THE UNBOUNDED DENOMINATORS CONJECTURE

FRANK CALEGARI, VESSELIN DIMITROV, AND YUNQING TANG

Abstract. We prove the unbounded denominators conjecture in the theory of noncongruence modular forms for finite index subgroups of $\mathrm{SL}_2(\mathbf{Z})$. Our result includes also Mason's generalization of the original conjecture to the setting of vector-valued modular forms, thereby supplying a new path to the congruence property in rational conformal field theory. The proof involves a new arithmetic holonomicity bound of a potential-theoretic flavor, together with Nevanlinna's second main theorem, the congruence subgroup property of $\mathrm{SL}_2(\mathbf{Z}[1/p])$, and a close description of the Fuchsian uniformization $D(0,1)/\Gamma_N$ of the Riemann surface $\mathbf{C} \smallsetminus \mu_N$.

Contents

1. Introduction

We prove the following:

Theorem 1.0.1 (Unbounded Denominators Conjecture)**.** *Let N be any positive integer, and let $f(\tau) \in \mathbf{Z}[\![q^{1/N}]\!]$ for $q = \exp(\pi i \tau)$ be a holomorphic function on the upper half plane. Suppose there*

F.C. was supported in part by NSF Grant DMS-2001097.

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

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

and suppose that $f(\tau)$ is meromorphic at the cusps, that is, locally extends to a meromorphic function near every cusp in the compactification of $\mathbf{H}/\Gamma$. Then $f(\tau)$ is a modular form for a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$.

The contrapositive of this statement is equivalent to the following, which explains the name of the conjecture: if $f(\tau) \in \mathbf{Q}[\![q^{1/N}]\!]$ is a modular form which is *not* modular for some congruence subgroup, then the coefficients of $f(\tau)$ have *unbounded denominators*. The corresponding statement remains true if one replaces $\mathbf{Q}$ by any number field (see Remark 6.3.1).

Let $\lambda(\tau)$ be the modular lambda function (Legendre's parameter):

$$\frac{\lambda(\tau)}{16} = \left(\frac{\eta(\tau/2)\eta(2\tau)^2}{\eta(\tau)^3}\right)^8 = q\prod_{n=1}^{\infty}\left(\frac{1+q^{2n}}{1+q^{2n-1}}\right)^8 = q - 8q^2 + \cdots \tag{1.0.2}$$

with $q = e^{\pi i\tau}$ and $\eta(\tau/2) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n)$. (Historic conventions force one to use q for both $e^{\pi i\tau}$ and $e^{2\pi i\tau}$ — we use the first choice unless we expressly state otherwise.) On replacing the weight k form f by the weight zero form $f(\tau)(\lambda(\tau)/16\eta(\tau/2)^2)^k$, we may (and do) assume that $k = 0$. The function f is then an *algebraic* function of λ, with branching only at the three punctures $\lambda = 0, 1, \infty$ of the modular curve $Y(2) \cong \mathbf{P}^1 \smallsetminus \{0, 1, \infty\}$. Thus another reading of our result states that the Belyĭ maps (étale coverings)

$$\pi : U \to \mathbf{CP}^1 \smallsetminus \{0,1,\infty\} := \operatorname{Spec}\mathbf{C}[\lambda, 1/\lambda, 1/(1-\lambda)] = Y(2)$$

possessing a formal Puiseux branch in $\mathbf{Z}[\![\lambda(\tau/m)/16]\!] \otimes \mathbf{C}$ for some $m \in \mathbf{N}$ are exactly the congruence coverings $Y_\Gamma = \mathbf{H}/\Gamma \to \mathbf{H}/\Gamma(2) = Y(2)$, with Γ ranging over all *congruence* subgroups of $\Gamma(2)$. The reverse implication follows from [Shi71, Theorem 3.52], and reflects the fact that the q-expansions of eigenforms on congruence subgroups are determined by their Hecke eigenvalues (see also [Kat73, §1.2]).

We refer the reader to Atkin and Swinnerton-Dyer [ASD71] for the roots of the unbounded denominators conjecture, and to Birch's article [Bir94] as well as to Long's survey [Lon08, § 5] for an introduction to this problem and its history. For the vector-valued generalization, see §7.3 and its references below. The cases of relevance to the partition and correlation functions of rational conformal field theories (of which the tip of the iceberg is the example (1.0.3) discussed below) were resolved in a string of works [DR18, DLN15, SZ12, NS10, Xu06, Ban03, Zhu96, AM88], by the modular tensor categories method. Some further sporadic cases of the unbounded denominators conjecture have been settled by mostly *ad hoc* means [FF22, FM16b, LL12, KL08, KL09].

To give some simple examples, the integrality property $\sqrt[8]{1-x} \in \mathbf{Z}[\![x/16]\!]$ corresponds to the fact that the modular form $(\lambda/16)^{1/8} = q^{1/8}\prod_{n=1}^{\infty}(1+q^{2n})(1+q^{2n-1})^{-1}$ and the affine Fermat curve $x^8 + y^8 = 1$ are congruence; whereas a simple non-example [Lon08, § 5.5] is the affine Fermat curve $x^n + y^n = 1$ for $n \notin \{1, 2, 4, 8\}$, for which the fact that its Fuchsian group is a noncongruence arithmetic group is detected arithmetically by the calculation $\sqrt[n]{1-x} \notin \mathbf{Z}[\![x/16]\!] \otimes \mathbf{C}$. This recovers a classical theorem of Klein [KF17, page 534]. To include an example related to two-dimensional rational conformal field theories, consider the following function (with $q = e^{2\pi i\tau}$):

$$j(\tau)^{1/3} = q^{-1/3}\,\frac{1 + 240\sum_{n=1}^{\infty}\sigma_3(n)q^n}{\prod_{n=1}^{\infty}(1-q^n)^8} = q^{-1/3}(1 + 248q + 4124q^2 + 34752q^3 + \cdots). \tag{1.0.3}$$

The resulting Fourier coefficients are closely linked to the dimensions of the irreducible representations of the exceptional Lie group $E_8(\mathbf{C})$, and in particular they are integers. To be more precise: the modular function $j^{1/3}$ coincides with the graded dimension of the level one highest-weight representation of the affine Kac–Moody algebra $E_8^{(1)}$; see Gannon's book [Gan06, page 196] for a broad view on this topic and its relation to mathematical physics. The unbounded denominators conjecture (Theorem 1.0.1) now implies that $j^{1/3}$ must be a modular function on a *congruence* subgroup, and indeed one may easily confirm that it is a Hauptmodul for the level 3 subgroup which is the kernel of the composite $\mathrm{PSL}_2(\mathbf{Z}) \to \mathrm{PSL}_2(\mathbf{F}_3) = A_4 \to \mathbf{Z}/3\mathbf{Z}$. One final example is the function

$$h := \frac{\lambda(\tau)(1-\lambda(\tau))}{16} = \left(\frac{\eta(\tau/2)\eta(2\tau)}{\eta(\tau)^2}\right)^{24}$$

of level $\Gamma^0(2) \supset \Gamma(2)$; here the complete list of n for which $h^{1/n}$ is either congruence modular or has bounded denominators are the divisors of 24. One may even compute that the abelianization of the congruence completion $\widehat{\Gamma^0(2)}$ of $\Gamma^0(2)$ (where $\Gamma^0(2)$ is considered as a subgroup of $\mathrm{PSL}_2(\mathbf{Z})$) is $\mathbf{Z}/24\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$. (The other $\mathbf{Z}/2\mathbf{Z}$ extension corresponds to $\sqrt{1-64h} = 1-2\lambda$.)

In a similar vein pertaining to the examples from the representation theory of vertex operator algebras, we prove in our closing section 7 the natural generalization of Theorem 1.0.1 to components of vector-valued modular forms for $\mathrm{SL}_2(\mathbf{Z})$, in particular resolving — in a sharper form, in fact — Mason's unbounded denominators conjecture [Mas12, KM08] on generalized modular forms.

1.1. **A sketch of the main ideas.** Our proof of Theorem 1.0.1 follows a broad Diophantine analysis path known in the literature (see [Bos04, Bos13] or [Bos20, Chapter 10]) as the *arithmetic algebraization* method.

1.1.1. *The Diophantine principle.* The most basic antecedent of these ideas is the following easy lemma:

Lemma 1.1.1. *A power series $f(x) = \sum_{n=0}^{\infty} a_n x^n \in \mathbf{Z}[\![x]\!]$ which defines a holomorphic function on $D(0,R)$ for some $R > 1$ is a polynomial.*

Lemma 1.1.1 follows upon combining the following two obsevations, fixing some $1 > \eta > R^{-1}$:

(1) The coefficients a_n are either 0 or else ≥ 1 in magnitude.
(2) The Cauchy integral formula gives a uniform upper bound $|a_n| = o(\eta^n)$.

We shall refer to the first inequality as a *Liouville lower bound*, following its use by Liouville in his proof of the lower bound $|\alpha - p/q| \gg 1/q^n$ for algebraic numbers $\alpha \neq p/q$ of degree $n \geq 1$. We shall refer to the second inequality as a *Cauchy upper bound*, following the example above where it comes from an application of the Cauchy integral formula. The first non-trivial generalization of Lemma 1.1.1 was Émile Borel's theorem [Bor94]. Dwork famously used a p-adic generalization of Borel's theorem in his p-adic analytic proof of the rationality of the zeta function of an algebraic variety over a finite field (see Dwork's account in the book [DGS94, Chapter 2]). The simplest non-trivial statement of Borel's theorem is that an integral formal power series $f(x) \in \mathbf{Z}[\![x]\!]$ must already be a rational function as soon as it has a *meromorphic* representation as a quotient of two convergent complex-coefficients power series on some disc $D(0,R)$ of a radius $R > 1$. The subject of arithmetic algebraization blossomed at the hands of many authors, including most prominently Carlson, Pólya, Robinson, Salem, Cantor, D. & G. Chudnovsky, Bertrandias, Zaharjuta, André, Bost, Chambert-Loir [CL02, BCL09, Ami75], [And04, § I.5], [And89, § VIII]. A simple milestone that we further develop in our § 2 is André's algebraicity criterion [And04, Théorème 5.4.3], stating in a particular case that an integral formal power series $f(x) \in \mathbf{Z}[\![x]\!]$ is algebraic as soon as the two formal functions x and f admit a simultaneous analytic uniformization — that means an analytic map $\varphi : (D(0,1),0) \to (\mathbf{C},0)$ such that the composed formal function germ $f(\varphi(z)) \in \mathbf{C}[\![z]\!]$ is also analytic on the full disc $D(0,1)$, and such that φ is sufficiently large in terms of *conformal size*, namely: $|\varphi'(0)| > 1$. For example, for any integer m, the algebraic power series $f = (1-m^2x)^{1/m} \in \mathbf{Z}[\![x]\!]$ admits the simultaneous analytic uniformization $x = (1-e^{Mz})m^{-2}$ and $f = e^{Mz/m}$, where the conformal size $|\varphi'(0)| = M/m^2$ can clearly be made arbitrarily large by making a suitable choice of M.

A common theme of all these generalizations of Lemma 1.1.1 is that they come down to a tension between a Liouville lower bound and a Cauchy upper bound. For example, in the proof of Borel's theorem ([Ami75, Ch 5.3]), the Liouville lower bound is applied not to the coefficients a_n themselves but rather to Hankel determinants $\det|\alpha_{i,j}|$ with $\alpha_{i,j} = a_{i+j+n}$. To consider a more complicated example (much closer in both spirit and in details to our own analysis), to prove André's algebraicity criterion [And04, Théorème 5.4.3], one wants to prove that certain powers of a formal function $f(\mathbf{x}) \in \mathbf{Z}[\![\mathbf{x}]\!]$ are linearly dependent over the polynomial ring $\mathbf{Z}[\mathbf{x}]$. (It will be advantageous to consider functions in several complex variables $\mathbf{x} = (x_1, \ldots, x_d)$.) The idea is now to consider a certain $\mathbf{Z}[\mathbf{x}]$ linear combination $F(\mathbf{x})$ of powers of $f(\mathbf{x})$ chosen such that they vanish to high order at $\mathbf{0}$ but yet the $\mathbf{Z}[\mathbf{x}]$ coefficients $p(\mathbf{x})$ are themselves not too complicated — the existence of such a choice follows from the classical Siegel's lemma. Now the Liouville lower bound is applied to a lowest order non-zero coefficient of $F(\mathbf{x}) \in \mathbf{Z}[\![\mathbf{x}]\!]$. Note that such a coefficient must exist or else the equality $F(\mathbf{x}) = 0$ realizes $f(\mathbf{x})$ as algebraic. The Cauchy upper bound in this case once again follows by an application of the Cauchy integral formula.

In our setting, the Liouville lower bound ultimately comes down to the integrality ("bounded denominators") hypothesis on the Fourier coefficients of $f(\tau)$, while the Cauchy upper bound comes down to studying the mean growth behavior $m(r, \varphi) := \int_{|z|=r} \log^+ |\varphi| \, \mu_{\mathrm{Haar}}$ of the largest (universal covering) analytic map $\varphi : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$ avoiding the N-th roots of unity. These are clearly distinguished in our abstract arithmetic algebraization work of § 2 as the steps (2.2.2) and (2.2.3), respectively. (We also refer to (2.4.8) and (2.4.6), resp. (2.5.23) and (2.5.22), in our alternative treatments.) Our Theorem 2.0.1 is effectively a quantitative refinement of André's algebraicity criterion to take into account the degree of algebraicity over $\mathbf{Q}(x)$, and still more precisely a certain holonomy rank over $\mathbf{Q}(x)$. Foreshadowing a key technical point (to be discussed in more detail later in the introduction), our Cauchy upper bound is given in terms of a mean (integrated) growth term rather than a supremum term, and this improvement is essential to our approach.

1.1.2. *Modularity and simultaneous uniformizations of f and λ.* Let us now explain the relevance of arithmetic holonomy rank bounds to the unbounded denominator conjecture. After reducing to weight $k = 0$ as above, the functions $f = f(\tau)$ and $x := \lambda(\tau)/16 \in q + q^2\mathbf{Z}[\![q]\!]$ are algebraically dependent and share both (we assume) the property of integral Fourier coefficients at the cusp $q = 0$. Let us assume for the purpose of this sketch that $f(\tau) \in \mathbf{Z}[\![q]\!]$ with $q = e^{\pi i \tau}$, i.e. that the cusp $i\infty$ has width dividing 2. Then the formal inverse series expansion

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

of (1.0.2) has integer coefficients, expressing the identity $\mathbf{Z}[\![q]\!] = \mathbf{Z}[\![x]\!]$ of formal power series rings, and that formal substitution turns our integral Fourier coefficients hypothesis into an *algebraic* power series with integer coefficients: henceforth in this introductory sketch we switch to writing, by a mild and harmless notational abuse, simply $f(x) \in \mathbf{Z}[\![x]\!]$ in place of $f(\tau)$ and $\lambda(q)$ in place of (1.0.2). In the general case of arbitrary cusp width, which we need anyhow for the inner workings of our proof even if one is ultimately interested in the $\mathbf{Z}[\![q]\!]$ case, we will only have $f \in \mathbf{Z}[1/N][\![x]\!]$ when we write out $f(\tau)$ as a power series in $x := \sqrt[N]{\lambda/16}$ to accommodate the Puiseux series — but there is still a hidden integrality property which we can exploit. That leads to some mild technical nuance with the power series (2.0.2) — think of $t = q^{1/N}$, $x(t) = \sqrt[N]{\lambda(t^N)/16}$ and $p(x) = x^N$ — in our refinement (2.0.3) of André's theorem.

The complex analysis enters by way of a linear ODE in the following way. To start with, we have, just *by fiat*, the simultaneous analytic uniformization of the two functions $x := \lambda/16$ and f by the complex unit q-disc $|q| < 1$. In this way, the tautological choice $\varphi(z) := \lambda(z)/16$ turns our algebraic power series $f(x) \in \mathbf{Z}[\![x]\!]$ into a boundary case (unit conformal size $\varphi'(0) = 1$) of André's criterion. Another boundary case, but this time transcendental and incidentally demonstrating the sharpness of the qualitative André algebraicity criterion even in the *a priori* holonomic situation (see [And04, Appendix, A.5] for a discussion), is provided by the Gauss hypergeometric function

$$F(x) := {}_2F_1\left[\begin{matrix} 1/2 \;\; 1/2 \\ 1 \end{matrix}; 16x\right] = \sum_{n=0}^{\infty} \binom{2n}{n}^2 x^n \in \mathbf{Z}[\![x]\!],$$

whose unit-radius simultaneous analytic uniformization with $x = \lambda/16$ is given again by the analytic q coordinate, and the classical Jacobi formula

$${}_2F_1\left[\begin{matrix} 1/2 \;\; 1/2 \\ 1 \end{matrix}; \lambda(q)\right] = \Big(\sum_{n \in \mathbf{Z}} q^{n^2}\Big)^2$$

which transforms this hypergeometric series into a weight one modular form for the congruence group $\Gamma(2)$. The existence of such transcendental $\mathbf{Z}[\![x]\!]$ holonomic functions on $\mathbf{C} \smallsetminus \{0, 1/16\}$ recovers—by André's algebraicity criterion—a classical "1/16 theorem" of Carathéodory [Car54, (412.8) on page 198]. (See also Goluzin [Gol69, § III.1, Theorem 1].)

1.1.3. *A finite local monodromy leads to an overconvergence.* It turns out, and this is the key to our method and already answers André's question in [And04, Appendix, A.5], that a different choice of $\varphi(z)$ allows one to arithmetically distinguish between these two cases (algebraic and transcendental), and to have the algebraicity of $f(x)$ recognized by André's Diophantine criterion by way of an "overconvergence." The common feature of these two functions $f(x)$ and $F(x)$ — coming respectively out of modular forms of weights 0 and 1 — is that they both vary holonomically in $x \in \mathbf{C} \smallsetminus \{0, 1/16\}$: they satisfy linear ODEs with coefficients in $\mathbf{Q}[x]$ and no singularities[1] apart from the three punctures $x = 0, 1/16, \infty$ of $Y(2) = \mathbf{H}/\Gamma(2)$. The difference feature is that their respective local monodromies around $x = 0$ are finite for the case of $f(x)$ (a quotient of $\mathbf{Z}/N$, with the order N equal to the LCM of cusp widths, or Wohlfahrt level [Woh64] of $f(x)$); and infinite for the case of $F(x)$ (isomorphic to $\mathbf{Z}$, corresponding more particularly to the fact that this particular hypergeometric function acquires a $\log x$ term after an analytic continuation around a small circle enclosing $x = 1/16$). If now we perform the variable change $x \mapsto x^N$, redefaulting to $x := \sqrt[N]{\lambda(q^N)/16}$, that resolves the N-th root ambiguity in the formal Puiseux branches of $f(x)$ at $x = 0$, and the resulting algebraic power series $f(x^N) \in \mathbf{Z}[\![x^N]\!] \subset \mathbf{Z}[\![x]\!]$ has turned holonomic on $\mathbf{P}^1 \smallsetminus \{16^{-1/N}\mu_N, \infty\}$: singularities only at $16^{-1/N}\mu_N \cup \{\infty\}$ (but not at $x = 0$: this key step of exploiting arithmetic algebraization is the same as in Ihara's arithmetic connectedness theorem [Iha94, Theorem 1], which together with Bost's extension [Bos99] to arithmetic Lefschetz theorems have in equal measure been inspirational for our whole approach to the unbounded denominators conjecture). Since $\lambda : D(0, 1) \to \mathbf{C} \smallsetminus \{1\}$ has fiber $\lambda^{-1}(0) = \{0\}$, the function $\varphi(z) := \sqrt[N]{\lambda(z^N)/16} : D(0, 1) \to \mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ is still holomorphic on the unit disc $|z| < 1$, and under this tautological choice, both functions $f(x^N)$ and $F(x^N)$ continue to be at the borderline of André's algebraicity criterion: $|\varphi'(0)| = 1$.

But if instead of the tautological simultaneous uniformization we take $\varphi : D(0, 1) \to \mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ to be the universal covering map (pointed at $\varphi(0) = 0$), then either by a direct computation with monodromy, or by Cauchy's analyticity theorem on the solutions of linear ODEs with analytic coefficients and no singularities in a disc, we have both function germs $x := \varphi(z)$ and $f(x) := f(\varphi(z))$ holomorphic, hence convergent, on the full unit disc $D(0, 1)$. In contrast, now $F(\varphi(z))$ converges only up to the "first" nonzero fiber point in $\varphi^{-1}(0) \smallsetminus \{0\}$, giving a certain radius rather smaller than 1. We must have the strict lower bound $|\varphi'(0)| > 1$, because the preceding unit-radius holomorphic map $\sqrt[N]{\lambda(z^N)/16} : D(0, 1) \to \mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ has to factorize *properly* through the universal covering map. Indeed in Theorem 5.1.4, using an explicit description by hypergeometric functions of the multivalued inverse of the universal covering map of $\mathbf{C} \smallsetminus \mu_N$ based on Poincaré's ODE approach [Hem88] to the uniformization of Riemann surfaces, we find an exact formula for this uniformization radius in terms of the Euler Gamma function.[2] Hence the algebraicity of $f(x)$ gets witnessed by André's criterion; and the formal new result that we get already at this opening stage (see Theorem 7.2.1) is that *any* integral formal power series solution $f(x) \in \mathbf{Z}[\![x]\!]$ to a linear ODE $L(f) = 0$ without singularities on $\mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\}$ is in fact algebraic as soon as the linear differential operator L has a finite local monodromy $\mathbf{Z}/N$ around the singular point $x = 0$. More than this: the quantitative Corollary 2.0.5 proves that the totality of such $f(x) \in \mathbf{Z}[\![x]\!]$ at a given N span a finite-dimensional $\mathbf{Q}(x)$-vector space, and gives an upper bound on its dimension as a function of the Wohlfahrt level parameter N. Now since a (noncongruence) counterexample $f(\tau) \in \mathbf{Z}[\![q]\!]$ to Theorem 1.0.1 would not exist on its own but spawn a whole sequence $f(p\tau) \in \mathbf{Z}[\![q]\!]$ of $\mathbf{Q}(x)$-linearly independent counterexamples at growing

[1] With nontrivial local monodromy. The precise definition is in 2.0.4.

[2] André pointed out to us that this explicit formula has previously been obtained by Kraus and Roth, see [KR16, Remark 5.1]. See also [Gol69, §III.1].

Wohlfahrt level $N \mapsto Np$, our idea is to measure up the supply of these putative (fictional) counterexamples alongside the congruence supply at a gradually increasing level until together they break the quantitative bound (2.0.3) supplied by our arithmetic holonomy Theorem 2.0.1.

1.1.4. *The dimension bound can be leveraged with growing level N.* We have the congruence supply of dimension $[\Gamma(2):\Gamma(2N)] \gg N^3$, and then as a glance at our shape (2.0.7) of holonomy rank bound readily reveals, it seems a fortuitous piece of luck that the conformal size (Riemann uniformization radius at 0) of our relevant Riemann surface $\mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ turns out to have the matching asymptotic form $1+\zeta(3)/(2N^3)+O(N^{-5})$. We "only" have to prove that the numerator (growth) term in the holonomy rank bound (2.0.7) inflates at a slower rate than our extrapolating putative counterexamples $f(\tau) \mapsto f(p\tau)$!

The meaning of the requisite inflation rate is clarified in § 4, with Proposition 4.3.5 and Remark 4.3.8. It turns out that the logarithmically inflated holonomy rank (dimension) bound by $O(N^3 \log N)$ is sufficient for the desired proof by contradiction (but an $O(N^{3+1/\log\log N})$ or worse form of bound would not suffice); and this is what we ultimately prove. Getting to this degree of precision creates however some additional challenges. A straightforward elaboration of André's original argument in [And89, Criterium VIII 1.6], taking the number of variables $d \to \infty$ and involving the $\sup_{|z|=r} \log|\varphi|$ growth term of *loc.cit.* in place of our mean (integrated) growth term $m(r,\varphi)$ in (2.0.7), leads quite easily to an $O(N^5)$ dimension bound; and by a further work explicitly with the cusps of the Fuchsian uniformization $D(0,1)/\Gamma_N \cong \mathbf{C} \smallsetminus \mu_N$, and an appropriate Riemann map precomposition, it is possible to further reduce that down to an $O(N^4)$. See Remark 5.2.18. This does not suffice to conclude the proof. Going further requires an intrinsic improvement into André's dimension bound itself: the reduction of the supremum term to the integrated term in the numerator of (2.0.7).

We give three proofs of this improvement, all being based on the same auxiliary construction scheme of § 2.1. Our default treatment §§2.1, 2.2, 2.3 is based on Nevanlinna's canonical factorization of meromorphic functions of bounded characteristic. Additionally, we also include in § 2.5 our original argument based on equidistribution ideas, and a simplified alternative path § 2.4 proposed to us by André and based on plurisubharmonicity and a lexicographic induction. The former variation has a potential-theoretic flavor familiar from the proof of Bilu's theorem [Bil97] (see also § 2.5.4), but it is in the cross-variables $d \to \infty$ asymptotic aspect and hence different than the well-established link (see [Bos99, BCL09, Bos04]) of arithmetic algebraization to adelic potential theory.

1.1.5. *Nevanlinna theory for Fuchsian groups.* Everything is thus reduced to establishing a uniform integrated growth bound of the form

$$m(r, F_N^N) := \int_{|z|=r} \log^+ |F_N^N|\, \mu_{\mathrm{Haar}} = O\Big(\log \frac{N}{1-r}\Big), \tag{1.1.2}$$

where $N \geq 2$ and $F_N : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$ is the universal covering map based at $F_N(0) = 0$. Heuristically this is supported by the idea that the renormalized function $F_N(q^{1/N})^N$ "converges" in some sense to the modular lambda function $\lambda(q)$, as $N \to \infty$. These functions do indeed converge as q-expansions as $N \to \infty$ on any ball around the origin of radius strictly less than 1. The problem is that this convergence is not in any way uniform as $r \to 1$, but we need to use (1.1.2) with a radius as large as $r = 1 - 1/(2N^3)$. The growth of the map F_N is governed by the growth of the cusps of the (N,∞,∞) triangle Fuchsian group Γ_N, and studying these directly, for instance by comparing them to the cusps of the limit (∞,∞,∞) triangle group $\Gamma(2)$, proves to be difficult.

Surprisingly perhaps, we are instead able in § 6 to prove the requisite mean growth bound (1.1.2) on the abstract grounds of Nevanlinna's value distribution theory for general meromorphic functions. For any universal covering map $F : D(0,1) \to \mathbf{C} \smallsetminus \{a_1,\ldots,a_N\}$ of a sphere with $N+1 \geq 3$ punctures, one has the mean growth asymptotic $m(r,F) = \int_{|z|=r} \log^+|F|\,\mu_{\mathrm{Haar}} \sim \frac{1}{N-1}\log\frac{1}{1-r}$ under $r \to 1^-$, providing extremal examples of Nevanlinna's defect inequality with $N+1$ full deficiencies on the disc [Nev70, page 272]. Contrast this with the qualitatively exponentially larger growth behavior $\sup_{|z|=r}\log|F| \asymp \frac{1}{1-r}$ of the crude supremum term. In our particular situation of $\{a_1, \dots, a_N\} = \mu_N$ for the puncture points, we are able to exploit the fortuitous relation $\prod_{i=1}^N (x - a_i) \sum_{i=1}^N \frac{1}{x-a_i} = N x^{N-1}$ particular to the partial fractions decomposition (6.2.3) to get to the uniformity precision of (1.1.2) with the method of the logarithmic derivative in Theorem 6.0.1.

2. The arithmetic holonomicity theorem

Our proof relies on the following dimension bound which is an extension of André's arithmetic algebraicity criterion [And04, Théorème 5.4.3]. We state and prove our result here in a particular case suited to our needs, beginning with the abstract form. We denote by $\mathcal{O}(\overline{D(0,1)}) \subset \mathbf{C}[\![z]\!]$ the ring of holomorphic function germs that converge on some open neighborhood of the closed unit disc $|z| \leq 1$. Throughout our paper, we will use the notation

$$\mathbf{T} := \{e^{2\pi i\theta} \, : \, \theta \in [0,1)\} \subset \mathbf{C}^\times$$

for the unit circle, the Cartesian power

$$\mathbf{T}^d := \{(e^{2\pi i\theta_1}, \dots, e^{2\pi i\theta_d}) \, : \, \theta_1, \dots, \theta_d \in [0,1)\} \subset \mathbf{G}_m^d(\mathbf{C})$$

for the unit d-torus, and

$$\mu_{\mathrm{Haar}} := d\theta_1 \cdots d\theta_d$$

for the normalized Haar measure of this compact group.

Theorem 2.0.1. *Consider the following data:*

(i) *a nonconstant rational function $p(x) \in \mathbf{Q}(x) \smallsetminus \mathbf{Q}$ without pole at $x = 0$,*

(ii) *a formal power series*

$$x(t) \in t + t^2\mathbf{Q}[\![t]\!] \tag{2.0.2}$$

pulling back p into an integral coefficients power series $x^*p := p(x(t)) \in \mathbf{Z}[\![t]\!]$ in the new variable t,

(iii) *and a holomorphic mapping $\varphi : \overline{D(0,1)} \to \mathbf{C}$ taking $\varphi(0) = 0$ with $|\varphi'(0)| > 1$, and pulling back p into a holomorphic function $\varphi^*p \in \mathcal{O}(\overline{D(0,1)})$ on some neighborhood of the closed unit disc.*

Suppose the formal power series $f_1, \dots, f_m \in \mathbf{Q}[\![x]\!]$ are $\mathbf{Q}(p(x))$-linearly independent and satisfy the following integrality and analyticity properties like in (ii) and (iii):

$$x^*f_1, \dots, x^*f_m \in \mathbf{Z}[\![t]\!], \quad \text{and} \quad \varphi^*f_1, \dots, \varphi^*f_m \in \mathcal{O}(\overline{D(0,1)}).$$

Then $f_1, \dots, f_m \in \mathbf{Q}[\![x]\!]$ are algebraic (i.e., all $f_i \in \overline{\mathbf{Q}(x)}$), and

$$m \leq e \cdot \frac{\int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}}}{\log |\varphi'(0)|}, \tag{2.0.3}$$

where $e = 2.718\dots$ is Euler's constant.

The novel point of the bound (2.0.3) is the integrated term in the numerator instead of a supremum term. This is critical for our proof of the unbounded denominators conjecture to have the numerator in (2.0.3), which measures the growth of φ, expressed as a Nevanlinna mean proximity function (or a characteristic function).

This abstract dimension bound (2.0.3) will be used more concretely as a holonomy rank bound. To state the relevant corollary, let us introduce an algebra of holonomic power series with integral coefficients and restricted singularities.

Definition 2.0.4. *For $U \subset \mathbf{C}$ an open subset, $R \subset \mathbf{C}$ a subring with fraction field $F := \mathrm{Frac}(R)$, and $x(t) \in t\mathbf{Q}[\![t]\!]$ a formal power series, we define $\mathcal{H}(U, x(t), R)$ to be the ring of formal power series $f(x) \in F[\![x]\!]$ whose t-expansion $f(x(t)) \in R[\![t]\!]$, and such that there exists a nonzero linear differential operator L over $\overline{\mathbf{Q}}(x)$ with $L(f) = 0$ and having a trivial local monodromy around all of its singular points that belong to U.*

Further, we let $\mathcal{V}(U, x(t), R)$ to be the $F(x)$-vector space spanned by $\mathcal{H}(U, x(t), R)$.

For $x(t) = t$, we more simply denote the $R[x]$-algebra $\mathcal{H}(U, t, R)$ by $\mathcal{H}(U, R)$ and the $F(x)$-vector space $\mathcal{V}(U, t, R)$ by $\mathcal{V}(U, R)$.

Here by *trivial local monodromy around* $x = \alpha$ we mean that there exist a complex neighborhood $U_\alpha \ni \alpha$ and meromorphic functions $g_1, \dots, g_n \in \mathcal{M}(U_\alpha)$ on U_α, where n is the order of L, such that $g_1, \dots, g_n$ form a $\mathbf{C}$-basis of the solution space of $L(f) = 0$ on $U_\alpha \setminus \{\alpha\}$. This is the case if $x = \alpha$ is not a singular point of L, and apart from that the typical examples singular at $x = 0$ include $L_n = x\frac{d}{dx} - n$ for $n \in \mathbf{Z} \setminus \{0\}$, of solution space $\ker L_n = \mathbf{C} \cdot x^n$.

Our holonomy bound is now a straightforward combination of Theorem 2.0.1 and Cauchy's analyticity theorem on the solutions of linear differential equations with analytic coefficients.

Corollary 2.0.5. *Let $0 \in U \subset \mathbf{C}$ be an open subset containing the origin. If the uniformization radius of the pointed Riemann surface $(U, 0)$ is strictly greater than 1, then the algebra $\mathcal{V}(U, \mathbf{Z})$ is finite dimensional as a $\mathbf{Q}(x)$-vector space.*

More precisely, let $p(x) \in \mathbf{Q}(x) \setminus \mathbf{Q}$ be a non-constant rational function without poles in U, and let $\varphi(z) : \overline{D(0,1)} \to U$ a holomorphic map taking $\varphi(0) = 0$ with $|\varphi'(0)| > 1$. If

$$x(t) \in t + t^2\mathbf{Q}[\![t]\!] \tag{2.0.6}$$

has $p(x(t)) \in \mathbf{Z}[\![t]\!]$, then the following dimension bound holds on $\mathcal{V}(U, x(t), \mathbf{Z})$ over $\mathbf{Q}(p(x))$:

$$\dim_{\mathbf{Q}(p(x))} \mathcal{V}(U, x(t), \mathbf{Z}) \leq e \cdot \frac{\int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}}}{\log |\varphi'(0)|}. \tag{2.0.7}$$

Proof. The pulled back space $\varphi^*\mathcal{H}(U, x(t), \mathbf{Z}) \subset \varphi^*\mathcal{H}(U, \mathbf{C})$ lies in the ring of formal power series fulfilling linear differential equations with analytic coefficients and no singularities with nontrivial local monodromies on the closed disc $\overline{D(0,1)}$. Hence, for any such function $f \in \mathcal{H}(U, x(t), \mathbf{Z})$, there exists a nonzero $g(x) \in \mathbf{Q}[p(x)] \setminus \{0\}$ such that for any singular point $\alpha \in U$ of the linear operator L in Definition 2.0.4, and for any local solution $h(x)$ in a small punctured neighborhood of α, the product function $g(x)h(x)$ is holomorphic at $x = \alpha$. (The singularities of L and thus $h(x)$ in U are all at algebraic points, hence such a $g(x)$ exists.) Cauchy's theorem then gives that $\varphi^*(gf)$ is a holomorphic function on $\overline{D(0,1)}$, and we conclude by Theorem 2.0.1. $\square$

Theorem 2.0.1 is modeled on André's Diophantine approximation method [And89, §VIII], [And04, §5]. We include as many as three proofs, all sharing a common basic framework § 2.1 and relying crucially on a $d \to \infty$ limit for the number of auxiliary variables in the auxiliary function constructed by Lemma 2.1.2 below. Our original treatment was based on equidistribution and is in §§2.1, 2.5, and an alternative approach proposed to us by André and based on plurisubharmonicity is in §§2.1, 2.4. Firstly we give a shorter proof based on Nevanlinna's canonical factorization §2.3 and the following intermediate form of Theorem 2.0.1.

