GENERALIZATION OF SELBERG'S 3/16 THEOREM AND AFFINE SIEVE

JEAN BOURGAIN, ALEX GAMBURD, AND PETER SARNAK

Dedicated to the memory of Atle Selberg

1. Introduction

A celebrated theorem of Selberg [32] states that for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ there are no exceptional eigenvalues below 3/16. We prove a generalization of Selberg's theorem for infinite index "congruence" subgroups of $\mathrm{SL}_2(\mathbb{Z})$. Consequently we obtain sharp upper bounds in the affine linear sieve, where in contrast to [4] we use an archimedean norm to order the elements.

Let Λ be a finitely generated non-elementary subgroup of $\mathrm{SL}_2(\mathbb{Z})$; let $X_\Lambda = \Lambda\backslash\mathbb{H}$ be the corresponding hyperbolic surface (which is of infinite volume if Λ is of infinite index in $\mathrm{SL}(2,\mathbb{Z})$). Let $\delta(\Lambda)$ denote the Hausdorff dimension of the limit set of Λ. The generalization of Selberg's theorem splits into two cases: $\delta(\Lambda) > \frac{1}{2}$ and $0 < \delta(\Lambda) \leq \frac{1}{2}$.

In the case that $\delta(\Lambda) > \frac{1}{2}$ the spectrum of the Laplace-Beltrami operator on $L^2(X_\Lambda)$ consists of finite number of points in $[0, \frac{1}{4})$ (see [17]). We denote them by

$$0 \leq \lambda_0(\Lambda) < \lambda_1(\Lambda) \leq \cdots \leq \lambda_{\max}(\Lambda) < \frac{1}{4}.$$

The assumption that $\delta(\Lambda) > \frac{1}{2}$ is equivalent to $\lambda_0(\Lambda) < \frac{1}{4}$, and in this case $\delta(1-\delta) = \lambda_0$ [24].

The following extension of Selberg's theorem is proved in section 2.

Theorem 1.1. *Let Λ be a finitely generated subgroup of $\mathrm{SL}(2,\mathbb{Z})$ with $\delta(\Lambda) > \frac{1}{2}$. For $q \geq 1$ let $\Lambda(q)$ be the "congruence" subgroup $\{x \in \Lambda : x \equiv I \mod q\}$. There is $\varepsilon = \varepsilon(\Lambda) > 0$ such that*

$$\lambda_1(\Lambda(q)) \geq \lambda_0(\Lambda(q)) + \varepsilon,$$

for all square-free $q \geq 1$ (note that $\lambda_0(\Lambda(q)) = \lambda_0(\Lambda)$).

The first author was supported in part by the NSF. The second author was supported in part by DARPA, NSF, and the Sloan Foundation. The third author was supported in part by the NSF.

In [11] an explicit and stronger version of Theorem 1.1 is proven under the assumption that $\delta(\Lambda) > \frac{5}{6}$. See [29] for the sharpest known bounds towards Selberg's $\frac{1}{4}$ Conjecture as well as bounds towards the Ramanujan Conjectures for more general groups.

Theorem 1.1 is a consequence of Theorem 1.2 in [4] and the following result, which is of independent interest.

Theorem 1.2. *Let $\Lambda = \langle S \rangle$ be a finitely generated subgroup of* $\mathrm{SL}(2,\mathbb{R})$ *with $\delta(\Lambda) > \frac{1}{2}$. Let $\{N_i\}$ be a family of finite index normal subgroups of Λ. Then the following are equivalent*

1. *The Cayley graphs $\mathcal{G}(\Lambda/N_i, S)$ form a family of expanders.*
2. *There is $\varepsilon = \varepsilon(\Lambda) > 0$ such that $\lambda_1(\Lambda/N_i) \geq \lambda_0(\Lambda/N_i) + \varepsilon$.*

The argument in section 2 establishes that $1 \Rightarrow 2$; the implication $2 \Rightarrow 1$ is proved using Fell's continuity of induction in section 7 of [11]. Theorem 1.2 generalizes results of Brooks [5] and Burger [6, 7] who proved it in the case of co-compact Λ.

Combining Theorem 1.1 with Lax-Phillips theory of asymptotic distribution of lattice points [17] we obtain the following result, which is the crucial ingredient in the execution of the affine linear sieve in the archimedean norm.

Theorem 1.3. *Let Λ be a finitely generated subgroup of* $\mathrm{SL}(2,\mathbb{Z})$ *with $\delta(\Lambda) > \frac{1}{2}$. Assume that q is square free and $(q, q_0) = 1$, where q_0 is provided by the strong approximation theorem* [19]. *There is $\varepsilon_1 > 0$ depending on Λ such that for any $g \in \mathrm{SL}_2(q)$ we have*

$$\begin{aligned} &|\{\gamma \in \Lambda \mid \|\gamma\| \leq T \text{ and } \gamma \equiv g \bmod q\}| \\ &= \frac{c_\Lambda T^{2\delta}}{|\mathrm{SL}_2(q)|} + O(q^3 T^{2\delta - \varepsilon_1}). \end{aligned} \tag{1.1}$$

We now turn to the discussion of the case $\delta(X) \leq \frac{1}{2}$. In this case there is no discrete L^2 spectrum and its natural replacement is furnished by the resonances of X, which are given as the poles of the meromorphic continuation of the resolvent $R_X(s) = (\Delta_X - s(1-s))^{-1}$. By the result of Patterson [24] and Sullivan [34] $R_X(s)$ is analytic for $\Re(s) > \delta$; Mazzeo and Melrose [20] proved that $R_X(s)$ has a meromorphic continuation to the entire plane. In [25] Patterson proved that $R_X(s)$ has a simple pole at $s = \delta$ and no further poles on the line $\Re(s) = \delta$; his proof is based on ideas from ergodic theory related to Ruelle zeta-function. Using further development of these ideas due to Dolgopyat [8], Naud [21] has recently established that $R_X(s)$ is holomorphic (with the exception of simple pole at $s = \delta$) for $\Re(s) > \delta - \varepsilon$, with ε depending on X. The following result, giving a resonance-free region for congruence resolvent, is proved in section 11.

Theorem 1.4. *Let Λ be a finitely generated subgroup of $\mathrm{SL}(2,\mathbb{Z})$ with $\delta(\Lambda) \le \frac{1}{2}$. For $q \ge 1$ square free let $\Lambda(q)$ be the "congruence" subgroup $\{x \in \Lambda : x \equiv I \mod q\}$; let $X(q) = \Lambda(q)\backslash\mathbb{H}$. There is $\varepsilon = \varepsilon(\Lambda) > 0$ such that $R_{X(q)}(s)$ is holomorphic(with the exception of simple pole at $s = \delta$) for $\Re(s) > \delta - \varepsilon \min\left(1, \frac{1}{\log(1+|\Im s|)}\right)$.*

When $\delta \le \frac{1}{2}$ we cannot apply the expansion property [4] directly; instead, to prove theorem 1.4 we use a dynamical treatment and invoke a generalization of the underlying result on measure convolution ("L^2-flattening lemma"): see Lemmas 1 and 2 in section 7. It is likely that by combining our methods with the extension of Dolgopyat's result [8] to vector-valued functions, analyticity of $R_{X(q)}(s)$ can be established for $\Re(s) > \delta - \varepsilon$ — in complete analogy[1] in with Theorem 1.1.

Using methods of Lalley [16] we obtain the following analogue of Theorem 1.3, which is sufficient for sieving applications.

Theorem 1.5. *Let Λ be a finitely generated subgroup of $\mathrm{SL}(2,\mathbb{Z})$ with $0 < \delta(\Lambda) \le \frac{1}{2}$. Assume that q is square free and $(q, q_0) = 1$, where q_0 is provided by the strong approximation theorem* [19]*. There is $\varepsilon_1 > 0$, $C > 0$ depending on Λ such that for any $g \in \mathrm{SL}_2(q)$ we have*

$$(1.2)\qquad \begin{aligned} &|\{\gamma \in \Lambda \,|\, \|\gamma\| \le T \text{ and } \gamma \equiv g \bmod q\}| \\ &= \frac{c_\Lambda T^{2\delta}}{|\mathrm{SL}_2(q)|}\left(1 + O\left(T^{-\frac{1}{\log\log T}}\right)\right) + O\left(q^C T^{2\delta - \varepsilon_1}\right). \end{aligned}$$

We turn to applications to affine linear sieve [4]. Consider the standard action on the two by two integer matrices by multiplication on the left, and take the orbit $\mathcal{O}$ of I (the identity matrix) under Λ. Set $|x| = \left(\sum_{i,j} x_{ij}^2\right)^{\frac{1}{2}}$ where $x = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$. Set $N_\Lambda(T) = |\{x \in \Lambda : |x| \le T\}|$ and let $\delta(\Lambda)$ be the Hausdorff dimension of the limit set of an orbit

[1]The analogy between Theorem 1.1 and Theorem 1.4 becomes clearer when their assertions are expressed in terms of the Selberg zeta function [31]. If Λ is a finitely generated subgroups of $\mathrm{SL}_2(\mathbb{R})$ the Selberg zeta function $Z_X(s)$ associated with $X = \Lambda\backslash\mathbb{H}$ is known to be an entire function, whose non-trivial zeros are given by the resonances and the finite point spectrum [12, 26]. Consequently, Theorem 1.1 is equivalent to the assertion that when $\delta(\Lambda) > \frac{1}{2}$ there is $\varepsilon(\Lambda) > 0$ such that $Z_{X(q)}(s)$ is analytic and non-vanishing on the set $\{\Re(s) > \delta - \varepsilon\}$, except at $s = \delta$ which is a simple zero, while Theorem 1.4 is equivalent to the assertion that when $\delta(\Lambda) \le \frac{1}{2}$ there is $\varepsilon(\Lambda) > 0$ such that $Z_{X(q)}(s)$ is analytic and non-vanishing on the set $\left\{\Re(s) > \delta - \varepsilon \min\left(1, \frac{1}{\log(1+|\Im s|)}\right)\right\}$, except at $s = \delta$ which is a simple zero.

$\Lambda z \subset \mathbb{H} \cup \{\infty\} \cup \mathbb{R}$, where $\mathbb{H}$ is the hyperbolic plane, $z \in \mathbb{H}$ and Λ act by linear fractional transformations. By the results of Lax-Phillips [17] and Lalley [16] we have that $N_\Lambda(T) \sim c_\Lambda T^{2\delta(\Lambda)}$, as $T \to \infty$. Let $f \in \mathbb{Q}[x_{ij}]$ be integral on $\mathcal{O}$ and assume that it is weakly primitive for $\mathcal{O}$, that is $\gcd\{f(x) : x \in \mathcal{O}\}$ is 1. If f is not weakly primitive then $\frac{1}{N} f$ is, where $N = \gcd f(\mathcal{O})$, and we can represent any weakly primitive f as $\frac{1}{N} g$ with $g \in \mathbb{Z}[x_{ij}]$ and $N = \gcd(\mathcal{O})$.

The coordinate ring $\mathbb{Q}[x_{ij}]/(\det(x_{ij}) - 1)$ is a unique factorization domain [28] and we can factor f into $t = t(f)$ irreducibles $f_1 f_2 \dots f_t$ in this ring. Set

$$\pi_{\Lambda,f}(T) = |\{x \in \Lambda; |x| \le T, f_j(x) \text{ is prime for } j = 1, \dots, t\}|.$$

For $f \in \mathbb{Z}[x_{ij}]$ weakly primitive with $t(f)$ irreducible factors our conjectured asymptotics is of the form:

$$\pi_{\Lambda,f}(T) \sim \frac{c(\Lambda, f) N_\Lambda(T)}{(\log T)^{t(f)}}, \text{ as } T \to \infty, \tag{1.3}$$

where $c(\Lambda, f)$ can be expressed as a product of local densities; see [10, 30] for an example of explicit computation and numerical experiments. In section 13 we establish the following sharp upper bound for $\pi_{\Lambda,f}(T)$.

Theorem 1.6. *Let Λ be a subgroup of $\mathrm{SL}(2,\mathbb{Z})$ which is Zariski dense in SL_2 and let $f \in \mathbb{Z}[x_{ij}]$ be weakly primitive with $t(f)$ irreducible factors. Then*

$$\pi_{\Lambda,f}(T) \ll \frac{N_\Lambda(T)}{(\log T)^{t(f)}}. \tag{1.4}$$

We also obtain the following lower bound for the number of points $x \in \Lambda$ for which f has at most a fixed number of prime factors.

