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*

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):

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

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)

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

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 [FF20, 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 $q = e^{2\pi i \tau}$):

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

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, 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 $f = (1 - m^2 x)^{1/m} \in \mathbf{Z}[\![x]\!]$ 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 $m(r, \varphi) := \int_{|z|=r} \log^+ |\varphi| \, \mu_{\text{Haar}}$ 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 (2.4.1) and (2.3.10), respectively. Our Theorem 2.0.2 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 $f \in \mathbf{Z}[1/N][\![x]\!]$ — but there is still a hidden integrality property which we can exploit. That leads to some mild technical nuance with the power series (2.0.3) — think of $t = q^{1/N}$, $x(t) = \sqrt[N]{\lambda(t^N)/16}$ and $p(x) = x^N$ — in our refinement (2.0.4) 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

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

which transforms this hypergeometric series into a weight one modular form for the congruence group $\Gamma(2)$.

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 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 5.3.8, 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.[1] 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 Theorem 2.0.2 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 entail a whole sequence $f(\tau/p) \in \mathbf{Z}[\![q^{1/p}]\!]$ of $\mathbf{Q}(x)$-linearly independent counterexamples at growing 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 (2.0.4) supplied by our arithmetic holonomy Theorem 2.0.2.

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 (2.0.4) of holonomy rank bound readily reveals, it seems a fortuitous piece of luck that the conformal size (Riemann

---

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

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 (2.0.4) inflates at a slower rate than our extrapolating putative counterexamples $f(\tau) \mapsto f(\tau/p)$!

The meaning of the requisite inflation rate is clarified in § 4, with Proposition 4.2.5 and Remark 4.2.8. 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 (2.0.4), 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 5.5.8. 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 (2.0.4). We achieve this in § 2.2 by an asymptotic equidistribution idea (all under $d \to \infty$), of a familiar potential-theoretic flavor similar to the proof of Bilu's equidistribution theorem [Bil97]. The resulting potential-theoretic connection is in the cross-variables asymptotic aspect and 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

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

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 $m(r, F) = \int_{|z|=r} \log^+ |F| \, \mu_{\text{Haar}} \sim \frac{1}{N-1} \log \frac{1}{1-r}$ 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 $\{a_1, \ldots, a_N\} = \mu_N$ for the puncture points, we are able to exploit the fortuitous relation $\prod_{i=1}^N (x - a_i) \sum_{i=1}^N \frac{1}{x - a_i} = Nx^{N-1}$ particular to the partial fractions decomposition (6.1.2) 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. Firstly we introduce an algebra of holonomic power series with integral coefficients and restricted singularities.

Definition 2.0.1. *For $U \subset \mathbf{C}$ an open subset, $R \subset \mathbf{C}$ a subring, 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 $f(x) \in (R \otimes_{\mathbf{Z}} \mathbf{Q})[\![x]\!]$ whose t-expansion $f(x(t)) \in R[\![t]\!]$, and such that there exists a nonzero linear differential operator L over $\mathbf{C}(x)$, without singularities on U, with $L(f) = 0$.*
  For $x(t) = t$, we simply denote this ring by $\mathcal{H}(U, R)$.

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

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

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.

Our holonomy bound is the following. The novel point is an equidistribution argument, of a potential-theoretic flavor, for achieving the integrated term instead of a supremum term in the bound (2.0.4) below. This is critical for our proof of the unbounded denominators conjecture to have the growth term in (2.0.4) expressed as a Nevanlinna mean proximity function (or a characteristic function).

Theorem 2.0.2. *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 $\mathcal{H}(U, \mathbf{Z}) \otimes \mathbf{Q}(x)$ has a finite dimension as a $\mathbf{Q}(x)$-vector space.*
  More precisely, let $p(x) \in \mathbf{Q}(x) \smallsetminus \mathbf{Q}$ 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

$$(2.0.3) \qquad\qquad x(t) \in t + t^2 \mathbf{Q}[\![t]\!]$$

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

$$(2.0.4) \qquad \dim_{\mathbf{Q}(p(x))}\left(\mathcal{H}(U, x(t), \mathbf{Z}) \otimes \mathbf{Q}(p(x))\right) \leq e \cdot \frac{\int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}}}{\log |\varphi'(0)|},$$

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

Proof. Our approach rests on the remark that $\varphi^* \mathcal{H}(U, x(t), \mathbf{Z}) \subset \varphi^* \mathcal{H}(U, \mathbf{C})$ lies in the ring of formal power series fulfilling <u>linear</u> differential equations with analytic coefficients and no singularities on the closed disc $\overline{D(0,1)}$ and hence, by Cauchy's theorem, it is contained by the ring $\mathcal{O}(\overline{D(0,1)})$ of holomorphic functions on $\overline{D(0,1)}$.

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

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

We consider $f_1, \ldots, f_m \in \mathbf{Q}[\![x]\!]$ an m-tuple of $\mathbf{Q}(p(x))$-linearly independent functions in the space $\mathcal{H}(U, x(t), \mathbf{Z})$, and proceed to establish the holonomy bound

$$(2.1.1) \qquad\qquad m \leq e \cdot \frac{\int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}}}{\log |\varphi'(0)|}.$$

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*

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

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

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

(2) all $a_{\mathbf{i},\mathbf{k}} \in \mathbf{Z}$ are rational integers bounded in absolute value by $\exp\left(\kappa C \alpha + o(\alpha)\right)$ for some constant $C \in \mathbf{R}$ depending only on $\max_{u \notin U} 1/|u|$ 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 $x(t) \in t + t^2 \mathbf{Z}[\![t/M]\!]$ for some $M \in \mathbf{N}$ bounded in terms of the degree and height of the rational function $p(x)$. The result then follows from the classical Siegel lemma [BG06, Lemma 2.9.1], with the degree parameter choice

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

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 $f_1, \ldots, f_m \in \mathbf{Q}[\![x]\!]$ are linearly independent over $\mathbf{Q}(p(x))$. $\square$

2.2. Equidistribution.

The key idea now 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 $dD \int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\text{Haar}}$. It is this Monte Carlo (randomized numerical integration) principle that makes possible the integrated growth term — as opposed to the cruder and rather more straightforward supremum growth term, compare to [And89, VIII 1.6] — in our holonomy rank bound (2.0.4). Having $\int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\text{Haar}}$ instead of $\sup_{|z|=1} \log |p \circ \varphi|$ in (2.0.4) is a critical step in our proof of the unbounded denominators conjecture.

To carry out this program, we employ in § 2.4 André's method [And89, And04] of extrapolating with the function

(2.2.1) $$F(x(\mathbf{t})) := F(x(t_1), \ldots, x(t_d)) \in \mathbf{Z}[\![\mathbf{t}]\!] \smallsetminus \{0\},$$

whose integrality of coefficients follows from Lemma 2.1.2 and our defining assumptions that $p(x) \in \mathbf{Z}[\![t]\!]$ while all $f_i(x(t)) \in \mathbf{Z}[\![t]\!]$. This integrality is the key to the Liouville lower bound. For the Cauchy upper bound, in counterpoint, we would need 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 here the problem is that 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 (2.2.8) below), the peaks at the biased part of $\mathbf{T}^d$ get overwhelmingly large, and a direct extrapolation from (2.2.1) still only leads to a dimension bound with $\sup_{|z|=1} \log |p \circ \varphi|$.

To improve the supremum term to the mean term $\int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\text{Haar}}$, our new idea is to 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

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

By applying the Hadamard volume inequality to the Vandermonde determinant in (2.2.2), we recover the following classical result of Fekete, crucial for our approach.

Lemma 2.2.3 (Fekete). *The supremum of $|V(\mathbf{z})| = \prod_{1 \leq i < j \leq d} |z_i - z_j|$ 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 2.2.9 below). This plays off against the $d^{d/2} = \exp(o(d^2))$ bound of Lemma 2.2.3 to sift out the equidistributed points in our pointwise upper bound in the Cauchy integral formula when we extrapolate in § 2.4 below. Liouville's Diophantine lower bound still succeeds like in André [And04], thanks to the chain rule and the integrality property of coefficients of (2.2.2), 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 (2.2.10)) 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 2.2.3) uniformly throughout the whole polycircle $\mathbf{T}^d$.