Lemma 2.0.8. *In the setting of Theorem 2.0.1, consider furthermore an arbitrary holomorphic function $h : \overline{D(0,1)} \to \mathbf{C}$ with $h(0) = 1$. Then*

$$m \leq e \, \frac{\max\left\{ \sup_{\mathbf{T}} \log |h|, \, \sup_{\mathbf{T}} \log |h \cdot \varphi^* p| \right\}}{\log |\varphi'(0)|}, \tag{2.0.9}$$

and $f_1, \dots, f_m \in \overline{\mathbf{Q}(x)} \cap \mathbf{Q}[\![x]\!]$.

Remark 2.0.10. We will see in §2.3 that Lemma 2.0.8 is in fact equivalent to Theorem 2.0.1. See however Remark 2.3.3 about a sketch of an improvement.

For a complete proof of the unbounded denominators conjecture, we invite the reader on a first pass to proceed directly to §3 after §2.3.

2.1. **The auxiliary construction.** We will make a use of a Diophantine approximation construction in a high number $d \to \infty$ of variables $\mathbf{x} := (x_1, \dots, x_d)$. We will write

$$\mathbf{x}^{\mathbf{j}} := x_1^{j_1} \cdots x_d^{j_d}, \quad p(\mathbf{x}) := (p(x_1), \dots, p(x_d)).$$

Since φ maps $(D(0,1), 0) \to (\mathbf{C}, 0)$ with nonzero derivative, the inverse function theorem gives a positive radius $\rho > 0$ such that

$$\varphi \, : \, \varphi^{-1}(D(0,\rho))_0 \xrightarrow{\cong} D(0,\rho) \tag{2.1.1}$$

is an analytic isomorphism from the connected component $\varphi^{-1}(D(0,\rho))_0$ containing 0 of $\varphi^{-1}(D(0,\rho))$.

Lemma 2.1.2. *Let $d, \alpha \in \mathbf{N}$ and $\kappa \in (0,1)$ be parameters. Asymptotically in $\alpha \to \infty$ as d and κ are held fixed, there exists a* nonzero *d-variate formal function $F(\mathbf{x})$ of the form*

$$F(\mathbf{x}) = \sum_{\substack{\mathbf{i} \in \{1,\dots,m\}^d \\ \mathbf{k} \in \{0,\dots,D-1\}^d}} a_{\mathbf{i},\mathbf{k}}\, p(\mathbf{x})^{\mathbf{k}} \prod_{s=1}^{d} f_{i_s}(x_s) \in \mathbf{Q}[\![\mathbf{x}]\!] \smallsetminus \{0\}, \tag{2.1.3}$$

vanishing to order at least α at $\mathbf{x} = \mathbf{0}$, with

(1)
$$D \leq \frac{1}{(d!)^{1/d}} \frac{1}{m} \Big(1 + \frac{1}{\kappa}\Big)^{\frac{1}{d}} \alpha + o(\alpha);$$

(2) *all $a_{\mathbf{i},\mathbf{k}} \in \mathbf{Z}$ are rational integers bounded in absolute value by $\exp\big(\kappa C \alpha + o(\alpha)\big)$ for some constant $C \in \mathbf{R}$ depending only on the radius ρ from* (2.1.1) *and on the degree and height of the rational function $p(x) \in \mathbf{Q}(x)$.*

Proof. We expand our sought-for formal function in (2.1.3) into a formal power series in $\mathbf{Q}[\![\mathbf{x}]\!]$ and solve $\binom{\alpha+d}{d} \sim \alpha^d/d!$ linear equations in the $(mD)^d$ free parameters $a_{\mathbf{i},\mathbf{j}}$. By the formal inverse function expansion, the integrality condition $p(x(t)) \in \mathbf{Z}[\![t]\!]$ entails that $x(t) \in t + (t^2/M)\mathbf{Z}[\![t/M]\!]$ for some $M \in \mathbf{N}$ bounded in terms of the degree and height of the rational function $p(x)$. Therefore also the inverse series $t(x) \in x + (x^2/M)\mathbf{Z}[\![x/M]\!]$, and so $f_i(x(t)) \in \mathbf{Z}[\![t]\!]$ entails $f_i(x) \in \mathbf{Z}[\![x/M]\!]$ for all $i = 1, \dots, m$. Furthermore, by (2.1.1), every power series $f_i(x) \in \mathbf{Q}[\![x]\!]$ is convergent on the archimedean disc $|x| < \rho$. The result then follows from the classical Siegel lemma [BG06, Lemma 2.9.1], with $e^C := M/\rho$ and the degree parameter choice

$$D \sim \frac{1}{m (d!)^{1/d}} \Big(1 + \frac{1}{\kappa}\Big)^{\frac{1}{d}} \alpha,$$

that brings in a Dirichlet exponent $\sim \kappa$ as $\alpha \to \infty$. That $F \not\equiv 0$ follows since at least one $a_{\mathbf{i},\mathbf{j}} \neq 0$ and the formal functions $f_1, \dots, f_m \in \mathbf{Q}[\![x]\!]$ are linearly independent over $\mathbf{Q}(p(x))$. □

2.2. **Extrapolation and proof of Lemma 2.0.8.** We consider the nonzero formal function

$$H(\mathbf{z}) := h(z_1)^D \cdots h(z_d)^D \cdot F(\varphi(z_1), \dots, \varphi(z_d)) \in \mathbf{C}[\![\mathbf{z}]\!] \smallsetminus \{0\}. \tag{2.2.1}$$

By construction, it vanishes at $\mathbf{z} = \mathbf{0}$ to order at least α, and it is holomorphic in a neighborhood of the closed unit polydisc because all the split-variables constituents

$$\varphi^* p; \quad \varphi^* f_1, \dots, \varphi^* f_m \in \mathcal{O}(\overline{D(0,1)}).$$

Let $\beta \geq \alpha$ be the exact order of vanishing of $F(\mathbf{x}) \in \mathbf{Q}[\![\mathbf{x}]\!] \smallsetminus \{0\}$ at $\mathbf{x} = \mathbf{0}$, and consider $c\,\mathbf{x}^{\mathbf{n}}$ any nonzero monomial of that lowest order $\beta = |\mathbf{n}|$. Since $x(t) \in t + t^2\mathbf{Q}[\![t]\!]$, the term $c\,\mathbf{t}^{\mathbf{n}}$ is a lowest order monomial in the formal power series $F(x(\mathbf{t})) \in \mathbf{Z}[\![\mathbf{t}]\!]$, and so $c \in \mathbf{Z} \smallsetminus \{0\}$. Thus we have the Liouville lower bound:

$$|c| \geq 1. \tag{2.2.2}$$

On the other hand, (2.2.1) and the normalizations $h(z) \in 1 + z\,\mathbf{C}[\![z]\!]$ and $\varphi(z) \in \varphi'(0)z + z^2\,\mathbf{C}[\![z]\!]$ exhibit $c\varphi'(0)^\beta\,\mathbf{z}^{\mathbf{n}}$ as a lowest order monomial in $H(\mathbf{z})$. Since the $\mathbf{z}^{\mathbf{n}}$ coefficient is also computed by Cauchy's integral formula $\int_{\mathbf{T}^d} \frac{H(\mathbf{z})}{\mathbf{z}^{\mathbf{n}}}\, \mu_{\mathrm{Haar}}(\mathbf{z})$, we have the Cauchy upper bound:

$$|c| \cdot |\varphi'(0)|^\alpha \leq |c| \cdot |\varphi'(0)|^\beta \leq \sup_{\mathbf{T}^d} |H|. \tag{2.2.3}$$

To estimate the last supremum under the asymptotic $\alpha \to \infty$ for fixed d and κ, we note that the sum (2.1.3) is comprised of $(mD)^d = \exp\big(o(\alpha)\big)$ terms, all of which are of the form

$$\prod_{j=1}^{d} h(z_j)^{D-k_j} \prod_{j=1}^{d} (h(z_j) \cdot p(\varphi(z_j)))^{k_j} \cdot f_{i_1}(\varphi(z_1)) \cdots f_{i_d}(\varphi(z_d)),$$

for some $k_1, \dots, k_d \in \{0, \dots, D-1\}$, and with coefficients bounded in magnitude by $\exp\big(\kappa C \alpha + o(\alpha)\big)$.

Every such term is bounded in magnitude on $\mathbf{T}^d$ by

$$e^{\kappa C\alpha+o(\alpha)} \cdot \max\big\{\sup_{\mathbf{T}}|h|, \sup_{\mathbf{T}}|h\cdot\varphi^*p|\big\}^{dD}\cdot \max_{1\le i\le m}\sup_{\mathbf{T}}|\varphi^* f_i|^d.$$

By the triangle inequality, we have in the $\alpha\to\infty$ asymptotic—with respect to a fixed d—the supremum bound

$$\sup_{\mathbf{T}^d}\log|H| \le dD\cdot\max\big\{\sup_{\mathbf{T}}\log|h|, \sup_{\mathbf{T}}\log|h\cdot\varphi^*p|\big\} + \kappa C\alpha + o(\alpha).$$

Combining with (2.2.2) and (2.2.3), we get the asymptotic bound

$$\alpha\log|\varphi'(0)| \le \frac{d}{(d!)^{1/d}}\Big(1+\frac{1}{\kappa}\Big)^{\frac{1}{d}}\cdot\frac{\alpha}{m}\max\big\{\sup_{\mathbf{T}}\log|h|, \sup_{\mathbf{T}}\log|h\cdot\varphi^*p|\big\}+\kappa C\alpha+o(\alpha)$$

as $\alpha\to\infty$ with respect to the other parameters.

This proves the dimension bound

$$m \le \inf_{\substack{d\in\mathbf{N}\\ 0<\kappa<(\log|\varphi'(0)|)/|C|}}\left\{\frac{\frac{d}{(d!)^{1/d}}\Big(1+\frac{1}{\kappa}\Big)^{\frac{1}{d}}\cdot\max\big\{\sup_{\mathbf{T}}\log|h|, \sup_{\mathbf{T}}\log|h\cdot\varphi^*p|\big\}}{\log|\varphi'(0)|-\kappa C}\right\}$$

contingent on the denominator being positive. Lemma 2.0.8 now follows by firstly letting $d\to\infty$ and then $\kappa\to 0$, and observing that in that limit

$$\frac{d}{(d!)^{1/d}}\Big(1+\frac{1}{\kappa}\Big)^{\frac{1}{d}}\to e \qquad \text{while} \qquad \kappa C\to 0,$$

by Stirling's asymptotic and the key point that the constant C depends only on the datum $\{p;\varphi\}$, but not on d nor κ.

The algebraicity of f_i is *a fortiori* by the finite dimension bound (2.0.3) as all powers of f_i satisfy $x^* f_i^N \in \mathbf{Z}[[t]]$ and $\varphi^* f_i^N \in \mathcal{O}(\overline{D(0,1)})$, for any $N\in\mathbf{N}$. $\square$

2.3. **Canonical factorization and proof of Theorem 2.0.1.** At this point Theorem 2.0.1 comes as the immediate combination of Lemma 2.0.8 and the following classical lemma of Nevanlinna.

Lemma 2.3.1 (Nevanlinna [Nev70])**.** *Consider a holomorphic function $g:\overline{D(0,1)}\to\mathbf{C}$, and let $\varepsilon>0$. Then there exists a quotient representation*

$$g=\frac{hg}{h},$$

where $h:\overline{D(0,1)}\to\mathbf{C}$ is holomorphic with

$$h(0)=1 \quad \textit{and} \quad \max\big\{\sup_{\mathbf{T}}\log|h|, \sup_{\mathbf{T}}\log|hg|\big\}\le\int_{\mathbf{T}}\log^+|g|\,\mu_{\mathrm{Haar}}+\varepsilon.$$

Proof. By a Blaschke product, we are reduced to the case that $g(z)\in\mathcal{O}^\times(D(0,1))$ is a *functional unit*: no zeros or poles inside the unit disc $D(0,1)$. In that case, the function $\log|g|$ is harmonic on $D(0,1)$, and the Poisson kernel formula—see (6.1.8) below—together with the canonical decomposition $\log=\log^+-\log^-$ into positive and negative parts gives the quotient representation

$$g(z)=\frac{\exp\Big(-\int_{\mathbf{T}}\log^+\frac{1}{|g(w)|}\cdot\frac{w+z}{w-z}\,\mu_{\mathrm{Haar}}(w)\Big)}{\exp\Big(-\int_{\mathbf{T}}\log^+|g(w)|\cdot\frac{w+z}{w-z}\,\mu_{\mathrm{Haar}}(w)\Big)}. \tag{2.3.2}$$

In this factorization, the top and bottom both are holomorphic functions of $z\in D(0,1)$, and they both are bounded in absolute value by ≤ 1, because the Poisson kernel satisfies $\Re\Big(\frac{w+z}{w-z}\Big)>0$ for $1=|w|>|z|$. Furthermore, the bottom in (2.3.2) takes the value $\exp\Big(-\int_{\mathbf{T}}\log^+|g|\,\mu_{\mathrm{Haar}}\Big)$ at $z=0$.

Thus, in the functional unit case $g \in \mathcal{O}^\times(D(0,1))$ and on $D(0,1)$ rather than $\overline{D(0,1)}$, we conclude the proof of the lemma by selecting

$$h(z) := \exp\Big(\int_{\mathbf{T}} \log^+ |g| \, \mu_{\mathrm{Haar}} - \int_{\mathbf{T}} \log^+ |g(w)| \cdot \frac{w+z}{w-z} \, \mu_{\mathrm{Haar}}(w) \Big).$$

The proof of the general case is completed by the Blaschke product clearance to a functional unit on the open unit disc, and then by shrinking the radius a little bit. The details are, for example, in Nevanlinna [Nev70, §VII.1.4] or Goluzin [Gol69, §VII.5]. ☐

Remark 2.3.3. Conversely, for any holomorphic map $h : \overline{D(0,1)} \to \mathbf{C}$ with $h(0) = 1$, we have the lower bound

$$\max \big\{ \sup_{\mathbf{T}} \log |h|, \sup_{\mathbf{T}} \log |hg| \big\} \geq \int_{\mathbf{T}} \log^+ |g| \, \mu_{\mathrm{Haar}},$$

as one sees immediately from integrating the pointwise identity

$$\max\{\log |h|, \log |hg|\} = \log |h| + \log^+ |g|$$

over $\mathbf{T}$ and using $\int_{\mathbf{T}} \log |h| \geq \log |h(0)| = 0$ from subharmonicity. This shows the necessity of the ε in Lemma 2.3.1. It also shows that Theorem 2.0.1—our final goal of the current §2, which at this point is fully proved—is in fact equivalent with the intermediate form Lemma 2.0.8.

On the other hand, with a bit more work based on the Law of Large Numbers for restricting all the sums $\sum_{i=1}^d k_i/D$ involved in Lemma 2.1.2 to be concentrated around the mean $d/2$ under our $d \to \infty$ asymptotic, the proof method of Lemma 2.0.8 does give the finer bound

$$m \leq (e/2) \inf_{h\colon h(0)=1} \Big\{ \frac{\sup_{\mathbf{T}} \log |h| + \sup_{\mathbf{T}} \log |h \cdot \varphi^* p|}{\log |\varphi'(0)|} \Big\},$$

where the infimum is taken over all holomorphic mappings $h : \overline{D(0,1)} \to \mathbf{C}$ subject to the normalizing constraint $h(0) = 1$. We will not need this improvement here.

2.4. **A first alternative proof.** In this section, we complete an idea proposed to us by André as an alternative to our original proof of Theorem 2.0.1 (itself recounted in § 2.5 further down), based on plurisubharmonicity and a lexicographic induction instead of on Cauchy's formula. We invite the reader at this point to skip ahead directly to §3 on a first pass, as the arithmetic holonomy bound (2.0.7)—the algebraization ingredient that we need for the unbounded denominators conjecture—has already been proved.

2.4.1. *Lemma on the lexicographically lowest coefficient.* The extrapolation step will now be based on the following analytic lemma, to be applied with $G(\mathbf{z}) = F(\varphi(z_1), \ldots, \varphi(z_d))$, where $F(\mathbf{x})$ is our auxiliary function from Lemma 2.1.2. The lemma reflects the plurisubharmonic property of the multivariable complex functions of the form $\log |H(\mathbf{z})|$ with $H(\mathbf{z})$ holomorphic, used inductively on the number of variables d.

Lemma 2.4.1. *Consider $G(\mathbf{z}) \in \mathbf{C}[\![\mathbf{z}]\!] \smallsetminus \{0\}$ holomorphic on the closed unit polydisc $\{\mathbf{z} \mid \max_{i=1}^d |z_i| \leq 1\}$, and let $c\,\mathbf{z}^{\mathbf{n}}$ be the lexicographically minimal monomial. Then*

$$\log |c| \leq \int_{\mathbf{T}^d} \log |G| \, \mu_{\mathrm{Haar}}. \tag{2.4.2}$$

Proof. We induct on the number of variables d. For $d = 1$, the bound (2.4.2) follows directly from Jensen's formula, or from the subharmonic property of the function $u(z) := \log |z^{-n} G(z)|$, which entails

$$\log |c| = u(0) \leq \int_{\mathbf{T}} u \, \mu_{\mathrm{Haar}} = \int_{\mathbf{T}} \log |G| \, \mu_{\mathrm{Haar}}.$$

The last equality uses that the functions $u(z) = \log |z^{-n} G(z)|$ and $\log |G(z)|$ have the same restriction on the unit circle $\mathbf{T}$.

For the induction step, we write $\mathbf{z} = (z_1, \mathbf{z}')$ and

$$G(\mathbf{z}) = z_1^{n_1} H(\mathbf{z}),$$

where $H \in \mathbf{C}[\![\mathbf{z}]\!]$ is holomorphic by our lexicographic minimality assumption. For any fixed $\mathbf{z}' \in \mathbf{T}^{d-1}$, by the same argument as the $d = 1$ case above, we have

$$\log |H(0, \mathbf{z}')| \leq \int_{\mathbf{T}} \log |H(z_1, \mathbf{z}')| \, \mu_{\mathrm{Haar}}(z_1) = \int_{\mathbf{T}} \log |G(z_1, \mathbf{z}')| \, \mu_{\mathrm{Haar}}(z_1). \tag{2.4.3}$$

By assumption, the lexicographically minimal monomial in $H(0, \mathbf{z}') \in \mathbf{C}[\![\mathbf{z}']\!]$ is $c\,\mathbf{z}'^{\mathbf{n}'}$, where $\mathbf{n} = (n_1, \mathbf{n}')$. Therefore the induction hypothesis gives

$$\log |c| \leq \int_{\mathbf{T}^{d-1}} \log |H(0, \mathbf{z}')| \, \mu_{\mathrm{Haar}}(\mathbf{z}'). \tag{2.4.4}$$

We complete the induction by integrating the inequality (2.4.3) over $\mathbf{z}' \in \mathbf{T}^{d-1}$. □

2.4.2. *Extrapolation and first alternative proof of Theorem 2.0.1.* We apply Lemma 2.4.1 to the φ-pullback of our d-variate auxiliary function:

$$G(z_1, \dots, z_d) := F(\varphi(z_1), \dots, \varphi(z_d)) \in \mathbf{C}[\![\mathbf{z}]\!] \smallsetminus \{0\}. \tag{2.4.5}$$

This is holomorphic in a neighborhood of the closed unit polydisc, because all the split-variables constituents

$$\varphi^* p; \quad \varphi^* f_1, \dots, \varphi^* f_m \in \mathcal{O}(\overline{D(0,1)}).$$

Thus, with $c\,\mathbf{z}^{\mathbf{n}}$ the lexicographically lowest monomial in $G(\mathbf{z})$, we get from (2.4.2) and Lemma 2.1.2 our Cauchy upper bound:

$$\begin{aligned} \log |c| &\leq \int_{\mathbf{T}^d} \log |F(\varphi(z_1), \dots, \varphi(z_d))| \, \mu_{\mathrm{Haar}} \\ &\leq dD \int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}} + \kappa C \, \alpha + o(\alpha), \end{aligned} \tag{2.4.6}$$

asymptotically as $\alpha \to \infty$ with regard to the other parameters. Here, we used the pointwise triangle inequality bound

$$\log |F(x_1, \dots, x_d)| \leq D \sum_{i=1}^{d} \log^+ |p(x_i)| + \kappa C \alpha + o(\alpha)$$

for $x_i := \varphi(z_i)$ (note that the sum in (2.1.3) is comprised of $(mD)^d = \exp(o(\alpha))$ terms), and integrated this pointwise bound over the unit polycircle $\mathbf{z} \in \mathbf{T}^d$.

The Liouville lower bound comes down to the integrality property

$$F(x(t_1), \dots, x(t_d)) \in \mathbf{Z}[\![\mathbf{t}]\!] \tag{2.4.7}$$

inherited from our respective assumptions

$$x^* p; \quad x^* f_1, \dots, x^* f_m \in \mathbf{Z}[\![t]\!]$$

on the split-variables constituents in (2.1.3). By our normalizations $x(t) \in t + t^2 \mathbf{Q}[\![t]\!]$ and $\varphi(z) \in \varphi'(0) z + z^2 \mathbf{C}[\![z]\!]$, the lexicographically lowest term of $G(\mathbf{z}) \in \mathbf{C}[\![\mathbf{z}]\!]$ is equal to $\varphi'(0)^{\beta}$ times the lexicographically lowest term of $F(x(\mathbf{t})) \in \mathbf{Z}[\![\mathbf{t}]\!]$, where $\beta := |\mathbf{n}| = n_1 + \cdots + n_d \geq \alpha$ is the common total degree of these lexicographically lowest terms in $G(\mathbf{z})$ and $F(x(\mathbf{t}))$. By (2.4.7), this entails that the nonzero coefficient

$$c \in \varphi'(0)^{\beta} \, \mathbf{Z} \smallsetminus \{0\},$$

and hence *a fortiori* that

$$\log |c| \geq \beta \log |\varphi'(0)| \geq \alpha \log |\varphi'(0)|. \tag{2.4.8}$$

We get our requisite dimension bound (2.0.3) on combining the degree bound (1) of Lemma 2.1.2 with the Cauchy upper bound (2.4.6) and the Liouville lower bound (2.4.8), and letting firstly $\alpha \to \infty$, then $d \to \infty$, and finally $\kappa \to 0$.

This completes another proof of Theorem 2.0.1. □

2.5. **A second alternative proof.** The remainder of §2 presents our original argument for Theorem 2.0.1, with the thought that it could still be useful for other settings including potential theory (see 2.5.4). Like § 2.2 and unlike § 2.4, it is based on the leading order jet rather than the overall lexicographically lowest monomial in $F(\mathbf{x})$, and on the pointwise Cauchy integral formula instead of on plurisubharmonicity. Contrastingly to both, it employs a cross-variables equidistribution idea.

2.5.1. *Equidistribution.* We start out the same way with Lemma 2.1.2, but now aim to extrapolate based directly on the pointwise Cauchy bound. The key idea here is that upon substituting $x_j = \varphi(z_j)$ into (2.1.3), the $d \to \infty$ equidistribution on the circle of the uniform independent and identically distributed points $z_1, \ldots, z_d$ will normally get the constituent monomials in (2.1.3) to grow at most at the integrated exponential rate of $dD \int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}}$. The problem with directly applying the Cauchy bound as in [And89, VIII 1.6] is that it involves a pointwise upper bound on the intervening functions $|p(\varphi(\mathbf{z}))^{\mathbf{k}}|$ on the unit polycircle $\mathbf{z} \in \mathbf{T}^d$, and while the Monte Carlo heuristic applies on the majority of $\mathbf{T}^d$ under $d \to \infty$, with a probability tending to 1 roughly speaking at a rate exponential in $-d$ (this follows by Hoeffding's concentration inequality with (2.5.7) below), the peaks at the biased part of $\mathbf{T}^d$ get overwhelmingly large, and a direct extrapolation with (2.1.3) in this way still only leads to a dimension bound with $\sup_{|z|=1} \log |p \circ \varphi|$.

To improve the supremum term to the mean term $\int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}}$, we dampen the size at the peaks by firstly multiplying (2.1.3) by a suitably chosen power $V(\mathbf{z})^M$ of the Vandermonde polynomial

$$V(\mathbf{z}) := \prod_{i<j} (z_i - z_j) = \det \begin{bmatrix} 1 & z_1 & z_1^2 & \cdots & z_1^{d-1} \\ 1 & z_2 & z_2^2 & \cdots & z_2^{d-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & z_d & z_d^2 & \cdots & z_d^{d-1} \end{bmatrix} \in \mathbf{Z}[z_1, \ldots, z_d] \smallsetminus \{0\}. \tag{2.5.1}$$

By applying the Hadamard volume inequality to the Vandermonde determinant in (2.5.1), we recover the following classical result of Fekete, crucial for the present approach.

Lemma 2.5.2 (Fekete). *The supremum of $|V(\mathbf{z})| = \prod_{1 \le i < j \le d} |z_i - z_j|$ over the unit polycircle $\mathbf{z} \in \mathbf{T}^d$ is equal to $d^{d/2}$, with equality if and only if the points $z_1, \ldots, z_d$ are the vertices of a regular d-gon.*

The idea for sifting out the equidistributed tuples $(z_1, \ldots, z_d)$ is the following. If the points $z_1, \ldots, z_d$ are poorly distributed in the uniform measure of the circle, the quantity $|V(\mathbf{z})|$ is uniformly exponentially small in $-d^2$ (Lemma 2.5.8 below). This plays off against the $d^{d/2} = \exp(o(d^2))$ bound of Lemma 2.5.2 to sift out the equidistributed points in our pointwise upper bound in the Cauchy integral formula when we extrapolate in § 2.4.2 below. Liouville's Diophantine lower bound still succeeds like in André [And04, §5], thanks to the chain rule and the integrality property of coefficients of (2.5.1), but at the Cauchy upper bound we are now aided by the fact that $V(\mathbf{z})^M$ is extremely small (an exponential in $-Md^2$, see (2.5.9)) at the peaks of the pointwise Cauchy bound, where the point $(z_1, \ldots, z_d)$ is poorly distributed, while still not too large (subexponential in Md^2, thanks to Lemma 2.5.2) uniformly throughout the whole polycircle $\mathbf{T}^d$.

In the remainder of the current subsection, we spell out the notion of 'well-distributed' and 'poorly distributed', and supply the key equidistribution property for the numerical integration step. The following is the standard notion of discrepancy theory.

Definition 2.5.3. *The (normalized, box)* **discrepancy function** $D : \mathbf{T}^d \to (0, 1]$ *is the supremum over all circular arcs $I \subset \mathbf{T}$ of the defect between the normalized arclength of I and the proportion of points falling inside I:*

$$D(z_1, \ldots, z_d) := \sup_{I \subset \mathbf{T}} \left| \mu_{\mathrm{Haar}}(I) - \frac{1}{d} \#\{i \, : \, z_i \in I\} \right|.$$

We also recall the basic properties of the total variation functional on the circle. In our situation, all that we need is that $\log^+|h|$ is of bounded variation for an arbitrary C^1 function $h : \mathbf{T} \to \mathbf{R}$. Then Koksma's estimate permits us to integrate numerically. All of this can be alternatively phrased in the qualitative language of weak-$*$ convergence.

Definition 2.5.4. *The* **total variation** $V(g)$ *of a function* $g : \mathbf{T} \to \mathbf{R}$ *is the supremum over all partitions* $0 \le \theta_1 < \cdots < \theta_n < 1$ *of* $\sum_{j=1}^{n-1} |g(e^{2\pi\sqrt{-1}\theta_{j+1}}) - g(e^{2\pi\sqrt{-1}\theta_j})|$.

Thus, for $g \in C^1(\mathbf{T})$, we have the simpler formula

$$V(g) = \int_{\mathbf{T}} |g'(z)| \, \mu_{\mathrm{Haar}}(z), \quad g \in C^1(\mathbf{T}). \tag{2.5.5}$$

We have $V(\log^+|h|) < \infty$ for $h \in C^1(\mathbf{T})$, and Koksma's inequality (see for example Drmota–Tichy [DT97, Theorem 1.14]):

$$\Big| \frac{1}{d} \sum_{j=1}^{d} g(z_j) - \int_{\mathbf{T}} g \, \mu_{\mathrm{Haar}} \Big| \le V(g) D(z_1, \ldots, z_d). \tag{2.5.6}$$

In practice the discrepancy function is conveniently estimated by the Erdös–Turán inequality (cf. Drmota–Tichy [DT97, Theorem 1.21]):

$$D(z_1, \ldots, z_d) \le 3 \Big(\frac{1}{K+1} + \sum_{k=1}^{K} \frac{1}{k} \Big| \frac{z_1^k + \cdots + z_d^k}{d} \Big| \Big) \quad \forall K \in \mathbf{N}, \tag{2.5.7}$$

in terms of the power sums. Here we note in passing that, by (2.5.7) and the Chernoff tail bound or the Hoeffding concentration inequality (see, for example, Tao [Tao12, Theorem 2.1.3 and Ex. 2.1.4]), we have that for any fixed $\varepsilon > 0$, the probability of the event $D(z_1, \ldots, z_d) \ge \varepsilon$ decays to 0 exponentially in $-d$ as $d \to \infty$. This last remark has purely a heuristic value for our next step, and is not used in the estimates in itself (but rather shows that these estimates are sharp).

Thus we introduce another parameter $\varepsilon > 0$, which in the end will be let to approach 0 but only after $d \to \infty$, and we divide the points $\mathbf{z} \in \mathbf{T}^d$ into two groups according to whether $D(z_1, \ldots, z_d) < \varepsilon$ (the well-distributed points) or $D(z_1 \ldots, z_d) \ge \varepsilon$ (the poorly distributed points). For the well-distributed group we use the Koksma inequality (2.5.6), and for the poorly distributed group we take advantage of the overwhelming damping force of the Vandermonde factor.

The following is essentially Bilu's equidistribution theorem [Bil97], in a mild disguise.

Lemma 2.5.8. *There are functions* $c(\varepsilon) > 0$ *and* $d_0(\varepsilon) \in \mathbf{R}$ *such that, for every* $\varepsilon \in (0, 1]$*, if* $d \ge d_0(\varepsilon)$ *and* $(z_1, \ldots, z_d) \in \mathbf{T}^d$ *is a d-tuple with discrepancy* $D(z_1, \ldots, z_d) \ge \varepsilon$*, then*

$$|V(z_1, \ldots, z_d)| = \prod_{1 \le i < j \le d} |z_i - z_j| < e^{-c(\varepsilon) d^2}. \tag{2.5.9}$$

Proof. Since the qualitative result suffices for our purposes here, we give a soft proof based on compactness. The following argument borrows from Bombieri and Gubler's exposition [BG06, page 103] of Bilu's equidistribution theorem. Suppose to the contrary that there is an $\varepsilon \in (0, 1]$ and an infinite sequence $(z_1^{(d)}, \ldots, z_d^{(d)}) \in \mathbf{T}^d$ such that

$$\lim_{d \to \infty} \frac{1}{\binom{d}{2}} \sum_{1 \le i < j \le d} \log \frac{1}{|z_i^{(d)} - z_j^{(d)}|} \le 0, \tag{2.5.10}$$

but

$$\forall d, \qquad D(z_1^{(d)}, \ldots, z_d^{(d)}) \ge \varepsilon. \tag{2.5.11}$$

By the Banach–Alaoglu theorem of the compactness of the weak-$*$ unit ball of $C(\mathbf{T})^*$, we may extract a subsequence of the sequence of normalized Dirac masses $\delta_{\{z_1^{(d)}, \ldots, z_d^{(d)}\}}$ that converges weak-$*$ to some limit probability measure μ of the unit circle. By continuity of the discrepancy functional, (2.5.11) implies that the limit discrepancy

$$D(\mu) := \sup_{I \subset \mathbf{T}} \left|\mu_{\mathrm{Haar}}(I) - \mu(I)\right| \geq \varepsilon.$$

In particular, μ is not the uniform measure μ_{Haar}.