Theorem 1.7. *Let Λ be a subgroup of $\mathrm{SL}(2,\mathbb{Z})$ which is Zariski dense in SL_2 and let $f \in \mathbb{Z}[x_{ij}]$ be weakly primitive with $t(f)$ irreducible factors. Then there is an $r < \infty$, which can be given explicitly in terms of $\varepsilon(\Lambda)$ in theorems 1.1 and 1.4, such that*

(1.5)

$$|\{x \in \Lambda; |x| \le T, \text{ and } f(x) \text{ has at most } r \text{ prime factors}\}| \gg \frac{N_\Lambda(T)}{(\log T)^{t(f)}}.$$

2. Generalization of Selberg's 3/16 theorem when $\delta > 1/2$

Being a subgroup of finite index in $\Lambda(1)$, $\Lambda(q)$ has the same bottom of the spectrum, $\lambda_0(\Lambda(q)\backslash\mathbb{H}) = \lambda_0(\Lambda(1)\backslash\mathbb{H})$. As in section 2 of [11], we have that for q large enough $\Lambda(1)/\Lambda(q) \cong \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$. Let $S = \{A_1, \dots, A_k\}$, and let S_q be the natural projection of S modulo q.

Theorem 1.7 in [4] implies that if $\Lambda = \langle S \rangle$ is non-elementary, then $\mathcal{G}_q = \mathcal{G}(\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z}), S_q)$ is a family of expanders. Consider the space $H(q)$ of vector-valued functions F on $\mathcal{F}(1)$, satisfying

$$F(\gamma z) = R_q(\gamma) F(z), \tag{2.1}$$

for $\gamma \in \Lambda(1)/\Lambda(q) \cong \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$, where $R_q(\gamma)$ denotes the regular representation of $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$; we denote by $\langle , \rangle$ the inner product on this space and by $\| \ \|$ the associated norm. Denoting by $H_0(q)$ the subspace of functions in $H(q)$ orthogonal to φ_0, the eigenfunction corresponding to λ_0, the assertion of Theorem 1.1 is equivalent to existence of $c > 0$ such that

$$\frac{\int_{\mathcal{F}} \|\nabla F\|^2 d\mu}{\int_{\mathcal{F}} \|F\|^2 d\mu} \geq \lambda_0 + c \tag{2.2}$$

for all $F \in H_0(q)$.

Applying Theorem 1.7 in [4] for each $z \in \mathcal{F}(1)$ implies that there is $\varepsilon > 0$, depending *only* on S, such that for all $F \in H_0(q)$, we have

$$\|F(\gamma z) - F(z)\| \geq \varepsilon \|F(z)\| \quad \text{for some } \gamma \in S. \tag{2.3}$$

Let $f = \|F\|$, and decompose it as

$$f = a\varphi_0(z) + b(z), \tag{2.4}$$

where

$$\int_{\mathcal{F}} \varphi_0(z) \overline{b(z)} d\mu(z) = 0 \tag{2.5}$$

and

$$\int_{\mathcal{F}} |f|^2 d\mu = a^2 + \int_{\mathcal{F}} |b|^2 d\mu = 1. \tag{2.6}$$

Write

$$F(z) = (F_1(z), \ldots, F_N(z)),$$

where $N = |\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})|$. Since

$$\nabla \left(\sum_{j=1}^{N} |F_j(z)|^2 \right)^{\frac{1}{2}} = \begin{cases} \frac{\sum_{j=1}^{N} F_j(z) \overline{\nabla F_j(z)}}{\left(\sum_{j=1}^{N} |F_j(z)|^2\right)^{\frac{1}{2}}} & \text{if } \sum_{j=1}^{N} |F_j(z)|^2 \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

we have

$$\|\nabla F\|^2(z) \geq |\nabla \|F\||^2(z) = |\nabla f|^2(z).$$

Consequently we obtain:

(2.7)

$$\frac{\int_{\mathcal{F}} \|\nabla F\|^2 d\mu}{\int_{\mathcal{F}} \|F\|^2 d\mu} \geq \frac{\int_{\mathcal{F}} |\nabla \|F\||^2 d\mu}{\int_{\mathcal{F}} \|F\|^2 d\mu} = \frac{\int_{\mathcal{F}} |\nabla f|^2 d\mu}{\int_{\mathcal{F}} |f|^2 d\mu} =$$
$$\frac{\int_{\mathcal{F}} \langle \Delta f, f \rangle d\mu}{\int_{\mathcal{F}} |f|^2 d\mu} = \int_{\mathcal{F}} \langle a\lambda_0\varphi_0 + \Delta b, a\varphi_0 + b \rangle d\mu =$$
$$a^2\lambda_0 + \langle \Delta b, b \rangle \overset{(2.5)}{\geq} a^2\lambda_0 + \lambda_1 \int_{\mathcal{F}} |b|^2 d\mu \geq \lambda_0 + (\lambda_1 - \lambda_0) \int_{\mathcal{F}} |b|^2 d\mu.$$

By a theorem of Lax and Phillips [17] there are only finitely many discrete eigenvalues of Λ in $[0, \frac{1}{4}]$; consequently

$$\lambda_1 - \lambda_0 \geq c_1 > 0. \tag{2.8}$$

Therefore, as soon as $\int_{\mathcal{F}} |b|^2 d\mu > \varepsilon_1 > 0$, we have that

$$\frac{\int_{\mathcal{F}} |\nabla F|^2 d\mu}{\int_{\mathcal{F}} |F|^2 d\mu} \geq \lambda_0 + c_1\varepsilon_1. \tag{2.9}$$

Now consider the case of $\int_{\mathcal{F}} |b|^2 d\mu = 0$. We can assume that $a = 1$ and write $F(z) = u(z)\varphi_0(z)$, with $u(z) = (u_1(z), \ldots, u_N(z))$, where $N = |\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})|$. Now

$$\|u(z)\| = \sum_{j=1}^{N} |u_j|^2(z) = 1 \tag{2.10}$$

implies

$$\sum_{j=1}^{N} u_j \frac{\partial u_j}{\partial x} = \sum_{j=1}^{N} u_j \frac{\partial u_j}{\partial x} = 0, \tag{2.11}$$

and since

$$\frac{\partial(\varphi_0 u_j)}{\partial x} = u_j \frac{\partial \varphi_0}{\partial x} + \varphi_0 \frac{\partial u_j}{\partial x},$$

$$\frac{\partial(\varphi_0 u_j)}{\partial y} = u_j \frac{\partial \varphi_0}{\partial y} + \varphi_0 \frac{\partial u_j}{\partial y},$$

we have that

$$\|\nabla\varphi_0 u\|^2 = \left(\frac{\partial\varphi_0}{\partial x}\right)^2\sum_{j=1}^{N} u_j^2 + \varphi_0^2\sum_{j=1}^{N}\left(\frac{\partial u_j}{\partial x}\right)^2 + \left(\frac{\partial\varphi_0}{\partial y}\right)^2\sum_{j=1}^{N} u_j^2 + \varphi_0^2\sum_{j=1}^{N}\left(\frac{\partial u_j}{\partial y}\right)^2 + 2\varphi_0\frac{\partial\varphi_0}{\partial x}\sum_{j=1}^{N} u_j\frac{\partial u_j}{\partial x} + 2\varphi_0\frac{\partial\varphi_0}{\partial y}\sum_{j=1}^{N} u_j\frac{\partial u_j}{\partial y} = |\nabla\varphi_0|^2 + \varphi_0^2\|\nabla u\|^2. \tag{2.12}$$

Consequently,

$$\frac{\int_{\mathcal{F}}\|\nabla F\|^2 d\mu}{\int_{\mathcal{F}}\|F\|^2 d\mu} = \frac{\int_{\mathcal{F}}|\nabla\varphi_0|^2 + \varphi_0^2\|\nabla u\|^2 d\mu}{\int_{\mathcal{F}}|\varphi_0|^2 d\mu} = \frac{\int_{\mathcal{F}}|\nabla\varphi_0|^2 d\mu}{\int_{\mathcal{F}}|\varphi_0|^2 d\mu} + \frac{\int_{\mathcal{F}}\varphi_0^2\|\nabla u\|^2 d\mu}{\int_{\mathcal{F}}|\varphi_0|^2 d\mu} \geq \lambda_0 + \frac{\int_{\mathcal{F}}\varphi_0^2\|\nabla u\|^2 d\mu}{\int_{\mathcal{F}}|\varphi_0|^2 d\mu}. \tag{2.13}$$

Our aim now is to show that

$$\frac{\int_{\mathcal{F}}\varphi_0^2\|\nabla u\|^2 d\mu}{\int_{\mathcal{F}}|\varphi_0|^2 d\mu} \geq c_2 > 0. \tag{2.14}$$

To that end we assume that that

$$\frac{\int_{\mathcal{F}}\varphi_0^2\|\nabla u\|^2 d\mu}{\int_{\mathcal{F}}|\varphi_0|^2 d\mu} < \kappa \tag{2.15}$$

and will obtain a contradiction for sufficiently small κ (κ_j below are of the form $a_j \cdot \kappa$ for suitable constants a_j). Consider the fundamental domain $\mathcal{F}(1) = \Lambda(1)\backslash\mathbb{D}$. Its boundary, $\partial\mathcal{F}(1)$, consists of finitely many geodesic arcs $\{l_i\}$ splitting into pairs l_j, l'_j in such a way that there is $\gamma_j \in S$ so that $l_j = \gamma_j l'_j$; γ_j are distinct and generate $\Lambda(1)$. Further, we have decomposition of the following form:

$$\mathcal{F}(1) = \mathcal{K}(1) \cup \bigcup_{i\in Cu(1)} \text{cusp}_i \cup \bigcup_{j\in Fl(1)} \text{flare}_j$$

where

(1) $\mathcal{K}(1)$ is relatively compact in $\mathbb{D}$

(2) $Cu(1)$ is a set of cusps of $\mathcal{F}(1)$. Each cusp$_i$ is isometric to a standard cuspidal fundamental domain $P(Y_i)$ of the form

$$P(Y) = \{z = x + iy \mid 0 < x < 1, y > Y\},$$

based on a horocycle

$$h_Y = \{x + iy \mid y = Y\}.$$

(3) $Fl(1)$ is a set of flares of $F(1)$. Each flare$_j(\alpha)$ is isometric to a standard hyperbolic fundamental domain $F(\alpha)$ of the form

$$F(\alpha) = \{z : 1 < |z| < \exp(\beta); 0 < \arg(z) < \alpha\},$$

where $\alpha < \frac{\pi}{2}$.

Since $\varphi_0 \in L^2(\mathcal{F}(1))$, we have that

$$\int_{\mathcal{K}} |\varphi_0|^2 d\mu \geq c_3 \int_{\mathcal{F}} |\varphi_0|^2 d\mu \text{ for some } c_3 > 0, \tag{2.16}$$

and therefore (2.15) implies that

$$\frac{\int_{\mathcal{K}} \varphi_0^2 \|\nabla u\|^2 d\mu}{\int_{\mathcal{K}} |\varphi_0|^2 d\mu} \leq \kappa_1. \tag{2.17}$$

We recall the definition of Fermi coordinates. Let η be the geodesic in the hyperbolic plane parameterized with the unit speed in the form

$$t \rightarrow \eta(t) \in \mathbb{H}^2 \quad t \in \mathbb{R}.$$

Then η separates $\mathbb{H}^2$ into two half-planes: a left hand side and a right hand side of η. For each $p \in \mathbb{H}^2$ we have the directed distance ρ from p to η. There exists a unique t such that the perpendicular from p to η meets η at $\eta(t)$. Now (ρ, t) is a pair of Fermi coordinates of p with respect to η. In these coordinates the metric tensor is

$$ds^2 = d\rho^2 + \cosh^2 \rho dt^2. \tag{2.18}$$

Introduce Fermi coordinates based on the bounding geodesics l_j, and use them to foliate $\mathcal{K}$. By compactness, using (2.17), we can find $z \in \mathcal{K}$ and $\delta > 0$ such that

$$\int_{B(z,\delta)} |\varphi_0|^2 > c_4 > 0, \tag{2.19}$$

and for all $j = 1, \dots, k$

$$\frac{\int_{T_j(\delta)} \varphi_0^2 \|\nabla u\|^2 d\mu}{\int_{T_j(\delta)} |\varphi_0|^2 d\mu} < \kappa_2, \tag{2.20}$$

where $T_j(\delta)$ is a tube lying in $\mathcal{K}$ and containing $B(z, \delta)$ along the perpendicular to l_j.