In the remainder of the current section, 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 2.2.4. *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:*

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

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 2.2.5. *The* **total variation** $V(g)$ *of a function $g : \mathbf{T} \to \mathbf{R}$ is the supremum over all partitions $0 \leq \theta_1 < \cdots < \theta_n < 1$ 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

$$(2.2.6) \qquad V(g) = \int_{\mathbf{T}} |g'(z)| \, \mu_{\text{Haar}}(z), \quad g \in C^1(\mathbf{T}).$$

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]):

$$(2.2.7) \qquad \left| \frac{1}{d} \sum_{j=1}^{d} g(z_j) - \int_{\mathbf{T}} g \, \mu_{\text{Haar}} \right| \leq V(g) D(z_1, \ldots, z_d).$$

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

$$(2.2.8) \qquad D(z_1, \ldots, z_d) \leq 3 \left(\frac{1}{K+1} + \sum_{k=1}^{K} \frac{1}{k} \left| \frac{z_1^k + \cdots + z_d^k}{d} \right| \right) \quad \forall K \in \mathbf{N},$$

in terms of the power sums. Here we note in passing that, by (2.2.8) 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 $D(z_1, \ldots, z_d) \geq \varepsilon$ 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 $D(z_1 \ldots, z_d) \geq \varepsilon$ (the poorly distributed points). For the well-distributed group we use the Koksma inequality (2.2.7), 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 2.2.9. *There are functions $c(\varepsilon) > 0$ and $d_0(\varepsilon) \in \mathbf{R}$ such that, for every $\varepsilon \in (0,1]$, if $d \geq d_0(\varepsilon)$ and $(z_1, \ldots, z_d) \in \mathbf{T}^d$ is a d-tuple with discrepancy $D(z_1, \ldots, z_d) \geq \varepsilon$, then*

$$(2.2.10) \qquad |V(z_1, \ldots, z_d)| = \prod_{1 \leq i < j \leq d} |z_i - z_j| < e^{-c(\varepsilon)d^2}.$$

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

$$(2.2.11) \qquad \lim_{d \to \infty} \frac{1}{\binom{d}{2}} \sum_{1 \leq i < j \leq d} \log \frac{1}{|z_i^{(d)} - z_j^{(d)}|} \leq 0,$$

but

$$(2.2.12) \qquad \forall d, \qquad D(z_1^{(d)}, \ldots, z_d^{(d)}) \geq \varepsilon.$$

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, (2.2.12) implies that the limit discrepancy

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

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 $\mu_{\mathbf{T}} = \mu_{\text{Haar}}$, and since $I(\mu_{\text{Haar}}) = 0$, but $\mu \neq \mu_{\text{Haar}}$, we have the strict inequality

$$(2.2.13) \qquad I(\mu) = \iint_{\mathbf{T} \times \mathbf{T}} \log \frac{1}{|z - w|} \, \mu(z) \, \mu(w) > 0.$$

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 (2.2.13) contradicts (2.2.11) 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 (2.2.11) 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 (2.2.13) 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 (2.2.11) 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 (2.2.11) on letting $\eta \to 0$. $\qquad \square$

2.3. **Damping the Cauchy estimate.** We combine Lemmas 2.2.3 and 2.2.9 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

$$(2.3.1) \qquad \log|F(\varphi(z_1), \ldots, \varphi(z_d))| \leq D \sum_{j=1}^d \log^+ |p(\varphi(z_j))| + \kappa C \alpha + o(\alpha).$$

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 (2.2.7) yields, on the well-distributed part $\mathbf{z} \in \mathbf{T}^d$, the uniform pointwise upper bound

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

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

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 $D(z_1, \ldots, z_d) \geq \varepsilon$, the sum in (2.3.1) can get as large as $d \sup_{|z|=1} \log |p \circ \varphi|$. This trivial bound gives

$$(2.3.2) \qquad \forall \mathbf{z} \in \mathbf{T}^d, \quad \log|F(\varphi(z_1), \ldots, \varphi(z_d))| \leq dD \sup_{|z|=1} \log^+ |p \circ \varphi| + \kappa C \alpha + o(\alpha).$$

We now impose the condition

$$(2.3.3) \qquad d \geq d_0(\varepsilon), \qquad \text{for the function } d_0(\varepsilon) \text{ in Lemma 2.2.9,}$$

for the remainder of the proof of Theorem 2.0.2 (at the end we will firstly take $d \to \infty$, and only then $\varepsilon \to 0$), and we select the Vandermonde exponent

$$(2.3.4) \qquad M := \left\lfloor \frac{\sup_{|z|=1} \log^+ |p \circ \varphi|}{c(\varepsilon)} \frac{D}{d} \right\rfloor,$$

with $c(\varepsilon)$ the function from Lemma 2.2.9. 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 $D(z_1, \ldots, z_d) \geq \varepsilon$, Lemma 2.2.9 with (2.3.2), (2.3.3) and (2.3.4) gives

$$(2.3.5) \qquad \sup_{\mathbf{z} \in \mathbf{T}^d : D(z_1, \ldots, z_d) \geq \varepsilon} \log|V(\mathbf{z})^M F(\varphi(\mathbf{z}))| \ll \kappa \alpha.$$

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

$$(2.3.6) \qquad \sup_{\mathbf{z} \in \mathbf{T}^d : D(z_1, \ldots, z_d) \leq \varepsilon} \log|V(\mathbf{z})^M F(\varphi(\mathbf{z}))|$$

$$\leq dD \int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\text{Haar}} + \kappa C \alpha + O_{p,\varphi}(\varepsilon \alpha) + O_{\varepsilon,p,\varphi}\left(\frac{\log d}{d} \alpha\right) + o(\alpha).$$

by (2.3.4) and Lemma 2.2.3.

Consider the holomorphic function

$$(2.3.7) \qquad H(\mathbf{z}) := V(\mathbf{z})^M F(\varphi(z_1), \ldots, \varphi(z_d)) =: \sum_{\mathbf{n} \in \mathbf{N}_0^d} c(\mathbf{n}) \, \mathbf{z}^{\mathbf{n}} \in \mathbf{C}[\![\mathbf{z}]\!],$$

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

$$(2.3.8) \qquad c(\mathbf{n}) = \int_{\mathbf{T}^d} \frac{H(\mathbf{z})}{\mathbf{z}^{\mathbf{n}}} \mu_{\text{Haar}}(\mathbf{z}),$$

entailing the Cauchy upper bound

$$(2.3.9) \qquad \forall \mathbf{n} \in \mathbf{N}_0^d, \qquad |c(\mathbf{n})| \leq \sup_{\mathbf{z} \in \mathbf{T}^d} |H(\mathbf{z})|.$$

On combining the bounds (2.3.6), on the well-distributed part of $\mathbf{T}^d$, and (2.3.5), on the poorly distributed part of $\mathbf{T}^d$, we arrive at our damped Cauchy estimate:

$$(2.3.10) \qquad \log |c(\mathbf{n})| \leq dD \int_{\mathbf{T}} \log^+ |p \circ \varphi| \, \mu_{\mathrm{Haar}}$$

$$+ O(\kappa \alpha) + O_{p,\varphi}(\varepsilon \, \alpha) + O_{\varepsilon, p, \varphi}\Big(\frac{\log d}{d} \alpha \Big) + o(\alpha),$$

asymptotically under $\alpha \to \infty$.

2.4. **The extrapolation.** We can now easily finish the proof of Theorem 2.0.2 by combining the degree estimate (1) of Lemma 2.1.2 with the Cauchy bound (2.3.10) and the integrality properties of the functions $F(x(\mathbf{t})) \in \mathbf{Z}[\![\mathbf{t}]\!]$ of (2.2.1) and $V(\mathbf{z}) \in \mathbf{Z}[\mathbf{z}]$ of (2.2.2).

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 (2.0.3) 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 property (2.2.1) 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 (2.3.9) is supplemented with the Liouville lower bound

$$(2.4.1) \qquad \exists \, \mathbf{n} \in \mathbf{N}_0^d \quad : \quad \log |c(\mathbf{n})| \geq \beta \log |\varphi'(0)| \geq \alpha \log |\varphi'(0)|.$$

We get the requisite holonomy rank bound (2.1.1) on combining the degree bound (1) of Lemma 2.1.2 with the Cauchy upper bound (2.3.10) and the Liouville lower bound (2.4.1), and letting firstly $\alpha \to \infty$, then $d \to \infty$, then $\kappa \to 0$, and finally $\varepsilon \to 0$.

This completes the proof of Theorem 2.0.2. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 2.4.2. André pointed out to us an alternative, and arguably simpler path to the proof of Theorem 2.0.2. His variation is to replace our use of the maximum modulus principle for $V(\mathbf{z})^M F(\varphi(\mathbf{z}))$ by the plurisubharmonic property of $\log |z_1^{-n_1} F(\varphi(\mathbf{z}))|$, with n_1 being the lowest partial degree of x_1 in $F(\mathbf{x})$. Such a procedure turns out to work well if we consider, in place of the leading order jet, instead the *overall* (possibly of a higher total degree $|\mathbf{n}| > \beta$) lexicographically lowest monomial $c\,\mathbf{x}^{\mathbf{n}}$ in $F(\mathbf{x})$. In this way, the equidistribution part of the proof can be bypassed, and the rest of the argument goes through unchanged. We plan to give further details in a subsequent work and explore the added flexibility of the lexicographically lowest monomial. (We have kept the original argument with the thought that it may have ideas useful in other settings.)

Remark 2.4.3. The result is more general, 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 (2.0.4).

3. SOME REDUCTIONS

The following encapsulates our argument:

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 ring R_{2N} of modular functions with bounded denominators and cusp widths dividing $2N$ has dimension at most $CN^3 \log N$ over the ring 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 2.0.2 with $U := \mathbf{C} \smallsetminus 16^{-1/N} \mu_N$, $p(x) := x^N$ and

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

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 ring R_{2N} entail that for every $f \in R_{2N}$, if $m \gg 1$ is large enough so that $\lambda^m f$ is holomorphic at all cusps in $\Gamma(2) \cdot i\infty$, then the formal expansion $x^{Nm} \cdot f(x) \in \mathbf{Q}[\![x]\!]$ belongs to the ring $\mathcal{H}(U, x(t), \mathbf{Z}) \otimes_{\mathbf{Z}} \mathbf{Q}$. It thus suffices to bound the $\mathbf{Q}[x^N]$-dimension of the latter ring 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

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

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

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

Theorem 2.0.2 now upper-bounds our requisite dimension as

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

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

We prove both of the required statements in Theorems 5.3.8 and Theorem 6.0.1 respectively. We then put up the $CN^3 \log N$ dimension bound supplied by Proposition 3.0.1 for the ring R_N 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 4.2.2). Hence we have:

3.1. A summary. This lays out exactly what we need to prove in order to at least get the unbounded denominators conjecture up to a "small error" (a logarithmic gap $O(\log N)$ in every level N). The following is then a summary of the rest 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 congruence ring, by considering $f(q^{1/p}) \in \mathbf{Z}[\![q^{1/Np}]\!]$ 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 5.3.8), 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 $m(r, F_N) = \int_{|z|=r} \log^+ |F_N| \, \mu_{\text{Haar}}$.