On the other hand, it is a well-known theorem from potential theory that every compact $K \subset \mathbf{C}$ admits a unique probability measure μ_K, called the equilibrium measure, that minimizes the Dirichlet energy integral

$$I(\nu) := \iint_{K \times K} \log \frac{1}{|z-w|} \, \nu(z) \, \nu(w)$$

across all probability measures ν supported by K. By symmetry, we have $\mu_{\mathbf{T}} = \mu_{\mathrm{Haar}}$, and since $I(\mu_{\mathrm{Haar}}) = 0$, but $\mu \neq \mu_{\mathrm{Haar}}$, we have the strict inequality

$$I(\mu) = \iint_{\mathbf{T} \times \mathbf{T}} \log \frac{1}{|z-w|} \, \mu(z) \, \mu(w) > 0. \tag{2.5.12}$$

If the measure μ is continuous (that is, the measure of a point is 0, or equivalently the diagonal of $\mathbf{T} \times \mathbf{T}$ has $\mu \times \mu$ measure 0), then the positive energy (2.5.12) contradicts (2.5.10) by weak-$*$ convergence. In more detail, take a continuous function $\phi : [0,\infty) \to [0,\infty)$ to have $\phi|_{[0,1/2]} \equiv 0$ and $\phi|_{[1,\infty)} \equiv 1$, and let $\phi_\eta(t) := \phi(t/\eta)$ for $0 < \eta \leq 1$. Then, since $\phi_\eta(t) < 1$ implies $\log(1/t) > 0$ while $\phi_\eta(t) \leq 1$ always, assumption (2.5.10) implies

$$\lim_{d \to \infty} \frac{1}{\binom{d}{2}} \sum_{1 \leq i < j \leq d} \phi_\eta\left(|z_i^{(d)} - z_j^{(d)}|\right) \log \frac{1}{|z_i^{(d)} - z_j^{(d)}|} \leq 0$$

leading by weak-$*$ convergence to the non-positivity

$$\iint_{\mathbf{T} \times \mathbf{T}} \phi_\eta(|x-y|) \log \frac{1}{|z-w|} \, \mu(z) \, \mu(w) \leq 0,$$

for every $\eta \in (0,1]$. Since the diagonal has measure 0, this runs in contradiction with (2.5.12) upon letting $\eta \to 0$.

If instead the measure μ is not continuous, then there is a point $a \in \mathbf{T}$ and a positive constant $c > 0$ such that, for any $\eta > 0$, and any $d \gg_\eta 1$ sufficiently large, there are at least cd points among $\{z_1^{(d)}, \ldots, z_d^{(d)}\}$ in the neighborhood $|z-a| < \eta/2$. The contribution to (2.5.10) from all these pairs of points is alone $\geq c^2 \log(1/\eta)$, and since the total contribution from any subset of the points is in any case $\geq -\log 2$, we get again in contradiction with (2.5.10) on letting $\eta \to 0$. ☐

2.5.2. *Damping the Cauchy estimate.* We combine Lemmas 2.5.2 and 2.5.8 for our choice of the damping term $V(\mathbf{z})^M$. In the following, all asymptotics are taken under $\alpha \to \infty$ with respect to all other parameters.

By Lemma 2.1.2 and our defining assumption that all $f_i(\varphi(z))$ are holomorphic on some neighborhood of the closed unit disc $|z| \leq 1$, we have uniformly on the polycircle $\mathbf{z} \in \mathbf{T}^d$ the pointwise bound

$$\log|F(\varphi(z_1), \ldots, \varphi(z_d))| \leq D \sum_{j=1}^{d} \log^+ |p(\varphi(z_j))| + \kappa C \alpha + o(\alpha). \tag{2.5.13}$$

Since the function $\log^+|p \circ \varphi| : \mathbf{T} \to \mathbf{R}$ is of finite variation $V(\log^+|p \circ \varphi|) < \infty$, Koksma's estimate (2.5.6) yields, on the well-distributed part $\mathbf{z} \in \mathbf{T}^d$, the uniform pointwise upper bound

$$D(z_1, \ldots, z_d) < \varepsilon \quad \Longrightarrow$$

$$\log|F(\varphi(z_1), \ldots, \varphi(z_d))| \leq dD \int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}} + \kappa C \alpha + O_{p,\varphi}(\varepsilon \, dD) + o(\alpha).$$

The implicit constant in $O_{p,\varphi}(\varepsilon \, dD)$ can be taken as the total variation $V(\log^+|p \circ \varphi|)$; that this error term is $o_{\varepsilon \to 0}(dD) = o_{\varepsilon \to 0}(\alpha)$ is all that matters to us in the asymptotic argument.

On the poorly distributed but exceptional part $D(z_1,\dots,z_d)\geq\varepsilon$, the sum in (2.5.13) can get as large as $d\sup_{|z|=1}\log|p\circ\varphi|$. This trivial bound gives

$$\forall \mathbf{z}\in\mathbf{T}^d,\quad \log|F(\varphi(z_1),\dots,\varphi(z_d))|\leq dD\sup_{|z|=1}\log^+|p\circ\varphi|+\kappa C\alpha+o(\alpha). \tag{2.5.14}$$

We now impose the condition

$$d\geq d_0(\varepsilon),\qquad \text{for the function } d_0(\varepsilon) \text{ in Lemma 2.5.8,} \tag{2.5.15}$$

for the remainder of the proof of Theorem 2.0.5 (at the end we will firstly take $d\to\infty$, and only then $\varepsilon\to 0$), and we select the Vandermonde exponent

$$M:=\Big\lfloor\frac{\sup_{|z|=1}\log^+|p\circ\varphi|}{c(\varepsilon)}\frac{D}{d}\Big\rfloor, \tag{2.5.16}$$

with $c(\varepsilon)$ the function from Lemma 2.5.8. We are now in a position to usefully estimate the supremum of $|V(\mathbf{z})^M F(\varphi(\mathbf{z}))|$ uniformly across the unit polycircle $\mathbf{z}\in\mathbf{T}^d$, by separately examining the well-distributed and the poorly distributed cases of $\mathbf{z}$.

On the poorly distributed part $D(z_1,\dots,z_d)\geq\varepsilon$, Lemma 2.5.8 with (2.5.14), (2.5.15) and (2.5.16) gives

$$\sup_{\mathbf{z}\in\mathbf{T}^d:\,D(z_1,\dots,z_d)\geq\varepsilon}\log|V(\mathbf{z})^M F(\varphi(\mathbf{z}))|\ll\kappa\alpha. \tag{2.5.17}$$

On the well-distributed part $D(z_1,\dots,z_d)\leq\varepsilon$, we have

$$\begin{aligned}&\sup_{\mathbf{z}\in\mathbf{T}^d:\,D(z_1,\dots,z_d)\leq\varepsilon}\log|V(\mathbf{z})^M F(\varphi(\mathbf{z}))|\\ &\leq dD\int_{\mathbf{T}}\log^+|p\circ\varphi|\,\mu_{\mathrm{Haar}}+\kappa C\alpha+O_{p,\varphi}(\varepsilon\alpha)+O_{\varepsilon,p,\varphi}\Big(\frac{\log d}{d}\alpha\Big)+o(\alpha).\end{aligned} \tag{2.5.18}$$

by (2.5.16) and Lemma 2.5.2.

Consider the holomorphic function

$$H(\mathbf{z}):=V(\mathbf{z})^M F(\varphi(z_1),\dots,\varphi(z_d))=:\sum_{\mathbf{n}\in\mathbf{N}_0^d}c(\mathbf{n})\,\mathbf{z}^{\mathbf{n}}\in\mathbf{C}[\![\mathbf{z}]\!], \tag{2.5.19}$$

convergent on the closed unit disc $\|\mathbf{z}\|\leq 1$. For each $\mathbf{n}\in\mathbf{N}_0^d$, the $\mathbf{z}^{\mathbf{n}}$ coefficient of $H(\mathbf{z})$ is given by the Cauchy integral formula

$$c(\mathbf{n})=\int_{\mathbf{T}^d}\frac{H(\mathbf{z})}{\mathbf{z}^{\mathbf{n}}}\,\mu_{\mathrm{Haar}}(\mathbf{z}), \tag{2.5.20}$$

entailing the Cauchy upper bound

$$\forall\mathbf{n}\in\mathbf{N}_0^d,\qquad |c(\mathbf{n})|\leq\sup_{\mathbf{z}\in\mathbf{T}^d}|H(\mathbf{z})|. \tag{2.5.21}$$

On combining the bounds (2.5.18), on the well-distributed part of $\mathbf{T}^d$, and (2.5.17), on the poorly distributed part of $\mathbf{T}^d$, we arrive at our damped Cauchy estimate:

$$\begin{aligned}\log|c(\mathbf{n})|&\leq dD\int_{\mathbf{T}}\log^+|p\circ\varphi|\,\mu_{\mathrm{Haar}}\\ &+O(\kappa\alpha)+O_{p,\varphi}(\varepsilon\,\alpha)+O_{\varepsilon,p,\varphi}\Big(\frac{\log d}{d}\alpha\Big)+o(\alpha),\end{aligned} \tag{2.5.22}$$

asymptotically under $\alpha\to\infty$.

2.5.3. *The extrapolation.* Finally we combine the degree estimate (1) of Lemma 2.1.2 with the Cauchy bound (2.5.22) and the integrality properties of the functions $F(x(\mathbf{t})) \in \mathbf{Z}[\![\mathbf{t}]\!]$ of (2.1.3) and $V(\mathbf{z}) \in \mathbf{Z}[\mathbf{z}]$ of (2.5.1).

Let $\beta \geq \alpha$ be the exact order of vanishing of $F(\mathbf{x})$ at the origin $\mathbf{x} = \mathbf{0}$. Among the nonvanishing monomials $c\,\mathbf{x}^{\mathbf{n}}$ of this minimal order $|\mathbf{n}| = \beta$, choose the one whose degree vector $\mathbf{n}$ has the highest lexicographical ordering. By the chain rule and the minimality of $|\mathbf{n}|$, the normalization condition (2.0.2) on the formal substitution $x(t)$ entails that $c\,\mathbf{t}^{\mathbf{n}}$ is a minimal order term in the t-expansion $F(x(\mathbf{t}))$. Hence the integrality $f(x(\mathbf{t})) \in \mathbf{Z}[\![\mathbf{t}]\!]$ gives that $c \in \mathbf{Z} \smallsetminus \{0\}$ is a nonzero rational integer.

Consider now our product function $H(\mathbf{z}) = V(\mathbf{z})^M F(\varphi(\mathbf{z})) \in \mathbf{C}[\![\mathbf{z}]\!]$. In the factor $V(\mathbf{z})^M$, it is $z_1^{(d-1)M} z_2^{(d-2)M} \cdots z_{d-1}^M$ that has the highest lexicographical ordering. Consequently, by the chain rule again,

$$c\,\varphi'(0)^\beta\, z_1^{n_1+(d-1)M} z_2^{n_2+(d-2)M} \cdots z_d^{n_d}$$

exhibits a monomial in $V(\mathbf{z})^M F(\varphi(\mathbf{z}))$ of the minimal order $\beta + M\binom{d}{2}$; for $\big(n_1 + (d-1)M, n_2 + (d-2)M, \ldots, n_d\big)$ has the strictly highest lexicographical ordering across all monomials of degree $\beta + M\binom{d}{2}$ in $V(\mathbf{z})^M F(\varphi(\mathbf{z}))$.

We have thus found a nonzero coefficient of $H(\mathbf{z}) \in \mathbf{C}[\![\mathbf{z}]\!]$ that belongs to the $\mathbf{Z}$-module $\varphi'(0)^\beta \mathbf{Z}$, where $\beta \geq \alpha$. Thus the Cauchy upper bound (2.5.21) is supplemented with the Liouville lower bound

$$\exists\, \mathbf{n} \in \mathbf{N}_0^d \quad : \quad \log|c(\mathbf{n})| \geq \beta \log|\varphi'(0)| \geq \alpha \log|\varphi'(0)|. \tag{2.5.23}$$

We get the requisite holonomy rank bound (2.0.3) on combining the degree bound (1) of Lemma 2.1.2 with the Cauchy upper bound (2.5.22) and the Liouville lower bound (2.5.23), and letting firstly $\alpha \to \infty$, then $d \to \infty$, then $\kappa \to 0$, and finally $\varepsilon \to 0$.

This concludes also our original proof of Theorem 2.0.1. □

2.5.4. *A potential-theoretic generalization.* The path with §§2.1,2.5 leads straightforwardly to an extension in potential theory, which we formulate without detailing a proof. Consider a compact subset $K \subset \overline{D(0,1)}$ with transfinite diameter $d(K) := \mathrm{cap}(K, [\infty])$. The *equilibrium measure* μ_K is the unique probability measure supported by K that attains the lowest possible logarithmic energy of

$$\iint_{K \times K} \log \frac{1}{|x-y|}\, \mu_K(x)\, \mu_K(y) = -\log d(K).$$

If

$$\log|\varphi'(0)| + \log d(K) > 0,$$

then under the hypotheses of Corollary 2.0.5 we have the holonomy rank bound

$$\dim_{\mathbf{Q}(p(x))} \mathcal{V}(U, x(t), \mathbf{Z}) \leq e\, \frac{\int_K \log^+ |p \circ \varphi|\, \mu_K}{\log|\varphi'(0)| + \log d(K)}. \tag{2.5.24}$$

The cases $K = \overline{D(0,1)}$ or $K = \mathbf{T}$ both recover Corollary 2.0.5.

Remark 2.5.25. The result is still more general than 2.5.4, and the restriction here to $\mathbf{Z}[\![t]\!]$ expansions was chosen as minimal for our application to noncongruence modular forms. In a sequel work we will generalize our integrated holonomy rank bound, in particular to the case of $\mathbf{Q}[\![t]\!]$ formal functions, and study its applications to transcendence theory. With regard to the latter, it is of some interest to inquire about the optimal numerical constant that could take the place of the coefficient e in (2.5.24).

In these optics, Bost and Charles [BC22, Corollary 8.3.5] have very recently refined our Theorem 2.0.1 to the cleaner form

$$m \leq \frac{\iint_{\mathbf{T}^2} \log|p(\varphi(z)) - p(\varphi(w))|\, \mu_{\mathrm{Haar}}(z) \mu_{\mathrm{Haar}}(w)}{\log|\varphi'(0)|}.$$

In particular, on replacing p by p^k with using the elementary inequality $\log|x-y| \le \log^+|x| + \log^+|y| + \log 2$ and taking $k \to +\infty$, their result improves our coefficient e in (2.5.24) to the value 2. We do not know whether or not this is the best-possible constant.

3. Our approach to the Unbounded Denominators Conjecture

In this section, we lay out our main approach to the unbounded denominators conjecture. This will reduce the proof to a number of independent results in group theory, complex geometry, and complex analysis which we take up in sections §4, §5, and §6. Our main idea is to use our arithmetic holonomicity theorems to prove the following:

Proposition 3.0.1. *Let $F_N : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$ be an analytic universal covering map sending 0 to 0. Suppose that:*

1. *The conformal radius $|F_N'(0)|$ of F_N is asymptotically at least*
$$16^{1/N}\left(1+\frac{A}{N^3}\right)$$
for some constant $A > 0$.
2. *For a fixed $B > 0$, the following mean value bound holds on the circle $|z| = 1 - BN^{-3}$:*
$$\int_{|z|=1-BN^{-3}} \log^+|F_N|\,\mu_{\mathrm{Haar}} \ll_B \frac{\log N}{N}.$$

Then the $\mathbf{Q}(\lambda)$-vector space R_{2N} generated by the modular functions with Fourier coefficients in $\mathbf{Q}$ and bounded denominators at the cusp $\zeta = i\infty$, and having cusp widths dividing $2N$ at all cusps $\zeta \in \mathbf{P}^1(\mathbf{Q})$, has dimension at most $CN^3 \log N$ over the field $\mathbf{Q}(\lambda)$ of modular functions of level $\Gamma(2)$, for some absolute constant C.

Proof. Let $t := q^{1/N} = e^{\pi i \tau/N}$. We use Theorem 2.0.5 with $U := \mathbf{C} \smallsetminus 16^{-1/N}\mu_N$, $p(x) := x^N$ and

$$x := (\lambda(\tau)/16)^{1/N} \in t + t^2\mathbf{Z}[1/N][\![t]\!], \tag{3.0.2}$$

with the Kummer integrality condition $p(x) = x^N \in \mathbf{Z}[\![q]\!] = \mathbf{Z}[\![t^N]\!] \subset \mathbf{Z}[\![t]\!]$ being in place.

The integrality and cusp widths conditions in the definition of the $\mathbf{Q}(\lambda)$-vector space R_{2N} entail a basis of R_{2N} made of elements of the ring $\mathcal{H}(U, x(t), \mathbf{Z}) \otimes_{\mathbf{Z}} \mathbf{Q}$. More precisely, for a modular function f with Fourier expansion at $i\infty$ lying in $\mathbf{Z}[\![q^{1/N}]\!] \otimes_{\mathbf{Z}} \mathbf{Q}$, by our choice of $t = q^{1/N}$ and $x(t) = (\lambda(q)/16)^{1/N}$ we have $x^*f \in \mathbf{Z}[\![t]\!] \otimes_{\mathbf{Z}} \mathbf{Q}$, on defining x^*f as the formal x-expansion of $f(q) = f(q(x)) \in \mathbf{Z}[\![q^{1/N}]\!] \otimes_{\mathbf{Z}} \mathbf{Q} \subset \mathbf{Q}[\![x]\!]$. As f is a regular function on some affine modular curve Y over $\overline{\mathbf{Q}}$ which admits a finite étale map to $Y(2)_{\overline{\mathbf{Q}}}$, there exists a (minimal) nonzero linear differential operator L over $\overline{\mathbf{Q}}(\lambda)$ such that $L(f) = 0$ and all singularities $\lambda \neq 0, 1, \infty$ of L in $\mathbf{C} \smallsetminus \{0,1\}$ have trivial local monodromy. Moreover, our assumption on the cusp widths dividing $2N$ implies that a local coordinate in a small neighborhood of each cusp of Y above $\lambda = 0$ can be chosen to be the lift of some (positive integer) power of $x = (\lambda/16)^{1/N}$. This means that the pullback of L to $U \smallsetminus \{0\} = \mathbf{C}^\times \smallsetminus 16^{-1/N}\mu_N$ admits a full set of meromorphic solutions in some sufficiently small neighborhood of $x = 0$, i.e. has a trivial local monodromy around $x = 0$. Therefore $f \in \mathcal{H}(U, x(t), \mathbf{Z}) \otimes_{\mathbf{Z}} \mathbf{Q}$, and $R_{2N} \subset \mathcal{V}(U, x(t), \mathbf{Z})$.

It thus suffices to bound $\dim_{\mathbf{Q}(x^N)} \mathcal{V}(U, x(t), \mathbf{Z})$ by $CN^3 \log N$. We take $r := 1 - AN^{-3}/2$ and

$$\varphi(z) := 16^{-1/N} F_N(rz) \quad : \quad \overline{D(0,1)} \to U.$$

By (1) and the choice of radius $r = 1 - AN^{-3}/2$, we have

$$\log|\varphi'(0)| > \log(1 + A/N^3) + \log r = AN^{-3}/2 + O_A(N^{-6}). \tag{3.0.3}$$

Thus, with $c := A/3$, we get for $N \gg 1$ sufficiently large that

$$\log|\varphi'(0)| > cN^{-3}. \tag{3.0.4}$$

Corollary 2.0.5 now upper-bounds our requisite dimension as

$$\dim_{\mathbf{Q}(x^N)} \mathcal{V}(U, x(t), \mathbf{Z}) \le e \cdot \frac{\int_{|z|=1-A/(2N^3)} \log^+|F_N^N|\,\mu_{\mathrm{Haar}}}{cN^{-3}}.$$

The requisite bound by $O(N^3 \log N)$ now results follows from (2) with the choice $B := A/2$. $\square$

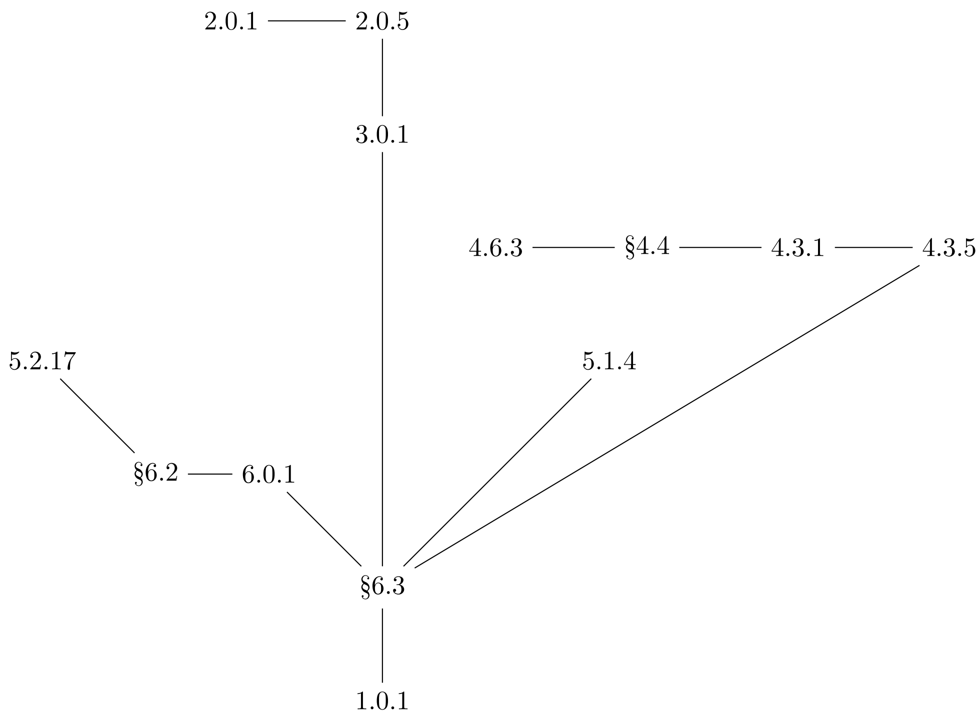
Remark 3.0.5. We may also prove this proposition by using Theorem 2.0.1 directly. Using the notation in the proof, let Y' denote the modular curve Y with all the cusps above $0 \in Y(2) \cup \{0\}$ filled in. The fiber product $Y' \times_{Y(2)_{\overline{\mathbf{Q}}} \cup \{0\}} U$ with its natural map to U is a covering map (one can check this claim locally; the assumption on cusp widths is used to prove that $0 \in U$ is not ramified). Therefore, the universal covering map $D(0,1) \to U$ factors through $Y' \times_{Y(2)_{\overline{\mathbf{Q}}} \cup \{0\}} U$ and thus we obtain a map $D(0,1) \to Y' \times_{Y(2)_{\overline{\mathbf{Q}}} \cup \{0\}} U \to Y'$ (the second map is the natural map) such that its composition with $Y' \to Y(2)_{\overline{\mathbf{Q}}} \cup \{0\}$ is the map $D(0,1) \to U \to Y(2)_{\overline{\mathbf{Q}}} \cup \{0\}$. Thus $f \circ \varphi$ is also given by the natural pullback of f from Y' to $D(0,1)$ and thus it is holomorphic over $D(0,1)$ as far as f is holomorphic at all cusps in Y', which can be achieved by multiplying f with a suitable power of λ. Thus we verify the analyticity property in Theorem 2.0.1. The rest of the proof is the same as the proof above.

3.1. **A guide to the proof of the main theorem.** We prove both of the assumptions of Proposition 3.0.1 hold in Theorems 5.1.4 and Theorem 6.0.1 respectively. This provides a $CN^3 \log N$ dimension bound for the vector space R_{2N} of all modular functions against the obvious $\gg N^3$ lower bound for the subring of the *congruence* examples from the fact that $[\Gamma(2) : \Gamma(2N)] \gg N^3$ (see Equation 4.3.3). We then need to provide an additional argument to overcome this "small error" (a logarithmic gap $O(\log N)$ in every level N) between the lower and upper bounds.

The following is a guide to what we do in the next few sections of our paper:

(1) In §4, we prove that the logarithmic gap between the ring of modular forms with bounded denominators and the ring of congruence modular forms can be leveraged to prove the full unbounded denominators conjecture. The main idea here is that given a noncongruence modular form $f(q) \in \mathbf{Z}[\![q^{1/N}]\!]$, one can construct many more such forms independent over the ring of congruence forms by considering $f(q^p) \in \mathbf{Z}[\![q^{1/N}]\!]$ for primes p.
(2) In §5, we study the properties of the function F_N. It turns out more or less to be related to a Schwarzian automorphic function on a (generally non-arithmetic) triangle group. This allows us to compute the conformal radius of F_N exactly (see Theorem 5.1.4), and indeed it has the form $16^{1/N}\big(1 + (\zeta(3)/2)N^{-3} + \cdots\big)$.
(3) In §5, we also study the maximum value of $|F_N|$ on the circle $|z| = R$, uniformly in both N and $R < 1$. The main idea here is that a normalized variant function $G_N(q) = F_N(q^{1/N})^N$ "converges" to the modular λ function $\lambda(q) = 16q - 128q^2 + \cdots$. Approximating the region where F_N is large by the corresponding region for $\lambda(q)$ one predicts a growth rate of the desired form. However, the problem is that the convergence of $G_N(q)$ to $\lambda(q)$ is not in any way uniform, especially in the neighbourhoods of the cusps of F_N which certainly vary with N.
(4) In §6 we solve this uniformity problem on the abstract grounds of Nevanlinna theory. We combine the crude growth bound on $|F_N|$ with a version of Nevanlinna's lemma on the logarithmic derivative to prove our requisite uniform upper estimate on the mean proximity function $m(r, F_N) = \int_{|z|=r} \log^+ |F_N| \, \mu_{\mathrm{Haar}}$.
(5) Putting all the pieces together, the proof of Theorem 1.0.1 is then completed in §6.3.

The following leitfaden gives an abbreviated summary of how the argument is laid out:



4. Noncongruence forms

4.1. **Wohlfahrt Level.** We begin by recalling a notion of level for noncongruence subgroups due to Wohlfahrt [Woh64]. Let $G \subset \mathrm{SL}_2(\mathbf{Z})$ be a finite index subgroup. (Many of the arguments of this section do not require this hypotheses but since it is satisfied for our applications we assume it to avoid unnecessary distractions.) The group consisting of the two matrices $\pm I$, where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, will be denoted by E. The group $\mathrm{SL}_2(\mathbf{Z})$ acts via Möbius transformations both on the upper half plane $\mathbf{H}$ and the extended upper half plane $\mathbf{H}^* = \mathbf{H} \cup \mathbf{P}^1(\mathbf{Q})$. The action of $\mathrm{SL}_2(\mathbf{Z})$ on $\mathbf{P}^1(\mathbf{Q})$ is transitive. It follows that if a non-trivial element $\gamma \in G$ fixes an element $\zeta \in \mathbf{P}^1(\mathbf{Q})$, then γ has the form $\pm MU^m M^{-1}$, where $M \in \mathrm{SL}_2(\mathbf{Z})$, $M\infty = \zeta$, and

$$U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{4.1.1}$$

We call such a $\zeta \in \mathbf{P}^1(\mathbf{Q})$ a *cusp* of G. If $\zeta \in \mathbf{P}^1(\mathbf{Q})$, then $MU^m M^{-1} = (MUM^{-1})^m \in G$ for some m because G has finite index in $\mathrm{SL}_2(\mathbf{Z})$, and hence every element of $\mathbf{P}^1(\mathbf{Q})$ is a cusp of G. The stabilizer in G of a cusp $\zeta \in \mathbf{P}^1(\mathbf{Q})$ is either isomorphic to $E \times \mathbf{Z} = \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}$ or $\mathbf{Z}$, depending on whether $E \subset G$ or not. For each ζ, there is a minimal positive integer m such that $\pm MU^m M^{-1} \in G$, and we say that m is the *width* of the cusp ζ. The action of G on $\mathbf{P}^1(\mathbf{Q})$ has finitely many orbits, and the cusp width only depends on the orbit of the cusp under G. Geometrically, the complex structure on $\mathbf{H}$ imbues the quotient $X(G) = \mathbf{H}^*/G$ with the structure of an algebraic curve. From this point of view, the equivalence classes of cusps of G (up to the action of G) are in bijection with the pre-images of ∞ under the projection $X(G) \to X(\mathrm{SL}_2(\mathbf{Z})) = \mathbf{P}^1_j$, and the cusp widths are exactly the ramification indices of this map at $j = \infty$.

Definition 4.1.2 ([Woh64])**.** *The level $L(G)$ of G is the lowest common multiple of all the cusp widths of G.*

We begin with some elementary properties concerning this definition. We typically only consider groups containing $E = \langle -I \rangle$ since we are generally interested in stabilizers of functions under Möbius transformations.

Lemma 4.1.3. *Let G and H be finite index subgroups of $\mathrm{SL}_2(\mathbf{Z})$ both containing E. Suppose that $L(G)$ and $L(H)$ both divide N, then any cusp of $G \cap H$ also has cusp width dividing N.*

Proof. The stabilizer of a cusp inside any subgroup of $\mathrm{SL}_2(\mathbf{Z})$ containing E is $E \times \mathbf{Z}$. In particular, if G contains the group $E \times a\mathbf{Z}$ and H contains $E \times b\mathbf{Z}$ then $G \cap H$ contains $E \times \mathrm{lcm}(a, b)\mathbf{Z}$, and the result follows. □

Lemma 4.1.4. *Let $G \subset \mathrm{SL}_2(\mathbf{Z})$ be a finite index subgroup containing E with Wohlfahrt level N. Let $N(G)$ be the largest normal subgroup of $\mathrm{SL}_2(\mathbf{Z})$ contained in G. Then $N(G)$ has finite index in $\mathrm{SL}_2(\mathbf{Z})$ and $L(N(G)) = N$.*

Proof. Since G has finite index in $\mathrm{SL}_2(\mathbf{Z})$, the group $N(G)$ is the intersection of the finitely many conjugates of G by $\mathrm{SL}_2(\mathbf{Z})$. Hence $N(G)$ has finite index and $L(N(G)) = N$ by Lemma 4.1.3. □

Definition 4.1.5. *Let A denote the following matrix:*

$$A := \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.1.6}$$

(We use this notation so as to be consistent with that of Serre in [Tho89] which we follow below.) We now prove the following lemma concerning how the level of a subgroup changes under conjugation by A.

Lemma 4.1.7. *Let $H \subset \mathrm{SL}_2(\mathbf{Z})$ be a finite index subgroup containing E such that $L(H) = N$. Then $L(A^{-1}HA \cap \mathrm{SL}_2(\mathbf{Z}))$ divides Np.*

Proof. Let us write $\widetilde{H} := A^{-1}HA \cap \mathrm{SL}_2(\mathbf{Z})$. Note that $\widetilde{H}$ contains $A^{-1}EA = E$. In particular, the stabilizer of any cusp of $\widetilde{H}$ has the form $E \times \mathbf{Z}$, where the $\mathbf{Z}$ is generated by a unipotent element $\widetilde{h}$ of $\widetilde{H}$ conjugate in $\mathrm{SL}_2(\mathbf{Z})$ to U^m for some positive integer m, and we want to show that m divides Np.

Any unipotent element $\widetilde{h}$ in $\widetilde{H}$ has the form $\widetilde{h} = A^{-1}hA$ for some unipotent element $h \in H$. The element $h \in H$ will stabilize some cusp $\zeta \in \mathbf{P}^1(\mathbf{Q})$. The stabilizer of ζ in H has the form $E \times \mathbf{Z}$ where $\mathbf{Z}$ is generated by a unipotent element $\gamma \in H$. It follows that h will be the smallest power of γ which lies in $\widetilde{H}$, or equivalently in $\mathrm{SL}_2(\mathbf{Z})$. Since $L(H) = N$, we may write

$$\gamma = B \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} B^{-1}$$

with $n|N$, and $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$. We define

$$\begin{aligned} \widetilde{\gamma} := A^{-1}\gamma A = A^{-1}B \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} B^{-1}A \\ &= (A^{-1}BA) \left(A^{-1} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} A \right) (A^{-1}B^{-1}A) \\ &= (A^{-1}BA) \begin{pmatrix} 1 & n/p \\ 0 & 1 \end{pmatrix} (A^{-1}BA)^{-1}, \end{aligned}$$

where $(A^{-1}BA) = \begin{pmatrix} a & b/p \\ cp & d \end{pmatrix}$. Since h is a power of γ, we deduce that $\widetilde{h} = A^{-1}hA$ is a power of $\widetilde{\gamma}$, although $\widetilde{\gamma}$ need not be in $\widetilde{H}$ since it is not necessarily integral. We consider two cases:

(1) Suppose that $(a, p) = 1$. Since $(a, c) = 1$ we have $(a, pc) = 1$, and thus there exist $r, s \in \mathbf{Z}$ with $as - rpc = 1$, and hence

$$C = \begin{pmatrix} a & r \\ pc & s \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}).$$

For such a C, we have, with $t = bs - dpr$, the identity

$$A^{-1}BA = C \begin{pmatrix} 1 & t/p \\ 0 & 1 \end{pmatrix}.$$

But since $\begin{pmatrix} 1 & t/p \\ 0 & 1 \end{pmatrix}$ commutes with $\begin{pmatrix} 1 & n/p \\ 0 & 1 \end{pmatrix}$, it follows that we may write

$$\widetilde{\gamma} = (A^{-1}BA) \begin{pmatrix} 1 & n/p \\ 0 & 1 \end{pmatrix} (A^{-1}BA)^{-1} = C \begin{pmatrix} 1 & n/p \\ 0 & 1 \end{pmatrix} C^{-1},$$

We now have

$$(\widetilde{\gamma})^p = C \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} C^{-1} \in \mathrm{SL}_2(\mathbf{Z})$$

and thus in $\widetilde{H}$. Hence either $\widetilde{h} = \widetilde{\gamma}$ if $\widetilde{\gamma}$ lies in $\mathrm{SL}_2(\mathbf{Z})$ or $\widetilde{h} = (\widetilde{\gamma})^p$. In particular, the cusp width at this cusp is either n or n/p and certainly divides N and hence also Np.

(2) Suppose that $p|a$, so p does not divide c, so a/p and c are co-prime integers. Now take

$$C = \begin{pmatrix} a/p & b \\ c & pd \end{pmatrix} = (A^{-1}BA) \begin{pmatrix} p & 0 \\ 0 & 1/p \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}).$$

Then

$$\widetilde{\gamma} = (A^{-1}BA) \begin{pmatrix} 1 & n/p \\ 0 & 1 \end{pmatrix} (A^{-1}BA)^{-1} = C \begin{pmatrix} 1 & np \\ 0 & 1 \end{pmatrix} C^{-1},$$

and hence $\widetilde{h} = \widetilde{\gamma}$ and the cusp width at this cusp is np which divides Np. □

4.2. **Modular Forms.** For an even integer N, we will consider the following spaces of modular functions with rational coefficients generated by forms with bounded denominators, that is, subspaces of $\mathbf{Q}(\!(q^{1/N})\!) = \mathbf{Q}[\![q^{1/N}]\!][1/q]$ generated by elements of $\mathbf{Z}[\![q^{1/N}]\!] \otimes \mathbf{Q}$ as a $\mathbf{Q}(\lambda)$-vector space.