Each T is of the form $[-\delta,\delta]\times[\rho_{1,j},\rho_{2,j}]$ in the appropriate Fermi coordinates. Rewriting (2.20) in Fermi coordinates (2.18), and using the fact that

$$\left(\frac{\partial u_j}{\partial \rho}\right)^2+\left(\frac{\partial u_j}{\partial t}\right)^2\geq\left(\frac{\partial u_j}{\partial \rho}\right)^2,$$

by Fubini's Theorem we obtain

$$\int_{T_j(\delta)}\varphi_0^2\|\nabla u\|^2 d\mu\geq 2\delta\int_{\rho_{1,j}}^{\rho_{2,j}}\varphi_0^2\|u'(\rho)\|^2\cosh(\rho)d\rho. \tag{2.21}$$

Let L denote the maximal length of the tubes T_j. Using the fact that if $|u(\rho_1)-u_(\rho_2)|\geq C$ and $\rho_1-\rho_2\leq L$ then $\int_{\rho_1}^{\rho_2}|u'(\rho)|^2 d\rho>C^2/L$ (since

$$C^2\leq\left(\int_{\rho_1}^{\rho_2}u'(\rho)d\rho\right)^2\leq\left(\int_{\rho_1}^{\rho_2}1d\rho\right)\left(\int_{\rho_1}^{\rho_2}|u'(\rho)|^2 d\rho\right)),$$

we obtain that (2.20) implies that for all $j=1,\dots,k$ we have

$$\int_{B(z,\delta)}\varphi_0^2\|u(\gamma_j z)-u(z)\|d\mu(z)<\kappa_3\int_{B(z,\delta)}\varphi_0^2 d\mu(z). \tag{2.22}$$

On the other hand, since $F(z)=u(z)\varphi_0(z)$ and $\varphi_0(\gamma z)=\varphi_0(z)$ for all $\gamma\in\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$, (2.3) implies that there is $\varepsilon(S)>0$ independent of q, such that

$$\|u(\gamma z)-u(z)\|>\varepsilon(S)\quad\text{for some }\gamma\in S. \tag{2.23}$$

Applying mean-value theorem, we see that (2.22) implies a contradiction with (2.23) once κ is small enough depending on $\varepsilon(S)$; consequently we have proved the validity of (2.14) and the proof of Theorem 1.1 is complete.

The adaption of the preceding argument to proving the implication $1\Rightarrow 2$ of theorem 1.2 is straightforward, as is the generalization of this result to higher dimensional hyperbolic spaces: the theorem of Lax and Phillips, of which we made crucial use in the first part of the argument, holds for geometrically finite subgroups of $\mathrm{SO}(n,1)$ with Hausdorff dimension of the limit set greater than $n/2$; the second part of the argument proceeds as above by restricting to compact part of the fundamental domain and foliating it using Fermi coordinates. In particular, by combining the $\mathbb{H}^3$ analogue of Theorem 1.2 with [37] and Theorem 6.3 in [4] we obtain the following theorem which has applications to integral Apollonian packings [10, 15, 30].

Theorem 2.1. *Let Λ be a geometrically-finite subgroup of* $\mathrm{SL}_2(\mathbb{Z}[\sqrt{-1}])$ *with $\delta(\Lambda) > 1$ and such that the traces of elements of Λ generate the field $\mathbb{Q}(\sqrt{-1})$. There is $\varepsilon = \varepsilon(\Lambda) > 0$ such that*

$$\lambda_1(\Lambda(\mathcal{A})) \geq \lambda_0(\Lambda(\mathcal{A})) + \varepsilon$$

as $\mathcal{A}$ varies over squarefree ideals in $\mathbb{Z}[\sqrt{-1}]$.

3. Counting lattice points for $\delta > \frac{1}{2}$

Recall that the Poincaré upper half-plane model is the following subset of the complex plane $\mathbb{C}$:

$$\mathbb{H}^2 = \{z = x + iy \in \mathbb{C} \mid y > 0\},$$

with the hyperbolic metric

$$ds^2 = \frac{1}{y^2}(dx^2 + dy^2). \tag{3.1}$$

The distance function on $\mathbb{H}^2$ is explicitly given by

$$\rho(z, w) = \log \frac{|z - \bar{w}| + |z - w|}{|z - \bar{w}| - |z - w|}. \tag{3.2}$$

We will use the following expression:

$$\cosh \rho(z, w) = 1 + 2u(z, w), \tag{3.3}$$

where

$$u(z, w) = \frac{|z - w|^2}{4\Im z \Im w}. \tag{3.4}$$

The ring $M_2(\mathbb{R})$ of two by two real matrices is a vector space with inner product given by

$$\langle g, h \rangle = \mathrm{trace}(gh^t).$$

One easily checks that $\|g\| = \langle g, g\rangle^{\frac{1}{2}}$ is norm in $M_2(\mathbb{R})$ and that

$$\|g\|^2 = a^2 + b^2 + c^2 + d^2 \quad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{3.5}$$

By taking $z = w = i$ in (3.4) we obtain that

$$\|g\|^2 = a^2 + b^2 + c^2 + d^2 = 4u(gi, i) + 2. \tag{3.6}$$

Now the result of Lax and Phillips [17] is as follows. Let

$$N_\Lambda(T; z, w) = \#\{\gamma \in \Lambda \,:\, \rho(z, \gamma w) \leq T\}. \tag{3.7}$$

Suppose $\delta > \frac{1}{2}$ and write $\lambda_j = \delta_j(1-\delta_j)$; $\delta_0 = \delta$. Denoting the eigenfunctions corresponding to λ_j by φ_j we have

$$|N(T;z,w) - \sum_j c_j \varphi_j(z)\overline{\varphi_j(w)} e^{\delta_j T}| = O(T^{5/6} e^{T/2}). \tag{3.8}$$

Turning to congruence subgroups $\Lambda(q)$ we have, using the methods of [17], that

$$|N_{\Lambda(q)}(T;z,w) - \sum_j c_j \varphi_{j,q}(z)\overline{\varphi_{j,q}(w)} e^{\delta_{j,q} T}| = O(q^3 T^{5/6} e^{T/2}), \tag{3.9}$$

where implied constant is independent of q.

The base eigenfunction for $\Lambda(q)$, normalized to have L^2 norm one, is given by

$$\varphi_{0,q} = \frac{1}{|\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})|}\varphi_{0,1}. \tag{3.10}$$

Combining Theorem 1.1 with (3.9) and (3.2), (3.6) we obtain $\varepsilon_1 > 0$ depending on Λ such that for any $g \in \mathrm{SL}_2(q)$ we have

$$|\{\gamma \in \Lambda \,|\, \|\gamma\| \le T \,\mathrm{and}\, \gamma \equiv g \bmod q\}| = \frac{c_\Lambda T^{2\delta}}{|\mathrm{SL}_2(q)|} + O(q^3 T^{2\delta - \varepsilon_1}),$$

establishing Theorem 1.3.

4. Shifts and thermodynamic formalism

When $\delta \le \frac{1}{2}$ the L^2 spectral theory of Lax and Phillips [17] is not available and we use symbolic dynamics approach, in particular the work of Lalley [16]. In this section we review the key necessary notions and results pertaining to shifts of finite type.

A shift of finite type is defined as follows. Let A be an irreducible, aperiodic $l \times l$ matrix of zeroes and ones, called the transition matrix. Define Σ to be the space of all sequences taking values in the alphabet $\{1, 2, , \ldots, l\}$ with transitions allowed by A, that is

$$\Sigma = \{x \in \prod_{n=0}^{\infty}\{1, \ldots, l\} \,:\, A(x_n, x_{n+1}) = 1 \;\forall n\}.$$

The space Σ is compact and metrizable in the product topology. Define the forward shift $\sigma : \Sigma \to \Sigma$ by $(\sigma x)_n = x_{n+1}$ for $n \ge 0$.

Let $C(\Sigma)$ be the space of continuous, complex-valued functions on Σ. For $f \in C(\Sigma)$ and $0 < \rho < 1$ define

$$\mathrm{var}_n f = \sup\{|f(x) - f(y)| \,:\, x_j = y_j \,\mathrm{for}\, 0 \le j \le n\};$$

$$|f|_\rho = \sup_{n\geq 0} \mathrm{var}_n(f)/\rho^n;$$

$$\mathcal{F}_\rho = \{f \in C(\Sigma) : |f|_\rho < \infty\}$$

Elements of $\mathcal{F}_\rho$ are called Holder continuous functions. The space $\mathcal{F}_\rho$, when endowed with the norm $\|\cdot\|_\rho = |\cdot|_\rho + \|\cdot\|_\infty$ is a Banach space.

For $f, g \in C(\Sigma)$ define the transfer operator $\mathcal{L}_f g \in C(\Sigma)$ by

$$\mathcal{L}_f g(x) = \sum_{y:\sigma y = x} e^{f(y)} g(y).$$

For each $\rho \in (0,1)$ and $f \in \mathcal{F}_\rho$, $\mathcal{L}_f : \mathcal{F}_\rho \to \mathcal{F}_\rho$ is a continuous linear operator; if f is real-valued then $\mathcal{L}_f$ is positive.

Denoting

$$S_n f = f + f\cdot\sigma + \ldots f\cdot\sigma^{n-1},$$

we have

$$(\mathcal{L}_f^n g)(x) = \sum_{y:\sigma^n y = x} e^{S_n f(y)} g(y).$$

The following result is due to Ruelle; a proof can be found in [23] or [27].

Theorem 4.1. *Let $f \in \mathcal{F}_\rho$ be a real-valued function.*

1. *There is a simple eigenvalue $\lambda_f > 0$ of $\mathcal{L}_f : \mathcal{F}_\rho \to \mathcal{F}_\rho$ with strictly positive eigenfunction h_f.*
2. *The rest of the spectrum of $\mathcal{L}_f$ is contained in $\{z \in \mathbb{C} : |z| \leq \lambda_f - \varepsilon\}$ for some $\varepsilon > 0$.*
3. *There is a Borel probability measure ν_f on Σ such that $\mathcal{L}_f^* \nu_f = \lambda_f \nu_f$.*
4. *If h_f is normalized so that $\int h_f d\nu_f = 1$ then for every $g \in C(\Sigma)$*

$$\lim_{n\to\infty} \|\lambda_f^{-n}\mathcal{L}_f^n g - (\int g d\nu_f) h_f\|_\infty = 0$$

5. *There exist constants C_1, ε_1 such that for all $g \in \mathcal{F}_\rho$ and for all n*

$$\|\lambda_f^{-n}\mathcal{L}_f^n g - (\int g d\nu_f) h_f\|_\rho \leq C_1(1-\varepsilon_1)^n \|g\|_\rho.$$

The pressure functional is defined by

$$P(f) = \sup_\nu \int f d\nu + H_\nu(\sigma),$$

where the supremum is taken over the set of σ-invariant probability measures and $H_\nu(\sigma)$ is the measure theoretic entropy of σ with respect to ν. We have (see [23] or [27])

$$P(f) = \log \lambda_f.$$

A measure μ is called the equilibrium state or the Gibbs measure with the potential f if

$$\int f d\mu + H_\mu(\sigma) = P(f).$$

For $f \in \mathcal{F}_\rho$ Gibbs measure μ_f is the unique σ- invariant probability measure on Σ for which there exist constants $0 < C_1 \leq C_2 < \infty$ such that

$$C_1 \leq \frac{\mu_f\{y \in \Sigma : y_i = x_i, 0 \leq i < n\}}{\lambda_f^{-n} \exp\{S_n f(x)\}} \leq C_2.$$

As will become clear in the next section, the analyticity properties of the map $z \to \mathcal{L}_{zf}$, $z \in \mathbb{C}$ will play crucial role in the proof. For $f \in \mathcal{F}_\rho$ fixed, real-valued function, such that $S_m f$ is strictly positive for some m, the quantities $\mathcal{L}_{zf}$, λ_{zf}, h_{zf}, ν_{zf} will be abbreviated by $\mathcal{L}_z$, λ_z, h_z, ν_z.

5. Resolvent of transfer operator and lattice count problem

Let Λ be a Fuchsian group with no parabolic elements (this condition is automatically satisfied in the case $\delta(\Lambda) \leq \frac{1}{2}$ – see [17]), generated by k elements $g_1, \ldots, g_k \subset \mathrm{SL}_2(\mathbb{Z})$. We identify Λ with Σ_*, defined as the set of finite sequences in the alphabet $\{g_1, g_1^{-1}, \ldots, g_k, g_k^{-1}\}$ $(l = 2k)$ with admissible transitions. According to Series [33], this may be done so as to obtain a shift of finite type.

Let $w \in \mathbb{H}(= \mathbb{D})$, and suppose it is not a fixed point for any $\gamma \in \Lambda$; let d_H denote hyperbolic distance. For $x = x_1, \ldots, x_m \in \Sigma_*(= \Lambda)$ define

$$\tau(x) = d_H(0, x_1 \ldots x_m w) - d_H(0, x_2 \ldots x_m w). \tag{5.1}$$

The left shift σ on Σ corresponds to the Nielsen map (see [33]) $F : L \to L$, where L denotes the limit set of Λ.