(5) Putting all the pieces together, the proof of Theorem 1.0.1 is then completed in §6.3.

4. Noncongruence forms

4.1. **Wohlfahrt Level.** We being by recalling a notion of level for noncongruence subgroups due to Wohlfahrt [Woh64]. Let $G \subset \mathrm{SL}_2(\mathbf{Z})$. If $\gamma \in G$ is any parabolic element, then γ is conjugate in $\mathrm{SL}_2(\mathbf{Z})$ to a power U^m of the matrix

$$(4.1.1) \qquad\qquad U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Definition 4.1.2 ([Woh64])**.** *The level $L(G)$ of G is the lowest common multiple of m as γ ranges over all parabolic γ which generate the stabilizer of some cusp of G — equivalently, over all parabolic γ which are not non-trivial powers of some other parabolic element.*

We begin with some elementary properties concerning this definition. We typically only consider groups containing $-I$ since we are secretly interested in subgroups of $\mathrm{PSL}_2(\mathbf{Z})$.

Lemma 4.1.3. *If G and H both contain $-I$ and have $L(G)$ and $L(H)$ dividing 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 $-I$ is $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}$. In particular, if G contains the group $a\mathbf{Z}$ and H contains $b\mathbf{Z}$ then $G \cap H$ contains $\mathrm{lcm}(a,b)\mathbf{Z}$, and the result follows. $\qquad\square$

Corollary 4.1.4. *Let $G \subset \mathrm{SL}_2(\mathbf{Z})$ be a finite index subgroup containing $-I$ with Wohlfahrt level N. Let $\widetilde{G}$ be the intersection of the conjugates of G by $\mathrm{SL}_2(\mathbf{Z})$, so $\widetilde{G}$ is a normal subgroup of $\mathrm{SL}_2(\mathbf{Z})$ contained in G. Then $\widetilde{G}$ has level N.*

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

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

(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 $H \subset \mathrm{SL}_2(\mathbf{Z})$ have $L(H) = N$. Then $\widetilde{H} := A^{-1}HA \cap \mathrm{SL}_2(\mathbf{Z})$ has $L(\widetilde{H})$ dividing Np.*

Proof. Any unipotent element in $\widetilde{H}$ is conjugate by A to a unipotent element in H, which is a power of a minimal unipotent element $\gamma \in H$. 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 have

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

where $(A^{-1}BA) = \begin{pmatrix} a & b/p \\ cp & d \end{pmatrix}$. We consider two cases:

(1) Suppose that $(a, p) = 1$. Since $(a, c) = 1$ we have $(a, pc) = 1$. Hence we may write

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

where

$$C = \begin{pmatrix} a & e \\ cp & f \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \text{ such that } A^{-1}BA = C \begin{pmatrix} 1 & h/p \\ 0 & 1 \end{pmatrix}, h \in \mathbf{Z},$$

for any suitable choice of e, f and h. 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 $\widetilde{H}$, and hence the cusp width at this cusp 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 so the cusp width divides Np. $\qquad\square$

4.2. **Modular Forms.** For an even integer N, we are considering the following rings of modular functions with rational coefficients and bounded denominators, that is, subrings of $\mathbf{Q} \otimes \mathbf{Z}[\![q^{1/N}]\!]$.

Definition 4.2.1.

(1) *Let M_N denote the ring of holomorphic modular functions on $Y(N)$ with coefficients in $\mathbf{Q}$.*
(2) *Let R_N denote the ring of holomorphic modular functions with coefficients in $\mathbf{Q}$, bounded denominators, and cusp widths dividing N.*

Our goal is to prove that $R_N = M_N$. We can and do assume that N is even. We have $M_2 = \mathbf{Q}[\lambda^{\pm 1}, (1 - \lambda)^{\pm 1}]$ with function field $\mathbf{Q}(\lambda)$. The $\mathbf{Q}(\lambda)$-vector space of functions generated by elements of R_N is finite dimensional. (We will later verify the assumptions of Proposition 3.0.1; note in any case that the crude finiteness $[R_N : M_2] < \infty$ has already been proved from Theorem 2.0.2 and the remark that the conformal radius of $\mathbf{C} \smallsetminus 16^{-1/N}\mu_N$ is strictly larger than 1.) It follows that all elements of R_N are invariant under a subgroup G_N of $\mathrm{SL}_2(\mathbf{Z})$ containing $-I$ which has finite index. In particular, R_N is a subring of the holomorphic functions on the corresponding curve and so finitely generated over M_2. If A is any finitely generated M_2-module, we write $[A : M_2] = [A \otimes \mathbf{Q}(\lambda) : \mathbf{Q}(\lambda)]$. Since M_2 has trivial Picard group, this is actually equivalent to A being free of this rank (although we do not use this fact).

There are injective algebra maps $M_2 \to M_N \to R_N$, where the degree of the first inclusion is given for $2|N$ by the explicit formula

$$(4.2.2) \qquad [M_N : M_2] = \frac{1}{2}[\Gamma(2) : \Gamma(N)] = \frac{N^3}{2[\mathrm{SL}_2(\mathbf{Z}) : \Gamma(2)]} \prod_{p|N} \left(1 - \frac{1}{p^2}\right) > \frac{N^3}{12\zeta(2)}.$$

Since modular forms on $Y(N)$ have bounded denominators, we furthermore have $[R_N : M_2] = [R_N : M_N] \cdot [M_N : M_2]$. It follows that we have a bound:

$$(4.2.3) \qquad\qquad\qquad [R_N : M_N] \leq \frac{12\zeta(2)[R_N : M_2]}{N^3}$$

for all N. The key leveraging step is contained in the following theorem:

Theorem 4.2.4. *Suppose that there exists an N such that $[R_N : M_N] > 1$. Then, for every prime p not dividing N, one has*

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

The proof of this theorem is in the next section 4.3 below. However, we now explain why Theorem 4.2.4 together with a sufficiently good bound on $[R_N : M_2]$ implies the unbounded denominators conjecture.

Proposition 4.2.5. *Suppose that there exists a constant C and a bound*

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

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 $R_N \neq M_N$. Let S denote the set of primes $< X$ which are co-prime to N. By induction, Theorem 4.2.4 implies for such an N that

$$(4.2.6) \qquad [R_{N \prod_{p \in S} p} : M_{N \prod_{p \in S} p}] \geq [R_N : M_N] 2^{\#S} > 2^{(1-\varepsilon)X/\log X},$$

by the prime number theorem. This quantity 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 (4.2.3), we obtain

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

Combining the bounds (4.2.6) and (4.2.7) gives

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

which (by some margin!) is a contradiction for sufficiently large X. $\qquad\qquad\square$

Remark 4.2.8. 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.

4.3. **The proof of Theorem 4.2.4.** We now put ourselves in the situation of Theorem 4.2.4. That is, we have R_N is strictly larger than M_N. Recall that forms in R_N are all invariant by a subgroup $G = G_N \subset \Gamma(N)$. By definition, all forms in R_N have cusp width dividing N, and thus G_N has level N in the sense of Wohlfahrt. We may assume that $-I \in G$. By Corollary 4.1.4, we may replace G by a normal subgroup G of finite index of $\mathrm{SL}_2(\mathbf{Z})$ containing $-I$ which also has finite index. Let p be a prime with $(N, p) = 1$.

The main idea of this section is to exploit the fact that if $f(\tau) \in \mathbf{Z}[\![q^{1/N}]\!] \in R_N$, then $f(\tau/p) \in \mathbf{Z}[\![q^{1/Np}]\!]$ is also a modular form with integer coefficients for any prime p. Since the form $f(\tau)$ is invariant under G, the form $f(\tau/p)$ is invariant under AGA^{-1} and thus also the group $AGA^{-1} \cap \mathrm{SL}_2(\mathbf{Z})$. Now, by Lemma 4.1.7, we know that this group has (Wohlfahrt) level dividing Np. In particular $f(\tau/p)$ has cusp width dividing Np at each cusp, and hence $f(\tau/p) \in R_{Np}$. Let $R_N M_{Np}$ denote the ring generated by forms in R_N and M_{Np}, which is contained in 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 forms in M_{Np} and R_N lie in disjoint Galois extensions of (the field of fractions of) M_N. Thus

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

is an integer, and it is either ≥ 2, in which case we have proven Theorem 4.2.4, or $R_{Np} = R_N M_{Np}$. In particular, in order to prove Theorem 4.2.4 we may assume that R_{Np} is generated by R_N and M_{Np}, and in particular that $f(\tau/p)$ is invariant under the group $G \cap \Gamma(Np)$. This implies

that $f(\tau/p)$ it is invariant under both $AGA^{-1} \cap \mathrm{SL}_2(\mathbf{Z})$ and $G \cap \Gamma(Np)$. We will exploit an idea due to Serre that shows G and AGA^{-1} are in some sense "far apart" when G is not a congruence subgroup. We first prove a preliminary lemma.

Lemma 4.3.1. *Let* $G \subset \mathrm{SL}_2(\mathbf{Z})$ *be a not-necessarily congruence subgroup of* $\mathrm{SL}_2(\mathbf{Z})$ *with level* N *in the sense of Definition 4.1.1. Let* H *be the group generated by* $AGA^{-1} \cap \mathrm{SL}_2(\mathbf{Z})$ *together with* $G \cap \Gamma(Np)$. *Then* H *contains* $G \cap \Gamma(N) \cap \Gamma^0(p)$.