Definition 4.2.1.

(1) *Let M_N denote the $\mathbf{Q}(\lambda)$-vector space generated by holomorphic modular functions on the modular curve $Y(N) = \mathbf{H}/\langle E, \Gamma(N)\rangle$ with coefficients in $\mathbf{Q}$ at the cusp $\zeta = i\infty$.*

(2) *Let R_N denote the $\mathbf{Q}(\lambda)$-vector space generated by holomorphic modular functions with coefficients in $\mathbf{Q}$, bounded denominators at the cusp $\zeta = i\infty$, and cusp widths dividing N at all cusps $\zeta \in \mathbf{P}^1(\mathbf{Q})$.*

(The vector space R_N was also defined in Proposition 3.0.1 but we repeat the definition here for convenience.) For example, the (weight 0) holomorphic modular forms on $Y(2)$ are given by $\mathbf{Q}[\lambda, 1/\lambda, 1/(1-\lambda)]$, and the $\mathbf{Q}(\lambda)$-vector space generated by such elements inside $\mathbf{Q}(\!(q^{1/N})\!)$ is $M_2 = \mathbf{Q}(\lambda)$.

Lemma 4.2.2. *There is a containment $M_N \subset R_N$, and M_N and R_N have finite dimensions over $M_2 = \mathbf{Q}(\lambda)$.*

Proof. Let f be a holomorphic modular function on $Y(N)$, that is, a meromorphic function on the compact modular curve $X(N)$ whose poles are all at the cusps. Assume also that f has coefficients in $\mathbf{Q}$ at the cusp $\zeta = i\infty$, Then the modular form $f\Delta(\tau)^m$ is holomorphic at the cusps for sufficiently large m. Moreover, $f\Delta(\tau)^m$ has coefficients in $\mathbf{Q}$. It follows from [Shi71, Theorem 3.52] that $f\Delta(\tau)^m$ has bounded denominators, and this implies the same for f, and thus there is a containment $M_N \subset R_N$.

The second claim follows from Corollary 2.0.5 and the remark (cf. the second paragraph of §1.1.3) that the conformal radius of $\mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ is strictly larger than 1. Indeed, $\sqrt[N]{\lambda(z^N)/16} : D(0,1) \to \mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ is a well-defined holomorphic map with unit derivative at the origin, and hence by Schwarz's lemma the universal covering $D(0,1) \to \mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ has derivative strictly larger than 1 in absolute value. (Later, in Theorem 5.1.4 below, we will exactly compute this latter derivative.) □

We have the following refinement of Lemma 4.2.2:

Lemma 4.2.3. *The vector spaces M_N and R_N are fields. The space M_N may be identified with the field of rational functions on the modular curve $Y(N)/\mathbf{Q}$. There are injective algebra maps*

$$M_2 \to M_N \to R_N.$$

The space R_N is invariant under a normal finite index subgroup $G_N \subset \langle E, \Gamma(N)\rangle \subset \mathrm{SL}_2(\mathbf{Z})$ containing E with $L(G_N) = N$.

Proof. Note that M_N and R_N are subspaces of $\mathbf{Q}((q^{1/N}))$, which is a domain. Hence if M_N and R_N are rings then they are also integral domains, and any integral domain which has finite dimension over a field is itself a field.

The curve $Y(N)$ has a standard model over $\mathbf{Q}$ (as a moduli space of elliptic curves C with a given symplectic isomorphism $C[N] \simeq \mathbf{Z}/N\mathbf{Z} \oplus \mu_N$) such that the cusp $\zeta = i\infty$ is defined over $\mathbf{Q}$, and the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the global sections of $Y(N)$ is compatible with the q-expansion map. It follows that the set of generators of M_N is closed under addition and multiplication and hence that M_N is a ring, and thus a field. Moreover, M_N contains the global sections of the (affine) curve $Y(N)/\mathbf{Q}$, and hence M_N must be the function field of $Y(N)/\mathbf{Q}$.

The vector space R_N is generated by holomorphic modular forms with bounded denominators at $\zeta = i\infty$. To show R_N is a ring, it suffices to show that the product of any two such generators g and h is also a generator. Certainly gh is a holomorphic modular form with rational coefficients and bounded denominators, so it suffices to show that the cusp width still divides N. But we may assume that g and h are invariant under finite index subgroup $G, H \subset \mathrm{SL}_2(\mathbf{Z})$ containing E, and thus gh is invariant under $G \cap H$. It follows from Lemma 4.1.3 that $L(G \cap H)$ also has Wohlfahrt level dividing N.

Since R_N is finite over M_2, it is generated by a finite number of basis elements each of which is invariant under some finite index subgroup $\Phi \subset \mathrm{SL}_2(\mathbf{Z})$ containing E with $L(\Phi)$ dividing N. The intersection of all these groups still has finite index and level N by Lemma 4.1.3, and then we take G_N to be the largest normal subgroup of $\mathrm{SL}_2(\mathbf{Z})$ contained in this intersection, which also has $L(G_N) = N$ by Lemma 4.1.4. □

4.3. **A leveraging argument.** Let us assume that there exists an (even) N such that R_N is strictly larger than M_N. Let $f(\tau) \in \mathbf{Z}[\![q^{1/N}]\!] \in R_N$ be an element which does not lie in M_N. Recall that all forms in R_N and thus in particular f is invariant by a subgroup $G = G_N \subset \langle E, \Gamma(N)\rangle$ which is normal with finite index in $\langle E, \Gamma(N)\rangle$ and has $L(G_N) = N$ by Lemma 4.2.3. The main idea of this section is to exploit the fact that $f(p\tau) \in \mathbf{Z}[\![q^{1/N}]\!]$ is also a modular form with integer coefficients for any prime p. Since the form $f(\tau)$ is invariant under G, the form $f(p\tau)$ is invariant under $A^{-1}GA$ and thus also the group $A^{-1}GA \cap \mathrm{SL}_2(\mathbf{Z})$. Now, by Lemma 4.1.7, we know that this group has (Wohlfahrt) level dividing Np. In particular $f(p\tau)$ has cusp width dividing Np at each cusp, and hence $f(p\tau) \in R_{Np}$.

Our main result is as follows:

Theorem 4.3.1. *Suppose that $(p, N) = 1$ is prime. Suppose that $f(\tau)$ is* not *invariant under a congruence subgroup. Then the form $f(p\tau)$ is not in the M_{Np}-algebra generated by R_N.*

That is, we can leverage one exception to the unbounded denominators to produce many examples. Before proving Theorem 4.3.1 (whose proof is deferred to §4.4), we first draw the following consequence:

Theorem 4.3.2. *Suppose that there exists an N such that $[R_N : M_N] > 1$. Then for any prime p not dividing N, one has*

$$[R_{Np} : M_{Np}] \geq 2[R_N : M_N].$$

Proof. Let $f(\tau)$ be a form in R_{Np} which is not in M_N. By Theorem 4.3.1, we deduce that $f(p\tau) \in R_{Np}$ is not in the M_{Np}-algebra $R_N M_{Np}$ generated by R_N (which is a subfield of R_{Np}). We have

$$[R_{Np} : M_N] = [R_{Np} : R_N M_{Np}][R_N M_{Np} : M_N] = [R_{Np} : R_N M_{Np}][R_N : M_N][M_{Np} : M_N],$$

because the intersection of M_{Np} and R_N is M_N. Thus

$$\frac{[R_{Np} : M_{Np}]}{[R_N : M_N]} = [R_{Np} : R_N M_{Np}]$$

is an integer which is ≥ 2, which implies Theorem 4.3.2. □

Our goal is to prove that $R_N = M_N$. We can and do assume that N is even. As noted in Lemma 4.2.3, $[R_N : M_2] < \infty$. The degree of M_N over M_2 is equal to the degree of the modular curve $Y(N)$ over $Y(2)$, and this is given by the explicit formula

$$\begin{aligned}[M_N : M_2] = \frac{1}{2}[\Gamma(2) : \Gamma(N)] &= \frac{N^3}{2[\mathrm{SL}_2(\mathbf{Z}) : \Gamma(2)]} \prod_{p|N} \left(1 - \frac{1}{p^2}\right) \\ &> \frac{N^3}{2[\mathrm{SL}_2(\mathbf{Z}) : \Gamma(2)]} \prod_{p} \left(1 - \frac{1}{p^2}\right) = \frac{N^3}{12\zeta(2)}.\end{aligned} \tag{4.3.3}$$

Since $[R_N : M_2] = [R_N : M_N] \cdot [M_N : M_2]$, it follows that we have a bound:

$$[R_N : M_N] \le \frac{12\zeta(2)[R_N : M_2]}{N^3} \tag{4.3.4}$$

for all N. We can now compare this bound against the one coming from Theorem 4.3.2

Proposition 4.3.5. *Suppose that there exists a constant C and a bound*

$$[R_N : M_2] \le CN^3 \log N$$

for all even integers N. Then $R_N = M_N$ for every N, that is, the unbounded denominators conjecture holds.

Proof. Assume there exists an N such that $R_N \ne M_N$. Let S denote the set of primes $< X$ which are co-prime to N. By induction, Theorem 4.3.2 implies for such an N that

$$[R_{N \prod_{p \in S} p} : M_{N \prod_{p \in S} p}] \ge 2^{\# S} > 2^{(1-\varepsilon)X/\log X}, \tag{4.3.6}$$

by the prime number theorem. The RHS certainly increases faster than any power of X. On the other hand, from the assumed bound on $[R_N : M_2]$ together with the bound (4.3.4), we obtain

$$\begin{aligned}[R_{N \prod_{p \in S} p} : M_{N \prod_{p \in S} p}] &\le 12C\zeta(2) \log \left(N \prod_{p \in S} p \right) \\ &= 12C\zeta(2) \log N + 12C\zeta(2) \sum_{p \in S} \log p < 12C\zeta(2)X(1+\varepsilon),\end{aligned} \tag{4.3.7}$$

where the last inequality follows from the prime number theorem. Combining the bounds (4.3.6) and (4.3.7) gives

$$2^{(1-\varepsilon)X/\log X} < 12C\zeta(2)X(1+\varepsilon),$$

which (by some margin!) is a contradiction for sufficiently large X. □

Remark 4.3.8. The argument still works with a bound weaker than $[R_N : M_2] \ll N^3 \log N$, although $[R_N : M_2] \ll N^{3+\varepsilon}$ would not be strong enough.

4.4. **Amalgams and a non-abelian version of Ihara's Lemma.** Since $G = G_N$ is normal and is contained in $\langle E, \Gamma(N) \rangle$, we may define a group S by taking $S = \langle E, \Gamma(N) \rangle / G$. By construction, the group S is finite. There is a natural projection:

$$f : \langle E, \Gamma(N) \rangle \to \langle E, \Gamma(N) \rangle / G = S.$$

We define two homomorphisms f_1 and f_2 from $\langle E, \Gamma(N) \rangle \cap \Gamma_0(p)$ to S as follows:

(1) The map f_1 is the restriction of f to $\langle E, \Gamma(N) \rangle \cap \Gamma_0(p)$ under the natural inclusion

$$\langle E, \Gamma(N) \rangle \cap \Gamma_0(p) \to \langle E, \Gamma(N) \rangle,$$

so $f_1(x) = f(x)$.

(2) Conjugation by A induces an isomorphism

$$\langle E, \Gamma(N) \rangle \cap \Gamma_0(p) \to \langle E, \Gamma(N) \rangle \cap \Gamma^0(p), \quad \gamma \longrightarrow A\gamma A^{-1}.$$

The map f_2 is the composition of this map composed with f, so $f_2(x) = f(AxA^{-1})$.

Lemma 4.4.1 (Serre, Berger)**.** *The map $(f_1, f_2) : \langle E, \Gamma(N) \rangle \cap \Gamma_0(p) \to S \times S$ is surjective.*

This is more or less precisely [Tho89, Theorem 3] with the addition of level structure as in [Ber94]. One may think of Lemma 4.4.1 as a non-abelian version of Ihara's Lemma, because (as explained below in the proof of Lemma 4.6.2) the case when S is a vector space over $\mathbf{F}_q$ reduces to the statement of Ihara's Lemma as proved by Ribet [Rib84]. (The proofs of both claims are very similar.)

Proof. The intersection of $\langle E, \Gamma(N)\rangle$ with $A\langle E, \Gamma(N)\rangle A^{-1}$ is the group $\langle E, \Gamma(N)\rangle \cap \Gamma_0(p)$. We proceed by contradiction. Assume that the map (f_1, f_2) is not surjective. By Goursat's lemma, there exists a non-trivial quotient Δ of S and projections $\pi_i : S \to \Delta$ such that the composites $\pi_1 \circ f_1$ and $\pi_2 \circ f_2$ agree. We define a map g_1 by the composite

$$g_1 : \langle E, \Gamma(N)\rangle \xrightarrow{\ f\ } S \xrightarrow{\ \pi_1\ } \Delta \tag{4.4.2}$$

and a map g_2 by the composite

$$g_2 : A^{-1}\langle E, \Gamma(N)\rangle A \longrightarrow \langle E, \Gamma(N)\rangle \xrightarrow{\ f\ } S \xrightarrow{\ \pi_2\ } \Delta \tag{4.4.3}$$

where the first map sends $x \to AxA^{-1}$. On the intersection

$$\langle E, \Gamma(N)\rangle \cap A^{-1}\langle E, \Gamma(N)\rangle A = \langle E, \Gamma(N)\rangle \cap \Gamma_0(p),$$

the restriction of g_1 is given by $\pi_1 \circ f_1$ and the restriction of g_2 is given by $\pi_2 \circ f_2$. By construction these maps coincide, and hence they induce a surjective map on the amalgam

$$\Phi := \langle E, \Gamma(N)\rangle \star_{\langle E,\Gamma(N)\rangle \cap \Gamma_0(p)} A^{-1}\langle E, \Gamma(N)\rangle A \to \Delta.$$

There are natural inclusions from $\langle E, \Gamma(N)\rangle$ and $A^{-1}\langle E, \Gamma(N)\rangle A$ to the congruence subgroup of $\mathrm{SL}_2(\mathbf{Z}[1/p])$ consisting of matrices congruent to $\pm I \bmod N$, and these inclusions induce a map from Φ to this congruence subgroup. This map is an isomorphism ([Ber94, p.919], using ideas of [Ser80] and following the proof of [Tho89, Theorem 3]). But the group $\mathrm{SL}_2(\mathbf{Z}[1/p])$ (and thus the congruence subgroup Φ) satisfies the congruence subgroup property [Men67, Ser70]. Hence the map $\Phi \to \Delta$ is a congruence map, and thus the same is true for the restriction to $\langle E, \Gamma(N)\rangle \subset \Phi$. This implies that the kernel $K \supseteq G$ of the map

$$\langle E, \Gamma(N)\rangle \to \langle E, \Gamma(N)\rangle/G = S \to \Delta$$

is a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$ containing E and strictly contained in $\langle E, \Gamma(N)\rangle$. But this contradicts the assumption that the Wohlfahrt level of G is N, because the smallest congruence subgroup of Wohlfahrt level N containing E is precisely $\langle E, \Gamma(N)\rangle$ by [Woh64, Theorem 2]. □

Let $B \subset \mathrm{SL}_2(\mathbf{F}_p)$ denote the Borel subgroup of upper triangular matrices. There is a natural surjection $\pi : \langle E, \Gamma(N)\rangle \cap \Gamma_0(p) \to B$ whose kernel is $\Gamma(Np)$. We have the following extension of Lemma 4.4.1.

Lemma 4.4.4. *The map $(f_1, f_2, \pi) : \langle E, \Gamma(N)\rangle \cap \Gamma_0(p) \to S \times S \times B$ is surjective.*

Proof. Let $\gamma = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}$ and $\eta = \begin{pmatrix} 1 & 0 \\ N & 1 \end{pmatrix}$. The assumption that $L(G) = N$ and G has finite index in $\mathrm{SL}_2(\mathbf{Z})$ impies that $\gamma, \eta \in G$. Since $A^{-1}\gamma^p A^{-1} = \gamma \in A^{-1}GA \cap \Gamma_0(p)$, we see that $\gamma \in \ker(f_1)$ and $\gamma \in \ker(f_2)$, and yet

$$\pi(\gamma) = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \in B$$

generates the normal unipotent subgroup $U \subset B$. By Goursat's Lemma, we can detect the failure of surjectivity coming from a map of $S \times S$ and B to some common quotient. Because the image contains $0 \times 0 \times \langle U\rangle$, this common quotient is a quotient of the abelian group $B/\langle U\rangle$. Thus, by Nakayama's Lemma, the failure of surjectivity can be detected by maps to $\mathbf{F}_q$ for primes q. Maps to $\mathbf{F}_q$ are determined by cohomology classes with coefficients in $\mathbf{F}_q$. Let $\mathrm{S}G := G \cap \Gamma(N)$. Since $E \in G$, we have

$$S \simeq \langle E, \Gamma(N)\rangle/G \simeq \Gamma(N)/\mathrm{S}G,$$

and so the map (f_1, f_2) remains surjective after restriction to $\Gamma(N) \cap \Gamma_0(p)$. The surjectivity of (f_1, f_2) implies the injectivity of the map

$$H^1(\Gamma(N)/SG, \mathbf{F}_q)^{\oplus 2} = H^1(S, \mathbf{F}_q)^2 \to H^1(\Gamma(N) \cap \Gamma_0(p), \mathbf{F}_q), \tag{4.4.5}$$

The assumption that $L(G) = N$ implies that

$$H^1(\Gamma(N)/SG, \mathbf{F}_q) \cap H^{1,\mathrm{cong}}(\Gamma(N), \mathbf{F}_q) = 0 \in H^1(\Gamma(N), \mathbf{F}_q), \tag{4.4.6}$$

where $H^{1,\mathrm{cong}}(\Gamma(N), \mathbf{F}_q) \subset H^1(\Gamma(N), \mathbf{F}_q)$ denotes the classes which vanish after restriction to a congruence subgroup (Definition 4.5.1). This is because the kernel of any non-trivial map in $H^{1,\mathrm{cong}}(\Gamma(N), \mathbf{F}_q)$ has level strictly divisible by N. The claim (4.4.5) follows from (4.4.6) as a consequence of Ihara's Lemma, as proved by Ribet [Rib84] (see Lemma 4.6.2). The maps $\langle E, \Gamma(N) \rangle \cap \Gamma_0(p) \to B/\langle U \rangle \to \mathbf{F}_q$ on the other hand come from the classes in $H^1(\Gamma(N) \cap \Gamma_0(p), \mathbf{F}_q)$ which restricts to zero on $H^1(\Gamma(N) \cap \Gamma_1(p), \mathbf{F}_q)$, and thus what is required is to upgrade the injection of (4.4.5) to an injection

$$H^1(\Gamma(N)/SG, \mathbf{F}_q)^{\oplus 2} = H^1(S, \mathbf{F}_q)^2 \to H^1(\Gamma(N) \cap \Gamma_1(p), \mathbf{F}_q), \tag{4.4.7}$$

which is dual to the desired claim that the map

$$\Gamma(N) \cap \Gamma_0(p) \to S^{\mathrm{ab}}/qS^{\mathrm{ab}} \times S^{\mathrm{ab}}/qS^{\mathrm{ab}} \times B/U$$

is surjective. But now we may invoke an enhanced version of Ihara's Lemma (Lemma 4.6.3) which we prove in §4.6, and the injectivity of (4.4.7) follows directly from (4.4.6). □

Remark 4.4.8. Because γ and E map to zero in $S \times S$ and $B/\langle E, U \rangle$ has order $(p-1)/2$, The proof of Lemma 4.4.4 is almost immediate if one imposes the additional hypothesis that $\left(\frac{p-1}{2}, |S|\right) = 1$. In particular, one would not have to appeal to the results in §4.5 and §4.6 (which are not used elsewhere in this paper). It turns out that proving Lemma 4.4.4 under this weaker hypothesis would suffice for the proof of the unbounded denominators conjecture. The key point is that if $\langle E, \Gamma(N) \rangle / G_N \simeq S$ and G_{Np} is the group given by the intersection of the three groups $\langle E, \Gamma(Np) \rangle$, G, and AGA^{-1}, then $\langle E, \Gamma(Np) \rangle / G_{Np} \simeq S \times S$. In particular, one can control the primes dividing S as one varies N. Then, in the argument of Proposition 4.3.5, instead of adding all primes $< X$ prime to N, one only includes primes in some arithmetic progression satisfying the congruence $\left(\frac{p-1}{2}, |S|\right) = 1$ for some fixed S. However, it seems more natural to prove Lemma 4.4.4 without such an ugly hypothesis. Additionally, sections §4.5 and §4.6 may be of independent interest.

Returning to the assumptions of Lemma 4.4.4, let $K = \langle \ker((f_1, \pi), \ker(f_2) \rangle$ be the group generated by $\ker((f_1, \pi))$ and $\ker(f_2)$. We deduce from Lemma 4.4.4 that the image of (f_1, f_2, π) contains the elements $(x, 0, z)$ and $(0, y, 0)$ for any triple $(x, y, z) \in S \times S \times B$. But the pre-images of these elements clearly lie in $\ker(f_2)$ and $\ker((f_1, \pi))$ respectively, and thus lie in K. But then the pre-image of any element lies in K, and we deduce that $K = \langle E, \Gamma(N) \rangle \cap \Gamma_0(p)$, or equivalently that

$$\langle E, G \cap \Gamma(Np), A^{-1}GA \cap \Gamma_0(p) \rangle = \langle E, \Gamma(N) \rangle \cap \Gamma_0(p). \tag{4.4.9}$$

We now complete the proof of Theorem 4.3.1 and hence the proof of Theorem 4.3.2 (as explained at the beginning of § 4.3).

Proof of Theorem 4.3.1 . Consider the function $f(p\tau)$. Assume that this lies in the algebra generated by $f(\tau)$ and M_{Np}. Then $f(p\tau)$ is invariant under both $A^{-1}GA \cap \mathrm{SL}_2(\mathbf{Z})$ and $G \cap \Gamma(Np)$. But from (4.4.9) we see that these groups together generate a congruence subgroup, and thus $f(p\tau)$ and $f(\tau)$ are congruence, a contradiction. □

4.5. **Invariant vectors.** We recall the following definition (cf. [CV19, §3.7]).

Definition 4.5.1. *Let $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ be a congruence subgroup. A* congruence class *$\eta \in H^1(\Gamma, \mathbf{F}_\ell)$ is a class that restricts to zero on some congruence subgroup $\Gamma' \subset \Gamma$. Denote the subgroup of congruence classes by*

$$H^{1,\mathrm{cong}}(\Gamma, \mathbf{F}_\ell) \subset H^1(\Gamma, \mathbf{F}_\ell).$$

If $\widehat{\Gamma}$ denotes the congruence completion of the group Γ, then $H^{1,\mathrm{cong}}(\Gamma, \mathbf{F}_\ell) \simeq H^1(\widehat{\Gamma}, \mathbf{F}_\ell)$. For a prime ℓ, one may define ([CE12]) the groups

$$\widetilde{H}^1(\mathbf{F}_\ell) := \lim_{N} H^1(\Gamma(N), \mathbf{F}_\ell), \quad \widetilde{H}^1(\mathbf{Q}/\mathbf{Z}) := \lim_{N} H^1(\Gamma(N), \mathbf{Q}/\mathbf{Z}).$$

over all levels N. The limit has an action of the group $\mathrm{SL}_2(\widehat{\mathbf{Z}}) = \prod_p \mathrm{SL}_2(\mathbf{Z}_p)$. The goal of this section is to prove:

Theorem 4.5.2. *The $\mathrm{SL}_2(\widehat{\mathbf{Z}})$-invariant subspace of $\widetilde{H}^1(\mathbf{F}_\ell)$ is trivial.*

It follows that the $\mathrm{SL}_2(\widehat{\mathbf{Z}})$-invariant subspace of $\widetilde{H}^1(\mathbf{Q}/\mathbf{Z})$ is also trivial. We shall use Theorem 4.5.2 in the following equivalent form.

Corollary 4.5.3. *Let N be an integer,and $\eta \in H^1(\Gamma(N), \mathbf{F}_\ell)$. If, for all $g \in \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$, the class $g\eta - \eta \in H^1(\Gamma(N), \mathbf{F}_\ell)$ is a congruence, then η is congruence.*

Proof. The assumptions impliy that the image of η in $\widetilde{H}^1(\mathbf{F}_\ell)$ is $\mathrm{SL}_2(\widehat{\mathbf{Z}})$-invariant, and thus zero. But the kernel of the map $H^1(\Gamma(N), \mathbf{F}_\ell) \to \widetilde{H}^1(\mathbf{F}_\ell)$ consists precisely of congruence classes. $\square$

Our first goal is to control the group $H^2(\widehat{\Gamma}(N), \mathbf{F}_\ell)$ for various N, in particular for $N = 1$, which we do in a sequence of steps.

Lemma 4.5.4. *We have $H_2(\mathrm{SL}_2(\mathbf{F}_p), \mathbf{Z}) = 0$ for all primes p.*

Proof. It suffices to prove the vanishing of $H_2(\Delta, \mathbf{Z})$ for any Sylow subgroup of $\mathrm{SL}_2(\mathbf{F}_p)$. For odd primes, the Sylow subgroup is cyclic and the cohomology of a cyclic group is only non-zero in even degree. The 2-Sylow subgroup is a generalized quaternion group, whose cohomology also vanishes in odd degree (as follows from [Hup67, Satz 25.3(a), p.643] and [Swa60, Theorem 2]). $\square$

Lemma 4.5.5. *For $n = 1$ and $n = 2$, we have:*

$$H^n(\mathrm{SL}_2(\mathbf{F}_p), \mathbf{F}_\ell)^\vee \simeq H_n(\mathrm{SL}_2(\mathbf{F}_p), \mathbf{F}_\ell) = \begin{cases} \mathbf{F}_\ell, & p = \ell \in \{2,3\} \\ 0, & \textit{otherwise.} \end{cases}$$

Proof. There is a short exact sequence:

$$0 \to H_2(G, \mathbf{Z})/\ell \to H_2(G, \mathbf{F}_\ell) \to H_1(G, \mathbf{Z})[\ell] \to 0,$$

and $H_1(G, \mathbf{F}_\ell) \simeq H_1(G, \mathbf{Z})/\ell$. Hence the result follows from combining Lemma 4.5.4 with the fact that $\mathrm{SL}_2(\mathbf{F}_p)^{\mathrm{ab}}$ is trivial for $p \geq 5$ and $\mathbf{Z}/p\mathbf{Z}$ for $p = 2$ and $p = 3$. $\square$

Lemma 4.5.6. *For $n = 1$, we have:*

$$H^1(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_\ell) = \begin{cases} \mathbf{F}_\ell, & p = \ell \in \{2,3\} \\ 0, & \textit{otherwise.} \end{cases}$$

For $n = 2$, we have $H^2(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_\ell) = 0$ unless $\ell = p$ and $p \leq 5$.

Remark 4.5.7. We shall compute the exceptional cases when $\ell = p \leq 5$ in Lemma 4.5.11 below as a consequence of Theorem 4.5.2.

Proof. Assume that $\ell \neq p$. By Hochschild–Serre, we have an isomorphism

$$H^*(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_\ell) \simeq H^*(\mathrm{SL}_2(\mathbf{F}_p), \mathbf{F}_\ell)$$

and thus the result follows from Lemma 4.5.5. Thus we may assume that $\ell = p$. Assume that $p > 2$. Let $G(p)$ be the p-congruence subgroup of $\mathrm{SL}_2(\mathbf{Z}_p)$. We may assume that $p > 2$. Then $G(p)$ is p-torsion free and p-powerful, so, with $M = M_0(\mathbf{F}_p)^\vee$ where $G(p)/G(p^2) \simeq M_0(\mathbf{F}_p)$, we deduce by Lazard's Theorem ([Laz65, Chapter V, 2.2.6.3 and 2.2.7.2, page 167]) that there are isomorphisms

$$H^1(G(p), \mathbf{F}_p) \simeq M, \quad H^2(G(p), \mathbf{F}_p) \simeq \wedge^2 M,$$

where the cup product map $\wedge^2 M : H^1 \wedge H^1 \to H^2$ is an isomorphism. Assuming $p \geq 3$, we find that $M \simeq M^\vee$ is self-dual as a $\mathrm{SL}_2(\mathbf{F}_p)$-module and so $\wedge^2 M \simeq M$. Moreover, we have $M^{\mathrm{SL}_2(\mathbf{F}_p)} = 0$. Consider the Hochschild–Serre spectral sequence:

$$H^i(\mathrm{SL}_2(\mathbf{F}_p), H^j(G(p), \mathbf{F}_p)) \Rightarrow H^{i+j}(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_p).$$

Since $H^1(\mathrm{SL}_2(\mathbf{F}_p), \mathbf{F}_p) = H^1(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_p)$ and $M^{\mathrm{SL}_2(\mathbf{F}_p)} = 0$, there is an exact sequence:

$$0 \to H^2(\mathrm{SL}_2(\mathbf{F}_p), \mathbf{F}_p) \to H^2(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_p) \to H^1(\mathrm{SL}_2(\mathbf{F}_p), M). \tag{4.5.8}$$

If $p \neq 5$, then $H^1(\mathrm{SL}_2(\mathbf{F}_p), M) = 0$ (see [DDT97, Lemma 2.48]) and the result follows from Lemma 4.5.5. $\square$

We deduce:

Lemma 4.5.9. *For every prime ℓ, there is an isomorphism $H^2(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_\ell) \simeq H^2(\mathrm{SL}_2(\mathbf{Z}_\ell), \mathbf{F}_\ell)$. If N is a power of ℓ and $G(N) \subset \mathrm{SL}_2(\mathbf{Z}_\ell)$ the corresponding principal congruence subgroup, then*

$$H^2(\widehat{\Gamma}(N), \mathbf{F}_\ell) \simeq H^2(G(N), \mathbf{F}_\ell).$$

If ℓ is odd and N is a non-trivial power of ℓ or $N \geq 8$ is a power of $\ell = 2$, then the map

$$H^2(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_\ell) \to H^2(\widehat{\Gamma}(N), \mathbf{F}_\ell)$$

is trivial.

Proof. Since $H^n(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_\ell) = 0$ for $n = 1$ and $n = 2$ unless $\ell = p$, the first two claims follow from the Künneth formula and Lemma 4.5.6. It remains to show that the map

$$H^2(\mathrm{SL}_2(\mathbf{Z}_p), \mathbf{F}_p) \to H^2(G(N), \mathbf{F}_p)$$

is the zero map for $N = p$ if p is odd and $N = 8$ if $p = 2$. For $p > 2$, we have $H^2(G(N), \mathbf{F}_p) \simeq M$ and $M^{\mathrm{SL}_2(\mathbf{F}_p)} = 0$. Since the source is $\mathrm{SL}_2(\mathbf{F}_p)$-invariant, the image must be trivial. For $p = 2$, we have $H^2(G(N), \mathbf{F}_2) \simeq \wedge^2(M) \simeq M$ whenever $N \geq 4$ (to ensure that $G(N)$ is 2-powerful). Unlike what happens for p odd, we have $M^{\mathrm{SL}_2(\mathbf{F}_2)} = \mathbf{F}_2$. However, the map

$$M = H^1(G(4), \mathbf{F}_2) \to H^1(G(8), \mathbf{F}_2) = M$$

is zero, and thus the induced map

$$\wedge^2 M = H^2(G(4), \mathbf{F}_2) \to H^2(G(8), \mathbf{F}_2) = \wedge^2 M$$

is also zero. $\square$

Now let us consider the following commutative diagram for $N \in \{3, 4, 5, 8, 16\}$ and ℓ dividing N coming from compatible Hochschild–Serre spectral sequences:

$$\begin{tikzcd}
0 \arrow[r] & H^1(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_\ell) \arrow[r] \arrow[d] & H^1(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_\ell) \arrow[r] \arrow[d] & (\widetilde{H}^1(\mathbf{F}_\ell))^{\mathrm{SL}_2(\widehat{\mathbf{Z}})} \arrow[r] \arrow[d, hook] & H^2(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_\ell) \arrow[r] \arrow[d] & H^2(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_\ell) \\
0 \arrow[r] & H^1(\widehat{\Gamma}(N), \mathbf{F}_\ell) \arrow[r] & H^1(\Gamma(N), \mathbf{F}_\ell) \arrow[r] & (\widetilde{H}^1(\mathbf{F}_\ell))^{\widehat{\Gamma}(N)} \arrow[r] & H^2(\widehat{\Gamma}(N), \mathbf{F}_\ell) \arrow[r] & 0.
\end{tikzcd}$$

Here $H^2(\Gamma(N), \mathbf{F}_\ell) = 0$ because $\Gamma(N)$ is a free group. The last vertical map is zero by the previous lemma if $N \in \{3, 5, 8, 16\}$, and thus the image of $(\widetilde{H}^1)^{\mathrm{SL}_2(\widehat{\mathbf{Z}})}$ in $(\widetilde{H}^1)^{\widehat{\Gamma}(N)}$ lands in the image of $H^1(\widehat{\Gamma}(N), \mathbf{F}_\ell)$ in these cases. But these are finite groups we can compute explicitly.

Lemma 4.5.10. *We have*

$$\dim H^1(\Gamma(3), \mathbf{F}_3)^{\mathrm{SL}_2(\mathbf{F}_3)} = 0,$$
$$\dim H^1(\Gamma(5), \mathbf{F}_3)^{\mathrm{SL}_2(\mathbf{F}_5)} = 0,$$
$$\dim H^1(\Gamma(4), \mathbf{F}_2)^{\mathrm{SL}_2(\mathbf{Z}/4\mathbf{Z})} = \dim H^1(\Gamma(8), \mathbf{F}_2)^{\mathrm{SL}_2(\mathbf{Z}/8\mathbf{Z})} = \dim H^1(\Gamma(16), \mathbf{F}_2)^{\mathrm{SL}_2(\mathbf{Z}/16\mathbf{Z})} = 1.$$

For $N = 3, 4, 5$, the same result holds even after considering the semi-simplifications of these modules.