Recalling that

$$S_n\tau = \tau + \tau \cdot \sigma + \cdots + \tau \cdot \sigma^{n-1}, \tag{5.2}$$

we have

$$S_n\tau(x) = d_H(0, x_1 \ldots x_n x_{n+1} \ldots w) - d_H(0, x_{n+1} \ldots w). \tag{5.3}$$

For $a \in \mathbb{R}$, $x \in \Sigma_*$, $\phi : \Sigma_* \to \mathbb{R}$, let

$$N(a,x) = \sum_{n=0}^{\infty} \sum_{y:\sigma^n y = x} \phi(y) 1_{\{S_n\tau(y) \le a\}}. \tag{5.4}$$

Clearly

$$N(a,x) = \sum_{y \in \Lambda, d_H(0,yxw) - d_H(0,xw) \le a} \phi(yx), \tag{5.5}$$

where in the summation y is restricted so as to make yx admissible.

In particular, for $\phi = 1$

$$N(a,x) = |\{\gamma \in \Lambda \,|\, d_H(i, \gamma xw) - d_H(i, xw) \le a\}|.$$

Returning to (5.4), one has the renewal equation

$$N(a,x) = \sum_{\sigma(x')=x} N(a - \tau(x'), x') + \phi(x) 1_{\{a \ge 0\}} \tag{5.6}$$

(cf. [16, (2.2)]).

The link with the transfer operator comes by taking the Laplace transform of (5.6). Defining for $\Re z < -C$

$$F(z,x) = \int_{-\infty}^{\infty} e^{az} N(a,x) da, \tag{5.7}$$

equation (5.6) gives the relation

$$F(z,x) = \sum_{\sigma(x')=x} e^{z\tau(x')} F(z,x') + \frac{\phi(x)}{z}. \tag{5.8}$$

Thus we have

$$(I - \mathcal{L}_z) F(z,x) = \frac{\phi(x)}{z}. \tag{5.9}$$

This leads us to the study of the resolvent $(I - \mathcal{L}_z)^{-1}$. Before stating the results of Lalley [16] and Naud [21] for $\mathcal{L}_z|_{\mathcal{F}_\rho(\Sigma)}$ (Theorem 5.1 below) we recall the following reinterpretation of the Hausdorff dimension of the limit set in terms of the pressure functional (see [23] or [27]). Because τ is eventually positive, the variational principle implies (see [23] or [27]) that the pressure functional $P(-x\tau)$ is strictly decreasing and has unique positive zero. Define δ by $P(-\delta\tau) = 0$, that is,

$$\lambda_{-\delta} = 1. \tag{5.10}$$

Theorem 5.1. (1) *There is $\varepsilon > 0$ such that for $\Re z < -\delta + \varepsilon$, $z \notin U$ (U a suitable neighborhood of $-\delta$), we have*

$$\|\mathcal{L}_z|_{\mathcal{F}_\rho(\Sigma)}\|_\rho \lesssim |\Im z|^2 e^{-\varepsilon n}. \tag{5.11}$$

(2) *For* $z \in U$ *decompose on* $\mathcal{F}_\rho(\Sigma)$

$$\mathcal{L}_z = \lambda_z(\nu_z \otimes h_z) + \mathcal{L}''_z, \tag{5.12}$$

(where $z \to \lambda_z, z \to h_z, z \to \nu_z$ *are holomorphic extensions to* U*, satisfying*

$$\mathcal{L}_z h_z = \lambda_z h_z, \quad \mathcal{L}^*_z \nu_z = \lambda_z \nu_z, \quad \int h_z d\nu_z = 1).$$

Then

$$\|(\mathcal{L}''_z)^n\|_\rho < e^{-\varepsilon n} \quad \text{for} \quad z \in U. \tag{5.13}$$

Part 1 follows from the discussion in Lalley [16, p. 25] (in the case when τ is non-lattice) and Theorem 2.3 in Naud [21] to provide (5.11) when $|\Im z|$ is large. Naud's work build crucially on the approach of Dolgopyat [8]. Note that Lalley does not give explicit estimates on $\|\mathcal{L}_z^n\|$ for $\Re z = -\delta$ and $\Im z \to \infty$ and certainly no bound of the strength of Theorem 2.3 in Naud [21].

Part 2 is Proposition 7.2. in [16]

6. Lattice count in congruence subgroups for $\delta \le \frac{1}{2}$

In this section we modify the setup discussed in the preceding two sections to the setting of congruence subgroups of Λ — the modification is analogous to the one preformed in the case $\delta > \frac{1}{2}$.

Fix modulus q such that $\pi_q(\Lambda) = \mathrm{SL}_2(q)$. Instead of considering functions on Σ (as in Lalley [16]) we consider functions on $\Sigma \times \mathrm{SL}_2(q)$.

For $f \in C(\Sigma \times \mathrm{SL}_2(q))$ define

$$\|f\|_\infty = \max_x \left(\sum_{g \in \mathrm{SL}_2(q)} |f(x,g)|^2 \right)^{\frac{1}{2}};$$

$$\mathrm{Var}_n f = \sup \left\{ \left(\sum_g |f(x,g) - f(y,g)|^2 \right)^{\frac{1}{2}} \mid x_j = y_j \text{ for } j \le n \right\};$$

$$|f|_\rho = \sup_n \frac{\mathrm{Var}_n(f)}{\rho^n}.$$

Let $\mathcal{F}_\rho = \mathcal{F}_\rho(\Sigma \times \mathrm{SL}_2(q))$ denote the space of ρ-Lipschitz continuous functions with the norm

$$\|\cdot\|_\rho = \|\cdot\|_\infty + |\cdot|_\rho.$$

Let $\tau : \Sigma_* \to \mathbb{R}$ be given by (5.1) and consider the "congruence transfer operator" $\mathcal{M}_z = \mathcal{M}_{z\tau}$ on $\mathcal{F}_\varrho(\Sigma \times \mathrm{SL}_2(q))$:

$$\mathcal{M}_z f(x,g) = \sum_{i=1}^{l} e^{z\tau(i,x)} f(i,x;g_i g), \tag{6.1}$$

where $z \in \mathbb{C}$ and the summation is restricted so as to make (i,x) admissible.

Thus our $\mathcal{M}_z$ differs from the one considered in [16] in that it acts on functions on $\Sigma \times \mathrm{SL}_2(q)$ rather than on functions on Σ: the reason behind this difference is the same as in the proof of the spectral gap when $\delta(\Lambda) > \frac{1}{2}$.

We have

$$\begin{aligned}\mathcal{M}_z^2 f(x,g) &= \sum_{i_1=1}^{l} e^{z\tau(i_1,x)} (\mathcal{M}_z f)((i_1,x), g_{i_1} g) \\ &= \sum_{i_1,i_2=1}^{l} e^{z\tau(i_1,x)+\tau(i_2 i_1,x))} f((i_2,i_1,x); g_{i_2} g_{i_1} g),\end{aligned}$$

and in general for the n-th iterate we have
(6.2)

$$\mathcal{M}_z^n f(x;g) = \sum_{i_1,\dots,i_n=1}^{l} e^{z(\tau(i_n,\dots,i_1,x)+\tau(i_{n-1},\dots,i_1,x)+\dots+\tau(i_1,x))} f((i_n \dots i_1,x); g_{i_n} \dots g_{i_1} g),$$

where again the summation is restricted to admissible words.

From Ruelle's theorem (Theorem 4.1) it follows that

$$\sum_{i_1,\dots,i_n=1}^{l} e^{\Re z(\tau(i_n,\dots,i_1,x)+\tau(i_{n-1},\dots,i_1,x)+\dots+\tau(i_1,x))} \sim \lambda_{\Re z}^n \tag{6.3}$$

for large n.

Let φ be a function on $\mathrm{SL}_2(q)$. Returning to (5.4), (5.5) we let

$$N(a,x) = \sum_{y\in\Lambda, d_H(0,yxw)-d_H(0,xw)\le a} \varphi(\pi_q(y)) \tag{6.4}$$

and $F(z,x)$ its Laplace transform defined by (5.7).

Then

$$F(z,x) = f(z,x,\pi_q(x)), \tag{6.5}$$

where $f(z,x,g)$ satisfies

$$(1-\mathcal{M}_z) f = \frac{1 \otimes \varphi}{z} \tag{6.6}$$

and $\mathcal{M}_z$ is the congruence transfer operator introduced above (note that obviously $1 \otimes \varphi$ is in $\mathcal{F}_\varrho(\Sigma \times \mathrm{SL}_2(q))$).

Our aim is to evaluate

$$N(a) = \sum_{y \in \Lambda, d_H(0,yw) - d_H(0,w) \leq a} \varphi(\pi_q(y)), \tag{6.7}$$

which gives the sum of φ on the $\bmod q$ reduction of the hyperbolic ball

$$\{y \in \Lambda \mid d_H(0,yw) - d_H(0,w) \leq a\}.$$

Our goal now is to obtain the appropriate extension of Theorem 5.1 to the setting of congruence subgroup. As is to be expected, it is at this point that expansion property will play a crucial role.

7. Expansion and L^2 flattening

Let μ be a symmetric measure on $G = \mathrm{SL}_2(p)$ (for the sake of exposition we first consider the simpler case of prime p) and consider the convolution map $T : L^2(G) \to L^2(G)$, given by $\varphi \to \mu * \varphi$. Decomposing the regular representation of G into irreducible representations if follows from the result of Frobenius [9] that each eigenvalue λ of the convolution restricted to $L^2_0(G)$ occurs with multiplicity at least $\frac{p-1}{2}$. Trace calculation yields therefore

$$|G|\|\mu\|_2^2 = \sum_{x \in G} \langle T^2 \delta_x, \delta_x \rangle \geq \frac{p-1}{2}\lambda^2.$$

Hence

$$|\lambda| \leq \sqrt{\frac{2}{p-1}}\|\mu\|_2 |G|^{\frac{1}{2}}. \tag{7.1}$$

Recall also the L^2- flattening lemma proven in [3]. Let $\mu \in \mathcal{P}(G)$ satisfy

$$\|\mu\|_\infty < p^{-\tau} \tag{7.2}$$

for some $\tau > 0$ and also

$$\mu(aG_1) < p^{-\tau} \tag{7.3}$$

for all cosets of proper subgroups G_1 of G. Given $\kappa > 0$ there is $l = l(\tau, x) \in \mathbb{Z}_+$ such that

$$\|\mu^{(l)}\|_2 < |G|^{-1/2+\kappa}. \tag{7.4}$$

Denote $\mu'(x) = \mu(x^{-1})$. Since $\mu * \mu'$ also satisfies (7.2), (7.3) we have by (7.4) that

$$\|(\mu' * \mu)^{(l)}\|_2 < |G|^{-1/2+\kappa}. \tag{7.5}$$

Consider the convolution operator $T\varphi = \mu' * \mu * \varphi$ and let λ be an eigenvalue of T on $L^2_0(G)$. Hence λ^l is an eigenvalue of T^l on $L^2_0(G)$ and applying (7.1) with μ replaced by $(\mu' * \mu)^{(l)}$ implies

$$|\lambda|^l \leq \sqrt{\frac{2}{p-1}} \|(\mu' * \mu)^{(l)}\|_2 |G|^{\frac{1}{2}} < \sqrt{\frac{2}{p-1}} |G|^{\kappa} < p^{-\frac{1}{4}}$$

if we take $\kappa < \frac{1}{4}$. Consequently

$$|\lambda| < p^{-\frac{1}{4l}}. \tag{7.6}$$

This means that if $\varphi \in L^2_0(G)$, then

$$\|\mu' * \mu * \varphi\|_2 \leq p^{-\frac{1}{4l}} \|\varphi\|_2$$

and hence

$$\|\mu * \varphi\|_2 \leq p^{-\frac{1}{8l}} \|\varphi\|_2. \tag{7.7}$$

We proved that if $\mu \in \mathcal{P}(G)$ satisfies (7.2), (7.3) for some $\tau > 0$, then

$$\|\mu * \varphi\|_2 \leq p^{-\tau'} \|\varphi\|_2, \quad \varphi \in L^2_0(G) \tag{7.8}$$

for some $\tau' > 0$. Therefore if $\mu \in \mathcal{M}_+(G)$ satisfies

$$\|\mu\|_\infty < p^{-\tau} \|\mu\|_1 \tag{7.9}$$

and

$$|\mu|(aG_1) < p^{-\tau} \|\mu\|_1 \tag{7.10}$$

for cosets of proper subgroups G_1 then

$$\|\mu * \varphi\|_2 \leq p^{-\tau'} \|\mu\| \|\varphi\|_2, \quad \varphi \in L^2_0(G). \tag{7.11}$$

More generally, let $\mu \in \mathcal{M}(G)$ and decompose $\mu = \mu_+ + \mu_-$. Estimate

$$\|\mu * \varphi\|_2 \leq \|\mu_+ * \varphi\|_2 + \|\mu_- * \varphi\|_2.$$

Assume μ satisfies (7.9) and (7.10). If $\|\mu_+\| > p^{-\frac{\tau}{2}} \|\mu\|$, we have

$$\|\mu_+\|_\infty \leq \|\mu\|_\infty < p^{-\frac{\tau}{2}} \|\mu_+\|_1 \quad \text{and} \quad \mu_+(aG_1) < p^{-\frac{\tau}{2}} \|\mu_+\|_1$$

and (7.11) implies that for $\varphi \in L^2_0(G)$ we have

$$\|\mu_+ * \varphi\|_2 \leq p^{-\tau'} \|\mu_+\| \|\varphi\|_2 \leq p^{-\tau'} \|\mu\| \|\varphi\|_2.$$

If $\|\mu_+\| \leq p^{-\frac{\tau}{2}} \|\mu\|$, then obviously

$$\|\mu_+ * \varphi\|_2 \leq \|\mu_+\| \|\varphi\|_2 \leq p^{-\frac{\tau}{2}} \|\mu\| \|\varphi\|_2.$$

Hence (7.11) holds.