Proof. There are maps

$$(4.3.2) \qquad \psi_G, \psi_H : G, H \to \mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/Np).$$

The kernel of ψ_G is equal to $G \cap \Gamma(Np)$, which is contained in H by definition. Hence to show that $G \cap \Gamma(N) \cap \Gamma^0(p)$ is contained in H it suffices to show that the image of ψ_H contains the image of $G \cap \Gamma(N) \cap \Gamma^0(p)$. Since G has finite index inside $\mathrm{SL}_2(\mathbf{Z})$ and level N, it contains the matrices

$$\begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ N & 1 \end{pmatrix},$$

and these generate the projection of $\mathrm{SL}_2(\mathbf{Z}/Np)$ onto $\mathrm{SL}_2(\mathbf{F}_p)$. Hence the image of $G \cap \Gamma(N)$ under ψ_G is precisely the matrices $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mod Np$ congruent to $I \mod N$. Now consider a matrix

$$(4.3.3) \qquad \begin{pmatrix} a & bp \\ c & d \end{pmatrix} \mod Np$$

inside $\mathrm{SL}_2(\mathbf{Z}/Np)$ which is congruent to $I \mod N$. These are precisely the matrices is in the image of $G \cap \Gamma(N) \cap \Gamma^0(p)$, and our task is to show such matrices also lies the image of H. We first observe that the matrix

$$(4.3.4) \qquad \begin{pmatrix} a & b \\ cp & d \end{pmatrix} \mod Np$$

also lies in $\mathrm{SL}_2(\mathbf{Z}/Np)$ (the determinant is unchanged) and is congruent to the identity modulo N since $(N,p) = 1$. Thus it lies in the image of G. Let γ denote any lift to G; clearly the lower left entry is divisible by p. Hence

$$A\gamma A^{-1} = \begin{pmatrix} a & bp \\ c & d \end{pmatrix} \in AGA^{-1}$$

lies in $\mathrm{SL}_2(\mathbf{Z})$, and thus in H. This shows that the image of H contains the image of $G \cap \Gamma(N) \cap \Gamma^0(p)$ and we are done. $\qquad \square$

Since G is normal, we may define the group S by taking $S = \widetilde{\Gamma}(N)/G$. By construction, the group S is finite. There are two homomorphisms f_1 and f_2 from $\widetilde{\Gamma}(N) \cap \Gamma_0(p)$ to S defined by:

(1) The inclusion map: $f_1 : \widetilde{\Gamma}(N) \cap \Gamma_0(p) \to \widetilde{\Gamma}(N) \to S$,
(2) The map $f_2 = f_1(AxA^{-1})$.

Lemma 4.3.5 (Serre). *The map* $(f_1, f_2) : \widetilde{\Gamma}(N) \cap \Gamma_0(p) \to S \times S$ *is surjective.*

Proof. This follows as in the proof of [Tho89, Theorem 3] with the addition of $\widetilde{\Gamma}(N)$ level structure (See [Ber94] for the modification in the presence of level structure). The main idea is as follows: if (f_1, f_2) is not surjective, there exists (by Goursat's lemma) 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. This implies that there exists a non-trivial map on the amalgamated product of $\widetilde{\Gamma}(N)$ and $A\widetilde{\Gamma}(N)A^{-1}$ along their intersection to Δ of S, and this amalgamated product may be identified with the level N congruence subgroup of $\mathrm{SL}_2(\mathbf{Z}[1/p])$. But this latter group has the congruence subgroup property [Men67, Ser70]. This implies that the kernel $K \supseteq G$ of the map map $\widetilde{\Gamma}(N) \to \widetilde{\Gamma}(N)/G \to \Delta$ is a proper congruence subgroup, but this contradicts the assumption that the level of G is N. $\quad \square$

Note that

(1) $\ker(f_1) = G \cap \Gamma_0(p)$.
(2) $\ker(f_2) = A^{-1}GA \cap \Gamma_0(p)$.

Let K be the group generated by $\ker(f_1)$ and $\ker(f_2)$. Since (f_1, f_2) is surjective, the image contains the elements $(x, 0)$ and $(0, y)$ for any $x \in S$. But the pre-images of these elements clearly lie in $\ker(f_2)$ and $\ker(f_1)$ respectively, and thus lie in K. But then the pre-image of any element lies in K, and we deduce that

$$(4.3.6) \qquad G \cap \Gamma_0(p) + A^{-1}GA \cap \Gamma_0(p) = \widetilde{\Gamma}(N) \cap \Gamma_0(p).$$

We now consider the function $f(\tau/p)$, which is obtained from $f(\tau)$ by acting by A^{-1}. As explained at the beginning of § 4.3, we deduce that $f(\tau/p)$ is invariant under both $AGA^{-1} \cap \mathrm{SL}_2(\mathbf{Z})$ and $G \cap \widetilde{\Gamma}(Np)$. But now conjugating by A^{-1} (to bring $f(\tau/p)$ back to $f(\tau)$) we deduce from Lemma 4.3.1 that $f(\tau)$ is invariant under the group

$$A^{-1}(G \cap \Gamma^0(p))A = A^{-1}GA \cap \Gamma_0(p).$$

But it is also invariant under G, and hence tautologically under $G \cap \Gamma_0(p)$. By equation (4.3.6) it is thus also invariant under $\widetilde{\Gamma}(N) \cap \Gamma_0(p)$, which is to say invariant under a congruence subgroup, which was to be shown.

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 $F_N : D(0, 1) \to \mathbf{C} \smallsetminus \mu_N$. André has pointed out to us that our two main results here, Theorem 5.3.8 and Lemma 5.5.7, 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 $\mathbf{C} \smallsetminus \mu_N$.

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 formula

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

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 $U(1, 1)$ whereas $\widetilde{\Gamma}_N$ denotes the corresponding lattice in $\mathrm{PSL}_2(\mathbf{R})$.

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:

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

The map $\widetilde{F}_N$ is unique up to the action of $\mathrm{PSL}_2(\mathbf{R})$ on the source — we pin down some precise choices in §5.1, but the analysis of this subsection will not depend on any such choices.

We have a companion uniformization map $1/\widetilde{F}_N : \mathbf{H} \to \mathbf{P}^1 \smallsetminus \{0, \mu_N\}$. We now derive the Schwarzian derivative of the inverse map (of $1/\widetilde{F}_N$ and then of $\widetilde{F}_N$) following [Hem88]. (The reason for first considering the reciprocal of $\widetilde{F}_N$ is that the standard form considered in [Hem88] is for maps to $\mathbf{P}^1 \smallsetminus S$ where S is a finite set of points which does not contain ∞.) Let m_k for $k = 1 \ldots N$ denote the accessory parameters at $z = p_k = \zeta^k = e^{2\pi i k/N}$, and let m_0 denote the accessory parameter at $z = p_0 = 0$. Exactly as in [Hem88, Example 1], the accessory parameters m_k

for $k \neq 0$ satisfy the symmetry $m_k = c \cdot \zeta^{-k}$ for some constant c. The accessory parameters are subject to three constraints, given by ([Hem88, Theorem 3.1])

$$(5.0.2) \qquad \sum_{k=0}^{N} m_k = 0, \quad \sum_{k=0}^{N} 2m_k p_k + 1 = 0, \quad \sum_{k=0}^{N} m_k p_k^2 + p_k = 0.$$

From the first constraint, we deduce from the fact that $\sum_{k=1}^{N} m_k = 0$ that $m_0 = 0$. But now from the second constraint in equation (5.0.2), we deduce that

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

and hence $c = -\frac{1}{2} - \frac{1}{2N}$. Writing $Z = 1/F_N(\tau)$, we deduce from [Hem88, Theorem 3.1] that the Schwarzian $\{\tau, 1/F_N\}$ is given by

$$\{\tau, 1/F_N\} = \{\tau(Z), Z\} := \left(\left(\frac{\tau''}{\tau'} \right)' - \frac{1}{2} \left(\frac{\tau''}{\tau'} \right)^2 \right)(Z) = \frac{1}{2} \sum_{k=0}^{N} \frac{1}{(Z - p_k)^2} + \sum_{k=0}^{N} \frac{m_k}{Z - p_k}$$

$$= \frac{1}{2Z^2} + \frac{1}{2} \sum_{k=1}^{N} \frac{1}{(Z - \zeta^k)^2} - \frac{(1 + N)}{2N} \sum_{k=1}^{N} \frac{\zeta^{-k}}{Z - \zeta^k} = \frac{(1 + (N^2 - 1)Z^N)}{2Z^2(Z^N - 1)^2}.$$

From the chain rule, we deduce that with $z = F_N = 1/Z$ the equality:

$$\{\tau, F_N\} = \frac{1}{z^4} \frac{(1 + (N^2 - 1)(1/z)^N)}{2(1/z)^2((1/z)^N - 1)^2} = \frac{(N^2 - 1)z^{N-2} + z^{2N-2}}{2(z^N - 1)^2}.$$

By [Hem88, Lemma 3.3], if η_1 and η_2 are solutions to the equation $y'' + \frac{1}{2}\{\tau, F_N\}y = 0$, or equivalently

$$(5.0.3) \qquad 4(z^N - 1)^2 y'' + ((N^2 - 1)z^{N-2} + z^{2N-2})y = 0$$

then the (locally analytic) inverse map ψ_N of $\widetilde{F}_N$ is the ratio η_1/η_2 up to a Möbius transformation (equivalently, up to the correct choice of linearly independent solutions). We find that equation (5.0.3) admits solutions

$$(5.0.4) \qquad \eta_1 = 1 - \frac{(N + 1)z^N}{4N} + \cdots \in \mathbf{Q}[\![z^N]\!],$$

$$(5.0.5) \qquad \eta_2 = z - \frac{(N - 1)z^{N+1}}{4N} + \cdots \in z \cdot \mathbf{Q}[\![z^N]\!],$$

the inclusions following in an elementary way from the Frobenius method applied to (5.0.3).