Proof. Recall that for $N = 3$, 4, and 5 that $X(N)$ has genus zero. Hence the cohomology $V = H^1(\Gamma(N), \mathbf{Z})$ is coming entirely from the the cusps, which correspond to the cosets of $\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$ in $\mathrm{PSL}_2(\mathbf{Z}/N\mathbf{Z})$. In particular, in the Grothendieck group,

$$V_{\mathbf{Q}} := \left[H^1(\Gamma(N), \mathbf{Q})\right] \simeq \mathbf{Q}\left[\mathrm{PSL}_2(\mathbf{Z}/N\mathbf{Z})/\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right] - [\mathbf{Q}].$$

Since $\Gamma(N)$ is free, this is enough to determine the semi-simplification of $V_{\mathbf{F}_\ell} := H^1(\Gamma(N), \mathbf{F}_\ell)$. We consider each case in turn.

1. For $N = 3$, we have $\mathrm{PSL}_2(\mathbf{F}_3) = A_4$, and $V_{\mathbf{Q}}$ is absolutely irreducible of dimension 3. The associated Brauer character is also irreducible and so $[V_{\mathbf{F}_\ell}]$ is also irreducible and has no invariants.
2. For $N = 5$, we have $\mathrm{PSL}_2(\mathbf{Z}/5\mathbf{Z}) \simeq A_5$, and $V_{\mathbf{Q}}$ decomposes as a sum of irreducibles of dimensions 3, 3, and 5. The corresponding Brauer characters are all still irreducible, so $[V_{\mathbf{F}_\ell}]$ does not contain the trivial representation.
3. For $N = 4$, we have $\mathrm{PSL}_2(\mathbf{Z}/4\mathbf{Z}) = S_4$, and $V_{\mathbf{Q}}$ is a sum of absolutely irreducible representations of dimensions 2 and 3. The group S_4 has two Brauer characters of dimension 1 and 2 respectively. The 2 dimensional representation remains irreducible and the semi-simplification of both the 3 dimensional representations has constituents of dimensions 1 and 2. Hence the dimension of the invariant space of $V^{\mathrm{ss}}_{\mathbf{F}_\ell}$ is 1-dimensional. But $H^1(\Gamma(4)/\Gamma(8), \mathbf{F}_2)$ is a direct sum of the 1 and 2 dimensional representations, so this one-dimensional constituent occurs as a sub-representation.
4. For $N = 8$ and $N = 16$, we resort to a less elegant calculation; the groups $\Gamma(N)$ are free (of ranks 33 and 257 respectively). The $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$-invariant part of cohomology over $\mathbf{F}_2$ can be determined as (the dual of) the quotient of this group by the relations $x^2 = (xy)^2 = e$ for each generator $x \in \Gamma(N)$ and the relations $gxg^{-1} = x$ for the generators g of $\mathrm{SL}_2(\mathbf{Z})$. In both cases, `magma` determines that the corresponding quotients have order 2.

□

Proof of Theorem 4.5.2. We now complete the proof of Theorem 4.5.2.

Let $v \in \widetilde{H}^1(\mathbf{F}_\ell)^{\mathrm{SL}_2(\widehat{\mathbf{Z}})}$. From the Künneth formula and Lemma 4.5.6, we deduce that there is an isomorphism $H^2(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_\ell) = 0$ for any prime $\ell > 5$. Thus, for $\ell > 5$, any class v lies in the image of $H^1(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_\ell) \simeq \mathbf{Z}/12\mathbf{Z} \otimes \mathbf{F}_\ell$, but one easily checks that the map

$$H^1(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_\ell) \to H^1(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_\ell)$$

is an isomorphism, and thus $v = 0$.

Suppose that $\ell = 3$ or $\ell = 5$. Since the map $H^2(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_\ell) \to H^2(\widehat{\Gamma}(\ell), \mathbf{F}_\ell)$ is zero by Lemma 4.5.9, the image of v lands in the $\mathrm{SL}_2(\mathbf{Z}/\ell\mathbf{Z})$-invariants of $H^1(\Gamma(\ell), \mathbf{F}_\ell)$. Thus it contributes to the $\mathrm{SL}_2(\mathbf{Z}/\ell\mathbf{Z})$-invariants of the semi-simplification of $H^1(\Gamma(\ell), \mathbf{F}_\ell)$ as a $\mathrm{SL}_2(\mathbf{Z}/\ell\mathbf{Z})$-module. But this space has dimension zero by Lemma 4.5.10.

Finally, let $\ell = 2$. By Lemma 4.5.9, the map

$$H^2(\mathrm{SL}_2(\widehat{\mathbf{Z}}), \mathbf{F}_2) \to H^2(\widehat{\Gamma}(8), \mathbf{F}_2)$$

is zero, and thus the element v lies in the image of $H^1(\Gamma(8), \mathbf{F}_2)$. Furthermore, it has the property that the $\mathrm{SL}_2(\mathbf{Z}/8\mathbf{Z})$-module generated by v is trivial after passing to the quotient by the congruence homology

$$H^1(\widehat{\Gamma}(8), \mathbf{F}_2) \simeq H^1(\widehat{\Gamma}(8)/\widehat{\Gamma}(16), \mathbf{F}_2).$$

But that means that the image of v in $H^1(\widehat{\Gamma}(16), \mathbf{F}_2)$ is genuinely invariant under $\mathrm{SL}_2(\mathbf{Z}/16\mathbf{Z})$. By Lemma 4.5.10, the space of such invariants is one dimensional. But this one dimensional space lands in the image of $H^1(\widehat{\Gamma}(16), \mathbf{F}_2)$, and thus v is trivial in $\widetilde{H}^1(\mathbf{F}_2)$. □

We note in passing that this result implies the following strengthening of Lemma 4.5.6:

Lemma 4.5.11. *For $n=1$ and $n=2$ we have:*

$$H^n(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{F}_\ell)=\begin{cases}\mathbf{F}_\ell, & p=\ell\in\{2,3\}\\ 0, & \textit{otherwise.}\end{cases}$$

We also have $H_2(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{Z})=0$ for all p.

Proof. From Theorem 4.5.6, it suffices to consider the case of $n=2$ and $\ell=p\in\{2,3,5\}$. There is an exact sequence:

$$0\to H_2(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{Z})/\ell\to H_2(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{F}_\ell)\to H_1(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{Z})[\ell]\to 0.$$

Since $H_1(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{Z})\simeq\mathbf{Z}/12\mathbf{Z}\otimes\mathbf{Z}_p$, this implies the desired lower bounds for in Lemma 4.5.11, with equality in all cases if and only if $H_2(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{Z})/\ell=0$ for all primes ℓ. Since the group $H_2(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{Z})$ is a finitely generated abelian group, to deduce that $H_2(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{Z})=0$ it suffices to prove the upper bounds.

For the upper bounds, we deduce from Theorem 4.5.2 that there is an injection

$$H^2(\mathrm{SL}_2(\mathbf{Z}_p),\mathbf{F}_p)\to H^2(\mathrm{SL}_2(\widehat{\mathbf{Z}}),\mathbf{F}_p)\to H^2(\mathrm{SL}_2(\mathbf{Z}),\mathbf{F}_p).$$

But for $n>0$ we have

$$H^n(\mathrm{SL}_2(\mathbf{Z}),\mathbf{Z})=\begin{cases}\mathbf{Z}/12\mathbf{Z}, & n\equiv 0 \bmod 2,\\ 0, & n\equiv 1\bmod 2\end{cases},$$

from which it follows that $H^2(\mathrm{SL}_2(\mathbf{Z}),\mathbf{F}_p)=\mathbf{F}_p$ if $p\in\{2,3\}$ and is zero otherwise. This completes the proof. □

4.6. **An enhancement of Ihara's Lemma.** We shall prove an enhanced version of Ihara's Lemma. We begin by recalling Ihara's Lemma. Let q be prime, let $N\ge 3$, and let $(N,p)=1$. There is a homomorphism

$$H^1(\Gamma(N),\mathbf{F}_q)^{\oplus 2}\to H^1(\Gamma(N)\cap\Gamma_0(p),\mathbf{F}_q), \tag{4.6.1}$$

given by the difference of the following two maps:

(1) The map sending $\psi:\Gamma(N)\to\mathbf{F}_q$ to its restriction to $\Gamma(N)\cap\Gamma_0(p)$.
(2) The twisted restriction map coming from viewing $\psi\in H^1(\Gamma(N),\mathbf{F}_q)$ as a map from $\Gamma(N)\to\mathbf{F}_q$ and then considering the map

$$A\psi:\Gamma(N)\cap\Gamma_0(p),\quad g\mapsto\psi(AgA^{-1}).$$

By abuse of notation we denote the restriction of ψ by ψ, so the map sends (ψ,ϕ) to $\psi-A\phi$.

Lemma 4.6.2 (Ihara's Lemma)**.** *The kernel of the map (4.6.1) lies inside $H^{1,\mathrm{cong}}(\Gamma,\mathbf{F}_q)^2$.*

Proof. This version of Ihara's Lemma was essentially proved by Ribet in [Rib84]. The proof is just an abelian version of Lemma 4.4.1. We recall some of the details. Let $\Phi\subset\Gamma(N)$ be the maximal normal subgroup whose quotient is an elementary q-abelian group T. Canonically, we have $\Gamma(N)/\Phi\simeq T$ and $H^1(\Gamma,\mathbf{F}_q)\simeq\mathrm{Hom}(T,\mathbf{F}_q)$. The kernel of the map

$$\mathrm{Hom}(T,\mathbf{F}_q)\times\mathrm{Hom}(T,\mathbf{F}_q)\to H^1(\Gamma(N)\cap\Gamma_0(p),\mathbf{F}_q)$$

is governed by the cokernel of the dual map

$$\Gamma(N)\cap\Gamma_0(p)\to T\times T.$$

Exactly as in the proof of Lemma 4.4.1, we deduce from Goursat's Lemma that the cokernel Δ arises from two maps from $\Gamma(N)$ to Δ which agree along $\Gamma(N)\cap\Gamma_0(p)$, and thus on their amalgam $\mathrm{SL}_2(\mathbf{Z}[1/p])(N)$. Since $\mathrm{SL}_2(\mathbf{Z}[1/p])(N)$ as the congruence subgroup property, it thus arises from a congruence quotient of this group at primes away from p. But that precisely means that the classes in $\mathrm{Hom}(T,\mathbf{F}_q)=H^1(\Gamma(N),\mathbf{F}_q)$ become trivial after passing to a congruence subgroup $\Gamma'\subset\Gamma(N)$, hence the claim. □

Using Corollary 4.5.3, we prove a slight enhancement of this claim.

Lemma 4.6.3 (Ihara's Lemma, enhanced)**.** *The kernel of the composite of the Ihara map (4.6.1) with the map*

$$H^1(\Gamma(N)\cap\Gamma_0(p),\mathbf{F}_q)\to H^1(\Gamma(N)\cap\Gamma_1(p),\mathbf{F}_q) \tag{4.6.4}$$

also lies inside $H^{1,\mathrm{cong}}(\Gamma,\mathbf{F}_q)^2$.

Proof. The map (4.6.1) is $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$-equivariant. But the kernel of the map (4.6.4) is also easily seen to be $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$-invariant. Hence, if (ψ,ϕ) lies in the kernel of the composite of (4.6.1) and (4.6.4), then $(\psi^g-\psi,\phi^g-\phi)$ lies in the kernel of (4.6.1), and thus lies in $H^{1,\mathrm{cong}}(\Gamma(N),\mathbf{F}_q)^2$ by Lemma 4.6.2. But then ψ and ϕ are themselves congruence classes by Corollary 4.5.3. □

5. The uniformization of $\mathbf{C}\smallsetminus\mu_N$

In this section we develop all the particular analytic properties that we need of the universal covering map $F_N : D(0,1)\to\mathbf{C}\smallsetminus\mu_N$ for $N\geq 2$. André has pointed out to us that our two main results here, Theorem 5.1.4 and Lemma 5.2.17, appear in work of Kraus and Roth [KR16, Remark 5.1 and Theorems 1.2 and 1.10]. Nevertheless, as our proofs are simplified to cover our current needs, and since the results of Kraus and Roth rely on some previous work of themselves and others, we keep our self-contained exposition as a convenience to the reader, and refer to [ASVV10, KRS11, KR16] and the references there for various further results and a more thorough study of the uniformization of $\mathbf{C}\smallsetminus\mu_N$. The reader will also benefit from the material in § III.1 in Goluzin's book [Gol69], which recovers F_N via an explicit computation of the Riemann map of a $\mathbf{Z}/N\mathbf{Z}$-rotationally symmetric circular N-gon, taking the case of zero angles and doing Schwarz reflections in the sides of the circular polygon.

Remark 5.0.1 (A word on notation). We denote by $\mathbf{H}$ the upper half plane and by $\mathbf{P}^1=\mathbf{C}\cup\{\infty\}$ the complex projective line or Riemann sphere. There is a conformal isomorphism from the disc $D(0,1)$ to $\mathbf{H}$ by the Cayley transform

$$z\mapsto i\cdot\frac{1+z}{1-z}.$$

This allows one to pass freely between uniformizations by $D(0,1)$ and $\mathbf{H}$. In this section, we choose notation so that the corresponding passage from $D(0,1)$ to $\mathbf{H}$ is marked by the addition of a tilde. Thus, for example, $\widetilde{F}_N$ constructed below denotes a map on $\mathbf{H}$ and F_N (Definition 5.1.1) is simply the pull-back of $\widetilde{F}_N$ to $D(0,1)$ via the map above. Similarly, Γ_N will denote a lattice in $\mathrm{PSU}(1,1)$ whereas $\widetilde{\Gamma}_N$ denotes the corresponding lattice in $\mathrm{PSL}_2(\mathbf{R})$.

Unless we expressly state otherwise, we reserve z,τ,x to denote respectively the coordinates on $D(0,1)$, $\mathbf{H}$, and $\mathbf{C}\smallsetminus\mu_N$.

5.1. **Schwarzians and the conformal radius.**

Let $N\geq 2$ be an integer. Then $\mathbf{C}\smallsetminus\mu_N=\mathbf{P}^1\smallsetminus\{\infty,\mu_N\}$ is the complement of at least 3 points, and thus admits a complex uniformization map:

$$\widetilde{F}_N:\mathbf{H}\to\mathbf{H}/\widetilde{\Gamma}_N=\mathbf{C}\smallsetminus\mu_N,$$

where $\widetilde{\Gamma}_N\subset\mathrm{PSL}_2(\mathbf{R})$ denotes the Fuchsian group of $\mathbf{C}\smallsetminus\mu_N$. The map $\widetilde{F}_N$ is unique up to the action of $\mathrm{PSL}_2(\mathbf{R})$ on the source.

We now pin down the definition of $\widetilde{F}_N$ with the following precise choices. Using the action of $\mathrm{PSL}_2(\mathbf{R})$ we may assume that $\widetilde{F}_N(i)=0$. The stabilizer of $i\in\mathbf{H}$ consists of Möbius transformations given by elements in $\mathrm{SO}_2(\mathbf{R})$, so by specifying $\widetilde{F}_N(i\infty)=1$ (here by abuse of notation, we also use $\widetilde{F}_N$ to denote the natural extension of $\widetilde{F}_N$ on $\mathbf{H}$ to cusps), we determine $\widetilde{F}_N$ uniquely.

Definition 5.1.1. *Define* $F_N:D(0,1)\to\mathbf{C}\smallsetminus\mu_N$ *by the formula*

$$F_N(z)=\widetilde{F}_N\left(i\cdot\frac{1+z}{1-z}\right).$$

Note that F_N is just the map $\widetilde{F}_N$ composed with the standard conformal isomorphism $D(0,1) \to \mathbf{H}$ sending 0 to i, and hence

$$F_N : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$$

is the universal covering map with $F_N(0) = 0$ and $F_N(1) = 1$.

Note that the statements of the main results of this section, Theorem 5.1.4 and Lemma 5.2.17, only depend on the normalization $F_N(0) = 0$ and do not depend on the choice $F_N(1) = 1$.

The following lemma gives the basic symmetric property of $\widetilde{F}_N$ and F_N.

Lemma 5.1.2. *Let $\zeta_N = \exp(2\pi i/N)$ and ζ be any Nth root of unity. Then $\zeta_N \widetilde{F}_N(\tau) = \widetilde{F}_N(\widetilde{r}_N \cdot \tau)$ and $F_N(\zeta x) = \zeta F_N(x)$, where*

$$\widetilde{r}_N = \begin{pmatrix} \cos(\pi/N) & -\sin(\pi/N) \\ \sin(\pi/N) & \cos(\pi/N) \end{pmatrix} \in \mathrm{SO}_2(\mathbf{R}). \tag{5.1.3}$$

Proof. Note that $\zeta_N \widetilde{F}_N$ is another covering map such that $\zeta_N \widetilde{F}_N(i) = 0$. Therefore $\zeta_N \widetilde{F}_N$ must differ from $\widetilde{F}_N$ by a Möbius transformation in the stabilizer of i; that is $\zeta_N \widetilde{F}_N(\tau) = \widetilde{F}_N(\widetilde{r}_N \cdot \tau)$ for some $\widetilde{r}_N \in \mathrm{SO}_2(\mathbf{R})$. We deduce that $\widetilde{F}_N(\widetilde{r}_N^N \cdot \tau) = \zeta_N^N \widetilde{F}_N(\tau) = \widetilde{F}_N(\tau)$, and thus $\widetilde{r}_N^N \in \mathrm{SO}_2(\mathbf{R})$ must also lie in $\widetilde{\Gamma}_N$. But $\widetilde{\Gamma}_N$ is a free group (due to the fact that $\widetilde{F}_N$ is a covering map with no ramification points), and hence $\widetilde{r}_N^N$ is trivial in $\mathrm{SO}_2(\mathbf{R})/\{\pm I\}$, and $\widetilde{r}_N$ is a hyperbolic rotation around i of order N.

The action of $\mathrm{SO}_2(\mathbf{R})$ on $D(0,1)$ under the pullback map is just given by rotation, and hence $\widetilde{r}_N$ acts on $D(0,1)$ by a rotation of order N. We deduce that $F_N(\zeta^m z) = \zeta F_N(z)$ for some $(m,N) = 1$. By taking the derivatives with respect to q of both sides at $q = 0$, we have $\zeta^m F_N'(0) = \zeta F_N'(0)$. Since F_N is a covering map, we must also have $F_N'(0) \neq 0$, and thus $m = 1$ and $F_N(\zeta z) = \zeta F_N(z)$. We thus also deduce (5.1.3) as this action corresponds to the rotation by degree $2\pi/N$ on $D(0,1)$. □

Our first main goal of this section is an explicit computation of the uniformization radius of $\mathbf{C} \smallsetminus \mu_N$. This formula has been previously proved by Kraus and Roth in [KR16, Remark 5.1].

Theorem 5.1.4. *The conformal size $|F_N'(0)|$ (Riemann uniformization radius of $\mathbf{C} \smallsetminus \mu_N$) is equal to*

$$|F_N'(0)| = \gamma_N := 16^{1/N} \frac{\Gamma\left(1 + \dfrac{1}{2N}\right)^2 \Gamma\left(1 - \dfrac{1}{N}\right)}{\Gamma\left(1 - \dfrac{1}{2N}\right)^2 \Gamma\left(1 + \dfrac{1}{N}\right)}. \tag{5.1.5}$$

We have an expansion for γ_N as follows:

$$\gamma_N = 16^{1/N} \left(1 + \frac{\zeta(3)}{2N^3} + \frac{3\zeta(5)}{8N^5} + O(N^{-6})\right), \tag{5.1.6}$$

where the remaining term $O(N^{-6})$ is a positive real number.

To prove this formula, we follow Hempel [Hem88] to get a second order linear ODE whose ratio of two linearly independent solutions gives the (local analytic) inverse of $\widetilde{F}_N$ (Lemma 5.1.8). The uniformization maps of Riemann surfaces—and their inverses—do not typically admit explicit solutions in terms of standard functions, but our particular case of interest turns out to be an exception due to the extra symmetries of $\mathbf{C} \smallsetminus \mu_N$. We use Lemma 5.1.2 to define a function G_N closely related to F_N (see Definition 5.1.13) and explicitly find two solutions of the associated linear ODE in terms of hypergeometric functions (Lemma 5.1.14). These solutions allow us to compute the explicit conformal radius for G_N and deduce the formula (5.1.5) for F_N.

Our computation here is very similar to the treatment by Goluzin in [Gol69, §III.1], who also gives the explicit formula for the inverse of G_N by hypergeometric functions. See the $q = 0$ case of equation (17) and the last paragraph on page 86 of *loc. cit.* Goluzin more generally computes the Riemann map for the $\mathbf{Z}/N\mathbf{Z}$-rotationally symmetric circular N-gon with angles πq, and explains [Gol69, § II.6] how the $q = 0$ case (formula (21) on page 86 of *loc. cit.*) by Schwarz reflections entails a description of G_N and F_N.

Definition 5.1.7. *Let ψ_N be the local analytic inverse of F_N such that $\psi_N(0) = 0$.*

This inverse exists and is unique in a small neighborhood of $z = 0$. As all we need is to compute $F_N'(0) = \psi_N'(0)^{-1}$, having ψ_N well-defined in a small neighborhood of $z = 0$ is enough for our purpose.

Lemma 5.1.8. *The local analytic inverse map ψ_N of F_N has the form $\psi_N = \eta_1/\eta_2$, where η_1 and η_2 satisfy the second order linear differential equation*

$$4(x^N - 1)^2 y'' + ((N^2 - 1)x^{N-2} + x^{2N-2})y = 0. \tag{5.1.9}$$

Remark 5.1.10. The equation (5.1.9) is more transparent in terms of the Schwarzian derivative $\{\tau, \widetilde{F}_N\}$ defined and computed in the proof below:

$$y'' + \frac{1}{2}\{\tau, \widetilde{F}_N\}y = 0.$$

Proof. First, since F_N is the composition of $\widetilde{F}_N$ and a Möbius transformation, we only need to prove the similar assertion for $\widetilde{F}_N$. We have a companion uniformization map $1/\widetilde{F}_N : \mathbf{H} \to \mathbf{P}^1 \smallsetminus \{0, \mu_N\}$. (The reason for first considering the reciprocal of $\widetilde{F}_N$ is that the standard form considered in [Hem88] is for maps from $\mathbf{H}$ to $\mathbf{P}^1 \smallsetminus S$ where S is a finite set of points which does not contain ∞.)

By [Hem88, Lemma 3.3], the analytic local inverse map of $1/\widetilde{F}_N$ (resp. $\widetilde{F}_N$) is, up to a Möbius transformation, the ratio of two linearly independent solutions of the differential equation equation $y'' + \frac{1}{2}\{\tau, 1/\widetilde{F}_N\}y = 0$ (resp. $y'' + \frac{1}{2}\{\tau, \widetilde{F}_N\}y = 0$), where $\{\tau, 1/\widetilde{F}_N\}$ and $\{\tau, \widetilde{F}_N\}$ denote the Schwarzian derivatives. More precisely, we write $X := 1/\widetilde{F}_N(\tau)$; for $X_0 \notin \{0, \mu_N\}$, once we pick a preimage of X_0 in $\mathbf{H}$, then we may view τ as an analytic function in X in a small neighborhood of X_0 with the value at X_0 being the chosen preimage and $\tau'(X_0) \neq 0$. The *Schwarzian derivative* of τ at X_0 is defined as $\{\tau, X_0\} := \left(\left(\frac{\tau''}{\tau'}\right)' - \frac{1}{2}\left(\frac{\tau''}{\tau'}\right)^2\right)(X_0)$, where all the derivatives are with respect to X (see [Hem88, (3.2)]). For different choices of the preimage of X_0, the local analytic function τ in X differs by a Möbius transformation and then by [Hem88, (3.4)], the value $\{\tau, X_0\}$ is independent of the choice. The function $\{\tau, \widetilde{F}_N\}$ is defined in a similar way. Therefore $\{\tau, X\}$ (resp. $\{\tau, \widetilde{F}_N\}$) is a well-defined analytic function away from $\{0, \mu_N\}$ (resp.$\{\infty, \mu_N\}$).

We now derive $\{\tau, 1/\widetilde{F}_N\}$ and then $\{\tau, \widetilde{F}_N\}$ following [Hem88, §3, §6].

Let $p_k = \zeta_N^k = e^{2\pi i k/N}$ for $k = 1, \ldots, N$ and let $p_0 = 0$. We deduce from [Hem88, Theorem 3.1] that the Schwarzian $\{\tau, 1/\widetilde{F}_N\}$ is given by

$$\{\tau, 1/\widetilde{F}_N\} = \frac{1}{2}\sum_{k=0}^{N} \frac{1}{(X - p_k)^2} + \sum_{k=0}^{N} \frac{m_k}{X - p_k}, \tag{5.1.11}$$

where the m_k for $k = 0, \ldots, N$ denote the so-called *accessory parameters* at $z = p_k$. The accessory parameters are notoriously hard to compute in general, but in our particular example we may find them using the $\mathbf{Z}/N\mathbf{Z}$ symmetry. Expressing the fact that (5.1.11) vanishes to order four at $X = \infty$, the accessory parameters are subject to the following three constraints [Hem88, Theorem 3.1] obtained by equating the $1/X, 1/X^2$ and $1/X^3$ coefficients to zero:

$$\sum_{k=0}^{N} m_k = 0, \quad \sum_{k=0}^{N} 2m_k p_k + 1 = 0, \quad \sum_{k=0}^{N} m_k p_k^2 + p_k = 0. \tag{5.1.12}$$

Since $\mathbf{C} \smallsetminus \mu_N$ is invariant under the action of μ_N, we deduce exactly as in [Hem88, §6, Example 1] that the accessory parameters m_k for $k \neq 0$ satisfy the symmetry $m_k = c \cdot \zeta_N^{-k}$ for some constant c. The constraint $\sum_{k=1}^{N} m_k = 0$ in (5.1.12) then gives $m_0 = 0$. Then the second constraint in (5.1.12) gives

$$\sum_{k=0}^{N} (2m_k \zeta_N^k + 1) = 1 + \sum_{k=1}^{N} (2c + 1) = 0,$$

and hence $c = -\frac{1}{2} - \frac{1}{2N}$. This determines all the m_k, and turns (5.1.11) (still with $X = 1/\widetilde{F}_N$) into

$$\{\tau, 1/\widetilde{F}_N\} = \frac{1}{2X^2} + \frac{1}{2}\sum_{k=1}^{N} \frac{1}{(X-\zeta_N^k)^2} - \frac{(1+N)}{2N}\sum_{k=1}^{N} \frac{\zeta_N^{-k}}{X-\zeta_N^k} = \frac{(1+(N^2-1)X^N)}{2X^2(X^N-1)^2}.$$

From the chain rule, we deduce that with $x = \widetilde{F}_N = 1/X$ the equality:

$$\{\tau, \widetilde{F}_N\} = \frac{1}{x^4}\frac{(1+(N^2-1)(1/x)^N)}{2(1/x)^2((1/x)^N-1)^2} = \frac{(N^2-1)x^{N-2}+x^{2N-2}}{2(x^N-1)^2},$$

and from this we find that the equation $y'' + \frac{1}{2}\{\tau, \widetilde{F}_N\}y = 0$ is given by (5.1.9). We then conclude the proof by [Hem88, Lemma 3.3]. □

Definition 5.1.13. *Let G_N denote the map $D(0,1) \to \mathbf{C} \smallsetminus \{1\}$ such that $G_N(z^N) = (F_N(z))^N$, or equivalently $G_N(z) = (F_N(z^{1/N}))^N$.*

The fact that G_N is well-defined is a formal consequence of the relation $F_N(\zeta z) = \zeta F_N(z)$ in Lemma 5.1.2.

The inverse map of G_N is closely related to the inverse map of F_N, and turns out to have a nicer form. We will give some geometric description of G_N in §5.2 in terms of triangle groups, which suggests an explicit description of the inverse of G_N in terms of hypergeometric functions.

Lemma 5.1.14. *Let φ_N denote the local inverse map of G_N around $x = 0$, normalized so that $\varphi_N(0) = 0$. The function φ_N has the form $\delta_N^{-1}(\phi_1/\phi_2)^N$, where ϕ_1 and ϕ_2 are the solutions to the differential equation:*

$$x(x-1)^2y'' + \left(1 - \frac{1}{N}\right)(x-1)^2y' + \left(\frac{1}{4} + \frac{x-1}{4N^2}\right)y = 0 \tag{5.1.15}$$

given explicitly by

$$\phi_1 = \sqrt{1-x}\cdot x^{1/N}\cdot {}_2F_1\left[\begin{matrix}\frac{N+1}{2N} & \frac{N+1}{2N} \\ 1+\frac{1}{N}\end{matrix}; x\right], \quad \phi_2 = \sqrt{1-x}\cdot {}_2F_1\left[\begin{matrix}\frac{N-1}{2N} & \frac{N-1}{2N} \\ 1-\frac{1}{N}\end{matrix}; x\right], \tag{5.1.16}$$

and $\delta_N = |G_N'(0)|$ denotes conformal radius of the map G_N.

Further, let $s_N(x)$ denote the function

$$s_N(x) := x^{1/N}\frac{{}_2F_1\left[\begin{matrix}\frac{N+1}{2N} & \frac{N+1}{2N} \\ 1+\frac{1}{N}\end{matrix}; x\right]}{{}_2F_1\left[\begin{matrix}\frac{N-1}{2N} & \frac{N-1}{2N} \\ 1-\frac{1}{N}\end{matrix}; x\right]}. \tag{5.1.17}$$

Then $\varphi_N(x) = \delta_N^{-1}s_N(x)^N$, $\psi_N(x) = |F_N'(0)|^{-1}s_N(x^N)$, and $\delta_N = |F_N'(0)|^N$.

Proof. By Definition 5.1.13, $G_N(z) = (F_N(z^{1/N}))^N$; and by the assumptions $\varphi_N(0) = 0$ and $\psi_N(0) = 0$, we obtain the formal identity $\varphi_N(x) = \psi_N(x^{1/N})^N$ (formally: $x = G_N(z)$).

Let η_1, η_2 denote the solutions of the differential equation in Lemma 5.1.8 such that $\eta_1(0) = 0, \eta_1'(0) = 1, \eta_2(0) = 1, \eta_2'(0) = 0$; then η_1, η_2 are linearly independent and $\eta_1(x)/\eta_2(x) = x + O(x^2)$. Since $\psi_N(x) = |F_N'(0)|^{-1}x + O(x^2)$, then by Lemma 5.1.8, we have $\psi_N = |F_N'(0)|^{-1}\,\eta_1/\eta_2$. We deduce

$$\varphi_N(x) = |F_N'(0)|^{-N}(\eta_1(x^{1/N})/\eta_2(x^{1/N}))^N.$$

Let $\phi_i(x) = \eta_i(x^{1/N})$. Then

$$\phi_i'(x) = N^{-1}x^{\frac{1}{N}-1}\eta_i'(x^{1/N})$$

and

$$\phi_i''(x) = \frac{1-N}{N^2}x^{\frac{1}{N}-2}\eta_i' + N^{-2}x^{\frac{2}{N}-2}\eta_i''(x^{1/N}).$$

From equation (5.1.9), we have

$$4(x-1)^2\eta_i''(x^{1/N}) + ((N^2-1)x^{1-\frac{2}{N}} + x^{2-\frac{2}{N}})\eta_i(x^{1/N}) = 0.$$

We rewrite this differential equation in terms of derivatives of ϕ_i using the above equations and then conclude that ϕ_1, ϕ_2 are solutions to (5.1.15).

In order to prove that ϕ_i are given by the explicit formula in (5.1.16), we first deduce from the second order differential equations satisfied by hypergeometric functions that both $\sqrt{1-x} \cdot x^{1/N} \cdot {}_2F_1\left[\begin{matrix} \frac{N+1}{2N} & \frac{N+1}{2N} \\ \multicolumn{2}{c}{1+\frac{1}{N}} \end{matrix}; x\right]$ and $\sqrt{1-x} \cdot {}_2F_1\left[\begin{matrix} \frac{N-1}{2N} & \frac{N-1}{2N} \\ \multicolumn{2}{c}{1-\frac{1}{N}} \end{matrix}; x\right]$ satisfy (5.1.15). (See, for instance, [Gol69, pp. 84-85] on how to adjust by some rational power of $1-x$ to obtain a hypergeometric differential equation and then obtain the two solutions.) Moreover, we conclude that these explicit solutions are exactly ϕ_1 and ϕ_2 by noticing that they have the same leading terms as η_1, η_2 (once we replace x by x^N).

The last assertion is just a summary of the above results. □

In order to prove Theorem 5.1.4, we need the following formula of the behavior of hypergeometric functions in Lemma 5.1.14 near $x = 1$.

Lemma 5.1.18 (See, for instance, [AS92, 15.3.10])**.** *Given $a \notin \mathbf{Z}_{\leq 0}$, for $|x| < 1$, we have*

$$ {}_2F_1\left[\begin{matrix} a & a \\ \multicolumn{2}{c}{2a} \end{matrix}; 1-x\right] = \frac{\Gamma(2a)}{\Gamma(a)^2} \sum_{k=0}^{\infty} \frac{(a)_k (a)_k}{k!^2} x^k \left(-\log x + 2(\psi(k+1) - \psi(k+a))\right), \tag{5.1.19}$$

where ψ denotes the digamma function*: the logarithmic derivative of the Gamma function. Note that the above hypergeometric function is multivalued around $x = 0$, but all different branches are accounted for by the branches of the logarithm.* □

Proof of Theorem 5.1.4. Since F_N is a covering map of $\mathbf{C} \smallsetminus \mu_N$, then local inverse ψ_N is naturally defined on $D(0,1) \subset \mathbf{C} \smallsetminus \mu_N$. Moreover, since $F_N(1) = 1$, we have that $\lim_{x\to 1} \psi_N(x) = 1$, where $x \in D(0,1)$ approaches 1. (*A priori*, we only conclude that $\lim_{x\to 1} \psi_N(x)$ approaches a cusp of $D(0,1)$, and thus that $|\lim_{x\to 1} \psi_N(x)| = 1$; this suffices for the rest of the proof. The more precise statement $\lim_{x\to 1} \psi_N(x) = 1$ follows from our assumption $\psi_N(0) = 0$, either by the description of the fundamental domain further down in Lemma 5.2.1, or more directly by the computation of the rest of the proof that follows, which shows that that $\lim_{x\to 1} \psi_N(x)$ is a positive real number.)