The same considerations apply for $\mu \in \mathcal{M}_{\mathbb{C}}(G)$; consequently we obtain the following result.

Lemma 1. *Given $\kappa > 0$, there is $\kappa' > 0$ such that if $\mu \in \mathcal{M}_{\mathbb{C}}(G)$ satisfies*

$$\|\mu\|_\infty < p^{-\kappa}\|\mu\|_1 \tag{7.12}$$

and

$$|\mu|(aG_1) < p^{-\kappa}\|\mu\|_1 \tag{7.13}$$

for cosets of proper subgroups G_1 of G, then

$$\|\mu * \varphi\|_2 \le Cp^{-\kappa'}\|\mu\|_1\|\varphi\|_2 \quad \text{for} \quad \varphi \in L^2_0(G). \tag{7.14}$$

Here p is assumed to be sufficiently large.

We have a similar result for $G = \mathrm{SL}_2(q)$ with q square-free (see [4] and [37]). We make the following decomposition of the space $L^2(\mathrm{SL}_2(q))$. For $q_1|q$, define E_{q_1} as the subspace of functions defined $\mod q_1$ and orthogonal to all functions defined $\mod q_2$ for some $q_2|q_1$, $q_2 \neq q_1$. Hence

$$L^2(\mathrm{SL}_2(q)) = \mathbb{R} \oplus \bigoplus_{q_1|q} E_{q_1}, \tag{7.15}$$

which is, in fact, the generalized Fourier-Walsh decomposition corresponding to the product representation

$$\mathrm{SL}_2(q) \cong \prod_{p|q} \mathrm{SL}_2(p).$$

Let P_1 (respectively P_{q_1}) be the projection operator on the constant functions (respectively E_{q_1}).

Lemma 2. *Let q be square free and $G = \mathrm{SL}_2(q)$. For $\mu \in \mathcal{M}_{\mathbb{C}}(G)$ and $q_1|q$ define $|||\pi_{q_1}(\mu)|||_\infty$ to be the maximum weight of $|\mu|$ over cosets of subgroups of $\mathrm{SL}_2(q_1)$ that have proper projection in each divisor of q_1. Given $\kappa > 0$, there is $\kappa' > 0$ such that if μ satisfies for all $q_1|q$*

$$|||\pi_{q_1}(\mu)|||_\infty < q_1^{-\kappa}\|\mu\|_1 \tag{7.16}$$

then

$$\|\mu * \varphi\|_2 \le Cq^{-\kappa'}\|\mu\|_1\|\varphi\|_2 \quad \text{for} \quad \varphi \in E_q. \tag{7.17}$$

8. Bounds for congruence transfer operator

Our goal is to obtain a bound for powers of transfer operator $\|\mathcal{M}_z^m\|_\rho$ for the family of congruence subgroups. Recall that $\mathcal{M}_z$ acts on functions on $\Sigma \times G$, so in order to apply Lemma 2 we need to decouple the variables. Returning to (6.2), fix $m \le r < n$ such that $m = n - r \sim \log q$ to be specified. Write

$$\mathcal{M}_z^n f(x, g) =$$

(8.1)

$$\sum_{i_1,\dots,i_n=1}^{l} e^{z(\tau(i_n,\dots,i_1,x)+\tau(i_{n-1},\dots,i_1,x)+\dots+\tau(i_1,x))} f((i_n\dots i_{n-r+1},0); g_{i_n}\dots g_{i_1}g)$$

(8.2) $$+O(\lambda^n_{\Re z}|f|_\rho \rho^r),$$

where the error term refers to the $L^\infty_{l^2(G)}(\Sigma)$-norm.

Fix then the matrices $i_n,\dots,i_{n-r+1}$ and consider the function φ on G defined by

$$\varphi(g) = f((i_n,\dots,i_{n-r+1},0), g_{i_n}\dots g_{i_{n-r+1}}g).$$

We assume $f(x,\cdot)\in E_q$ for each x; hence $\varphi\in E_q$.

Our aim is to apply Lemma 2 with

$$\mu = \sum_{i_1,\dots,i_{n-r}} e^{z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))}\delta_{g_{i_{n-r}}\cdots g_{i_1}}. \tag{8.3}$$

Thus by (6.3) we have

$$\|\mu\| \lesssim \lambda^m_{\Re z} e^{\Re z(\tau(i_n,\dots,i_{n-r+1},0)+\dots+\tau(i_{n-r+1},0))} e^{\Re z\frac{|\tau|_\rho}{1-\rho}}, \tag{8.4}$$

where we used the inequality

$$|\tau(i_n,\dots,i_{n-r+1},i_{n-r},\dots,i_1,x)-\tau(i_n,\dots,i_{n-r+1},\dots)| \le \tau_\rho|\rho|^r. \tag{8.5}$$

We bound $\|\mu\|_\infty$, which amounts to estimating

$$\frac{1}{(8.4)}\sum_{g_{i_m}\cdots g_{i_1}=g} e^{\Re z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))}, \tag{8.6}$$

where g is fixed. (We use here the fact that the relation $g_{i_m}\dots g_{i_1} = g \bmod q$ is equivalent to $g_{i_m}\dots g_{i_1}=g$ because of the restriction on m). Also because of the index restriction on the transition matrix A, the condition $g_{i_m}\dots g_{i_1}=g$ specifies $(i_m,\dots,i_1)\in\Sigma$ so that estimating (8.6) amounts to bounding

$$\begin{aligned}&\frac{1}{(8.4)} e^{\Re z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))}\\ &\sim \lambda^{-m} e^{\Re z(\tau(i_m,\dots,i_1,x)+\dots+\tau(i_1,x))}\end{aligned} \tag{8.7}$$

for fixed $(i_n,\dots i_1,x)\in\Sigma$. Thus

$$(8.7) = \lambda^{-m}\mathcal{M}^m_{\Re z}\delta_{(i_m,\dots,i_1,x)}(x) \le \lambda^{-m}\|\mathcal{M}^m_{\Re z}\phi\|_\infty$$

with ϕ any function on Σ satisfying $\phi(i_m,\dots,i_1,x)=1$.

By Ruelle's Theorem (Thm. 4.1),

$$\|\lambda^{-m}\mathcal{M}^m\phi - (\int\phi d\nu)h\|_\rho \lesssim (1-\varepsilon_1)^m\|\phi\|_\rho,$$

implying

$$\|\lambda^{-m}\mathcal{M}^m\phi\|_\infty \le c\left(\int \phi d\nu + (1-\varepsilon_1)^m\|\phi\|_\rho\right). \tag{8.8}$$

We may now choose ϕ suitably, so as to obtain an estimate

$$\lambda^{-m}\|\mathcal{M}^m\phi\|_\infty \lesssim e^{-cm}. \tag{8.9}$$

Hence,

$$(8.6), (8.7) \lesssim q^{-\kappa}. \tag{8.10}$$

More generally, we also need to evaluate $|||\pi_{q_1}(\mu)|||_\infty$ for $q_1|q$. It turns out that the issue reduces to the previous one, using the following observation (cf [3]). Let $H < G$ and $\pi_p(H) < \mathrm{SL}_2(p)$ proper for each $p|q_1$. Then we can assume the second commutator of $\pi_p(H)$ to be trivial if $p|q_1$ and hence the second commutator of H to be trivial $(\bmod\, q_1)$. Take $m_1 < m$, $m_1 \sim \log q_1$ so as to ensure that words of length $2m_1$ have norm less than q_1. Using properties of the free group (see [3]), it follows from the preceding that the number of $(i_{m_1}, \dots, i_1) \in \Sigma$ such that $g_{i_{m_1}} \dots g_{i_1} \in aH$ is bounded by $O(m_1^C)$ for some constant C. Hence we may invoke the estimate on (8.7) with m replaced by m_1 to obtain also

$$|||\pi_{q_1}(\mu)|||_\infty < q_1^{-\kappa} \cdot (8.4).$$

Applying Lemma 2, it follows that

$$\|\mu{*}\varphi\|_2 \le q^{-\kappa'}\|\mu\|_1\|\varphi\|_2 \le q^{-\kappa'}\lambda^m_{\Re z} e^{\Re z(\tau(i_n,\dots,i_{n-r+1},0)+\dots+\tau(i_{n-r+1},0))} e^{\Re z \frac{|\tau|_\rho}{1-\rho}}\|f\|_\infty$$

or

$$\Big\|\sum_{i_1,\dots,i_{n-r}} e^{z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))} f(i_n,\dots,i_{n-r+1},0; g_{i_n}\dots g_{i_1}g)\Big\|_{l^2(G)} \le q^{-\kappa'}\lambda^m_{\Re z} e^{\Re z(\tau(i_n,\dots,i_{n-r+1},0)+\dots+\tau(i_{n-r+1},0))} e^{\Re z\frac{|\tau|_\rho}{1-\rho}}\|f\|_\infty. \tag{8.11}$$

Summing (8.11) over $i_n, \dots, i_{n-r+1}$ implies by (6.3) again

$$\Big\|\sum_{i_1,\dots,i_n=1}^{l} e^{z(\tau(i_n,\dots,i_1,x)+\tau(i_{n-1},\dots,i_1,x)+\dots+\tau(i_1,x))} f((i_n\dots i_{n-r+1},0); g_{i_n}\dots g_{i_1}g)\Big\|_{l^2(G)} \lesssim q^{-\kappa'}\lambda^{m+r}\|f\|_\infty. \tag{8.12}$$

Therefore it follows that if $n > \log q$

$$\begin{aligned}\|\mathcal{M}_z^n f\|_{L^\infty_{l^2(G)}(\Sigma)} &\leq \lambda^n_{\Re z}(q^{-\kappa'}\|f\|_\infty + \rho^r|f|_\rho) \\ &\leq \lambda^n_{\Re z} q^{-\kappa'}(\|f\|_\infty + \rho^{\frac{n}{2}}|f|_\rho)\end{aligned} \tag{8.13}$$

for

$$f \in \mathcal{F}'_\rho = \mathcal{F}_\rho \cap \mathcal{C}_{E_q}(\Sigma). \tag{8.14}$$

Note that in (8.13) there is no restriction on $\Im z$.

We also need to estimate $|\mathcal{M}_z^n f|_\rho$.