5.1. **Normalizations of $\widetilde{F}_N$.** Using the action of $\mathrm{PSL}_2(\mathbf{R})$ we may assume that $\widetilde{F}_N(i\infty) = 1$. The stabilizer of ∞ consists of Möbius transformations $z \mapsto az + b$, so by specifying $\widetilde{F}_N(0) = 0$ we determine $\widetilde{F}_N$ uniquely. But note that if $\zeta_N = \exp(2\pi i/N)$, then $\zeta_N \widetilde{F}_N$ is another covering map which must thereby differ from $\widetilde{F}_N$ by a Möbius transformation; therefore $\zeta_N \widetilde{F}_N(\tau) = \widetilde{F}_N(r_N \cdot \tau)$ upon checking the derivative at $i \in \mathbf{H}$. We deduce that $\widetilde{F}_N(r_N^N \cdot \tau) = \zeta^N \widetilde{F}_N(\tau) = \widetilde{F}_N(\tau)$, and thus $r_N^N \in \mathrm{SO}_2(\mathbf{R})$ must also lie in $\widetilde{\Gamma}_N$. But $\widetilde{\Gamma}_N$ is a free group, and hence r_N^N is trivial, and r_N is a hyperbolic rotation around i of order N.

Definition 5.1.1. *Let us define $F_N : D(0, 1) \to \mathbf{C} \setminus \mu_N$ by the formula*

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

Note that this 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

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

is a universal covering map. The action of $\mathrm{SO}_2(\mathbf{R})$ on $D(0, 1)$ under the pullback map is just given by rotation, and hence r_N acts on $D(0, 1)$ by a rotation of order N. We deduce that $F_N(\zeta^m q) =$

$\zeta F(q)$ for some $(m, N) = 1$. Since F_N is a covering map and $F_N(0) = 0$, we must also have $F_N'(0) \neq 0$, and thus $F_N(\zeta q) = \zeta F_N(q)$ for any Nth root of unity ζ.

Definition 5.1.2. *Let ψ_N be the the inverse of F_N.*

The map ψ_N is well-defined up to the action of $\Gamma_N \subset \mathrm{PSU}(1,1)$ acting on $D(0,1)$. In a neighbourhood of $q = 0$ so $\psi_N(0) = 0$. If $\widetilde{\psi}_N$ is the local inverse of F_N around 0 so that $\widetilde{\psi}_N(0) = i$, then clearly $\psi_N(z)$ is a Möbius translate of ψ_N and hence also of η_2/η_1. But now from the equality $F_N(\zeta q) = \zeta F_N(q)$ we deduce (from the form of η_1 and η_2 in equations (5.0.4) and (5.0.5)) that

$$(5.1.3) \qquad \psi_N(z) = \gamma_N^{-1} \cdot \frac{\eta_2}{\eta_1} = \gamma_N^{-1}\left(z + \frac{z^{N+1}}{2N} + \cdots\right)$$

for some constant γ_N^{-1}. Moreover, we certainly also have

$$F_N(q) = \gamma_N q + O(q^2),$$

and hence the conformal radius of F_N is given by $|\gamma_N|$. We shall compute γ_N explicitly in Theorem 5.3.8 below.

Definition 5.1.4. *Let G_N denote the map $D(0,1) \to \mathbf{C} \smallsetminus \{1\}$ such that $G_N(q^N) = (F_N(q))^N$, or equivalently $G_N(q) = (F_N(q^{1/N}))^N$.*

The fact that G_N is well-defined is a formal consequence of the relation $F_N(\zeta q) = \zeta F_N(q)$. From the fact that F_N is the universal covering map, we also immediately deduce the following geometric description of G_N.

Lemma 5.1.5. *The map $G_N : D(0,1) \to \mathbf{C} \smallsetminus \{1\}$ is a covering map away from the point $0 \in \mathbf{C} \smallsetminus \{1\}$. The map G_N is locally an isomorphism in a neighbourhood of 0 in $D(0,1)$, but it is totally ramified of degree N at all other preimages of 0. Moreover, it is universal with respect to any such map.*

The inverse map of G_N is closely related to the inverse map of F_N, and turns out to have a nicer form:

Lemma 5.1.6. *Let φ_N denote the inverse map of G_N, normalized so that $\varphi_N(0) = 0$. The function φ_N has the form $\delta_N^{-1}(\phi_2/\phi_1)^N$, where ϕ_1 and ϕ_2 are the solutions to the differential equation:*

$$(5.1.7) \qquad z(z-1)^2 y'' + \left(1 - \frac{1}{N}\right)(z-1)^2 y' + \left(\frac{1}{4} + \frac{z-1}{4N^2}\right) y = 0$$

such that

$$(5.1.8) \qquad \phi_1 = 1 - \frac{(N+1)z}{4N} + \cdots, \qquad \phi_2 = z^{1/N}\left(1 - \frac{(N-1)z}{4N} + \cdots\right).$$

The conformal radius of the map G_N is $|\gamma_N^N|$.

Proof. From the formal identity $\varphi_N(x) = \psi_N(x^{1/N})^N$, we deduce

$$\varphi_N(z) = \gamma_N^{-N} \eta_2(z^{1/N})^N / \eta_1(z^{1/N})^N.$$

Letting $\phi_i = \eta_i(z^{1/N})$, we deduce the lemma from an elementary manipulation directly from equation (5.0.3). $\qquad\square$

5.2. **The geometry of $\widetilde{\Gamma}_N$ and $\widetilde{\Phi}_N$.** Recall that $\widetilde{F}_N(\tau|r_N) = \zeta\widetilde{F}_N(\tau)$, where $r_N \in \mathrm{SO}_2(\mathbf{R})$ corresponds in $D(0,1)$ to multiplication by ζ. Explicitly, we have

$$(5.2.1) \qquad r_N = \begin{pmatrix} \cos(\pi/N) & -\sin(\pi/N) \\ \sin(\pi/N) & \cos(\pi/N) \end{pmatrix}.$$

The element r_N acts transitively on the cusps which include $i\infty$, and hence N of the cusps are given by $\cot(\pi k/N)$ for any integer k. For example, if N is even, then we can take $k = (N/2)$ and see that 0 is a cusp.

Definition 5.2.2. *Let $\widetilde{\Phi}_N$ denote the group $\langle \widetilde{\Gamma}_N, r_N \rangle$.*

Since r_N normalizes $\widetilde{\Gamma}_N$, this contains $\widetilde{\Gamma}_N$ as a normal subgroup with $\Phi_N/\widetilde{\Gamma}_N \simeq \mathbf{Z}/N\mathbf{Z}$.

Theorem 5.2.3. *We have $\widetilde{\Phi}_N \simeq \mathbf{Z}/N\mathbf{Z} * \mathbf{Z}$ where r_N is as in equation (5.2.1), and the $\mathbf{Z}$ is the stabilizer of $i\infty$ in $\widetilde{\Gamma}_N$ which is generated by*

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

The group $\widetilde{\Gamma}_N$ is the free group on N generators generated by $t = t_N$ and its conjugates by powers of $r = r_N$.

Proof. The group $\widetilde{\Gamma}_N$ is certainly 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

$$(5.2.4) \qquad\qquad t := \begin{pmatrix} 1 & c_N \\ 0 & 1 \end{pmatrix},$$

then the stabilizers of the other cusps associated to μ_N are generated by the conjugates of t by r. But we know that $\widetilde{\Gamma}_N$ has finite covolume and is not compact. Consider the Dirichlet domain Ω_N associated to $\widetilde{\Gamma}_N$ around $z = 0$ in the Poincaré model. We can describe Ω_N as the region:

$$\Omega := z \in \mathbf{C} \text{ such that } d(gz, i) \geq d(z, i), \quad g, g^{-1} \in \{r^k t r^{-k}\}, k = 0, 1, \ldots, N-1.$$

The region such that $d(r^k t r^{-k} z, i) \geq d(z, i)$ and $d(r^k t^{-1} r^{-k} z, i) \geq d(z, i)$ in the Poincaré disc model is the region bounded by two geodesics starting at ζ^k going in opposite directions and intersecting the boundary at $\zeta^k e^{\pm i\theta}$ 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, and yet $\mathbf{H}/\widetilde{\Gamma}_N$ has $N+1$ cusps. Thus we must have $c_N = 2\cot(\pi/2N)$. $\qquad\square$

The region Ω_N has $2N$ cusps given by half-integer powers of $\zeta = \exp(2\pi i/N)$. It follows from this that the cusp width at this remaining cusp will be $Nc_N = 2N\cot(\pi/2N)$, and that all the remaining N cusps are in the same orbit of $\widetilde{\Gamma}_N$. As an example, note that $e^{k\pi i/N}$ in the Poincaré disc model corresponds to

$$(5.2.5) \qquad\qquad i \cdot \frac{1 + e^{-k\pi i/N}}{1 - e^{-k\pi i/N}} = \cot(\pi k/2N)$$

in $\partial\mathbf{H}$. But now we have (for example)

$$(5.2.6) \qquad\qquad r_N^m t_N r_N^{-m} \cdot \cot\left(\frac{\pi(2m-1)}{2N}\right) = \cot\left(\frac{\pi(2m+1)}{2N}\right).$$

A fundamental domain for $\widetilde{\Phi}_N$ is given by the hyperbolic quadrilateral with vertices $0, \zeta^{-1/2}, 1, \zeta^{1/2}$. 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.