By Lemma 5.1.14, we have $\psi_N(x) = |F_N'(0)|^{-1} s_N(x^N)$. In particular,

$$|F_N'(0)|^{-1} \lim_{x \in D(0,1), x\to 1} s_N(x^N) = 1.$$

Thus, by (5.1.17) and (5.1.19), we have

$$|F_N'(0)| = \lim_{x \in D(0,1), x\to 1} s_N(x^N) = \lim_{x \in D(0,1), x\to 1} \frac{{}_2F_1\left[\begin{matrix} \frac{N+1}{2N} & \frac{N+1}{2N} \\ \multicolumn{2}{c}{1+\frac{1}{N}} \end{matrix}; x\right]}{{}_2F_1\left[\begin{matrix} \frac{N-1}{2N} & \frac{N-1}{2N} \\ \multicolumn{2}{c}{1-\frac{1}{N}} \end{matrix}; x\right]} = \frac{\Gamma\left(\frac{N-1}{2N}\right)^2 \Gamma\left(1+\frac{1}{N}\right)}{\Gamma\left(\frac{N+1}{2N}\right)^2 \Gamma\left(1-\frac{1}{N}\right)}.$$

Basic properties of the Gamma function [AS92, 6.1.18] transform the latter expression into

$$\gamma_N = 2^{4/N} \frac{\Gamma\left(1+\frac{1}{2N}\right)^2 \Gamma\left(1-\frac{1}{N}\right)}{\Gamma\left(1-\frac{1}{2N}\right)^2 \Gamma\left(1+\frac{1}{N}\right)}.$$

Then by [AS92, 6.1.33], we also have

$$\log \gamma_N = \frac{\log 16}{N} + \sum_{k=1}^{\infty} \frac{(2^{2k}-1)}{2^{2k-1}(2k+1)} \cdot \frac{\zeta(2k+1)}{N^{2k+1}}.$$

We obtain (5.1.6) by taking the exponential of the above formula. □

Example 5.1.20. If $N=2$, then $\mathbf{C}\smallsetminus\{\pm 1\}$ is itself biregular to $Y(2)=\mathbf{P}^1\smallsetminus\{0,1,\infty\}$ and thus one can find a direct description of the uniformization $\mathbf{H}\to\mathbf{C}\smallsetminus\{\pm 1\}$ sending i to 0 by $2\lambda(\tau)-1$. In this case, the formulas above specialize to the standard identity $q=e^{-\pi K'/K}$ where the elliptic periods K' and K are directly related to hypergeometric functions. The only other such case of an incidental isomorphism $\mathbf{C}\smallsetminus\mu_N\cong Y(N)$ is $N=3$: this is [Hem88, § 6 Example 5].

5.2. **Geometry of Γ_N and a uniform growth estimate of F_N.** Our second aim in the present § 5 is the uniform supremum growth estimate Lemma 5.2.17 of the universal covering map $F_N: D(0,1)\to\mathbf{C}\smallsetminus\mu_N$ near the boundary, as both the circle $|z|=r$ and the level N vary. This result is subsumed by the more precise bound of Kraus and Roth [KR16, Theorems 1.2 and 1.10]; our treatment is self-contained. In § 6 we will refine this supremum growth bound (uniformly exponential in $\frac{N}{1-r}$) to an integrated growth bound (uniformly linear in $\frac{N}{1-r}$).

The idea of the proof is to use the symmetry of $\mathbf{C}\smallsetminus\mu_N$ and the action of the Fuchsian group $\widetilde{\Gamma}_N$ to reduce the question to the study of the asymptotic of $\widetilde{F}_N$ near the cusp $\tau=i\infty$ and we study this asymptotic using the explicit description of the inverse of $\widetilde{F}_N$ given in Lemma 5.1.14.

We begin in this subsection by explicitly describing the Fuchsian group $\widetilde{\Gamma}_N$.

Proposition 5.2.1. *The stabilizer of $i\infty$ in $\widetilde{\Gamma}_N$ is generated by*

$$\widetilde{t}_N:=\begin{pmatrix}1 & 2\cot(\pi/2N)\\ 0 & 1\end{pmatrix}.$$

The group $\widetilde{\Gamma}_N$ is the free group on N generators generated by $\widetilde{t}_N$ and its conjugates by powers of $\widetilde{r}_N$ in (5.1.3)*. A fundamental domain $\Omega\subset D(0,1)$ for Γ_N is given by the region with $2N$ cusps given by half-integer powers of $\zeta_N=\exp(2\pi i/N)$ and bounded by geodesics connecting adjacent cusps.*

Proof. Since $\widetilde{F}_N$ is the universal covering map to $\mathbf{C}\smallsetminus\mu_N$, then the Fuchsian group $\widetilde{\Gamma}_N$ is generated by the stabilizers of the cusps c with $\widetilde{F}_N(c)\in\mu_N$. If we denote the generator of the stabilizer of $i\infty$ by

$$\widetilde{t}:=\begin{pmatrix}1 & c_N\\ 0 & 1\end{pmatrix}, \tag{5.2.2}$$

then the stabilizers of the other cusps associated to μ_N are generated by the conjugates of $\widetilde{t}_N$ by $\widetilde{r}_N$ since $\zeta_N^k\widetilde{F}_N(\tau)=\widetilde{F}_N(\widetilde{r}_N^k\cdot\tau)$ by Lemma 5.1.2.

Consider the Dirichlet domain $\Omega_N\subset D(0,1)$ associated to Γ_N around $z=0$. More precisely, we can describe Ω_N as the region

$$\{z\in D(0,1)\,:\,d(gz,0)\geq d(z,0)\quad\forall g,g^{-1}\in\{r_N^k\,t\,r_N^{-k}\},k=0,1,\dots,N-1\}.$$

Here, d is the hyperbolic distance in $D(0,1)$. The region in $D(0,1)$ such that $d(gz,0)\geq d(z,0)$ and $d(g^{-1}z,0)\geq d(z,0)$ for $g=r_N^k\cdot t\cdot r_N^{-k}$ is the region bounded by two geodesics starting at ζ_N^k going in opposite directions and intersecting the boundary at $\zeta_N^k e^{\pm i\theta}$ where $c_N=2\cot(\theta/2)$. There are exactly $2N$ such arcs corresponding to the N generators and their inverses. In particular, if $\theta<\pi/N$ is too small, the fundamental region will have infinite volume, whereas if $\theta>\pi/N$ is too big, then the Dirichlet domain will only contain at most N cusps. Since $\mathbf{H}/\widetilde{\Gamma}_N$ has $N+1$ cusps and $\widetilde{\Gamma}_N$ has finite covolume, we must have $c_N=2\cot(\pi/2N)$ and these geodesics intersecting at $\zeta_N^{k+1/2}$, $k=0,\dots,N-1$. □

Now we consider the group associated to $\widetilde{F}_N^N$.

Definition 5.2.3. *Let $\widetilde{\Phi}_N$ denote the group $\langle\widetilde{\Gamma}_N,\widetilde{r}_N\rangle=\langle\widetilde{r}_N,\widetilde{t}_N\rangle$ and let Φ_N denote the corresponding lattice in* $\mathrm{PSU}(1,1)$.

Corollary 5.2.4. *The function $\widetilde{F}_N^N$ is invariant under $\widetilde{\Phi}_N$ and $\widetilde{\Phi}_N$ is the largest subgroup of* $\mathrm{PSL}_2(\mathbf{R})$ *with this property. A fundamental domain Ω'_N for Φ_N in $D(0,1)$ is given by the hyperbolic quadrilateral with vertices $0,\zeta_N^{-1/2},1,\zeta_N^{1/2}$. Translated to $\mathbf{H}$ this is bounded by geodesics from i to $\cot(\pi/2N)$ to $i\infty$ to $-\cot(\pi/2N)$ and back to i.*

Proof. The statements follow directly from Lemma 5.1.2 and Proposition 5.2.1. □

Lemma 5.2.5. *Let $\widetilde{s} \in \mathrm{PSL}_2(\mathbf{R})$ be a rotation of order $2N$ such that $\widetilde{s}^2 = \widetilde{r}_N$. Then*

$$1 - \widetilde{F}_N^N(\widetilde{s} \cdot \tau) = \frac{1}{1 - \widetilde{F}_N^N(\tau)}. \tag{5.2.6}$$

Proof. Recall our normalization of $\widetilde{F}_N$ that $\widetilde{F}_N(i) = 0$ and $\widetilde{F}_N(i\infty) = 1$. By Corollary 5.2.4, $\tau = \pm\cot(\pi/2N)$ (in the same $\widetilde{\Phi}_N$-orbit) is the other cusp of $\widetilde{\Phi}_N$ and thus $\widetilde{F}_N(\pm\cot(\pi/2N)) = \infty$. Moreover, both sides of (5.2.6) are uniformizers of $\mathbf{H}/\widetilde{\Phi}_N$ which take the value 0 at the cusp $\cot(\pi/2N)$ and ∞ at the cusp $i\infty$. This specifies them uniquely up to $x \mapsto \lambda x$ scalings. However, this last ambiguity is removed by noting that both sides are 1 at $\tau = i$. □

Remark 5.2.7. The group $\widetilde{\Phi}_N$ is contained with index two in the larger group $\widetilde{\Psi}_N = \langle \widetilde{s}, \widetilde{t}_N \rangle$. The group $\widetilde{\Psi}_N$ has a fundamental domain consisting of the points $0, 1, \zeta_N^{1/2}$. But this is none other than a hyperbolic triangle with angles $\{\alpha, \beta, \gamma\} = \{\pi/N, 0, 0\}$, whose conformal mapping from $\mathbf{H}$ is given by Schwarz triangle functions (see, for instance, [Car54, §404 on page 185]). This suggests that $\widetilde{F}_N$ should directly be related to Schwarz triangle functions, which leads to a direct description of the inverse functions ψ_N and φ_N in terms of hypergeometric functions in Lemma 5.1.14.

In order to study the behavior of $\widetilde{F}_N$ near $x = 1$, we use the explicit formula for its inverse ψ_N given in Lemma 5.1.14. The following lemma gives the asymptotic of the function s_N used in formula of ψ_N.

Lemma 5.2.8. *Fix a real constant $M_0 > 0$. For $M \geq M_0$ and $|x| < e^{-MN}$, we have the uniform estimate:*

$$\left| \frac{s_N(1-x)}{\gamma_N} - (1-x)^{1/N} \frac{-\log x + 2\gamma - 2\psi(1/2 + 1/2N)}{-\log x + 2\gamma - 2\psi(1/2 - 1/2N)} \right| \ll \frac{|x|}{N}, \tag{5.2.9}$$

where ψ is the digamma function as in Lemma 5.1.18, and the implicit coefficient depends on M_0 but not on N, M.

Proof. By Lemma 5.1.18, we have for $a = 1/2 \pm 1/2N$ and $|x| < 1$ the following equality:

$${}_2F_1\left[\begin{matrix} a \ a \\ 2a \end{matrix}; 1-x\right] = \frac{\Gamma(2a)}{\Gamma(a)^2} \sum_{k=0}^{\infty} \frac{(a)_k (a)_k}{k!^2} x^k \left(-\log x + 2(\psi(k+1) - \psi(k+a))\right). \tag{5.2.10}$$

We first prove that the coefficients in this power series are uniformly bounded. Since $|a| < 1$, we have $|(a)_k|/k! < 1$. Basic properties of the digamma function (cf. [AS92, 6.3.5, 6.3.14]) show that $\psi(x)$ is negative and strictly increasing for $0 < x < 1$, and $|\psi(k) - \psi(k+a)| > |\psi(k+1) - \psi(k+1+a)|$. Hence, $|\psi(k) - \psi(k+a)|$ is maximized when $k = 1$ and $a = 1/2 - 1/4$. This immediately leads to the uniform estimates

$$\left| \sum_{k=1}^{\infty} \frac{(a)_k (a)_k}{k!^2} x^k \right|, \quad \left| \sum_{k=1}^{\infty} \frac{(a)_k (a)_k}{k!^2} x^k \left(2(\psi(k+1) - \psi(k+a))\right) \right| \ll |x| < e^{-MN}$$

For $|x| < e^{-MN}$, we also have $\Re(\log x) \leq -MN$, and so in particular $|\log x| \geq MN$ regardless of the branch of logarithm. Combined with Lemma 5.1.14 and (5.2.10), this leads to the estimate

$$\frac{s_N(1-x)}{\gamma_N(1-x)^{1/N}} = \frac{-\log x + 2\psi(1) - 2\psi(1/2 + 1/2N) + O(x)}{-\log x + 2\psi(1) - 2\psi(1/2 - 1/2N) + O(x)}$$

where the implied constants are uniform in N, from which the result follows (using that $|\log x| \gg N$). □

Lemma 5.2.11. *Fix a pair of real positive numbers $M_0 > 0$ and $\epsilon > 0$. Consider any $M \geq M_0$, and let $\widetilde{\Omega}'_N \subset \mathbf{H}$ denote the fundamental domain for $\widetilde{\Phi}_N$ corresponding to Ω'_N in Corollary 5.2.4. If $\tau \in \widetilde{\Omega}'_N$ has*

$$\|\widetilde{F}_N(\tau)^N - 1\| < e^{-MN},$$

then

$$\Im(\tau) > \frac{2N^2M}{\pi^2}(1-\epsilon) \tag{5.2.12}$$

once $N \gg_{\epsilon,M_0} 1$, where the implicit coefficient depends only on M_0 and ϵ.

Proof. By Corollary 5.2.4, $\widetilde{\Omega}'_N$ is a fundamental domain of $\widetilde{F}_N^N$ and the only cusp where $\widetilde{F}_N^N = 1$ is at $\tau = i\infty$. So it suffices to consider $\widetilde{F}_N^N$ in a neighbourhood of the cusp $i\infty$.

For N sufficiently large, the inequality $\|\widetilde{F}_N(\tau)^N - 1\| < e^{-MN}$ implies that $|\widetilde{F}_N(\tau) - 1| < e^{-M(1-\varepsilon_0)N}$ for some ε_0 that tends to zero as N increases. Recall from the proof of Theorem 5.1.4 that $\psi_N(1) = 1$ and hence $\widetilde{\psi}_N(1) = i\infty$; then it suffices to bound the imaginary part of

$$\tau = \widetilde{\psi}_N(1-x), \qquad \text{for } |x| < e^{-M(1-\varepsilon_0)N}.$$

We may write this as

$$\tau = i \cdot \frac{1+\psi_N(1-x)}{1-\psi_N(1-x)} = i \cdot \frac{\gamma_N + s_N((1-x)^N)}{\gamma_N - s_N((1-x)^N)}. \tag{5.2.13}$$

Writing $1 - X := (1-x)^N$, then the same estimate as above implies $|X| < e^{-M(1-2\varepsilon_0)N}$ for sufficiently large N. Thus we reduce the lemma to the estimate of

$$\tau = i \cdot \frac{\gamma_N + s_N(1-X)}{\gamma_N - s_N(1-X)}, \quad |X| < e^{-M(1-2\varepsilon_0)N}. \tag{5.2.14}$$

Now by Lemma 5.2.8 and [AS92, 6.3.7], we have

$$\tau = \frac{i\cot(\pi/2N)}{\pi}\left(2\gamma - \log X - \psi(1/2 - 1/2N) - \psi(1/2 + 1/2N)\right) + O(1),$$

where the implicit constant only depends on M_0 and ε_0. The imaginary part of this does not depend on the choice of branch of $\log X$ and indeed only depends on $|X|$, and we deduce with this approximation that

$$\Im(\tau) \geq \frac{\cot(\pi/2N)}{\pi}\left(2\gamma + NM(1-2\varepsilon_0) - \psi(1/2 - 1/2N) - \psi(1/2 + 1/2N)\right) + O(1).$$

If we choose $\varepsilon_0 := \epsilon/2$, then for $N \gg_{\epsilon,M_0} 1$ this lower bound clearly exceeds the requisite

$$\frac{2N^2M(1-\epsilon)}{\pi^2}.$$

□

The region (5.2.12) is a *horoball* for the cusp $i\infty$ in the upper half plane model $\mathbf{H}$ of the hyperbolic plane. Recall that in the Poincaré disc model $D(0,1)$, the horoballs are the euclidean discs inside $D(0,1)$ which are tangent to the boundary circle. The following easy lemma describes how these horoballs transform under the hyperbolic isometry group.

Lemma 5.2.15. *Let $\mathcal{H}_D$ denote the image in the Poincaré disc model $D(0,1)$ of the horoball*

$$\{\tau \in \mathbf{H} \,:\, \Im(\tau) \geq D\}$$

in the upper half plane model $\mathbf{H}$. For $\gamma \in \mathrm{PSU}(1,1)$ with image $\widetilde{\gamma} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbf{R})$, the image $\gamma\mathcal{H}_D$ of $\mathcal{H}_D$ under γ in $D(0,1)$ is the disc with diameter

$$E(\gamma, D) := \frac{2}{1 + D(a^2+c^2)} \tag{5.2.16}$$

tangent to the boundary circle $\mathbf{T}$ at the point $(a-ic)/(a+ic)$. □

We close this section by using Lemma 5.2.11 to derive a coarse yet fairly uniform upper bound on $\sup_{|z|=r} \log|F_N(z)|$. Although not best-possible, it is enough as an input for the logarithmic error term in the Nevanlinna theory estimate in §6.

Lemma 5.2.17. *For $N \gg 1$ and $r \in (0,1)$, we have*

$$\sup_{|z|=r} \log|F_N(z)| \ll \frac{N}{1-r},$$

where the implicit constants are both absolute.

Proof. Set $S(M,N) := \{z \in D(0,1) : |F_N^N(z) - 1| < e^{-MN}\}$ and $M := \frac{N}{1-r}$. Since $M \geq 2$, we may take $M_0 = 2$ and $\epsilon = 1/2$ in Lemma 5.2.11 and conclude that for $N \gg 1$ (with absolute constant here), we have

$$S(M,N) \subset \bigcup_{\gamma \in \Phi_N} \gamma \mathcal{H}_D, \quad \text{where } D = \frac{N^2 M}{\pi^2}.$$

By Shimizu's Lemma (see, for example, [EGM98, Theorem 3.1]) and Proposition 5.2.1, we have $2(|a|+|c|)\cot(\pi/2N) \geq 1$ for all $\widetilde{\gamma} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \widetilde{\Gamma}_N$. Thus by Lemma 5.2.15,

$$E(\gamma, D) \leq 2D^{-1}(a^2+c^2) \ll N^{-2}N^{-1}(1-r)N^2 = \frac{1-r}{N},$$

where the implicit constant is absolute. Thus we have $E(\gamma, D) \leq 1-r$ for all $\gamma \in \Phi_N$ once N^{-1} times the implicit constant is less than 1.

By Lemma 5.2.5, the set $\{z \in D(0,1) : |F_N(z)| > e^M + 1\}$ is contained in $sS(M,N)$, which is contained in $D(0,1) \smallsetminus \overline{D(0,r)}$ by the above argument for $N \gg 1$. Thus we conclude that

$$\sup_{|z|=r} \log|F_N(z)| \leq \log(e^M + 1) \ll M = \frac{N}{1-r}. \qquad \square$$

Remark 5.2.18. A more refined bound is proved in Kraus–Roth [KR16, Theorems 1.2 and 1.10]. On the other hand, one can push our method further and prove, with rather more work but uniformly in $N \in \mathbf{N}$ and $M \in [1,\infty)$, that the supremum region $|F_N| < e^M$ is simply connected of conformal radius $1 - O(M^{-2}N^{-3})$ from the origin; this is a sharp estimate. But taking for φ in Theorem 2.0.5 the pullback of F_N by the Riemann map of some such region $|F_N| < e^M$, and ignoring thus the fine savings from the integrated bound (2.0.7) as opposed to the supremum, would only lead to an $O(N^4)$ holonomy rank bound in place of our requisite logarithmically inflated bound $O(N^3 \log N)$. In the next section we will see how to make the full use of the integrated holonomy bound, and use Nevanlinna's value distribution theory to supply our final piece of the proof of the unbounded denominators conjecture.

6. Nevanlinna theory and uniform mean growth near the boundary

For our application of Corollary 2.0.5, we prove in this section the following uniform growth bound. Throughout this section, we assume as we may that $N \geq 2$. Then the analytic map $F_N : D(0,1) \to \mathbf{P}^1$ omits the $N+1 \geq 3$ values $\mu_N \cup \{\infty\}$. In such a situation, we seek to exploit whatever growth constraints are imposed on the map by Nevanlinna's value distribution theory. A theorem of Tsuji [Tsu52, Theorem 11] gives the general asymptotic

$$\int_{|z|=r} \log^+|F| \, \mu_{\mathrm{Haar}} = \frac{1}{N-1} \log \frac{1}{1-r} + O_{a_1,\dots,a_N}(1),$$

for any universal covering map $F : D(0,1) \to \mathbf{C} \smallsetminus \{a_1, \dots, a_N\}$ based at $F(0) = 0$ (see also the discussion in Nevanlinna [Nev70, page 272]), however this is only asymptotically in $r \to 1^-$ for given punctures $\{a_i\}$ whereas we need a uniformity in both r and N. It is at the point (6.2.3) of the explicit partial fraction coefficients that our argument below makes a critical use of the special feature of the target set $\mu_N \cup \{\infty\}$ of omitted values.

Theorem 6.0.1. *For each of the choices*

$$p(x) \in \{x^N, x^N/(x^N-1), 1/(x^N-1)\},$$

we have uniformly in $N \in \mathbf{N}$ and $r \in (0,1)$ the mean growth bound

$$\int_{|z|=r} \log^+ |p \circ F_N| \, \mu_{\mathrm{Haar}} \ll \log \frac{N}{1-r}, \tag{6.0.2}$$

with some (effectively computable) absolute constant implicit coefficient.

6.1. **Preliminaries in Nevanlinna theory.** This section collects some standard material from Nevanlinna's value distribution theory. The reader should feel encouraged to skip this part on a first reading, and refer back as necessary.

6.1.1. *The Nevanlinna characteristic.* The left-hand side of (6.0.2) is known as the *mean proximity function at* ∞

$$m(r,f) = m(r,f;\infty) := \int_{|z|=r} \log^+ |f| \, \mu_{\mathrm{Haar}} \in [0,\infty).$$

It is complemented by the *counting function*

$$N(r,f) = N(r,f;\infty) := \sum_{\rho \,:\, 0<|\rho|<r} \mathrm{ord}_\rho^-(f) \log \frac{r}{|\rho|} + \mathrm{ord}_0^-(f) \, \log r,$$

where, in general for a meromorphic mapping $f : D(0,1) \to \mathbf{P}^1$, we let $\mathrm{ord}_\rho^-(f) := \mathrm{ord}^+(1/f) = \max(0, \mathrm{ord}(1/f))$ is the pole order (if ρ is a pole, and 0 if f is holomorphic at ρ).

The *Nevanlinna characteristic function*

$$T(r,f) := m(r,f) + N(r,f)$$

is the well-behaved quantity functorially.

Lemma 6.1.1. *For every meromorphic function $f : D(0,1) \to \mathbf{P}^1$ regular at 0 (that is: $f(0) \neq \infty$), and every $r \in (0,1)$, we have*

$$N(r,f) \geq 0,$$

with equality if and only if f is holomorphic (has no poles) throughout the disc $D(0,r)$.

The Nevanlinna characteristic function $T(r,f)$ satisfies for every $a \in \mathbf{C}$ the relation

$$|T(r,f) - T(r,1/(f-a)) - \log |c(f,a)|| \leq \log^+ |a| + \log 2, \tag{6.1.2}$$

where

$$c(f,a) := \lim_{z \to 0} (f(z)-a) z^{-\mathrm{ord}_0(f-a)}.$$

Proof. This is Rolf Nevanlinna's *first main theorem*, and is proved formally and straightforwardly from the Poisson–Jensen formula (see, for instance, [BG06, Proposition 13.2.6]), which we may rewrite as

$$T(r,f) - T(r,1/f) = \log |c(f,0)|, \tag{6.1.3}$$

and the triangle inequality relation

$$\big| \log^+ |f-a| - \log^+ |f| \big| \leq \log^+ |a| + \log 2. \tag{6.1.4}$$

See Hayman [Hay64, Theorem 1.2] or Bombieri–Gubler [BG06, Theorem 13.2.10] for the details. We note that $c(f,a) = f(0) - a$ when $a \neq f(0)$. □

6.1.2. *The lemma on the logarithmic derivative.* The *lemma on the logarithmic derivative*—a strong explicit form of which is cited in (6.1.12) below—is the centerpiece of R. Nevanlinna's original analytic proof of his *second main theorem* of value distribution theory. The logarithmic error feature of this sharp upper bound on the proximity function of a logarithmic derivative enables us to derive Theorem 6.0.1 from the relatively crude supremum growth bound in Lemma 5.2.17.

The reader willing to take (6.1.12) for granted may at this point proceed directly to §6.2. Nevertheless, since the proof simplifies considerably in the case that we need of a *functional unit* (a nowhere vanishing holomorphic function), we include our own self-contained treatment of a basic explicit case of the lemma on the logarithmic derivative.

Lemma 6.1.5. *Let $g : \overline{D(0,R)} \to \mathbf{C}^\times$ 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 $0 < r < R$,*

$$m\Big(r, \frac{g'}{g}\Big) < \log^+ \Big\{ \frac{m(R,g)}{r} \frac{R}{R-r} \Big\} + \log 2 + 1/e. \tag{6.1.6}$$

Proof. Our functional unit assumption means that the function $\log g(z)$ has a single valued holomorphic branch on a neighborhood of the closed disc $|z| \leq R$ with $\log g(0) = 0$. Its real part is the harmonic function $\log |g(z)|$. Poisson's formula on the harmonic extension of a continuous function from the boundary to the interior of a disc reads

$$\log |g(z)| = \int_{|w|=R} \log |g(w)| \cdot \Re\Big(\frac{w+z}{w-z}\Big)\, \mu_{\mathrm{Haar}}(w), \tag{6.1.7}$$

where $k(z,w) := \Re\Big(\frac{w+z}{w-z}\Big)$ is the *Poisson kernel*. This formula in fact upgrades to

$$\log g(z) = \int_{|w|=R} \log |g(w)| \cdot \frac{w+z}{w-z}\, \mu_{\mathrm{Haar}}(w), \tag{6.1.8}$$

because both sides are holomorphic in z, have identical real parts, and evaluate to zero at $z = 0$.

Differentiation in the integrand of (6.1.8) gives the reproducing formula

$$\frac{g'(z)}{g(z)} = \int_{|w|=R} \frac{2w}{(w-z)^2} \log |g(w)|\, \mu_{\mathrm{Haar}}(w), \quad \forall z \in D(0,R). \tag{6.1.9}$$

for the logarithmic derivative in the interior of the disc $|z| \leq R$ in terms of boundary values on the circle $|z| = R$. We have the elementary calculation

$$\int_{|z|=r} |w-z|^{-2}\, \mu_{\mathrm{Haar}}(z) = \frac{1}{R^2 - r^2} \qquad \text{for } |w| = R > r, \tag{6.1.10}$$

and thus the $|z| = r$ integral of (6.1.9) with the triangle inequality and interchanging the orders of the integrations and using $|\log |g|| = \log^+ |g| + \log^- |g| = \log^+ |g| + \log^+ |1/g|$ yields

$$\begin{aligned}
\int_{|z|=r} \Big| \frac{g'(z)}{g(z)} \Big|\, \mu_{\mathrm{Haar}} &\leq 2R \int_{|z|=r} \int_{|w|=R} |w-z|^{-2}\, \big| \log |g(w)| \big|\, \mu_{\mathrm{Haar}}(w)\, \mu_{\mathrm{Haar}}(z) \\
&= 2R \int_{|w|=R} \Big(\int_{|z|=r} |w-z|^{-2}\, \mu_{\mathrm{Haar}}(z) \Big) \big| \log |g(w)| \big|\, \mu_{\mathrm{Haar}}(w) \\
&= \frac{2R}{R^2 - r^2} \int_{|w|=R} \big| \log |g(w)| \big|\, \mu_{\mathrm{Haar}}(w) \\
&= \frac{2R}{R^2 - r^2} \Big(m(R,g) + m(R,1/g) \Big) = \frac{4R\, m(R,g)}{R^2 - r^2},
\end{aligned}$$

on using on the final line the harmonicity property again which implies

$$\int_{|w|=R} \log |g|\, \mu_{\mathrm{Haar}}(w) = \log |g(0)| = 0.$$

The final piece of the proof borrows from [BK01, section 4]. Let

$$E := \Big\{ z \,:\, |z| = r,\, |g'(z)/g(z)| > 1 \Big\},$$

a measurable subset of the circle $|z| = r$. Since the function $\log^+ |x|$ is concave on $x \in [1, \infty)$ where it coincides with $\log |x|$, Jensen's inequality gives

$$\begin{aligned}
\int_{|z|=r} \log^+ \Big| \frac{g'}{g} \Big|\, \mu_{\mathrm{Haar}} &\leq \mu_{\mathrm{Haar}}(E) \log^+ \Big(\frac{1}{\mu_{\mathrm{Haar}}(E)} \int_E \Big| \frac{g'(z)}{g(z)} \Big|\, \mu_{\mathrm{Haar}}(z) \Big) \\
&\leq \log^+ \int_{|z|=r} \Big| \frac{g'(z)}{g(z)} \Big|\, \mu_{\mathrm{Haar}}(z) + \sup_{t \in (0,1]} \big\{ t \log (1/t) \big\} \\
\leq \log^+ \Big\{ \frac{4R\, m(R,g)}{R^2 - r^2} \Big\} + \frac{1}{e} &\leq \log^+ \Big\{ \frac{m(R,g)}{r} \frac{R}{R-r} \Big\} + \log 2 + \frac{1}{e},
\end{aligned}$$

using $R^2 - r^2 = (R+r)(R-r) > 2r(R-r)$ on the final line. □

Remark 6.1.11. The case of arbitrary meromorphic functions $g : \overline{D(0,R)} \to \mathbf{P}^1$ is handled similarly by a differentiation in the general Poisson–Jensen formula, but with rather more work to estimate the finite sum over the zeros and poles of g. See for instance [Nev70, §IX.3.1, page 244, (3.2)] or [Hay64, Lemma 2.3 on page 36] for similar bounds. By using a technique due to Kolokolnikov for handling the sum over the zeros and poles, Goldberg and Grinshtein [GG76] obtained the general bound

$$m\Big(r, \frac{g'}{g}\Big) < \log^+ \Big\{ \frac{T(R,g)}{r} \frac{R}{R-r} \Big\} + 5.8501, \quad \text{for } g(0) = 1, \tag{6.1.12}$$

and proved that it is essentially best-possible in form apart for the value of the free numerical constant 5.8501 (that has since been somewhat further reduced in the literature, see Benbourenane–Korhonen [BK01]). The paper of Hinkkanen [Hin92] and the books of Cherry–Ye [CY01] and Ru [Ru21] discuss the implications to the structure of the error term in Nevanlinna's second main theorem, mirroring Osgood and Vojta's dictionary to Diophantine approximation and comparing to Lang's conjecture modeled on Khinchin's theorem.

6.2. **Proof of Theorem 6.0.1.** For $f : D(0,1) \to \mathbf{C}$ holomorphic, the polar divisor is empty, and so $N(r,f) = 0$ and $m(r,f) = T(r,f)$. Since by definition $F_N^N - 1$ is a unit in the ring of holomorphic functions on $D(0,1)$, our requisite bound (6.0.2) rewrites in Nevanlinna notation into

$$T(r, p \circ F_N) \ll \log \frac{N}{1-r}, \quad \text{for each of } p(x) \in \big\{x^N/(x^N-1),\, 1/(x^N-1),\, x^N\big\}. \tag{6.2.1}$$

6.2.1. *Equivalence of bounds for different* $p(x)$. By Lemma 6.1.1, the fact $x^N/(x^N-1) = 1 + 1/(x^N-1)$, and (6.1.4), the three cases for $p(x)$ are equivalent to one another. Here we give the explicit estimate in one direction, which will be used later:

$$\begin{aligned} T(r, p\circ F_N) &= m\Big(r, 1 + \frac{1}{F_N^N - 1}\Big) \geq m\Big(r, \frac{1}{F_N^N-1}\Big) - \log 2 \\ &= T\Big(r, \frac{1}{F_N^N - 1}\Big) - \log 2 = T(r, F_N^N - 1) - \log 2 \\ &\geq T(r, F_N^N) - 2\log 2 = N\,T(r, F_N) - \log 4, \end{aligned} \tag{6.2.2}$$

where we use $F_N(0) = 0$ and $F_N^N - 1$ is a unit in the ring of holomorphic functions on $D(0,1)$. In the rest of this subsection, we will prove Theorem 6.0.1 in the form $T(r, F_N^N) \ll \log \frac{N}{1-r}$ but pivoting around the choice

$$p(x) := \frac{x^N}{x^N - 1} = \frac{x}{N} \sum_{\zeta \in \mu_N} \frac{1}{x - \zeta}. \tag{6.2.3}$$

6.2.2. *Reduction to a logarithmic derivative.* By either the chain rule or the partial fractions decomposition, we see that the logarithmic derivative f'/f of the nowhere vanishing holomorphic function

$$f := 1 - F_N^N \quad : \quad D(0,1) \to \mathbf{C}^\times \tag{6.2.4}$$

is related to $p \circ F_N = F_N^N/(F_N^N - 1)$ by

$$p \circ F_N = \frac{F_N}{N F_N'} \frac{f'}{f}. \tag{6.2.5}$$

The idea then is that the piece f'/f in the decomposition (6.2.5) is small on average over circles by the lemma on the logarithmic derivative (Corollary 6.2.6 below), while—again by the lemma on the logarithmic derivative, in Corollary 6.2.8 below—the characteristic functions of $p \circ F_N$ and F_N/F_N' are equal respectively to $NT(r, F_N)$ and $T(r, F_N)$ up to a small error.