Let $x, y \in \Sigma$ be such that $x_i = y_i$ for $0 \leq i < l$. Estimate

$$|\mathcal{M}_z^n f(x,g) - \mathcal{M}_z^n f(y,g)| \leq$$

(8.15)
$$\sum_{i_1,\dots,i_n} e^{\Re z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))}|f(i_n,\dots,i_1 x; g_{i_n}\dots g_{i_1}g) - f(i_n,\dots,i_1 y; g_{i_n}\dots g_{i_1}g)|$$

(8.16)
$$+|\sum_{i_1,\dots,i_n}\left(e^{z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))} - e^{z(\tau(i_n,\dots,i_1,y)+\dots+\tau(i_1,y))}\right) f(i_n,\dots,i_1 y; g_{i_n}\dots g_{i_1}g)|.$$

Clearly for the first term we have

$$(8.15) \lesssim \lambda^n_{\Re z}|f|_\rho \rho^{n+l}. \tag{8.17}$$

To estimate (8.16) we repeat the argument leading to (8.13). Thus we bound (8.16) as follows

(8.18)
$$|\sum_{i_1,\dots,i_n}\left(e^{z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))} - e^{z(\tau(i_n,\dots,i_1,y)+\dots+\tau(i_1,y))}\right) f(i_n,\dots,i_{n-r+1},0; g_{i_n}\dots g_{i_1}g)|$$

(8.19)
$$+\rho^r|f|_\rho \sum_{i_1,\dots,i_n} |\left(e^{z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))} - e^{z(\tau(i_n,\dots,i_1,y)+\dots+\tau(i_1,y))}\right)|.$$

Estimate

(8.20)
$$\sum_{i_1,\dots,i_n} |\left(e^{z(\tau(i_n,\dots,i_1,x)+\dots+\tau(i_1,x))} - e^{z(\tau(i_n,\dots,i_1,y)+\dots+\tau(i_1,y))}\right)|$$

$$\leq \sum_{i_1,\dots,i_n} e^{\Re z(\tau(i_n,\dots,i_1,x)+\tau(i_1,x))}|1 - e^{z(\tau(i_n,\dots,i_1,y)+\dots+\tau(i_1,y)-\tau(i_n,\dots,i_1,x)-\dots-\tau(i_1,x))}|$$

$$\lesssim \lambda^n_{\Re z}(1+|\Im z|)|\tau|_\rho(\rho^{n+l}+\dots+\rho^{1+l}) < \lambda^n_{\Re z}\frac{1+|\Im z|}{1-\rho}|\tau|_\rho\rho^l;$$

therefore

$$(8.19) < \lambda^n \frac{1+|\Im z|}{1-\rho} |\tau|_\rho \rho^{l+r} |f|_\rho. \tag{8.21}$$

To bound (8.18) we apply again the convolution estimate on G from section 7. Consider the measure

$$\nu = \sum_{i_1,\dots,i_{n-r}} \left(e^{z(\tau(i_n,\dots,i_1,x)+\cdots+\tau(i_1,x))} - e^{z(\tau(i_n,\dots,i_1,y)+\cdots+\tau(i_1,y))} \right) \delta_{g_{i_{n-r}}\cdots g_{i_1}} \tag{8.22}$$

with $i_n, \dots i_{n-r+1}$ fixed.

Repeating (8.20) gives (with $m = n - r$)

$$\|\nu\| \lesssim \lambda^m (1+|\Im z|)|\tau|_\rho \frac{\rho^l}{1-\rho} e^{\Re z(\tau(i_n,\dots,i_{n-r+1},0)+\cdots+\tau(i_{n-r+1},0))}. \tag{8.23}$$

Also, as above, we have

$$\begin{aligned} &\frac{1}{(8.23)}\|\nu\|_\infty = \frac{1}{(8.23)} |e^{z(\tau(i_n,\dots,i_1,x)+\cdots+\tau(i_1,x))} - e^{z(\tau(i_n,\dots,i_1,y)+\cdots+\tau(i_1,y))}| \\ &\le \lambda^{-m} e^{\Re z(\tau(i_n,\dots,i_1,x)+\cdots+\tau(i_1,x))} \overset{(8.10)}{<} q^{-\kappa}. \end{aligned} \tag{8.24}$$

and

$$\frac{1}{(8.23)} |||\pi_{q_1}(\nu)|||_\infty < q_1^{-\kappa} \quad \text{for} \quad q_1 | q. \tag{8.25}$$

Therefore, by the results from section 7, we obtain (with φ defined as in section 7) that

$$\begin{aligned} &\|\nu * \varphi\|_{l^2(G)} \le q^{-\kappa'} (8.23) \|f\|_\infty \lesssim \\ &q^{-\kappa'} \lambda^m (1+|\Im z|)|\tau|_\rho \frac{\rho^l}{1-\rho} e^{\Re z(\tau(i_n,\dots,i_{n-r+1},0)+\cdots+\tau(i_{n-r+1},0))} \|f\|_\infty. \end{aligned} \tag{8.26}$$

Summation over $i_n, \dots, i_{n-r+1}$ gives then

$$\|(8.18)\|_{l^2(G)} \lesssim q^{-\kappa'} \lambda^n (1+|\Im z|)|\tau|_\rho \rho^l \|f\|_\infty. \tag{8.27}$$

From (8.17), (8.21), (8.27), it follows that

$$\|\mathcal{M}_z^n f(x,\cdot) - \mathcal{M}_z^n f(y,\cdot)\|_{l^2(G)} \lesssim \rho^l \lambda_{\Re z}^n \left(\rho^n |f|_\rho + \rho^r (1+|\Im z|)|f|_\rho + q^{-\kappa'} (1+|\Im z|)\|f\|_\infty \right). \tag{8.28}$$

Therefore, if $n > \log q$ we have

$$|\mathcal{M}_z^n f|_\rho \le C \lambda_{\Re z}^n q^{-\kappa'} \left(\|f\|_\infty + \rho^{n/2} |f|_\rho \right) (1+|\Im z|). \tag{8.29}$$

Take n such that

$$n \sim \log q + C \log(1 + |\Im z|) \tag{8.30}$$

for a suitable constant C. It follows from (8.13), (8.29) that

$$\|\mathcal{M}_z^n f\|_\infty + \rho^{\frac{n}{2}} |\mathcal{M}_z^n f|_\rho < \lambda_{\Re z}^n q^{-\kappa'} (\|f\|_\infty + \rho^{\frac{n}{2}} |f|_\rho). \tag{8.31}$$

Iterating (8.31) shows that if $f \in \mathcal{F}'_\rho$, then for all $m \in \mathbb{Z}_+$

$$\|\mathcal{M}_z^{mn} f\|_\infty + \rho^{\frac{n}{2}} |\mathcal{M}_z^{mn} f|_\rho \leq \lambda_{\Re z}^{mn} q^{-m\kappa'} \|f\|_\rho, \tag{8.32}$$

and hence

$$\|\mathcal{M}_z^{mn} f\|_\rho < \lambda_{\Re z}^{mn} q^{-m\kappa'} q(1 + |\Im z|) \|f\|_\rho, \tag{8.33}$$

where n is given by (8.30). Thus for $m \geq 1$

$$\|\mathcal{M}_z^m|_{\mathcal{F}'_\rho}\|_\rho < \lambda_{\Re z}^m q^{-\frac{m}{n}\kappa'} q(1 + |\Im z|). \tag{8.34}$$

We distinguish two cases: $\log(1 + |\Im z|) \lesssim \log q$ and $\log q \ll \log(1 + |\Im z|)$. The conclusion is the following:

Lemma 3. *Notation being as above, there is $\varepsilon > 0$ such that*

$$\|\mathcal{M}_z^m|_{\mathcal{F}'_\rho}\|_\rho < q^C e^{-\varepsilon m} \lambda_{\Re z}^m \text{ if } |\Im z| \leq q \tag{8.35}$$

and

$$\|\mathcal{M}_z^m|_{\mathcal{F}'_\rho}\|_\rho < |\Im z|^C e^{-\varepsilon \frac{\log q}{\log |\Im z|} m} \lambda_{\Re z}^m \text{ if } |\Im z| > q. \tag{8.36}$$

9. Resolvent of congruence transfer operator

We now use Lemma 3 to estimate the resolvent $(I - \mathcal{M}_z)^{-1}$ on $\mathcal{F}'_\rho$. By (6.5), (6.6) this will provide us bounds on $F(z, x)$, assuming $\varphi \in E_q$.

Take $\Re z < -\delta + \varepsilon_1$ such that

$$\lambda_{-\delta+\varepsilon_1} < e^{\frac{\varepsilon}{2}} \tag{9.1}$$

with $\varepsilon > 0$ from (8.35). If $|\Im z| < q$, we obtain

$$\|(I - \mathcal{M}_z)^{-1}|_{\mathcal{F}'_\rho}\| < q^C \sum e^{-\varepsilon m} \lambda_{\Re z}^m \lesssim \frac{1}{\varepsilon} q^C. \tag{9.2}$$

If $|\Im z| \geq q$, we impose the restriction

$$\Re z < -\delta + \varepsilon_2 \frac{\log q}{\log |\Im z|} \tag{9.3}$$

with $\varepsilon_2 > 0$ small enough to ensure that

$$\lambda_{\Re z} e^{-\varepsilon \frac{\log q}{\log |\Im z|}} < e^{-\frac{\varepsilon}{2} \frac{\log q}{\log |\Im z|}}.$$

Under this restriction on z, we obtain from (8.36) that

$$\|(I - \mathcal{M}_z)^{-1}|_{\mathcal{F}'_\rho}\| < |\Im z|^C. \tag{9.4}$$

In summary, we proved the following

Theorem 9.1. *The resolvent* $(I - \mathcal{M}_z)^{-1}|_{\mathcal{F}'_\rho}$ *is holomorphic on the complex region* $D(q)$ *given by*

$$\Re z < -\delta + \varepsilon_2 \min\left(1, \frac{\log q}{\log(|\Im z| + 1)}\right) \tag{9.5}$$

(with ε_2 *independent of* q*) and satisfies the estimate*

$$\|(I - \mathcal{M}_z)^{-1}|_{\mathcal{F}'_\rho}\| < (q + |\Im z|)^C. \tag{9.6}$$

Returning to (6.5), (6.6), it follows that for $\varphi \in E_q$ the Laplace transform $F(z, x)$ of $N(a, x)$ is bounded by

$$|F(z, x)| \lesssim \frac{(q + |\Im z|)^C}{|z|} \|\varphi\|_2 \tag{9.7}$$

for z satisfying (9.5).

To extract information about $N(a)$ we apply Fourier inversion to (5.7), following the argument in [16, p. 31] (but with a different class of functions k).

Specify some smooth and compactly supported bump function k on $\mathbb{R}$. We get from (5.7)

$$\int_{-\infty}^{\infty} k(t) e^{-\delta t} N(a + t) dt = e^{\delta a} \int e^{-ia\theta} \hat{k}(-i\theta) F(-\delta + i\theta) d\theta, \tag{9.8}$$

where

$$\hat{k}(z) = \int e^{zt} k(t) dt$$

is an entire function.

Note that $|\hat{k}(i\theta)|$ is rapidly decaying since k is smooth.

In fact, proceeding more precisely, fix a small parameter $\gamma > 0$ (the localization of k) and consider functions

$$k_\gamma(t) = \frac{1}{\gamma} K\left(\frac{t}{\gamma}\right), \tag{9.9}$$

where K is a fixed smooth bump function such that

$$\int K = 1, \tag{9.10}$$

$$\text{supp} K \subset [-\frac{1}{2}, \frac{1}{2}], \tag{9.11}$$

$$|\hat{K}(\lambda)| \lesssim e^{-|\lambda|^{\frac{1}{2}}} \quad \text{for} \quad |\lambda| \to \infty. \tag{9.12}$$

Hence

$$|\hat{k}(z)| \lesssim e^{-|\gamma|z||^{\frac{1}{2}}} \quad \text{for} \quad |\Re z| < O(1). \tag{9.13}$$

Returning to (9.8), modify the line of integration $\Re z = 0$ to the curve

$$z(\theta) = w(\theta) + i\theta,$$

where

$$w(\tau) = \frac{1}{2}\varepsilon_2 \min\left(1, \frac{\log q}{\log(1+|\theta|)}\right), \tag{9.14}$$

so as to remain in the analyticity region given by Theorem 9.1.

We obtain

$$e^{\delta a}\int_{-\infty}^{\infty} e^{-az(\theta)}\hat{k}(z(\theta))F(-\delta+z(\theta))d\theta = e^{\delta a}\int_{-\infty}^{\infty} e^{-aw(\theta)-ia\theta}\hat{k}(w(\theta)+i\theta)F(-\delta+w(\theta)+i\theta)d\theta,$$

which is bounded by

$$\|\varphi\|_2 e^{\delta a}\int_{-\infty}^{\infty} e^{-aw(\theta)}e^{-(\gamma|\theta|)^{\frac{1}{2}}}(q+|\theta|)^C d\theta, \tag{9.15}$$

applying (9.13) and (9.7).

From the definition of $w(\theta)$ it is clear that

$$(9.8), (9.15) < e^{\delta a}q^C\gamma^{-C}\exp\left(-a\varepsilon_3 \min\left(1, \frac{\log q}{\log\frac{a}{\gamma}}\right)\right)\|\varphi\|_2. \tag{9.16}$$

This proves (replacing $k(t)$ by $e^{\delta t}k(t)$)

Proposition 1. *Let $\varphi \in E_q$ and $N(a)$ given by (6.7). Then*

$$\left|\int_{-\gamma/2}^{\gamma/2} k_\gamma(t)N(a+t)dt\right| < q^C\gamma^{-C}\exp\left(-a\varepsilon_3 \min\left(1, \frac{\log q}{\log\frac{a}{\gamma}}\right)\right)e^{\delta a}\|\varphi\|_2. \tag{9.17}$$

10. Bound for the error term

Next consider the case where in (5.4), $\phi = 1$ (the constant function).