5.3. The group $\widetilde{\Psi}_N$. The group $\widetilde{\Phi}_N = \langle r_N, t_N \rangle$ is contained with index two in the larger group $\widetilde{\Psi}_N = \langle s, t_N \rangle$ where $s^2 = r_N$ is a rotation of order $2N$. The function $\widetilde{F}_N^N$ is invariant under $\widetilde{\Phi}_N$ taking the value 1 at one cusp and ∞ at the other. From this we deduce:

Lemma 5.3.1. *There is an equality*

$$(5.3.2) \qquad\qquad 1 - \widetilde{F}_N^N(s \cdot \tau) = \frac{1}{1 - \widetilde{F}_N^N(\tau)}.$$

Proof. Both sides of the equation are uniformizers of $\mathbf{H}/\widetilde{\Phi}_N$ which take the value 1 at one cusp and ∞ at the other. This specifies them uniquely up to $z \mapsto \lambda z$ scalings. However, this last ambiguity is removed by noting that both sides have their (unique) zero at $\tau = i$. $\qquad\square$

We also deduce from this the equation

$$1 - G_N(-q) = \frac{1}{1 - G_N(q)}.$$

The group $\widetilde{\Psi}_N$ has a fundamental domain consisting of the points $0, 1, \zeta^{1/2}$. But this is none other than a hyperbolic triangle with angles $\{\alpha, \beta, \gamma\} = \{\pi/N, 0, 0\}$. But 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.

Definition 5.3.3. *Let $s_N(z)$ denote the function*

$$(5.3.4) \qquad s_N(z) := z^{1/N} \frac{{}_2F_1\left[\begin{matrix} \frac{N+1}{2N} & \frac{N+1}{2N} \\ 1 + \frac{1}{N} \end{matrix}; z\right]}{{}_2F_1\left[\begin{matrix} \frac{N-1}{2N} & \frac{N-1}{2N} \\ 1 - \frac{1}{N} \end{matrix}; z\right]}.$$

Lemma 5.3.5. *The solutions ϕ_1 and ϕ_2 of the ODE of equation (5.1.6):*

$$(5.3.6) \qquad z(z-1)^2 y'' + \left(1 - \frac{1}{N}\right)(z-1)^2 y' + \left(\frac{1}{4} + \frac{z-1}{4N^2}\right) y = 0$$

are given explicitly by

$$(5.3.7) \qquad \phi_1 = \sqrt{1-z} \cdot z^{1/N} \cdot {}_2F_1\left[\begin{matrix} \frac{N+1}{2N} & \frac{N+1}{2N} \\ 1 + \frac{1}{N} \end{matrix}; z\right], \quad \phi_2 = \sqrt{1-z} \cdot {}_2F_1\left[\begin{matrix} \frac{N-1}{2N} & \frac{N-1}{2N} \\ 1 - \frac{1}{N} \end{matrix}; z\right],$$

In particular, we have $\varphi_N(z) = (\gamma_N^{-1}\phi_2(z)/\phi_1(z))^N = (\gamma_N^{-1}s_N(z))^N$ and $\psi_N = \gamma_N^{-1}s_N(z^N)$.

Proof. This is elementary, since one can directly check that both sides satisfy the same differential equation then check the leading terms. □

Theorem 5.3.8. *The conformal radii of $F_N : D(0,1) \to \mathbf{C} \smallsetminus \mu_N$ and $G_N : D(0,1) \to \mathbf{C} \smallsetminus \{1\}$ are equal to γ_N and γ_N^N respectively, where*

$$(5.3.9) \qquad \gamma_N = \frac{\Gamma\left(\frac{N-1}{2N}\right)^2 \Gamma\left(1 + \frac{1}{N}\right)}{\Gamma\left(\frac{N+1}{2N}\right)^2 \Gamma\left(1 - \frac{1}{N}\right)}.$$

We have a (convergent) expansion for γ_N as follows:

$$(5.3.10) \qquad \gamma_N = 16^{1/N}\left(1 + \frac{\zeta(3)}{2N^3} + \frac{3\zeta(5)}{8N^5} + \cdots\right).$$

Proof. From equation (5.1.3), it suffices to compute the limit of ϕ_2/ϕ_1 as $z \to 1$, which we can do directly given the explicit form in terms of hypergeometric functions (see also (5.4.3)). Similarly, the expansion can be derived directly from the expression in terms of the Gamma function. More precisely, we can write

$$(5.3.11) \qquad \log \gamma_N = \frac{\log 16}{N} + \sum_{k=1}^{\infty} \frac{(2^{2k}-1)}{2^{2k-1}(2k+1)} \cdot \frac{\zeta(2k+1)}{N^{2k+1}}.$$

□

Example 5.3.12. 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].

5.4. **Uniform asymptotics for hypergeometric functions.**

Lemma 5.4.1. *Fix a real constant $M_0 > 0$. Let ψ be the digamma function. For $M \geq M_0$ and $|z| < e^{-MN}$, we have the uniform estimate:*

$$(5.4.2) \qquad \left| \frac{s_N(1-z)}{\gamma_N} - (1-z)^{1/N} \frac{\log z + 2\gamma - 2\psi(1/2 + 1/2N)}{\log z + 2\gamma - 2\psi(1/2 - 1/2N)} \right| \ll \frac{|z|}{N}$$

where the implied estimates depend on M_0 but not on N, M.

Proof. Let $a \notin \mathbf{Z}$. For $|z| < 1$, we have the following equality:

$$(5.4.3) \qquad {}_2F_1\!\left[\begin{matrix} a \; a \\ 2a \end{matrix}; 1-z \right] = \frac{\Gamma(2a)}{\Gamma(a)^2} \sum_{k=0}^{\infty} \frac{(a)_k(a)_k}{k!^2} z^k \left(\log(z) + 2(\psi(k+1) - \psi(k+a))\right).$$

Note that the hypergeometric function is multivalued around $z = 1$, but all different branches are accounted for by the branches of the logarithm. The function $s_N(1-z)$ is given by the ratio (up to a factor of $(1-z)^{1/N}$) by these functions for $a = 1/2 \pm 1/2N$. As $a \to 1/2$, the coefficients in this power series are uniformly bounded. For example, since $|a| < 1$ we have $|(a)_k|/k! < 1$, whereas $|\psi(k) - \psi(k+a)|$ is maximized when $k = 1$ and $a = 1/2 - 1/4$. This immediately leads to the uniform estimates:

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

For $|z| < e^{-MN}$, we also have $\operatorname{Re}(\log z) \leq -MN$, and so in particular $|\log z| \geq MN$ regardless of the branch of logarithm. This leads to the estimate

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

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

5.5. **The region $F_N^N \sim 1$ and $F_N \sim \infty$.**

Lemma 5.5.1. *Fix a real number $M_0 > 0$. Consider $M \geq M_0$ and let Ω_N be a fundamental domain for Φ_N. If*

$$\|F_N(\tau)^N - 1\| < e^{-MN}$$

for $\tau \in \Omega_N$, then for any $\epsilon > 0$

$$(5.5.2) \qquad \operatorname{im}(\tau) > \frac{2N^2 M}{\pi^2}(1-\varepsilon)$$

for $N \gg 1$, where the implied constant depends only on M_0 and ϵ.

Proof. Note that F_N^N is $1:1$ on Ω_N and the only cusp where $F_N = 1$ is at ∞, so it suffices to consider F_N^N in a neighbourhood of the cusp $i\infty$. We may also take the branch of F_N on this domain so that $F_N(i\infty) = 1$. For N sufficiently large, the inequality $\|F_N(\tau)^N - 1\| < e^{-MN}$ implies that $|F_N(\tau) - 1| < e^{-M(1-\varepsilon)N}$ for some ε that tends to zero as N increases. Hence, replacing M by $M(1-\varepsilon)$, we may equivalently assume that, for the branch of ψ_N such that $\psi_N(1) = 1$ and $\widetilde{\psi}_N(1) = i\infty$, that

$$\tau = \widetilde{\psi}_N(1-z), \qquad |z| < e^{-MN}.$$

We may write this as

$$(5.5.3) \qquad \tau = i \cdot \frac{1 + \psi_N(1-z)}{1 - \psi_N(1-z)} = i \cdot \frac{\gamma_N + s_N((1-z)^N)}{\gamma_N - s_N((1-z)^N)}.$$

Using once more the estimate $(1-z)^N \sim 1 - Z$ where $Z < e^{M(1+\varepsilon)N}$ for sufficiently large N, we may consider

$$(5.5.4) \qquad \tau = i \cdot \frac{\gamma_N + s_N(1-Z)}{\gamma_N - s_N(1-Z)}, \qquad |Z| < e^{-MN}.$$

Now by Lemma 5.4.1 the leading term of this is

$$\tau \sim -\frac{i\cot(\pi/2N)}{\pi}\left(2\gamma + \log x - \psi(1/2 - 1/2N) - \psi(1/2 + 1/2N)\right)$$

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

$$\operatorname{im}(\tau) \geq -\frac{\cot(\pi/2N)}{\pi}\left(2\gamma - NM + \psi(1/2 - 1/2N) + \psi(1/2 + 1/2N)\right) \sim \frac{2N^2M}{\pi^2}$$

as required. $\square$

We use this to derive a coarse upper bound on $\sup_{|z|=r}\log|F_N|$, which is not optimal but is enough as an input for the logarithmic error term in the Nevanlinna theory estimate in §6.