6.2.3. *Two corollaries of the lemma on the logarithmic derivative.*

Corollary 6.2.6. *For $f = 1 - F_N^N$, we have*

$$m\Big(r, \frac{f'}{f}\Big) \ll \sup_{|z|=(1+r)/2} \log^+ \log |F_N| + \log \frac{N}{1-r}. \tag{6.2.7}$$

Proof. By applying Lemma 6.1.5 to f and the outer radius choice $R := 1-(1-r)/2 = (1+r)/2$, and using (cf. [BG06, Corollary 13.2.14]) that $m(r, f'/f) = T(r, f'/f)$ is a monotone increasing function of r, we find the mean growth bound

$$\begin{aligned} m\Big(r, \frac{f'}{f}\Big) &\ll \log^+ T\Big(\frac{1+r}{2}, f\Big) + \log \frac{e}{1-r} \\ &= \log^+ m\Big(\frac{1+r}{2}, 1 - F_N^N\Big) + \log \frac{e}{1-r} \\ &\ll \log^+ m\Big(\frac{1+r}{2}, F_N^N\Big) + \log \frac{e}{1-r} \\ &\ll \log^+ m\Big(\frac{1+r}{2}, F_N\Big) + \log \frac{N}{1-r} \\ &\ll \sup_{|z|=(1+r)/2} \log^+ \log |F_N| + \log \frac{N}{1-r}, \end{aligned}$$

where in the last step we have estimated a mean proximity function trivially by a supremum function. $\square$

Corollary 6.2.8. *We have*

$$m\Big(r, \frac{F_N}{F_N'}\Big) \le T(r, F_N) + O\Big(\log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|\Big). \tag{6.2.9}$$

The idea of the proof is to combine Lemma 6.1.5 applied to the functional unit $1 - F_N$ and the standard chain of implications based on Jensen's formula in the reduction of the second main theorem to the lemma on the logarithmic derivative (see, for example, [Hay64, pages 33–34]).

Proof. By (6.1.3) for the function F_N'/F_N, and the fact that F_N is holomorphic on the disc $D(0,1)$ with $F_N(0) = 0$ and $F_N'(0) \neq 0$, we have:

$$\begin{aligned} m\Big(r, \frac{F_N}{F_N'}\Big) &= m\Big(r, \frac{F_N'}{F_N}\Big) + N\Big(r, \frac{F_N'}{F_N}\Big) - N\Big(r, \frac{F_N}{F_N'}\Big) - \log c(F_N'/F_N, 0) \\ &= m\Big(r, \frac{F_N'}{F_N}\Big) + N\Big(r, 1/F_N\Big) - N\Big(r, F_N\Big) - N\Big(r, 1/F_N'\Big) + N\Big(r, F_N'\Big) \end{aligned}$$

$$= m\Big(r, \frac{F_N'}{F_N}\Big) + N\Big(r, 1/F_N\Big) - N\Big(r, 1/F_N'\Big) = m\Big(r, \frac{F_N'}{F_N}\Big) + N\Big(r, 1/F_N\Big). \tag{6.2.10}$$

Here for the last equality we recall that $F_N : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$ is an étale analytic mapping, hence the derivative F_N' is nowhere vanishing.

We continue to estimate with the triangle inequality (for the second and third lines) and then (6.1.3), noting that $|F_N'(0)| > 1$ (for the inequality in the fourth line):

$$\begin{aligned} m\Big(r, \frac{F_N}{F_N'}\Big) &= m\Big(r, \frac{F_N'}{F_N}\Big) + N\Big(r, 1/F_N\Big) \\ &\le m\Big(r, \frac{F_N'}{1-F_N}\Big) + m\Big(r, \frac{1-F_N}{F_N}\Big) + N\Big(r, 1/F_N\Big) \\ &\le m\Big(r, \frac{F_N'}{1-F_N}\Big) + \log 2 + m\Big(r, \frac{1}{F_N}\Big) + N\Big(r, 1/F_N\Big) \\ = m\Big(r, \frac{(1-F_N)'}{1-F_N}\Big) + T(r, 1/F_N) + \log 2 &\le m\Big(r, \frac{(1-F_N)'}{1-F_N}\Big) + T(r, F_N) + \log 2 \\ &\le T(r, F_N) + O\Big(\log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|\Big), \end{aligned}$$

upon again using Lemma 6.1.5 with $R := (1-r)/2$ but now for the functional unit $g = 1 - F_N$, and a similar argument as in the proof of Corollary 6.2.6. $\square$

6.2.4. *Completing the proof from the crude supremum bound in Lemma 5.2.17.* At this point the key identity (6.2.5) allows to combine the estimates (6.2.7) and (6.2.9), arriving at the uniform bound

$$\begin{aligned} T(r, p \circ F_N) = m(r, p\circ F_N) &\leq m\Big(r, \frac{f'}{f}\Big) + m\Big(r, \frac{F_N}{F_N'}\Big) \\ &\leq T(r, F_N) + O\Big(\log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|\Big). \end{aligned} \tag{6.2.11}$$

We leverage the upper bound (6.2.11) on $T(r, p\circ F_N) = N\, T(r, F_N) + O(1)$ against the lower bound (6.2.2) and get a uniform upper bound on $T(r, F_N)$:

$$(N-1)T(r, F_N) \ll \log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|. \tag{6.2.12}$$

Upon doubling the implicit absolute coefficient, plainly for $N \geq 2$ this is equivalent to

$$T(r, F_N^N) = NT(r, F_N) \ll \log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|,$$

uniformly in all $N \geq 2$ and $r \in (0,1)$.

Hence Theorem 6.0.1 follows from Lemma 5.2.17 upon replacing r there with $(1+r)/2$.

6.2.5. *A historical note.* The bound (6.2.12) can be compared to the well-known particular case for entire holomorphic functions of the classical Nevanlinna second main theorem (whose method of proof we emulate here), stating that for any entire function $g : \mathbf{C} \to \mathbf{C}$, and any N-tuple of pairwise distinct points $a_1, \ldots, a_N \in \mathbf{C}$, the Nevanlinna characteristic $T(r,g) = m(r,g) = \int_{|z|=r} \log^+ |g| \, \mu_{\mathrm{Haar}}$ satisfies the upper bound

$$(N-1)T(r,g) + N_{\mathrm{ram}}(r,g) \leq \sum_{i=1}^{N} N(r, a_i) + O(\log T(r,g)) + O(\log r) \tag{6.2.13}$$

outside of an exceptional set of radii $r \in E \subset [0, \infty)$ of finite Lebesgue measure: $m(E) < \infty$. Here $N_{\mathrm{ram}}(r,g) = N(r, 1/g')$ is a ramification term, which is always nonnegative and vanishes if the map g is étale. This is Nevanlinna's quantitative strengthening of Picard's theorem on at most one omitted value for a nonconstant entire function, for if each of $a_1, \ldots, a_N$ is omitted then all counting terms $N(r, a_i) = 0$ vanish on the right-hand side of (6.2.13), leading if $N \geq 2$ to an $O(\log r)$ upper bound on the growth $T(r,g)$ of g. The idea is that we similarly have a holomorphic map F_N omitting the N values $a_h = \exp(2\pi i h/N)$, except F_N is on a disc rather than the entire plane, and that (6.2.13) largely extends as a growth bound for holomorphic maps on a disc. For such completely quantitative results we refer the reader to Hinkkanen [Hin92, Theorem 3] or Cherry–Ye [CY01, Theorem 4.2.1 or Theorem 2.8.6]. We cannot directly apply these general theorems in their verbatim forms as they only lead to a bound of the form $m(r, F_N) \ll \frac{1}{N} \log \frac{1}{1-r} + \log N$ in place of the required $m(r, F_N) \ll \frac{1}{N} \log \frac{N}{1-r}$; cf. the term $(q+1)\log(q/\delta)$ in [Hin92, line (1.24)], where $q = N$ signifies the number of targets a_i. But fortuitously we were able to modify their proofs by making an additional use of the key pivot relation (6.2.3) particular to our situation of $\{a_1, \ldots, a_q\} = \mu_N$.

For our case of functions on the disc, we compare to [Hay64, Theorem 2.1]. For holomorphic f, we again have the *ramification term* $N_{\mathrm{ram}}(r,f) = N(r, 1/f')$ (this term is denoted by $N_1(r)$ in [Hay64]), which is always nonnegative. In (6.2.10), even without using the étaleness of F_N, one would drop the ramification term by positivity and still obtain the requisite bound $m\Big(r, \dfrac{\varphi}{\varphi'}\Big) \leq m\Big(r, \dfrac{\varphi'}{\varphi}\Big) + N\Big(r, 1/\varphi\Big)$. In this way, our treatment also recovers the bound $\int_{|z|=r} \log^+ |\varphi| \, \mu_{\mathrm{Haar}} \leq$

$\frac{1}{N-1}\log\frac{1}{1-r}+O_{\varphi}(1)$ for every holomorphic map $\varphi : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$ avoiding the N-th roots of unity (which is not necessarily the universal covering map).

6.3. **Proof of Theorem 1.0.1.** At this point we have established all the pieces for the proof of our main result. By Theorem 5.1.4, assumption (1) in Proposition 3.0.1 is indeed satisfied, with the sharp constant $A := \zeta(3)/2 > 0$. By Theorem 6.0.1 with the choices $p(x) := x^N$ and $r := 1 - BN^{-3}$, assumption (2) in Proposition 3.0.1 is also satisfied. In terms of the algebras of modular forms M_N and R_N at an even Wohlfahrt level N introduced in 4.2.1, the conclusion of Proposition 3.0.1 is thus an inequality $[R_N : M_2] \leq CN^3 \log N$, for some absolute constant $C \in \mathbf{R}$ independent of N. At this point Proposition 4.3.5 proves the equality $R_N = M_N$ for all $N \in \mathbf{N}$, which is the unbounded denominators conjecture.

The proof of Theorem 1.0.1 is thus completed. □

Remark 6.3.1. Our proof for Theorem 1.0.1 generalizes in the obvious way to establish that a modular form $f(\tau)$ having a Fourier expansion in $\overline{\mathbf{Z}}[\![q^{1/N}]\!]$ (algebraic integer Fourier coefficients) at one cusp, and meromorphic at all cusps, is a modular form for a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$. We include an indication of the details.

Since $f(\tau)$ is a modular form, we are reduced to the situation of a number field K such that $f(\tau) \in O_K[\![q^{1/N}]\!]$. We use R_{2N} to denote the $K(\lambda)$-algebra generated by modular functions with coefficients in K, bounded denominators at $\zeta = i\infty$, and cusp widths dividing $2N$ at all cusps $\zeta \in \mathbf{P}^1(\mathbf{Q})$ (similar to Definition 4.2.1). We follow the proof of Proposition 3.0.1 now on the case of the $K(\lambda)$-vector space $\mathcal{V}(U, x(t), O_K)$ from Definition 2.0.4. Then $R_{2N} \subset \mathcal{V}(U, x(t), O_K)$. Note that U is stable under the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, and thus $\mathcal{V}(U, x(t), O_K) = \mathcal{V}(U, x(t), \mathbf{Z}) \otimes_{\mathbf{Q}} K$ and $\dim_{K(\lambda)} \mathcal{V}(U, x(t), O_K) = \dim_{\mathbf{Q}(\lambda)} \mathcal{V}(U, x(t), \mathbf{Z})$. Thus by Corollary 2.0.5, we still have that R_{2N} has dimension at most $CN^3 \log N$ over $K(\lambda)$. The claimed extension to $\overline{\mathbf{Z}}[\![q^{1/N}]\!]$ Fourier expansions now follows upon remarking that the proof of Proposition 4.3.5 still persists when $\mathbf{Q}$ is replaced by K.

Remark 6.3.2. It is also possible to derive the $\bar{\mathbf{Z}}[\![q^{1/N}]\!]$ generalization directly from Theorem 1.0.1, by the following argument pointed out to us by John Voight. The absolute Galois group $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ acts on the q-expansions of modular forms. If $f(\tau) \in O_K[\![q^{1/N}]\!]$ is a modular form on a finite index subgroup of $\mathrm{SL}_2(\mathbf{Z})$, and $\alpha_1, \ldots, \alpha_d$ is a $\mathbf{Z}$-basis of O_K, then $f_i(\tau) := \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i f(\tau)) \in \mathbf{Z}[\![q^{1/N}]\!]$ for each $i = 1, \ldots, d$. Theorem 1.0.1 gives that each $f_i(\tau)$ is modular for some congruence subgroup, say Γ_i. At this point $f(\tau)$, being a K-linear combination of $f_1, \ldots, f_d$, is modular for the congruence subgroup $\Gamma_1 \cap \cdots \cap \Gamma_d$.

7. Generalization to vector-valued modular forms

7.1. **Generalized McKay–Thompson series with roots from Monstrous Moonshine.** Our argument also proves a vector generalization of the unbounded denominators conjecture, which has been conjectured by Mason [Mas12] (see also the earlier work of Kohnen and Mason [KM08] for a special case) to the setting of vector-valued modular forms of $\mathrm{SL}_2(\mathbf{Z})$, with motivation from the theory of vertex operator algebras and the Monstrous Moonshine conjectures. The weaker statement of algebraicity over the ring of modular forms was conjectured earlier by Anderson and Moore [AM88], within the context of the partition functions or McKay–Thompson series attached to rational conformal field theories. We refer also to André [And04, Appendix], for a discussion from the arithmetic algebraization point of view — the method that we build upon in our present paper — on the Grothendieck–Katz p-curvature conjecture. Eventually the more precise expectation crystallized, see Eholzer [Eho95, *Conjecture* on page 628], that all RCFT graded twisted characters are in fact classical modular forms for a *congruence* subgroup of $\mathrm{SL}_2(\mathbf{Z})$ (which is more precise than Anderson and Moore's conjectured algebraicity over the modular ring $\mathbf{Z}[E_4, E_6]$).

This conjecture became known as the *congruence property in conformal field theory*, and was proved in the eponymous paper of Dong, Lin and Ng [DLN15], after landmark progresses from many authors (for some history, including notably Bantay's solution [Ban03] under a certain heuristic assumption, the *orbifold covariance principle* [Ban00, Ban02, Xu06], we refer the reader to the introduction of [DLN15]). Finally, the congruence property for the McKay–Thompson series in the full equivariant setting (orbifold theory) V^G of a finite group G of automorphisms of a rational, C_2-cofinite vertex operator algebra V (the prime example being the Fischer–Griess Monster group operating on the Moonshine module of Frenkel–Lepowski–Meurman [FLM88]) was proved by Dong and Ren [DR18] by a reduction to the special case $G = \{1\}$ that is [DLN15].

Our paper, via Theorem 7.3.3 below for the vector valued extension of the congruence property, inherits a new proof of these modularity theorems. The connection was engineered by Knopp and Mason [KM03a], with their formalization of generalized modular forms for $\mathrm{SL}_2(\mathbf{Z})$, and fine tuned by Kohnen and Mason [KM08, § 4], who brought forward the idea of a purely arithmetic approach — based on the integrality properties of the Fourier coefficients, that record a graded dimension and are hence integers — for a part of Borcherds's theorem [Bor92] (the Conway–Norton "Monstrous Moonshine" conjecture). Namely, suppressing the Hauptmodul property, for the classical modularity — under a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$ — of all the various McKay–Thompson series for the Monster group over the Moonshine module $V^\natural$. Whereas Borcherds's proof, based on his own generalized Kac–Moody algebras that go outside of the general framework of vertex operator algebras, is rather particular to the Monster vertex algebra and genus 0 arithmetic groups, Kohnen and Mason proposed that an arithmetic abstraction from the integrality of Fourier coefficients might open up a window on the modularity and congruence properties to apply just as well in the equivariant setting to any rational C_2-cofinite vertex operator algebra — this theorem, eventually proved in [DLN15, DR18] by other means, was an open problem at the time of [KM08].

It is precisely this arithmetic scheme that we are able to complete with our paper.

7.2. **Unbounded denominators for the solutions of certain ODEs.** In the language of Anderson–Moore [AM88, page 445], the functions occurring below are said to be *quasi-automorphic* for the modular group $\mathrm{PSL}_2(\mathbf{Z})$, while in Knopp–Mason [KM03b] or Gannon [Gan14], they arise as component functions of vector-valued modular forms for $\mathrm{SL}_2(\mathbf{Z})$. We firstly take up the holonomic viewpoint and give a yet another formulation, in the equivalent language of linear ODEs on the triply punctured projective line, where we think of x as the modular function $\lambda(\tau)/16 \in q + q^2\mathbf{Z}[\![q]\!]$, where $q = \exp(\pi i \tau)$, and of $\mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\}$ as the modular curve $Y(2) = \mathbf{H}/\Gamma(2)$. This answers the question raised in [And04, Appendix, A.5]. For simplicity of exposition, we only consider the case of a power series expansion $f(x) \in \mathbf{Z}[\![x]\!]$ here, as opposed to a general Puiseux expansion (see Remark 7.2.2).

Theorem 7.2.1. *Let $f(x) \in \mathbf{Z}[\![x]\!]$ be an integer coefficients formal power series solution of $L(f) = 0$, where L is a linear differential operator without singularities*[3] *on $\mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\}$. If the $x = 0$ local monodromy of L is finite, then $f(x)$ is algebraic, and more precisely, the function $f(\lambda(\tau)/16)$ on $\mathbf{H}$ is automorphic for some congruence subgroup $\Gamma(N)$ of $\mathrm{SL}_2(\mathbf{Z})$.*

Proof. Our condition is that the $x = 0$ local monodromy group is $\mathbf{Z}/N$ for some $N \in \mathbf{N}$. Then the formal function $g(x) := f(x^N)$ is in $\mathbf{Z}[\![x]\!]$ and fulfills a linear ODE on $\mathbf{P}^1 \smallsetminus \{16^{-1/N}\mu_N, \infty\}$. In our notation of Theorem 2.0.5, that means $g \in \mathcal{H}(\mathbf{C} \smallsetminus 16^{-1/N}\mu_N, \mathbf{Z})$. Hence, denoting again by $F_N : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$ the universal covering map taking $F_N(0) = 0$, recalling our exact uniformization radius formula in Theorem 5.1.4 giving in particular the strict lower bound

$$|F_N'(0)| = \sqrt[N]{16}\left(1 + \frac{\zeta(3)}{2N^3} + \frac{3\zeta(5)}{8N^5} + \cdots\right) > \sqrt[N]{16},$$

and letting then

$$\varphi(z) := 16^{-1/N} F_N(rz)$$

for some parameter r with $\sqrt[N]{16}/|F_N'(0)| < r < 1$, Theorem 2.0.5 implies that $g(x) \in \mathbf{Z}[\![x]\!]$ is an algebraic power series. Hence $f(x) = g(\sqrt[N]{x})$ is algebraic.

[3]Similarly to 2.0.4, the proof allows for singularities on $\mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\}$ provided their local monodromy is trivial.

At this point we know that $f(\lambda(\tau)/16)$ is automorphic for some finite index subgroup $\Gamma \subset \Gamma(2)$. Theorem 1.0.1 then upgrades this to automorphy under some congruence modular group $\Gamma(M)$, for some $M \equiv 0 \mod N$, and the result follows upon replacing N with M. $\square$

Remark 7.2.2. To include Puiseux series $f(x) \in \mathbf{C}[\![x^{1/m}]\!]$, the statement and proof apply verbatim on replacing the integrality condition $f(x) \in \mathbf{Z}[\![x]\!]$ by $f(\lambda(\tau)/16) \in \mathbf{Z}[\![\lambda(\tau/m)/16]\!] \otimes \mathbf{C}$.

Remark 7.2.3. The condition in Theorem 7.2.1 that the linear differential operator L has a finite local monodromy at $x = 0$ is essential for algebraicity. The canonical and explicit transcendental example, which is given in [And04, Appendix, A.5] and we have already mentioned in our introduction § 1.1, is the Gauss hypergeometric series or complete elliptic integral of the first kind

$$\frac{2}{\pi}K(x) := {}_2F_1\left[\begin{matrix}1/2 \ 1/2\\ 1\end{matrix}; 16x\right] = \sum_{n=0}^{\infty}\binom{2n}{n}^2 x^n,$$

that is the Hadamard square of $(1-4x)^{-1/2}$ and has the Jacobi theta function parametrization making

$$ {}_2F_1\left[\begin{matrix}1/2 \ 1/2\\ 1\end{matrix}; \lambda(q)\right] = \Big(\sum_{n\in\mathbf{Z}} q^{n^2}\Big)^2 \tag{7.2.4}$$

a weight one modular form for the congruence group $\Gamma(2)$. The modularity streak is not an accident: more generally, to get $\mathbf{Z}[\![x]\!]$ holonomic functions on $\mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\}$ with infinite $x = 0$ local monodromy, we may reversely start with any congruence modular form of a weight $k > 0$, such as for instance Ramanujan's (discriminant) weight 12 modular form $\Delta(\tau) = q\prod_{n=1}^{\infty}(1-q^n)^{24} \in q\mathbf{Z}[\![q]\!]$, and express it formally into a power series in $x := \lambda(\tau)/16$, using $\mathbf{Z}[\![q]\!] = \mathbf{Z}[\![x]\!]$ as in §1.1. It is then a classical fact, cf. Stiller [Sti84] or Zagier [Zag08, § 5.4], that the resulting formal power series fulfills a linear ODE on a finite étale cover of $\mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\} \cong Y(2)$, of order $k+1$ and monodromy group commensurable with $\mathrm{Sym}^k \mathrm{SL}_2(\mathbf{Z}) \hookrightarrow \mathrm{SL}_{k+1}(\mathbf{Z})$.

It remains to us an open question whether a complete description of all integral solutions $f \in \mathbf{Z}[\![x]\!]$ on dropping the $x = 0$ finite local monodromy condition in Theorem 7.2.1 should arise in this way from a classical congruence modular form expressed into a holonomic function in $x = \lambda/16$. We formulate the precise statement in Question 7.4.3 below.

7.3. **Vector-valued modular forms.** We close our paper by another formulation of Theorem 7.2.1, translated now over to the language of vector-valued modular forms. The following definition is a special case of vector-valued modular forms studied in [FM16a, §2, Definition 1]

Definition 7.3.1. *A* vector-valued modular form *of weight $k \in \mathbf{Z}$ and dimension n for $\mathrm{SL}_2(\mathbf{Z})$ is a pair (F, ρ) made of a holomorphic mapping $F = (F_1, \ldots, F_n) : \mathbf{H} \to \mathbf{C}^n$ and an n-dimensional complex representation*

$$\rho : \mathrm{SL}_2(\mathbf{Z}) \to \mathrm{GL}_n(\mathbf{C})$$

obeying the following properties:

- *For all $\gamma \in \mathrm{SL}_2(\mathbf{Z})$,*
$$F^{\mathrm{t}}\,|_k\gamma = \rho(\gamma)F^{\mathrm{t}};$$
- *The matrix*
$$\rho\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} \in \mathrm{GL}_n(\mathbf{C})$$
is semisimple.
- *All components $F_j : \mathbf{H} \to \mathbf{C}$ have* moderate growth in vertical strips*: for all $a < b$ and $C > 0$, there exist $A, B > 0$ such that*
$$\forall \tau \in \mathbf{H}, \quad a \le \operatorname{Re}\tau \le b, \quad \operatorname{Im}\tau \ge C \quad \Longrightarrow \quad |F_j(\tau)| \le Ae^{B\operatorname{Im}\tau}.$$

Here, as usual, $|_k$ is used to denote the componentwise right action of $\gamma = \begin{pmatrix}a & b\\ c & d\end{pmatrix}$ via the usual automorphy factor $j_k(\gamma,\tau) = (c\tau+d)^{-k}$:

$$f(\tau)\,|_k\gamma := j_k(\gamma,\tau)f(\gamma\tau) = (c\tau+d)^{-k}f(\gamma\tau).$$

Remark 7.3.2. Taken together, see [AM88, § 2.A], the semisimplicity and moderate growth conditions are equivalent to the existence of generalized Puiseux formal expansions (except in general with irrational exponents: but without $\log q$ terms, due to semisimplicity) of each component function $F_j(\tau)$ at the cusp $q = 0$. More precisely, via a change of basis (see the equivalent notion in [FM16a]), we may assume that $\rho\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ is a diagonal matrix. If F_j is a λ-eigenvector of $\rho\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$, then $F_j = \sum_{n\in\mathbf{Z}_{\geq n_0}} a_{n,j}q^{n+\mu}$ for some $n_0 \in \mathbf{Z}$, where $q = e^{2\pi i\tau}$ and we choose a $\mu \in \mathbf{C}$ such that $\lambda = e^{2\pi i\mu}$.

Thus, the classical (scalar-valued) modular forms $M_k(\Gamma(1), \chi)$ attached to a finite-order character $\chi : \Gamma(1) \to U(1)$ are precisely the special case $n = 1$ of one-dimensional vector-valued modular forms and a unitary character ρ. In a reverse direction, any classical (scalar-valued) modular form for a finite index subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbf{Z})$ can be considered as the first component of a vector-valued modular form for $\mathrm{SL}_2(\mathbf{Z})$ of dimension $[\Gamma(1) : \Gamma]$. From that point of view, there is no loss of generality in Definition 7.3.1 to limit to the representations of the ambient group $\mathrm{SL}_2(\mathbf{Z})$.

Knopp and Mason's *generalized modular forms* [KM03a] are the case, intermediate in generality, where the representation ρ is *monomial*: that is, induced from a linear character $\chi : \Gamma \to \mathbf{C}^\times$ on a finite index subgroup $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$. If that character χ is unitary, then in fact it has finite image and all components of F are classical modular forms of weight k for a finite index subgroup [KM03a]. The general (non-unitary) case does come up for the partition function and correlation functions of a rational conformal field theory [KM03a], to which the point of contact is supplied by Zhu's modularity theorem [Zhu96] (see also Codogni [Cod20] for a recent different proof and a generalization), and its extension to the equivariant setting by Dong, Li and Mason [DLM00].

To make the connection to Theorem 7.2.1, note upon restricting the representation ρ to the free subgroup $\mathbf{Z} * \mathbf{Z} \cong \Gamma(2) \subset \Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ that the case of weight $k = 0$ and finite-order element $\rho\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ is equivalent to exactly the situation of 7.2.1: a local system on the triply-punctured projective line $Y(2) \cong \mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\}$ *that has a finite local monodromy around the puncture* $x = 0$. We refer the reader to [BG07, Gan14] regarding the bridge between these two equivalent points of view.

Our general result on unbounded denominators for components of vector-valued modular forms is the following.

Theorem 7.3.3. *Let (F, ρ) be a vector-valued modular form for $\mathrm{SL}_2(\mathbf{Z})$ of dimension n and weight k. Suppose that some component function $F_j(\tau) : \mathbf{H} \to \mathbf{C}$ of $F = (F_1, \ldots, F_n) : \mathbf{H} \to \mathbf{C}^n$ has at $\tau = i\infty$ a formal Fourier expansion lying in $\mathbf{Z}[\![q]\!] = \mathbf{Z}[\![e^{\pi i\tau}]\!]$. Then that component $F_j(\tau)$ is a classical modular form of weight k on a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$.*

Proof. After some standard theorems from the theory of G-functions to reduce to the case that the semisimple matrix $\rho\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} \in \mathrm{GL}_n(\mathbf{C})$ is in fact of finite order, this is an equivalent expression of Theorem 7.2.1.

The transition is as follows. By taking the componentwise product $F(\tau)g(\tau)(\lambda(\tau)/16\Delta(\tau/2))^{\frac{k+k'}{12}}$, where we choose a non-zero scalar-valued modular form $g(\tau) \in \mathbf{Z}[[q]]$ of weight k' such that $12 \mid k + k'$, and restricting the full modular group to its subgroup $\Gamma(2)$, we reduce to the case $k = 0$ of local systems on $Y(2) \cong \mathbf{P}^1 \smallsetminus \{0, 1/16, \infty\}$. Without loss of generality upon passing to a factor, we may assume that local system to be irreducible. The holomorphic vector bundle with integrable connection admitting F for its horizontal sections is indeed meromorphic at the cusps of $Y(2)$ due to the existence of the q-expansion of F. Hence F is a solution of a rank-n system of first-order linear homogeneous ODEs over $\mathbf{Q}[\lambda, 1/\lambda, 1/(1-\lambda)]$. By the theorem of the cyclic vector, see [DGS94, § III.4], there is an irreducible linear differential operator L over $\mathbf{Q}(\lambda)$ without singularities on $Y(2) = \mathrm{Spec}\,\mathbf{Z}[\lambda, 1/\lambda, 1/(1-\lambda)]$ and such that all n component functions $F_1, \ldots, F_n$ are formal solutions of the linear homogeneous ODE $L(f) = 0$. Since one of these (namely, F_j) has a $\lambda = 0$ formal expansion in $\mathbf{Z}[\![q]\!] = \mathbf{Z}[\![\lambda/16]\!]$, Chudnovsky's theorem [DGS94, Theorem VIII.1.5] implies that L satisfies the Galočkin (finite global operator height $\sigma(L) < \infty$) condition [DGS94, VII.2.(2.3) on page 227], hence by the Bombieri–André theorem [DGS94, Theorem VII.2.1], L satisfies the Bombieri (finite generic global inverse radius $\rho(L) < \infty$) condition [DGS94, VII.2.(2.1) on page 226], and is therefore globally nilpotent. At this point Katz's local monodromy theorem [Kat70] (see also [DGS94, Theorem III 2.3 (ii)]) proves that L has quasi-unipotent local monodromies. Now by the semisimplicity condition on $\rho\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ in Definition 7.3.1, it follows that in fact the $x = 0$ local monodromy of L has finite order. The result then follows on applying Theorem 7.2.1 to $f(x) = F_j(\tau)$. □

Corollary 7.3.4 (Mason's conjecture)**.** *If all components of a vector-valued modular form (F, ρ) for $\mathrm{SL}_2(\mathbf{Z})$ have Fourier expansions with bounded denominators, then the representation ρ has a finite image, and more precisely $\ker(\rho) \supseteq \Gamma(N)$ for some $N \in \mathbf{N}$.*

7.4. **Some questions and concluding remarks.**

7.4.1. *A brief survey of the literature on Mason's conjecture.* Mason's conjecture, as discussed in [Mas12, KM08, KM12], concerned the stronger condition in Corollary 7.3.4, namely that *all* components $F_1, \ldots, F_n$ have bounded denominators. These are the cases emerging in conformal field theories, and apart from Gottesman's result [Got20, Theorem 1.7] resolving a strong form of the conjecture for a class of two-dimensional vector-valued modular forms on $\Gamma_0(2)$, the literature on the vector-valued case has focused on the stronger assumption for the full vector of components F. We review some of this work here.

Originally Kohnen and Mason [KM08, KM12] focused on the particular case (*GMF*) that the representation ρ is monomial (induced from a one-dimensional character on a finite index subgroup of $\mathrm{SL}_2(\mathbf{Z})$). They used the Rankin–Selberg method to prove the conjecture in the case of a generalized modular function (weight 0) without any zeros or poles on the extended upper-half plane [KM08, Theorem 1]. In fact Selberg's paper [Sel65] that they used here had already considered vector-valued modular forms for the purpose of extending the Rankin–Selberg estimate into the noncongruence case. Kohnen and Mason [KM08, Theorem 2], again based on the Rankin–Selberg L-function method but now with a finer input from the Eichler–Shimura–Weil bound on Fourier coefficients of congruence cusp forms in weight 2, also proved that when ρ is induced from a linear character of a *congruence* subgroup of $\mathrm{SL}_2(\mathbf{Z})$, the same result on generalized modular function units also holds if the condition on integer coefficients is relaxed to S-integer coefficients: a case that goes beneath the scope of our results here.

In a sequel work [KM12], Kohnen and Mason used the Knopp–Mason canonical factorization [KM09] $f = f_0 f_1$ (over $\mathbf{C}$) of a parabolic generalized modular function f on a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$, where f_0 is a parabolic generalized modular function of a *unitary* character χ, while f_1 is a parabolic generalized modular function without zeros or poles on the extended upper-half plane [KM03a]. Combining to their earlier method from [KM08], they thus proved that the unbounded denominators conjecture for the case of *parabolic* GMF is equivalent to the algebraicity of the first "few" Fourier coefficients of the component f_1 in the canonical factorization of f. As an application they proved Mason's unbounded denominators conjecture for the case of a cuspidal parabolic GMF of weight 0 on a congruence group.

In the case $n = 2$ of two-dimensional representations, Mason's conjecture was settled by Franc and Mason [Mas12, FM14], and extended further by Franc, Gannon and Mason [FGM18] to the stronger sense of only requiring the p-adic boundedness of the coefficients for a full density set of primes p. Their proof relies on the special incidence that the rank-2 local systems on $\mathbf{P}^1 \setminus \{0, 1, \infty\}$ reduce to the Gauss hypergeometric equation, and the classical theory of hypergeometric functions. It is conceivable that the algebraicity part (over $\mathbf{Q}(x)$, respectively over the ring of classical modular forms) in Theorems 7.2.1 and 7.3.3 could likewise hold under a similar loosening of the integrality condition; but our proof does not yield to this. On the other hand, for representations of dimension $n \geq 3$, it is plain that the congruence property ceases to hold as in [FM14] if we relax $\mathbf{Z}[\![q]\!]$ to $\mathbf{Z}[1/S][\![q]\!]$. The hypergeometric method was extended to three-dimensional representations

($n = 3$) of $\mathrm{SL}_2(\mathbf{Z})$ by Franc–Mason [FM16a] and Marks [Mar15], and employed back in [FM16b] to derive certain cases of the original unbounded denominators conjecture.

7.4.2. *Logarithmic vector-valued modular forms.* If one drops the semisimplicity stipulation on $\rho\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ in the definition of a vector-valued modular form, the resulting structure has been named a *logarithmic vector-valued modular form* by Knopp and Mason [KM11]. They also do arise in conformal field theories, termed *logarithmic* (in place of rational). See, for example, Fuchs–Schweigert [FS19]. But now by Remark 7.2.3 the components of a weight zero vector-valued modular form with bounded denominators can certainly be transcendental over $\mathbf{C}(\lambda)$. Still the examples there are classical (congruence) modular forms, except of a higher weight. In Question 7.4.3 below we give an extension of the unbounded denominators problem over to the logarithmic setting. It remains outside the scope of our method as far as we could see. Before stating this question, we recall some basic facts concerning quasi-modular forms. Recall that the ring of *quasi-modular forms* $\widetilde{M}(\Gamma)$ for $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ may be identified with the ring generated by E_2 over the ring of classical holomorphic modular forms $M(\Gamma)$ of integral weight for Γ [Zag08, Prop 20(ii)], and by [Zag08, Prop 20(i)] it is stable under the operator

$$\theta = q \cdot \frac{d}{dq} = \frac{1}{\pi i} \cdot \frac{d}{d\tau}.$$

Let $M^!(\Gamma)$ denote the ring of weakly holomorphic modular forms for Γ; that is, the meromorphic modular forms which are holomorphic away from the cusps.