Here we consider simply the action of $\mathcal{L}_z$ on $\mathcal{F}_\rho(\Sigma)$ exactly as in [16], but we use the stronger estimates on $(I-\mathcal{L}_z)^{-1}$ following from (5.11), given by [21].

If $\Re z < -\delta + \varepsilon_4$ (with ε_4 small enough) and $z \notin U$ (some complex neighborhood of $-\delta$), (5.11) implies that

$$\|(I-\mathcal{L}_z)^{-1}\| \lesssim |\Im z|^2. \tag{10.1}$$

For $s \in U$, apply (5.12). Thus

$$\mathcal{L}_z^n = \lambda_z^n(\nu_z \otimes h_z) + (\mathcal{L}_z'')^n,$$

where $\|(\mathcal{L}_z'')^n\| < e^{\varepsilon n}$ by (5.13).

Hence for $z \in U$

$$(I - \mathcal{L}_z)^{-1} = \frac{1}{1-\lambda_z}(\nu_z \otimes h_z) + (I - \mathcal{L}_z'')^{-1} \tag{10.2}$$

with $(I - \mathcal{L}_z'')^{-1}$ holomorphic (this is Proposition 7.2 in [16]).

Combining (10.1), (10.2) we get

Proposition 2. *Consider $\mathcal{L}_z$ acting on $\mathcal{F}_\rho(\Sigma)$. Then for $\Re z < -\delta + \varepsilon_5$*

$$(I - \mathcal{L}_z)^{-1} - \frac{1}{1-\lambda_z}(\nu_z \otimes h_z) \tag{10.3}$$

is holomorphic and bounded by $C(|\Im z|^2 + 1)$.

Let $\mu_z = \nu_z \otimes h_z$. The function $\frac{1}{1-\lambda_z}$ has a pole at $z = -\delta$ with residue

$$-\frac{1}{(\frac{d}{dz}\lambda_z)|_{z=-\delta}} = \frac{1}{\int \tau d\mu_{-\delta}}.$$

Consequently

$$(I - \mathcal{L}_z)^{-1} - \frac{\nu_{-\delta} \otimes h_{-\delta}}{\int \tau d\mu_{-\delta}} \frac{1}{z+\delta} \tag{10.4}$$

is analytic for $\Re z < -\delta + \varepsilon_5$.

Letting

$$N(a) = \sum_{\substack{y \in \Lambda \\ d_H(0,yw) - d_H(0,w) \le a}} 1 \tag{10.5}$$

and

$$F(z) = \int e^{az} N(a) da, \tag{10.6}$$

it follows from (5.9) , (10.4) that

$$F(z) = \frac{1}{z}(I - \mathcal{L}_z)^{-1} 1 = \frac{h_{-\delta}(\xi \equiv w)}{\int \tau d\mu_{-\delta}} \frac{1}{z+\delta} + G(z), \tag{10.7}$$

where $G(z)$ is analytic on $\Re z < -\delta + \varepsilon_5$ and bounded by $C(|\Im z|^2 + 1)$.

As in section 9 we have

$$\begin{aligned} &\int k_\gamma(t) e^{-\delta t} N(a+t) dt = e^{\delta a} \int \hat{k}_\gamma(i\theta) F(-\delta + i\theta) e^{-ia\theta} d\theta \\ &= e^{\delta a} \left(C_0 \int_{PV} e^{-ia\theta} \hat{k}_\gamma(i\theta) \frac{1}{i\theta} d\theta + \int e^{-ia\theta} \hat{k}_\gamma(i\theta) G(-\delta + i\theta) \right), \end{aligned} \tag{10.8}$$

where

$$C_0 = \frac{h_{-\delta}(\xi \equiv w)}{\int \tau d\nu_{-\delta}}. \tag{10.9}$$

The second term in (10.8) is estimated by moving the line of integration $\Re z = 0$ to $\Re z = \frac{1}{2}\varepsilon_5$. We obtain by (9.13) and the assumption on G:

(10.10)
$$|\int e^{-ia\theta}\hat{k}_\gamma(i\theta)G(-\delta+i\theta)| \lesssim e^{-\frac{\varepsilon_5}{2}a}\int(1+\theta^2)e^{-(\gamma|\theta|)^{\frac{1}{2}}}d\theta < c\gamma^{-3}e^{-\frac{\varepsilon_5}{2}a}.$$

Also
$$\int_{PV} e^{-ia\theta}\hat{k}_\gamma(i\theta)\frac{1}{i\theta}d\theta = \int_0^\infty k_\gamma(t+a)dt \overset{(9.10)}{=} 1.$$

Therefore we obtain

Proposition 3. *Let $N(a)$ be given by (10.5). Then*

$$\int k_\gamma(t)N(a+t)dt = C_0 e^{\delta a} + o\left(\gamma^{-3}e^{(\delta-\varepsilon_6)a}\right) \tag{10.11}$$

for some $\varepsilon_6 > 0$. Here C_0 is a fixed constant.

11. Proof of Theorem 1.4

Let φ be a function on $\mathrm{SL}_2(q)$ and let $N(a,x)$ denote the counting function given as above by

$$N(a,x) = \sum_{y\in\Lambda, d_H(0,yxw)-d_H(0,xw)\le a}\varphi(\pi_q(y)).$$

What we proved in Theorem 9.1 is that for $\varphi \in E_q$ the Laplace transform of $N(a,x)$ in a (given by (5.7)) is holomorphic on $D(q)$ given by

$$D(q) = \left\{z\ :\ \Re z < -\delta + \varepsilon_2 \min\left(1, \frac{\log q}{\log(|\Im z|+1)}\right)\right\} \tag{11.1}$$

with ε_2 independent of q. Let us denote by $\mathcal{L}_z(q)$ the dynamical transfer operator on the congruence subgroup $\Lambda(q)$. Thus $\det(1-\mathcal{L}_z(q))$ is the dynamical (Ruelle's) zeta-function associated with the congruence subgroup $\Lambda(q)$. Using (5.9) we have that the Laplace transform of $N(a,x)$ is also obtained as the inverse of $(1-\mathcal{L}_z(q))$. Now considering the action of $\mathcal{L}_z(q)$ on $\mathcal{F}_\rho(\Sigma(q))$, recalling the decomposition of $L^2(\mathrm{SL}_2(q))$ given by (7.15), and applying theorem 9.1 to E_{q_1} for all $q_1|q$, and Proposition 2 to the constant function, we obtain that $1-\mathcal{L}_z(q)$ has holomorphic inverse (apart from $z=-\delta$) on

$$D = D(1)\cap\bigcap_{q_1|q}D(q_1),$$

where
$$D(1) = \{z\ :\ \Re z < -\delta + \varepsilon_5\}$$

by Proposition 2. Consequently D is given by

$$D = \left\{ z \,:\, \Re z < -\delta + \varepsilon_6 \min\left(1, \frac{1}{\log(|\Im z| + 1)}\right)\right\} \tag{11.2}$$

for some ε_6 independent of q, implying that dynamical zeta function $\det(1 - \mathcal{L}_z(q))$ has no zeros on D (apart from simple zero at $-\delta$).

Theorem 1.4 now follows from the equality of the dynamical zeta-function and Selberg zeta function (Theorem 15.8 in [2]), and the correspondence between the zeros of Selberg's zeta function and resonances (see [26] and Chapter 10 in [2]).

12. Proof of Theorem 1.5

Propositions 2 and 3 are our basic estimates used in the proof of Theorem 1.5. Note that what comes out Proposition 1 will only play the role of error terms.

Fix a modulus q, $(q, q_0) = 1$ (with q_0 given by the strong approximation property) and q square-free.

For some element $\xi \in \mathrm{SL}_2(q)$ we need to evaluate

$$N(a; q, \xi) = |\{y \in \Lambda; \pi_q(y) = \xi \quad \text{and} \quad d_H(0, yw) \leq a\}| \tag{12.1}$$

(replace a by $a + d_H(0, w)$ in (6.7)).

Recall the decomposition of the space $L^2(\mathrm{SL}_2(q))$ in (7.15). Writing

$$1_{g=\xi} = \frac{1}{|\mathrm{SL}_2(q)|} + \sum_{\substack{q_1|q \\ q_1 \neq 1}} P_{q_1}(1_{g=\xi}), \tag{12.2}$$

we get

$$N(a; q, \xi) = \frac{1}{|\mathrm{SL}_2(q)|} \sum_{\substack{y \in \Lambda \\ d_H(0, yw) \leq a}} 1 \tag{12.3}$$

$$+ \sum_{\substack{q_1|q \\ q_1 \neq 1}} \sum_{\substack{y \in \Lambda \\ d_H(0, yw) \leq a}} \varphi_{q_1}(\pi_{q_1}(y)) \tag{12.4}$$

with

$$\varphi_{q_1} = P_{q_1}(1_{g=\xi}) \in E_{q_1}.$$

Thus

$$\|\varphi_{q_1}\|_2 < \frac{|\mathrm{SL}_2(q_1)|^{\frac{1}{2}}}{|\mathrm{SL}_2(q)|}. \tag{12.5}$$

We use Proposition 3 to evaluate the right-hand side of (12.3) and Proposition 1 to bound terms (12.4). Hence, fixing some $\gamma > 0$

$$\int k_\gamma(t) N(a+t; q, \xi) dt = \frac{C_1}{|\mathrm{SL}_2(q)|} e^{\delta a} + o(\gamma^{-3} e^{-ca} e^{\delta a}) \tag{12.6}$$

$$+\gamma^{-C} \sum_{\substack{q_1 | q \\ q_1 \neq 1}} q_1^C \exp\left(-ca \min\left(1, \frac{\log q_1}{\log \frac{a}{\gamma}}\right)\right) \frac{|\mathrm{SL}_2(q_1)|^{\frac{1}{2}}}{|\mathrm{SL}_2(q)|} e^{\delta a}. \tag{12.7}$$

We estimate (12.7) as

$$\frac{\gamma^{-C} |a|^C}{|\mathrm{SL}_2(q)|} \left(\sum_{\substack{q_1 | q \\ 1 < q_1 < \frac{|a|}{\gamma}}} e^{-ca \frac{\log q_1}{\log \frac{a}{\gamma}}} \right) e^{\delta a} \tag{12.8}$$

$$+\gamma^{-C} q^C e^{(\delta - c)a} \tag{12.9}$$

and

$$\begin{aligned} \sum_{\substack{q_1 | q \\ q_1 \neq 1}} e^{-ca \frac{\log q_1}{\log \frac{a}{\gamma}}} &< \prod_{p | q} \left(1 + e^{-ca \frac{\log p}{\log \frac{a}{\gamma}}}\right) - 1 \\ &< \exp\left(\sum_{s=2}^{\infty} e^{-ca \frac{\log s}{\log \frac{a}{\gamma}}}\right) - 1 < e^{-c \frac{a}{\log \frac{a}{\gamma}}}, \end{aligned} \tag{12.10}$$

assuming

$$\log \frac{1}{\gamma} \ll a. \tag{12.11}$$

Therefore we proved the following result, of which Theorem 1.5 is an immediate consequence.

Proposition 4. *Notation being as above, we have*

$$\int k_\gamma(t) N(a+t; q, \xi) dt = \frac{e^{\delta a}}{|\mathrm{SL}_2(q)|} \left(C_1 + o(\gamma^{-C} e^{-c \frac{a}{\log \frac{a}{\gamma}}})\right) + \gamma^{-C} q^C e^{(\delta - c)a}, \tag{12.12}$$

where we assume

$$\log \frac{1}{\gamma} \ll \frac{a}{\log a}. \tag{12.13}$$

We remark that what is really required for sieving applications is a bound for the ratio

$$\frac{\int k_\gamma(t) N(a+t; q) dt}{\int k_\gamma(t) N(a+t) dt} \tag{12.14}$$

of the form

$$\frac{1}{|\mathrm{SL}_2(q)|} + O(e^{-\varepsilon a} q^C)$$

or

$$\frac{1}{|\mathrm{SL}_2(q)|}\Big(1 + O(e^{-c\frac{a}{\log a}})\Big) + O(e^{-\varepsilon a} q^C).$$

To bound the ratio (12.14) it suffices to use Proposition 1 (which builds crucially on the generalized expansion result given by Lemma 2) combined with the result of Lalley [16]. Of course, the results of Dolgopyat [8] and Naud [21], [22] are necessary to establish Theorem 1.4, which is of independent interest.