A horoball in the Poincaré disc model is a ball bounded by a circle inside the disc $D(0,1)$ which is tangent to the disc.

Lemma 5.5.5. *Let $\mathcal{H}_D$ denote the horoball in $D(0,1)$ which is the image of the subset in $\mathbf{H}$ consisting of $\tau \in \mathbf{H}$ with $\operatorname{im}(\tau) \geq D$. For $\gamma \in U(1,1)$ such that $\widetilde{\gamma} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbf{R})$, the image $\gamma\mathcal{H}_D$ of $\mathcal{H}_D$ under γ in $D(0,1)$ is a circle tangent to $(a-ic)/(a+ic)$ with diameter*

$$(5.5.6) \qquad\qquad\qquad E(\gamma, D) := \frac{2}{1 + D(a^2 + c^2)}$$

Lemma 5.5.7. *For $N \in \mathbf{N}$ and $r \in (0,1)$, we have*

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

where the implicit coefficient is absolute.

Proof. By Lemma 5.3.1, the set $\{z \in D(0,1) : |F_N(z)| > e^M + 1\}$ is contained in the rotation under s of the set $S(M, N) := \{z \in D(0,1) : |F_N^N(z) - 1| < e^{-MN}\}$. Thus to prove the lemma, it suffices to show that for $\frac{N}{1-r} \gg 1$, if we set $M = \frac{N}{1-r}$, then $S(M, N)$ is contained in $\{z \in D(0,1) : |z| > r\}$. (When $\frac{N}{1-r} \leq C$ for some large constant absolute C, then $\sup_{|z|=r}\log|F_N| \leq \sup_{|z|=1-1/C}\log|F_N|$, which is a finite number and there are only finitely many N to consider; thus the lemma follows trivially in this case.)

It follows from Lemma 5.5.1 that

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

Thus we only need to show that the diameters $E(\gamma, D) \leq 1 - r$ of all $\gamma\mathcal{H}_D$.

By Shimizu's Lemma (see, for example, [EGM98, Theorem 3.1]), we have $2|c|\cot(\pi/2N) \geq 1$ for all $\widetilde{\gamma} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \widetilde{\Gamma}_N$. Thus by Lemma 5.5.5,

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

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

Remark 5.5.8. 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 2.0.2 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 (2.0.4) 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 Theorem 2.0.2, 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

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

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 (6.1.2) 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*

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

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

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

with some (effectively computable) absolute constant implicit coefficient.

This quantity is known in Nevanlinna theory as the *mean proximity function at* ∞

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

It is complemented by the *counting function*

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

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

Lemma 6.0.3. *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

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

satisfies for every $a \in \mathbf{C}$ the relation

(6.0.4) $$|T(r,f) - T(r,1/(f-a)) - \log |c(f,a)|| \leq \log^+ |a| + \log 2,$$

where

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

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

(6.0.5) $$T(r,f) - T(r,1/f) = \log |c(f,0)|,$$

and the triangle inequality relation

$$(6.0.6) \qquad\qquad \left| \log^+ |f - a| - \log^+ |f| \right| \leq \log^+ |a| + \log 2.$$

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 $a \neq f(0)$. $\hfill\square$

6.1. **Reduction to a logarithmic derivative.** For $f : D(0, 1) \to \mathbf{C}$ holomorphic as opposed to meromorphic, the polar divisor is empty and $N(r, f) = 0$, so in that case $m(r, f) = T(r, f)$. Since by design $F_N^N - 1$ is a unit in the ring of holomorphic functions on $D(0, 1)$, our requisite bound (6.0.2) rewrites in Nevanlinna notation into

$$(6.1.1) \qquad T(r, p \circ F_N) \ll \log \frac{N}{1 - r}, \quad \text{for each of } p(x) \in \left\{ x^N/(x^N - 1),\ 1/(x^N - 1),\ x^N \right\},$$

and Lemma 6.0.3 using $x^N/(x^N - 1) = 1 + 1/(x^N - 1)$ and (6.0.6) shows that the later three cases for $p(x)$ are equivalent to one another. See the explicit inequality (6.1.5) below for one of these implications. 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

$$(6.1.2) \qquad\qquad p(x) := \frac{x^N}{x^N - 1} = \frac{x}{N} \sum_{\zeta \in \mu_N} \frac{1}{x - \zeta}.$$

By either the chain rule or the partial fractions decomposition, we see that the logarithmic derivative f'/f of the nowhere vanishing holomorphic function

$$(6.1.3) \qquad\qquad f := 1 - F_N^N \quad : \quad D(0, 1) \to \mathbf{C}^\times$$

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

$$(6.1.4) \qquad\qquad p \circ F_N = \frac{F_N}{N F_N'} \frac{f'}{f}.$$

We furthermore have, since $F_N(0) = 0$ and $F_N^N - 1$ is a unit in the ring of holomorphic functions on $D(0, 1)$:

$$T(r, p \circ F_N) = m\left(r, 1 + \frac{1}{F_N^N - 1}\right) \geq m\left(r, \frac{1}{F_N^N - 1}\right) - \log 2$$

$$= T\left(r, \frac{1}{F_N^N - 1}\right) - \log 2 = T(r, F_N^N - 1) - \log 2$$

$$(6.1.5) \qquad\qquad \geq T(r, F_N^N) - 2\log 2 = N\, T(r, F_N) - \log 4.$$

Our proof of Theorem 6.0.1 combines Lemma 5.5.1 with the centerpiece of R. Nevanlinna's original analytic proof — based on the lemma on the logarithmic derivative — of his *second main theorem* of value distribution theory. To bound the logarithmic derivative terms in (6.1.4), it is sufficient to cite [Nev70, Lemma IV.3.1 on page 244] or [Hay64, Lemma 2.3 on page 36]. For convenience to the reader, and particularly since the argument 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 6.1.6. *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 $0 < r < R$,*

$$(6.1.7) \qquad\qquad m\left(r, \frac{g'}{g}\right) < \log^+ \left\{ \frac{m(R, g)}{r} \frac{R}{R - r} \right\} + \log 2 + 1/e.$$

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

$$(6.1.8) \qquad\qquad \log |g(z)| = \int_{|w| = R} \log |g(w)| \cdot \Re\left(\frac{w + z}{w - z}\right) \mu_{\text{Haar}}(w),$$

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

$$(6.1.9) \qquad \log g(z) = \int_{|w|=R} \log|g(w)| \cdot \frac{w+z}{w-z} \, \mu_{\text{Haar}}(w),$$

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

Differentiation in the integrand of (6.1.9) gives the reproducing formula

$$(6.1.10) \qquad \frac{g'(z)}{g(z)} = \int_{|w|=R} \frac{2w}{(w-z)^2} \log|g(w)| \, \mu_{\text{Haar}}(w), \quad \forall z \in D(0, R).$$

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

$$(6.1.11) \qquad \int_{|z|=r} |w - z|^{-2} \, \mu_{\text{Haar}}(z) = \frac{1}{R^2 - r^2} \quad \text{for } |w| = R > r,$$

and thus the $|z| = r$ integral of (6.1.10) 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

$$\int_{|z|=r} \left|\frac{g'(z)}{g(z)}\right| \mu_{\text{Haar}} \leq 2R \int_{|z|=r} \int_{|w|=R} |w - z|^{-2} \big|\log|g(w)|\big| \, \mu_{\text{Haar}}(w) \, \mu_{\text{Haar}}(z)$$

$$= 2R \int_{|w|=R} \left(\int_{|z|=r} |w - z|^{-2} \, \mu_{\text{Haar}}(z)\right) \big|\log|g(w)|\big| \, \mu_{\text{Haar}}(w)$$

$$= \frac{2R}{R^2 - r^2} \int_{|w|=R} \big|\log|g(w)|\big| \, \mu_{\text{Haar}}(w)$$

$$= \frac{2R}{R^2 - r^2} \Big(m(R, g) + m(R, 1/g)\Big) = \frac{4R\, m(R, g)}{R^2 - r^2},$$

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

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

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

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

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

$$\int_{|z|=r} \log^+\left|\frac{g'}{g}\right| \mu_{\text{Haar}} \leq \mu_{\text{Haar}}(E) \log^+\left(\frac{1}{\mu_{\text{Haar}}(E)} \int_E \left|\frac{g'(z)}{g(z)}\right| \mu_{\text{Haar}}(z)\right)$$

$$\leq \log^+ \int_{|z|=r} \left|\frac{g'(z)}{g(z)}\right| \mu_{\text{Haar}}(z) + \sup_{t \in (0,1]} \big\{t \log(1/t)\big\}$$

$$\leq \log^+\left\{\frac{4R\, m(R, g)}{R^2 - r^2}\right\} + \frac{1}{e} \leq \log^+\left\{\frac{m(R, g)}{r} \frac{R}{R-r}\right\} + \log 2 + \frac{1}{e},$$

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

Remark 6.1.12. 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. In this way, by using a technique due to Kolokolnikov for handling the sum over the zeros and poles, Goldberg and Grinshtein [GG76] obtained the general bound

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

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 book of Cherry and Ye [CY01] 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.** On applying Lemma 6.1.6 to the nowhere vanishing holomorphic function (6.1.3) 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

$$m\Big(r, \frac{f'}{f}\Big) \ll \log^+ T\Big(\frac{1+r}{2}, f\Big) + \log^+ \frac{e}{1-r}$$

$$= \log^+ m\Big(\frac{1+r}{2}, 1 - F_N^N\Big) + \log^+ \frac{e}{1-r}$$

$$\ll \log^+ m\Big(\frac{1+r}{2}, F_N^N\Big) + \log^+ \frac{e}{1-r}$$

$$\ll \log^+ m\Big(\frac{1+r}{2}, F_N\Big) + \log^+ \frac{N}{1-r}$$

$$(6.2.1) \qquad \ll \sup_{|z|=(1+r)/2} \log^+ \log |F_N| + \log^+ \frac{N}{1-r},$$

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

We shall handle the term F_N/F_N' in (6.1.4) by employing Lemma 6.1.6 to the functional unit $g = 1 - F_N$, and the following 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]), beginning with (6.0.5) for the function F_N'/F_N, and using that our function F_N is holomorphic on the disc $D(0, 1)$ with $F_N(0) = 0$ and $F_N'(0) \neq 0$:

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

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

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

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 and the *ramification term* $N_{\mathrm{ram}}(r, F_N) = N(r, 1/F_N') = 0$ is actually zero. (In any event one could drop a ramification term $N_{\mathrm{ram}} = N_1 \geq 0$ by positivity, here or in [Hay64, Theorem 2.1]. See also the discussion in Remark 6.2.5 below.)

