v_2(n) \geqslant 4$ but is non-negative for $n \in \{2, 4, 8\}$ since $v_2(m!) = \sum_{k=1}^{\infty} \lfloor m/2^k \rfloor \leqslant m$. Finally, $v_p(a_m) \geqslant 0$ for all primes p not dividing n, as can be seen by choosing $r \geqslant v_p(m!)$ and replacing the 1s in the numerator of a_m with $1 = un + vp^r$ for some integers u, v.

[4] Here, $(\alpha)_m = \alpha(\alpha+1)\cdots(\alpha+m-1)$ denotes the Pochhammer symbol of a complex number α.