Definition 7.4.1. *The ring of* weakly holomorphic quasi-modular forms $\widetilde{M}^!(\Gamma)$ *for* $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ *is the ring* $\widetilde{M}(\Gamma)[1/\Delta]$.

Lemma 7.4.2. *The ring $\widetilde{M}^!(\Gamma)$ is the smallest ring which contains $M^!(\Gamma)$ and which is closed under θ.*

Proof. If $f \in M^!(\Gamma)$ then $f\Delta^m \in M(\Gamma) \subset \widetilde{M}(\Gamma)$ for some m and thus $M^!(\Gamma) \subset \widetilde{M}^!(\Gamma)$. Recall that $\theta\Delta = E_2\Delta$. If $g \in \widetilde{M}^!(\Gamma)$, then $h = \Delta^m g \in \widetilde{M}(\Gamma)$ for some m, and thus

$$\theta g = \theta\frac{h}{\Delta^m} = \frac{\theta h}{\Delta^m} - \frac{mE_2 h}{\Delta^m} \in \widetilde{M}^!(\Gamma),$$

and hence $\widetilde{M}^!(\Gamma)$ is closed under θ. Finally, any ring containing $M^!(\Gamma)$ and closed under θ contains both Δ^{-1} and $E_2 = \theta\Delta/\Delta$ and thus contains $\widetilde{M}^!(\Gamma)$. □

Question 7.4.3. *If a component $F_j(\tau)$ of a logarithmic vector-valued modular form for $\mathrm{SL}_2(\mathbf{Z})$ has a $\mathbf{Z}[\![q]\!]$ Fourier expansion, does $F_j(\tau)$ belong to the ring of weakly holomorphic quasi-modular forms $\widetilde{M}^!(\Gamma)$ for some* congruence *subgroup $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$?*

Recall the classical Jacobi theta functions:

$$\vartheta_2 = \sum_{n\in\mathbf{Z}} q^{(n+1/2)^2}, \quad \vartheta_3 = \sum_{n\in\mathbf{Z}} q^{n^2}, \quad \vartheta_4 = \sum_{n\in\mathbf{Z}} (-1)^n q^{n^2}.$$

By Jacobi's triple product identity, these functions ϑ_i have explicit representations in terms of the Dedekind η function:

$$\vartheta_2 = \frac{2\eta^2(2\tau)}{\eta(\tau)}, \qquad \vartheta_3 = \frac{\eta^5(\tau)}{\eta^2(\tau/2)\eta^2(2\tau)}, \qquad \vartheta_4 = \frac{\eta^2(\tau/2)}{\eta(\tau)},$$

and hence they are holomorphic modular forms of weight $1/2$ without any zeros on $\mathbf{H}$. Consequently, all the Laurent monomials $\vartheta_2^a\vartheta_3^b\vartheta_4^c$ (with $a,b,c \in \mathbf{Z}$) of an even degree $a+b+c$ belong to $\widetilde{M}(\Gamma)$ for some fixed congruence subgroup Γ (one can take $\Gamma = \Gamma(12)$, although some monomials are invariant under smaller groups, for example: $(\vartheta_2/\vartheta_3)^4 = \lambda$ by Equation (1.0.2)). Finally, we also have $2(\theta\vartheta_i)/\vartheta_i = (\theta\vartheta_i^2)/\vartheta_i^2 \in \widetilde{M}^!(\Gamma)$.

We now turn to some basic examples hinting towards a positive answer to question 7.4.3. Complementing Example 7.2.3 is the λ-pullback of the complete elliptic integral of the second kind:

$$\frac{2}{\pi}E(\lambda(q)) := {}_2F_1\left[\begin{matrix}1/2 & -1/2\\ & 1\end{matrix};\lambda(q)\right] = 1-4q+20q^2-64q^3+164q^4-392q^5+\cdots \in \mathbf{Z}[\![q]\!],$$

clearly a component of a logarithmic vector-valued modular form on $\mathrm{SL}_2(\mathbf{Z})$, whose q-expansion is in $\mathbf{Z}[\![q]\!]$. But one can indeed verify that

$$\frac{2}{\pi}E(\lambda(q)) = \frac{\vartheta_3\vartheta_4^4+4\theta\vartheta_3}{\vartheta_3^3}.$$

is also an element of $\widetilde{M}^!(\Gamma)$ for the congruence subgroup $\Gamma=\Gamma(12)$. One can express E in terms of K and its integral:

$$\frac{2}{\pi}E(16x) = (16x-1)\,\frac{2}{\pi}K(16x) - 8\int_0^x \frac{2}{\pi}K(16t)\,dt,$$

where one finds that

$$\int_0^x \frac{2}{\pi}K(16t)\,dt = \sum_{n=0}^{\infty}\frac{1}{n+1}\binom{2n}{n}^2 x^{n+1} \in \mathbf{Z}[\![x]\!],$$

the integrality of the coefficients now manifested by the Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n}\in\mathbf{Z}$. One further integration still has integer coefficients:

$$\int_0^x \frac{dy}{y}\int_0^y \frac{2}{\pi}K(16t)\,dt = \sum_{n=0}^{\infty}\frac{1}{(n+1)^2}\binom{2n}{n}^2 x^{n+1} = \sum_{n=0}^{\infty} C_n^2\,x^{n+1} \in \mathbf{Z}[\![x]\!],$$

and $\sum \frac{1}{(n+1)^2}\binom{2n}{n}^2(\lambda/16)^n \in \mathbf{Z}[\![q]\!]$ is a component of a logarithmic vector-valued modular form with a $\mathbf{Z}[\![q]\!]$ expansion. Zudilin has pointed out to us the formula

$$\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}\binom{2n}{n}^2\left(\lambda(q)/16\right)^n = \frac{4}{\vartheta_2^4}\left(4\vartheta_3^2\cdot\frac{\theta\vartheta_2}{\vartheta_2}+4\vartheta_3\cdot\theta\vartheta_3-\vartheta_3^4\right)$$

exhibiting this $\mathbf{Z}[\![q]\!]$ power series as an element of $\widetilde{M}^!(\Gamma)$ for the congruence subgroup $\Gamma=\Gamma(12)$, in accordance with Question 7.4.3.

7.4.3. *Some variations.* Our proof of Theorems 1.0.1 and 7.3.3 is readily refined to yield a further precision in two regards:

Firstly, the condition on $\mathbf{Z}[\![q^{1/N}]\!]$ Fourier coefficients can be relaxed to $\mathbf{Z}[\![q^{1/N}]\!]\otimes\mathbf{C}$ Fourier coefficients.

Secondly, the condition that the modular form $f(\tau)$, respectively the vector-valued modular form $F(\tau)$ are holomorphic on $\mathbf{H}$ can be relaxed to the condition of meromorphy on $\mathbf{H}$.

We leave it to the interested reader to fill in the details of these further extensions of our results.

7.4.4. *Beyond* $\mathrm{SL}_2(\mathbf{Z})$. Much less obvious is how to extend our results to arithmetic groups other than $\mathrm{SL}_2(\mathbf{Z})$. Here are two possible settings one could consider.

Firstly, the group $\mathrm{SL}_2(\mathbf{F}_q[t])$ in function field arithmetic and its attendant theory of Drinfeld–Goss modular forms. See Pellarin [Pel21] for a recent survey of this area. Here, in the analogy with $\mathrm{SL}_2(\mathbf{Z})$ where the congruence kernels of these two arithmetic groups are similarly large, it would be interesting to decide whether the modular forms on a finite index subgroup of $\mathrm{SL}_2(\mathbf{F}_q[t])$ that have (up to a $\mathbf{F}_q(t)^\times$ scalar multiple) a u-expansion [Pel21, § 4.7.1] with coefficients in $A=\mathbf{F}_q[t]$ are likewise the congruence modular forms.

Secondly, the mapping class groups $\Gamma_{g,n}=\mathrm{Mod}(S_{g,n})$ in signatures (g,n) other than $(1,1),(1,0)$ or $(0,4)$ that we have implicitly been limiting to. Recall that $\Gamma_{1,1}\cong\Gamma_{1,0}=\mathrm{Mod}(\mathbf{T}^2)=\mathrm{SL}_2(\mathbf{Z})$ and $\Gamma_{0,4}\cong\mathrm{PSL}_2(\mathbf{Z})\ltimes(\mathbf{Z}/2\times\mathbf{Z}/2)$, and correspondingly the discussion in the rational conformal field theory under §7.1 has been for the 1-loop partition function with a complex torus ($g=1$) as the worldsheet [Gan06]. In a more recent research stream in two-dimensional conformal field theory, a higher genus extension of Zhu's modularity theorem was recently obtained by Codogni [Cod20], on associating to any holomorphic vertex operator algebra a Teichmüller modular form in every signature (g, n): a section of a tensor power $\lambda^{\otimes(c/2)}$ of the Hodge bundle over $\overline{\mathcal{M}_{g,n}}$, where the (doubled) weight c is the *central charge* of the vertex algebra. This Teichmüller modular form is, up to the c-th power of a certain higher genus generalization [MT06] of the Dedekind eta function, equal to the partition function of the conformal field theory associated to the vertex algebra. At the very least, one could ask about extending the cruder algebraicity proviso of our Theorem 7.3.3 over to the more general setting of a component of a vector-valued Teichmüller modular form that has the appropriate integrality property.

7.4.5. *Algebraic fundamental groups.* Finally we return to our introductory outline § 1.1 where we acknowledged that our approach to the unbounded denominators conjecture has been particularly inspired by the papers of Ihara [Iha94] and Bost [Bos99] on arithmetic algebraization and Lefschetz theorems in Arakelov geometry. Our central overconvergence boost emerged from the isogeny $[N]$ of $\mathbf{G}_m$ to trade a Belyĭ map, or more generally a local system on $\mathbf{P}^1 \smallsetminus \{0, 1, \infty\}$ that has a $\mathbf{Z}/N$ local monodromy around $x = 0$, for a local system on $\mathbf{P}^1 \smallsetminus \{\mu_N \cup \infty\}$: the step of extending through the falsely apparent singularity at $x = 0$. This is directly inspired by Ihara's employment of an arithmetic rationality theorem of Harbater [Iha94, §1 Lemma] to derive π_1 results on certain arithmetic schemes, including for instance a Diophantine analysis proof of Saito's example of $\pi_1\big(\operatorname{Spec} \mathbf{Z}[x, 1/x, 1/(x-1)]\big) = \{1\}$. In a similar fashion, our Theorem 1.0.1 can be used to establish a π_1 result in the style of Bost [Bos99].

Theorem 7.4.4. *Let $N \in \mathbf{N}$, let $K/\mathbf{Q}(\mu_N)$ be a finite extension, and let $\pi : \mathcal{X}(N) \to \operatorname{Spec} O_K$ ("connected Néron model") be the connected component containing the cusp ∞ in the smooth part of the minimal regular model of $X(N)$ over $\operatorname{Spec} O_K$. Thus the cusp ∞ extends to a morphism $\varepsilon : \operatorname{Spec} O_K \to \mathcal{X}(N)$.*

Then, for every geometric point η of $\operatorname{Spec} O_K$, the maps of algebraic fundamental groups

$$\pi_* : \pi_1(\mathcal{X}(N), \varepsilon(\eta)) \to \pi_1(\operatorname{Spec} O_K, \eta)$$

and

$$\varepsilon_* : \pi_1(\operatorname{Spec} O_K, \eta) \to \pi_1(\mathcal{X}(N), \varepsilon(\eta))$$

are mutually inverse isomorphisms.

Proof (a sketch). This follows rather formally by the argument of [Iha94, § 4 on page 252] and [Iha94, proof of Theorem 1 *loc.cit.* on pages 248–249], upon replacing Ihara's function field $k(t)$ by the modular function field $K(X(N))$ and Ihara's formal power series ring $\mathfrak{O}[\![t]\!]$ by $O_K[\![\lambda(\tau/N)/16]\!]$, taking account of Remark 6.3.1, and on using our Theorem 1.0.1 in place of Harbater's arithmetic rationality input [Iha94, Claim 1A on page 248]. □

Remark 7.4.5. Very recently, Bost and Charles [BC22] have obtained new relative π_1 finiteness theorems for certain quasi-projective arithmetic surfaces $\mathcal{X} \to \operatorname{Spec} O_K$, including for the case [BC22, § 9.3.4] of the affine modular scheme $\mathcal{Y}(N)^{\mathrm{arith}} \to \operatorname{Spec} \mathbf{Z}$ that represents the functor "full level N structure" ($N \geq 3$) in the sense of isomorphisms $\iota : (\mu_N \times \mathbf{Z}/N\mathbf{Z})_S \xrightarrow{\simeq} \mathcal{E}[N]$ of finite flat group schemes over a test scheme S.

Remark 7.4.6. Another π_1 interpretation of the unbounded denominators conjecture, in terms of the Galois theory of the Tate curve and the congruence kernel of $\mathrm{SL}_2(\mathbf{Z})$, was given by Chen [Che18, Conjecture 5.5.10].

Similarly to our choice of the isogeny $[N] : \mathbf{G}_m \to \mathbf{G}_m$, one could perhaps more directly consider the modular covering $X(2N) \to X(2)$ and use that it is totally ramified of index N over the three cusps of $X(2)$. Thus a local system $(\mathcal{E}, \nabla)$ on the modular curve $Y(2) \cong \mathbf{P}^1 \smallsetminus \{0, 1, \infty\}$ that has $\mathbf{Z}/N$ local monodromies around the three singularities has its pullback $g^*\mathcal{E}$ under the modular covering $g : Y(2N) \to Y(2)$ extend through the cusps of $Y(2N)$ to a local system on the projective curve $X(2N)$. See also André [And04, II § 8.3], for a more general setting. Another natural approach to the unbounded denominators conjecture would then be to aim directly for rationality on the curve $X(2N)$, instead of for a tight algebraicity or holonomicity rank bound over $X(2)$. Certainly at least the algebraicity clause of Theorems 7.2.1 and 7.3.3 is also possible by this alternative higher genus route to an arithmetic algebraization.

It is tempting to approach Theorem 7.4.4 or the congruence property directly using the arithmetic rationality theorem of Bost and Chambert-Loir [BCL09], although we were unable to do so. In these optics, it may be of some interest to remark that the case of Theorem 7.4.4 with $N = 6$ and K a sufficiently large number field to attain semistable reduction is contained in [Bos99, Corollary 1.3 with Example 7.2.2 (i)]. Indeed, the modular curve $X(6)$ has genus 1 and turned into an elliptic curve using the cusp ∞ for the origin. Since this elliptic curve contains the automorphism $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ of order 6, it has j-invariant 0 and is analytically isomorphic with the complex torus $\mathbf{C}/\mathbf{Z}[\omega]$, $\omega = e^{\pi i/3} = \frac{1+\sqrt{-3}}{2}$, with complex multiplication by the Eisenstein integers $\mathbf{Z}[\omega]$, and in particular extending to a (smooth, proper) abelian scheme over $\operatorname{Spec} O_K$. Its Faltings height is

$$-\frac{1}{2}\log\Big\{\frac{1}{\sqrt{3}}\Big(\frac{\Gamma(1/3)}{\Gamma(2/3)}\Big)^3\Big\} = -0.749\ldots < -0.05\ldots = \frac{1}{2}\log\frac{\pi}{4\operatorname{Im}\omega},$$

by the Lerch–Chowla–Selberg formula making Bost's capacitary condition [Bos99, Corollary 1.3] apply, and this is the isolated minimum value of the Faltings height across all elliptic curves. In practice this means that this complex torus has a "large" *univalent* complex-analytic uniformization (in the sense of conformal size from the origin $[\infty]$ and potential theory), sufficient to place this particular case of Theorem 7.4.4 to within the framework of arithmetic *rationality* — as opposed to algebraicity or holonomicity — theorems [Bos99, BCL09] on the algebraic curve $X(N)$. Can such an approach be continued to all N?

8. Acknowledgments

We would like to thank Yves André for a number of insightful remarks on the first version of this manuscript, leading in particular to § 2.4.1 as an alternative and simplified approach to Theorem 2.0.1. We would also like to thank Michael Barz, Jean-Benoît Bost, François Charles, Pierre Deligne, Cameron Franc, Igor Frenkel, Javier Fresán, Kenz Kallal, Mark Kisin, Geoffrey Mason, Alex Smith, Richard Taylor, John Voight, and Wadim Zudilin for useful remarks, suggestions, and corrections.

References

[AM88] Greg Anderson and Greg Moore, *Rationality in conformal field theory*, Commun. Math. Physics **117** (1988), 119–136.

[Ami75] Yvette Amice, *Les nombres p-adiques*, Collection SUP: "Le Mathématicien", vol. 14, Presses Universitaires de France, Paris, 1975, Préface de Ch. Pisot.

[And89] Yves André, *G-Functions and Geometry*, Aspects of Mathematics, no. E13, Friedr. Vieweg Sohn, Braunschweig, 1989.

[And04] ———, *Sur la conjecture des p-courbures de Grothendieck–Katz et un problème de Dwork*, Geometric Aspects of Dwork Theory, vol. I, de Gruyter, Berlin, 2004, pp. 55–112.

[AS92] Milton Abramowitz and Irene A. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, Inc., New York, 1992, Reprint of the 1972 edition.

[ASD71] A.O.L. Atkin and H.P.F. Swinnerton-Dyer, *Modular forms on noncongruence subgroups*, Proc. Symposia Pure Math.: Combinatorics, vol. XIX, American Mathematical Society, 1971, pp. 1–26.

[ASVV10] G. D. Anderson, T. Sugawa, M. K. Vamanamurthy, and M. Vuorinen, *Twice-punctured hyperbolic sphere with a conical singularity and generalized elliptic integral*, Math. Z. **266** (2010), no. 1, 181–191.

[Ban00] Peter Bantay, *Frobenius–Schur indicators, the Klein-bottle amplitude, and the principle of orbifold covariance*, Phys. Lett. B **488** (2000), no. 2, 207–210.

[Ban02] ———, *Permutation orbifolds*, Nuclear Phys. B **633** (2002), no. 3, 365–378.

[Ban03] ———, *The kernel of the modular representation and the Galois action in RCFT*, Comm. Math. Phys. **233** (2003), no. 3, 423–438.

[BC22] Jean-Benoît Bost and François Charles, *Quasi-projective and formal-analytic arithmetic surfaces*, 2022, https://arxiv.org/abs/2206.14242, pp. 165+xx.

[BCL09] Jean-Benoît Bost and Antoine Chambert-Loir, *Analytic curves in algebraic varieties over number fields*, Algebra, arithmetic and geometry: in honor of Yu. I. Manin. Vol. I, Birkhäuser Boston, Boston, MA, 2009, pp. 69–124.

[Ber94] Gabriel Berger, *Hecke operators on noncongruence subgroups*, C. R. Acad. Sci. Paris Sér. I Math. **319** (1994), no. 9, 915–919.

[BG06] Enrico Bombieri and Walter Gubler, *Heights in Diophantine Geometry*, Cambridge New Mathematical Monographs, no. 4, Cambridge University Press, 2006.

[BG07] Peter Bantay and Terry Gannon, *Vector-valued modular functions for the modular groups and the hypergeometric equation*, Commun. Number Theory Phys. **1** (2007), 651–680.

[Bil97] Yuri Bilu, *Limit distribution of small points on algebraic tori*, Duke Math. J. **89** (1997), no. 3, 465–476.

[Bir94] Bryan Birch, *Noncongruence subgroups, covers and drawings*, The Grothendieck theory of dessins d'enfants (Luminy 1993, ed. L. Schneps), Cambridge University Press, Cambridge, 1994, London Math. Soc. Lecture Note Series, vol. **200**, pp. 25–46.

[BK01] Djamel Benbourenane and Risto Korhonen, *On the growth of the logarithmic derivative*, Comput. Methods Funct. Theory **1** (2001), no. 2, [On table of contents: 2002], 301–310.

[Bor94] Émile Borel, *Sur une application d'un théorème de M. Hadamard*, Bulletin des sciences mathématiques **18** (1894), 22–25.

[Bor92] Richard E. Borcherds, *Monstrous moonshine and monstrous Lie superalgebras*, Invent. Math. **109** (1992), 405–444.

[Bos99] Jean-Benoît Bost, *Potential theory and Lefschetz theorems for arithmetic surfaces*, Ann. Sci. École Norm. Sup. (4) **32** (1999), 241–312.

[Bos04] Jean-Benoît Bost, *Germs of analytic varieties in algebraic varieties: canonical metrics and arithmetic algebraization theorems*, Geometric aspects of Dwork theory. Vol. I, Walter de Gruyter, Berlin, 2004, pp. 371–418.

[Bos13] ———, *Algebraization, transcendence, and D-group schemes*, Notre Dame J. Form. Log. **54** (2013), no. 3-4, 377–434.

[Bos20] ———, *Theta invariants of Euclidean lattices and infinite-dimensional Hermitian vector bundles over arithmetic curves*, Progress in Mathematics, vol. 334, Birkhäuser/Springer, 2020.

[Car54] C. Carathéodory, *Theory of Functions of a Complex Variable. Vol. 2*, Chelsea Publishing Co., New York, 1954, Translated by F. Steinhardt.

[CE12] Frank Calegari and Matthew Emerton, *Completed cohomology—a survey*, Non-abelian fundamental groups and Iwasawa theory, London Math. Soc. Lecture Note Ser., vol. 393, Cambridge Univ. Press, Cambridge, 2012, pp. 239–257.

[Che18] William Yun Chen, *Moduli interpretations for noncongruence modular curves*, Math. Ann. **371** (2018), 41–126.

[CL02] Antoine Chambert-Loir, *Théorèmes d'algébricité en géométrie diophantienne (d'après J.-B. Bost, Y. André, D. & G. Chudnovsky)*, Séminaire Bourbaki. Astérisque **282** (2002), 175–209.

[Cod20] Giulio Codogni, *Vertex algebras and Teichmüller modular forms*, 2020, arXiv:1901.03079, p. 33.

[CV19] Frank Calegari and Akshay Venkatesh, *A torsion Jacquet-Langlands correspondence*, Astérisque (2019), no. 409, x+226.

[CY01] William Cherry and Zhuan Ye, *Nevanlinna's Theory of Value Distribution*, Springer Monographs in Mathematics, Springer Verlag, Berlin, 2001.

[DDT97] Henri Darmon, Fred Diamond, and Richard Taylor, *Fermat's last theorem*, Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993), Int. Press, Cambridge, MA, 1997, pp. 2–140.

[DGS94] Bernard Dwork, Giovanni Gerotto, and Francis J. Sullivan, *An introduction to G-functions*, Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, 1994.

[DLM00] Chongying Dong, Haisheng Li, and Geoffrey Mason, *Modular-invariance of trace functions in orbifold theory and generalized Moonshine*, Comm. Math. Phys. **214** (2000), 1–56.

[DLN15] Chongying Dong, Xingjun Lin, and Siu-Hung Ng, *Congruence property in conformal field theory*, Algebra and Number Theory **9** (2015), 2121–2166.

[DR18] Chongying Dong and Li Ren, *Congruence property in orbifold theory*, Proc. Amer. Math. Soc. **146** (2018), no. 2, 497–506.

[DT97] Michael Drmota and Robert F. Tichy, *Sequences, discrepancies and applications*, Lecture Notes in Mathematics, vol. 1651, Springer-Verlag, Berlin, 1997.

[EGM98] Jürgen Elstrodt, Fritz Grunewald, and Jens Mennicke, *Groups acting on hyperbolic space. Harmonic analysis and number theory*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998.

[Eho95] Wolfgang Eholzer, *On the classification of modular fusion algebras*, Commun. Math. Physics **172** (1995), 623–659.

[FF22] Andrew Fiori and Cameron Franc, *The unbounded denominators conjecture for the noncongruence subgroups of index 7*, J. Number Theory **240** (2022), 611–640.

[FGM18] Cameron Franc, Terry Gannon, and Geoffrey Mason, *On unbounded denominators and hypergeometric series*, J. Number Theory **192** (2018), 197–220.

[FLM88] Igor Frenkel, James Lepowsky, and Arne Meurman, *Vertex operator algebras and the Monster*, Pure and Applied Mathematics, vol. 134, Academic Press, Inc., Boston, MA, 1988.

[FM14] Cameron Franc and Geoffrey Mason, *Fourier coefficients of vector-valued modular forms of dimension* 2, Canad. Math. Bull. **57** (2014), 485–494.

[FM16a] ———, *Hypergeometric series, modular linear differential equations, and vector-valued modular forms*, Ramanujan J. **41** (2016), no. 1–3, 233–267.

[FM16b] ———, *Three-dimensional imprimitive representations of the modular group and their associated modular forms*, J. Number Theory **160** (2016), 186–214.

[FS19] Jürgen Fuchs and Christoph Schweigert, *Full logarithmic conformal field theory — an attempt at a status report*, Fortschr. Phys. (2019), no. 8–9, Special issue: Proceedings of the LMS/PESR Durham Symposium on Higher Structures in M-theory, 1910018, 12 pp.

[Gan06] Terry Gannon, *Moonshine beyond the Monster: The bridge connecting algebra, modular forms and physics*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2006.

[Gan14] ———, *The theory of vector-valued modular forms for the modular group*, Conformal Field Theory, Automorphic Forms and Related Topics, Springer-Verlag, 2014, pp. 247–286.

[GG76] A.A. Goldberg and V.A. Grinshtein, *The logarithmic derivative of a meromorphic function*, Math. Notes **19** (1976), 320–323.

[Gol69] G. M. Goluzin, *Geometric Theory of Functions of a Complex Variable*, Translations of Mathematical Monographs, Vol. 26, American Mathematical Society, Providence, R.I., 1969.

[Got20] Richard Gottesman, *The arithmetic of vector-valued modular forms on* $\Gamma_0(2)$, Int. J. Number Theory **16** (2020), no. 2, 241–289.

[Hay64] Walter K. Hayman, *Meromorphic Functions*, Oxford Mathematical Monographs, Clarendon Press, Oxford, 1964.

[Hem88] Joachim A. Hempel, *On the uniformization of the n-punctured sphere*, Bull. London Math. Soc. **20** (1988), no. 2, 97–115.

[Hin92] Aimo Hinkkanen, *A sharp form of Nevanlinna's second main theorem*, Invent. Math. **108** (1992), 549–574.

[Hup67] B. Huppert, *Endliche Gruppen. I*, Die Grundlehren der mathematischen Wissenschaften, Band 134, Springer-Verlag, Berlin-New York, 1967.

[Iha94] Yasutaka Ihara, *Horizontal divisors on arithmetic surfaces associated with Belyǐ uniformizations*, The Grothendieck theory of dessins d'enfants (Luminy 1993, ed. L. Schneps), Cambridge University Press, Cambridge, 1994, London Math. Soc. Lecture Note Series, vol. **200**, pp. 245–254.

[Kat70] Nicholas M. Katz, *Nilpotent connections and the monodromy theorem: Applications of a result of Turrittin*, Inst. Hautes Études Sci. Publ. Math. **39** (1970), 175–232.

[Kat73] ———, *p-adic properties of modular schemes and modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), 1973, pp. 69–190. Lecture Notes in Mathematics, Vol. 350.

[KF17] Felix Klein and Robert Fricke, *Lectures on the theory of elliptic modular functions. Vol. 1*, CTM. Classical Topics in Mathematics, vol. 1, Higher Education Press, Beijing, 2017, Translated from the German original by Arthur M. DuPre.

[KL08] Chris A. Kurth and Ling Long, *Modular forms for some noncongruence subgroups of* $\mathrm{SL}(2,\mathbf{Z})$, J. Number Theory **128** (2008), 1989–2009.

[KL09] ———, *On modular forms for some noncongruence subgroups of* $\mathrm{SL}(2,\mathbf{Z})$. *II.*, Bull. London Math. Soc. **41** (2009), 589–598.

[KM03a] Marvin Knopp and Geoffrey Mason, *Generalized modular forms*, J. Number Theory **99** (2003), 1–28.

[KM03b] ———, *On vector-valued modular forms and their Fourier coefficients*, Acta Arithmetica **110** (2003), 117–124.

[KM08] Winfried Kohnen and Geoffrey Mason, *On generalized modular forms and their applications*, Nagoya Math. J. **192** (2008), 119–136.

[KM09] Marvin Knopp and Geoffrey Mason, *Parabolic generalized modular forms and their characters*, Int. J. Number Theory **5** (2009), no. 5, 845–857.

[KM11] ———, *Logarithmic vector-valued modular forms*, Acta Arithmetica **147** (2011), 261–282.

[KM12] Winfried Kohnen and Geoffrey Mason, *On the canonical decomposition of generalized modular functions*, Proc. Amer. Math. Soc. **140** (2012), no. 4, 1125–1132.

[KR16] Daniela Kraus and Oliver Roth, *Sharp lower bounds for the hyperbolic metric of the complement of a closed subset of the unit circle and theorems of Schwarz-Pick-, Schottky- and Landau-type for analytic functions*, Constr. Approx. **43** (2016), no. 1, 47–69.

[KRS11] Daniela Kraus, Oliver Roth, and Toshiyuki Sugawa, *Metrics with conical singularities on the sphere and sharp extensions of the theorems of Landau and Schottky*, Math. Z. **267** (2011), no. 3-4, 851–868.

[Laz65] Michel Lazard, *Groupes analytiques p-adiques*, Inst. Hautes Études Sci. Publ. Math. (1965), no. 26, 389–603.

[LL12] W.-C.W. Li and L. Long, *Fourier coefficients of noncongruence cuspforms*, Bull. London Math. Soc. **44** (2012), no. 3, 591–598.

[Lon08] Ling Long, *Finite index subgroups of the modular group and their modular forms*, Fields Inst. Commun. **54** (2008), 83–102.

[Mar15] Christopher Marks, *Fourier coefficients of three-dimensional vector-valued modular forms*, Commun. Number Theory Phys. **9** (2015), 387–412.

[Mas12] Geoffrey Mason, *On the Fourier coefficients of 2-dimensional vector-valued modular forms*, Proc. Amer. Math. Socl. **140** (2012), no. 6, 1921–1930.

[Men67] Jens L. Mennicke, *On Ihara's modular group*, Invent. Math. **4** (1967), 202–228.

[MT06] Andrew R. McIntyre and Leon A. Takhtajan, *Holomorphic factorization of determinants of Laplacians on Riemann surfaces and a higher genus generalization of Kronecker's first limit formula*, Geom. Funct. Anal. **16** (2006), 1291–1323.

[Nev70] Rolf Nevanlinna, *Analytic Functions*, Die Grundlehren der mathematischen Wissenschaften, Springer-Verlag, New York-Berlin, 1970.

[NS10] Siu-Hung Ng and Peter Schauenburg, *Congruence subgroups and generalized Frobenius-Schur indicators*, Comm. Math. Phys. **300** (2010), no. 1, 1–46.

[Pel21] Federico Pellarin, *From the Carlitz exponential to Drinfeld modular forms*, Arithmetic and geometry over local fields, Lecture Notes in Math., vol. 2275, de Gruyter, Berlin, 2021, pp. 93–177.

[Rib84] Kenneth A. Ribet, *Congruence relations between modular forms*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983), PWN, Warsaw, 1984, pp. 503–514.

[Ru21] Min Ru, *Nevanlinna theory and its relation to Diophantine approximation*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2021, Second edition.

[Sel65] Atle Selberg, *On the estimation of Fourier coefficients of modular forms*, Proc. Sympos. Pure Math., American Mathematical Society, Providence, R.I., 1965, pp. 1–15.

[Ser70] Jean-Pierre Serre, *Le problème des groupes de congruence pour* SL_2, Ann. of Math. **2** (1970), 489–527.

[Ser80] ———, *Trees*, Springer-Verlag, Berlin-New York, 1980, Translated from the French by John Stillwell.

[Shi71] Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*, Kanô Memorial Lectures, No. 1, Publications of the Mathematical Society of Japan, No. 11. Iwanami Shoten, Publishers, Tokyo; Princeton University Press, Princeton, N.J., 1971.

[Sti84] Peter Stiller, *Special values of Dirichlet series, monodromy, and the periods of automorphic forms*, Mem. Amer. Math. Soc. **49** (1984), no. 299, iv+116.

[Swa60] Richard G. Swan, *The p-period of a finite group*, Illinois J. Math. **4** (1960), 341–346.

[SZ12] Yorck Sommerhäuser and Yongchang Zhu, *Hopf algebras and congruence subgroups*, Mem. Amer. Math. Soc. **219** (2012), no. 1028, vi+134.

[Tao12] Terence Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012.

[Tho89] J. G. Thompson, *Hecke operators and noncongruence subgroups*, Group theory (Singapore, 1987), de Gruyter, Berlin, 1989, Including a letter from J.-P. Serre, pp. 215–224.

[Tsu52] Masatsugu Tsuji, *Theory of Fuchsian groups*, Jpn. J. Math. **21** (1952), 1–27.

[Woh64] Klaus Wohlfahrt, *An extension of F. Klein's level concept*, Illinois J. Math. **8** (1964), 529–535.

[Xu06] Feng Xu, *Some computations in the cyclic permutations of completely rational sets*, Comm. Math. Phys. **267** (2006), no. 3, 757–782.

[Zag08] Don Zagier, *Elliptic modular forms and their applications*, The 1-2-3 of modular forms, Universitext, Springer, Berlin, 2008, pp. 1–103.

[Zhu96] Yongchang Zhu, *Modular invariance of characters of vertex operator algebras*, J. Amer. Math. Soc. **9** (1996), no. 1, 237–302.

Email address: `fcale@math.uchicago.edu`

The University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA

Email address: `vesselin.dimitrov@gmail.com`

Department of Mathematics, University of Toronto Room 6290, 40 St. George Street, Toronto, Ontario, Canada M5S 2E4

Email address: `yungqingt@math.princeton.edu`

Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544, USA