We continue to estimate with the triangle inequality (for the second and third lines) and then (6.0.5), noting that $|F_N'(0)| > 1$ (for the inequality in the fourth line):

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

$$\leq m\Big(r, \frac{F_N'}{1 - F_N}\Big) + m\Big(r, \frac{1 - F_N}{F_N}\Big) + N\Big(r, 1/F_N\Big)$$

$$\leq m\Big(r, \frac{F_N'}{1 - F_N}\Big) + \log 2 + m\Big(r, \frac{1}{F_N}\Big) + N\Big(r, 1/F_N\Big)$$

$$= m\Big(r, \frac{(1 - F_N)'}{1 - F_N}\Big) + T(r, 1/F_N) + \log 2 \leq m\Big(r, \frac{(1 - F_N)'}{1 - F_N}\Big) + T(r, F_N) + \log 2$$

$$(6.2.2) \qquad \leq T(r, F_N) + O\Big(\log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|\Big),$$

upon again using Lemma 6.1.6 with $R := (1-r)/2$ but now for the functional unit $g := 1 - F_N$, and a similar argument as in (6.2.1).

At this point the key identity (6.1.4) allows to combine the estimates (6.2.1) and (6.2.2), arriving at the uniform bound

$$T(r, p \circ F_N) = m(r, p \circ F_N) \leq m\left(r, \frac{f'}{f}\right) + m\left(r, \frac{F_N}{F_N'}\right)$$

$$\text{(6.2.3)} \qquad \leq T(r, F_N) + O\left(\log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|\right).$$

We leverage the upper bound (6.2.3) on $T(r, p \circ F_N) = N T(r, F_N) + O(1)$ against the lower bound (6.1.5) and get a uniform upper bound on $T(r, F_N)$:

$$\text{(6.2.4)} \qquad (N-1)T(r, F_N) \ll \log^+ \frac{N}{1-r} + \sup_{|z|=(1+r)/2} \log^+ \log |F_N|.$$

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

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

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

Hence Theorem 6.0.1 follows from Lemma 5.5.7 upon replacing r there with $(1+r)/2$.

Remark 6.2.5. The bound (6.2.4) 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 $T(r, g) = m(r, g) = \int_{|z|=r} \log^+ |g| \, \mu_{\text{Haar}}$ satisfies the upper bound

$$\text{(6.2.6)} \qquad (N-1)T(r, g) + N_{\text{ram}}(r, g) \leq \sum_{i=1}^{N} N(r, a_i) + O(\log T(r, g)) + O(\log r)$$

outside of an exceptional set of radii $r \in E \subset [0, \infty)$ of finite Lebesgue measure: $m(E) < \infty$. Here $N_{\text{ram}}(r, g) = N(r, 1/g')$ 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 (6.2.6), 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 (6.2.6) 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 $m(r, F_N) \ll \log \frac{N}{1-r}$ 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 (6.1.2) particular to our situation of $\{a_1, \ldots, a_q\} = \mu_N$.

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 5.3.8, 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 rings 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 4.2.5 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 ring of modular functions with coefficients in K, bounded denominators, and cusp widths dividing $2N$. We follow the proof of Proposition 3.0.1 now on the case of the ring $\mathcal{H}(U, x(t), O_K)$ from Definition 2.0.1: the ring of formal power series $f(x) \in K[\![x]\!]$ such that $f(x(t)) \in O_K[\![t]\!]$ and $L(f) = 0$ for some nonzero linear differential operator L over $\mathbf{Q}(x)$ without any singularities on U. Then $R_{2N} \subset \mathcal{H}(U, x(t), O_K) \otimes_{O_K} K$. Note that U is stable under the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, and thus $\mathcal{H}(U, x(t), O_K) = \mathcal{H}(U, x(t), \mathbf{Z}) \otimes_{\mathbf{Z}} O_K$ and $\dim_{K(p(x))} \mathcal{H}(U, x(t), O_K) \otimes K(p(x)) = \dim_{\mathbf{Q}(p(x))} \mathcal{H}(U, x(t), \mathbf{Z}) \otimes \mathbf{Q}(p(x))$. Thus by Theorem 2.0.2, we still have that R_{2N} has dimension at most $CN^3 \log N$ over $K[\lambda^{\pm}, (1-\lambda)^{\pm}]$. The claimed extension to $\overline{\mathbf{Z}}[\![q^{1/N}]\!]$ Fourier expansions now follows upon remarking that the proof of Proposition 4.2.5 still persists when $\mathbf{Q}$ is replaced by K.

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 $V^\sharp$. 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 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 2.0.2, 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 5.3.8 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 2.0.2 implies that $g(x) \in \mathbf{Z}[\![x]\!]$ is an algebraic power series. Hence $f(x) = g(\sqrt[N]{x})$ is algebraic.

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

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

a weight one modular form for the congruence group $\Gamma(2)$. The modularity streak is not an accident: more generally, 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 $\mathbf{P}^1 \setminus \{0, 1/16, \infty\} \cong Y(2)$, 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 7.4.1 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*
$$\forall \tau \in \mathbf{H}, \quad a \leq \mathrm{Re}\,\tau \leq b, \quad \mathrm{Im}\,\tau \geq C \quad \Longrightarrow \quad |F_j(\tau)| \leq A e^{B \, \mathrm{Im}\,\tau}.$$

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. 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 $\mathbf{P}^1 \smallsetminus \{0, 1, \infty\}$ 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. 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. The following is 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.

Question 7.4.1. *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 classical (congruence) quasi-modular forms of all weights and levels?*

Here, the *quasi-modular ring* is the smallest ring containing all classical modular forms and stable under the $q\frac{d}{dq}$ operator; we refer to [Zag08, § 5.3] for the precise definition. Complementing Example 7.2.3 is the λ-pullback of the complete elliptic integral of the second kind:

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

clearly a component of a logarithmic vector-valued modular form on $\mathrm{SL}_2(\mathbf{Z})$, whose q-expansion is in $\mathbf{Z}[\![q]\!]$ and a little less obvious than (7.2.4), but indeed one can verify that

$$\frac{2}{\pi}E(\lambda(q)) = \frac{\left(\sum_{n\in\mathbf{Z}} q^{n^2}\right)\left(\sum_{n\in\mathbf{Z}}(-1)^n q^{n^2}\right)^4 + 4q\frac{d}{dq}\left(\sum_{n\in\mathbf{Z}} q^{n^2}\right)}{\left(\sum_{n\in\mathbf{Z}} q^{n^2}\right)^3}$$

is also an element of the quasi-modular ring for the congruence level $\Gamma(2)$. A further apparent example is

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

where the integrality of the coefficients comes from the Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n} \in \mathbf{Z}$. But this example is not really new:

$$8\int \frac{2}{\pi}K(16x)\,dx = \frac{2}{\pi}E(16x) + (16x-1)\frac{2}{\pi}K(16x).$$

TODO: But what about $\sum \frac{1}{(n+1)^2}\binom{2n}{n}^2(\lambda/16)^n$, does that belong to the quasi-modular ring?

7.4.3. 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. 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. 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 $\pi_1\big(\mathrm{Spec}\,\mathbf{Z}[x, 1/x, 1/(x-1)]\big) = \{1\}$. 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 7.4.2. *Let $N \in \mathbf{N}$, let $K/\mathbf{Q}(\mu_N)$ be a finite extension, and let $\pi : \mathcal{X}(N) \to \mathrm{Spec}\,O_K$ ("connected Néron model") be the connected component containing the cusp ∞ in the smooth part of the minimal regular model of $X(N)$ over $\mathrm{Spec}\,O_K$. Thus the cusp ∞ extends to a morphism $\varepsilon : \mathrm{Spec}\,O_K \to \mathcal{X}(N)$.*

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

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

and

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

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]. $\qquad\square$

Remark 7.4.3. 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 7.4.2 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 7.4.2 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 $\mathrm{Spec}\,O_K$. Its Faltings height is

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

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 7.4.2 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

FC: To be
completed

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.

[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.

[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.

[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.

[CY01]   William Cherry and Zhuan Ye, *Nevanlinna's Theory of Value Distribution*, Springer Monographs in Mathematics, Springer Verlag, Berlin, 2001.

[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.

[FF20] Andrew Fiori and Cameron Franc, *The unbounded denominator conjecture for the noncongruence subgroups of index* 7, 2020, arXiv:2007.01336, p. 23.

[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.

[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.

[Iha94] Yasutaka Ihara, *Horizontal divisors on arithmetic surfaces associated with Belyĭ 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.

[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.

[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.

[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.

[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