13. Proof of Theorem 1.6

13.1. **Combinatorial sieve.** As in [4], we will make use of the simplest combinatorial sieve which is turn is based on the Fundamental Lemma in the theory of elementary sieve, see [14] and [13]. Our formulation is tailored for the applications below.

Let A denote a finite sequence a_n, $n \geq 1$ of nonnegative numbers. Denote by X the sum

$$\sum_n a_n = X. \tag{13.1}$$

X will be large, in fact tending to infinity. For a fixed finite set of primes B let z be a large parameter (in our applications z will be a small power of X and B will usually be empty). Let

$$P = P_z = \prod_{\substack{p \leq z \\ p \notin B}} p. \tag{13.2}$$

Under suitable assumptions about sums of A over n's in progressions with moderate-size moduli d, the sieve gives upper and lower estimates which are of the same order of magnitude for sums of A over the n's which remain after sifting out numbers with prime factors in P.

More precisely, let

$$S(A,P) := \sum_{(n,P)=1} a_n. \tag{13.3}$$

The assumptions on sums in progressions are as follows:

(A_0) For d square-free, and having no prime factors in B ($d < X$), we assume that the sums over multiples of d take the form

$$\sum_{n \equiv 0(d)} a_n = \beta(d) X + r(A,d), \tag{13.4}$$

where $\beta(d)$ is a multiplicative function of d and

$$\text{for}\, p \notin B, \beta(d) \leq 1 - \frac{1}{c_1} \text{ for a fixed}\, c_1.$$

The understanding being that $\beta(d)X$ is the main term and that the remainder $r(A,d)$ is smaller, at least on average (see the next axiom).

(A_1) A has level distribution $D = D(X)$, $(D < X)$ that is

$$\sum_{d \leq D} |r(d,A)| \ll \frac{X}{(\log X)^B} \text{ for all } B > 0.$$

(A_2) A has sieve dimension $t > 0$, that is for a fixed c_2 we have

$$\left| \sum_{\substack{w \leq p \leq z \\ p \notin B}} \beta(p) \log p - t \log \frac{z}{w} \right| \leq c_2$$

for $2 \leq w \leq z$.

In terms of these conditions (A_0), (A_1), (A_2) the elementary combinatorial sieve yields:

Theorem 13.1. *Assume (A_0), (A_1) and (A_2) for $s > 9t$ and $z = D^{1/s}$ and X large we have*

$$\frac{X}{(\log X)^t} \ll S(A, P_z) \ll \frac{X}{(\log X)^t}. \tag{13.5}$$

13.2. **Applying the sieve.** Now let Λ be a Zariski dense subgroup of $\mathrm{SL}(2,\mathbb{Z})$ and let $f \in \mathbb{Z}[x_{ij}]$ be weakly primitive with $t(f)$ irreducible factors. The key nonnegative sequence a_n to which we apply the combinatorial sieve is defined as follows: for $n \geq 0$ we let

$$a_n = a_n(T) = \sum_{\substack{\gamma \in \Lambda \\ |\gamma| \leq T \\ |f(\gamma)| = n}} 1. \tag{13.6}$$

The sums on progressions are then, for $d \geq 1$ square free

$$\sum_{n \equiv 0(d)} a_n(T) = \sum_{\substack{\gamma \in \Lambda; |\gamma| \leq T \\ f(\gamma) \equiv 0(d)}} 1 = \sum_{\substack{\rho \in \Lambda/\Lambda(d) \\ f(\rho) \equiv 0(d)}} \sum_{\substack{\gamma \in \Lambda(d) \\ |\gamma| \leq T}} 1. \tag{13.7}$$

Consider the case of $\delta < 1/2$; the case of $\delta > 1/2$ is similar and simpler. According to Theorem 1.5 we then obtain

(13.8)
$$\sum_{n\equiv 0(d)} a_n(T) = \sum_{\substack{\rho\in\Lambda/\Lambda(d)\\ f(\rho)\equiv 0(d)}} \frac{T^{2\delta}}{|\Lambda_d|}\left(1+O\left(T^{-\frac{1}{\log\log T}}\right)\right)+O\left(d^C T^{2\delta-\varepsilon_1}\right),$$
$$= X\frac{|\Lambda_d^f|}{|\Lambda_d|}+O\left(\frac{|\Lambda_d^f|}{|\Lambda_d|}X^{1-\frac{1}{2\delta\log\log X}}\right)+O\left(|\Lambda_d^f|d^C X^{1-\frac{\varepsilon_1}{2\delta}}\right).$$

where

$$X=\sum_{k\in\mathbb{N}} a_k(T)=\sum_{\substack{\gamma\in\Lambda\\ |\gamma|\le T}} 1, \tag{13.9}$$

Λ_d is the reduction of Λ mod d, and Λ_d^f is the subset of Λ_d at which $f(x)=0 \bmod d$.

Using strong approximation theorem [19] and Goursat lemma as in [4] we obtain that outside of finite set of primes $S(\Lambda)$ we have $\Lambda_p \cong \mathrm{SL}_2(\mathbb{F}_p)$ and $\Lambda\to\Lambda_{d_1}\times\Lambda_{d_2}$ is surjective for $(d_1,d_2)=1$ and d_1d_2 square free. Let

$$\beta(d)=\frac{|\Lambda_d^f|}{|\Lambda_d|}. \tag{13.10}$$

Using Lang-Weil theorem [18] as in [4], we obtain

$$|\Lambda_d^f| \ll d^2 \tag{13.11}$$

and

$$\frac{|\Lambda_d^f|}{|\Lambda_d|}=\frac{t(f)}{p}+O(p^{-\frac{3}{2}}). \tag{13.12}$$

Hence we have

$$\sum_{n\equiv 0(d)} a_n(T)=\beta(d)X+r(A,d), \tag{13.13}$$

with

$$r(A,d)\ll \frac{1}{d}X^{1-\frac{1}{2\delta\log\log X}}+d^{C+2}X^{1-\frac{\varepsilon_1}{2\delta}}. \tag{13.14}$$

Verification of (A_0) is completely analogous to Proposition 3.1 in [4]. Regarding the level distribution (A_1) we have that

$$\sum_{d\le D}|r(A,d)|\ll X^{1-\frac{1}{2\delta\log\log X}}+D^{C+3}X^{1-\frac{\varepsilon_1}{2\delta}}\ll\frac{X}{(\log X)^B} \tag{13.15}$$

for any $B > 0$ as long as

$$D \leq X^{\tau} \text{ with } \tau < \frac{2\delta}{(C+3)\varepsilon_1}. \tag{13.16}$$

Finally, to verify the third axiom concerning the sieve dimension, we have, using (13.12), that

(13.17)
$$\sum_{w \leq p \leq z} \beta(p) \log p = \sum_{w \leq p \leq z} \left(\frac{t \log p}{p} + O\left(\frac{\log p}{p^{3/2}}\right)\right) = t \log \frac{z}{w} + O(1),$$

which establishes (A_2) with the sieve dimension being t.

References

[1] A. F. Beardon, *The Geometry of Discrete Groups*, Springer, New York, 1983.

[2] D. Borthwick, *Spectral theory of infinite-area hyperbolic surfaces*, Birkhauser, Boston, 2007.

[3] J. Bourgain and A. Gamburd, *Uniform expansion bounds for Cayley graphs of* $\mathrm{SL}_2(\mathbb{F}_p)$, Annals of Mathematics, **167**, 2008, 625-642.

[4] J. Bourgain. A. Gamburd, P. Sarnak, *Affine linear sieve, expanders, and sum-product*, Inventiones Mathematicae, to appear.

[5] R. Brooks, *The spectral geometry of a tower of coverings*, J. Differential Geom. **23**, 1986), 97–107.

[6] M. Burger, *Estimation de petites valeurs propres du laplacien d'un revetement de variétés riemanniennes compactes*, C. R. Acad. Sci. Paris Sr. I Math. **302**, 1986, 191–194.

[7] M. Burger, *Spectre du laplacien, graphes et topologie de Fell*, Comment. Math. Helv. **63**, 1988, 226–252.

[8] D. Dolgopyat, *On decay of correlations in Anosov flows*, Ann. of Math., **147**, 1998, 357-390.

[9] G. Frobenius, *Über Gruppencharaktere*, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, 1896, 985-1021.

[10] E. Fuchs and K. Sanden, *Prime number and local to global gonjectures in Apollonian circle packings*, preprint.

[11] A. Gamburd, *Spectral gap for infinite index "congruence" subgroups of* $\mathrm{SL}_2(\mathbb{Z})$, Israel Journal of Mathematics **127**, (2002), 157-200.

[12] L. Guillopé, K. Lin, M. Zworski, *The Selberg zeta function for convex co-compact Schottky groups*, Comm. Math. Phys. **245**, 2004, 149-176.

[13] H. Halberstam and H. Richert, *Sieve methods*, Academic Press, 1974.

[14] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS, 2004.

[15] A. Kontorovich and H. Oh, *Apollonian circle packings and closed horospheres on hyperbolic* 3*-manifolds*, preprint.

[16] S. Lalley, *Renewal theorems in symbolic dynamics, with applications to geodesic flows, noneuclidean tesselations and their fractal limits*, Acta Math., **163**, 1989, 1-55.

[17] P.D. Lax and R. S. Phillips, *The asymptotic distribution of lattice points in Euclidean and non-Euclidean space*, Journal of Functional Analysis, **46**, 1982, 280-350.

[18] S. Lang and A. Weil, *Number of points of varieties in finite fields*, Amer. J. Math. **76**, 1954, 819–827.
[19] C. Matthews, L. Vaserstein and B. Weisfeiler, *Congruence properties of Zariski-dense subgroups*, Proc. London Math. Soc, **48**, 1984, 514-532.
[20] R. Mazzeo and R. Melrose, *Meromorphic extension of the resolvent on complete spaces with asymptotically constant negative curvature*, Journal of Functional Analysis, **75**, 1987, 260-310.
[21] F. Naud, *Expanding maps on Cantor sets and analytic continuation of zeta functions*, Ann. Sci. École Norm. Sup., **38**, 2005, 116–153.
[22] F. Naud, *Precise asymptotics of the length spectrum for finite-geometry Riemann surfaces*, Int. Math. Res. Not. 2005, no. 5, 299–310.
[23] W. Parry and M. Pollicott, *Zeta functions and periodic orbit structure of hyperbolic dynamics*, Astérisque **187-188**, 1990.
[24] S. J. Patterson, *The limit set of a Fuchsian group*, Acta. Math., **136**, 1975, 241-273.
[25] S. J. Patterson, *On a lattice point problem in hyperbolic space and related questions in spectral theory*, Ark. Mat. **26**, 1988, 167-172.
[26] S. J. Patterson and P. A. Perry, *The divisor of Selberg's zeta function for Kleinian groups*, Duke Math. J., **106**, 2001, 321-390, Appendix A by C. Epstein.
[27] D. Ruelle, *Thermodynamic Formalism.*
[28] J. J. Sansuc, *Groupe de Brauer et arithmétique des groupes algébriques linéaires sur un corps de nombres*, J. Reine Angew. Math., **327**, 1981, 12-80.
[29] P. Sarnak, *Notes on the generalized Ramanujan conjectures*, Clay Mathematics Proceedings, **4**, 2005, 659-685.
[30] P. Sarnak, *Integral Apollonian Packings*, 2009 MAA Lecture, available at `http://www.math.princeton.edu/sarnak/`
[31] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) **20**, 1956, 47-87.
[32] A. Selberg, *On the estimation of Fourier coefficients of modular forms*, Proc. Symp. Pure Math. **VII**, Amer. Math. Soc.(1965), 1-15.
[33] C. Series, *The infinite word problem and limit sets in Fuchsian groups*, Ergodic Theory Dynamical Systems, **1**, 1981, 337–360.
[34] D.Sullivan, *Discrete conformal groups and measurable dynamics*, Bull. Amer. Math. Soc., **6**, 1982, 57-73.
[35] T. Tao, *Product sets estimates for non-commutative groups*, Combinatorica, **28**, 2008, 547–594.
[36] T. Tao and V. Vu, *Additive combinatorics*, Cambridge University Press, 2006
[37] P. Varju, *Expansion in* $\mathrm{SL}_d(\mathcal{O}/\mathcal{I})$*,* $\mathcal{I}$ *square-free*, preprint.

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540
E-mail address: bourgain@math.ias.edu

Department of Mathematics, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064
E-mail address: agamburd@ucsc.edu

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540 and Department of Mathematics, Princeton University
E-mail address: sarnak@math.princeton.edu