ISRAEL JOURNAL OF MATHEMATICS 127 (2002), 157–200

ON THE SPECTRAL GAP FOR INFINITE INDEX "CONGRUENCE" SUBGROUPS OF $SL_2(\mathbf{Z})$

BY

ALEX GAMBURD*

Department of Mathematics, Stanford University
Stanford, CA 94305, USA
e-mail: agamburd@math.stanford.edu

ABSTRACT

A celebrated theorem of Selberg states that for congruence subgroups of $SL_2(\mathbf{Z})$ there are no exceptional eigenvalues below 3/16. Extending the work of Sarnak and Xue for cocompact arithmetic lattices, we prove a generalization of Selberg's theorem for infinite index "congruence" subgroups of $SL_2(\mathbf{Z})$. For such subgroups with a high enough Hausdorff dimension of the limit set we establish a spectral gap property and consequently solve a problem of Lubotzky pertaining to expander graphs.

1. Introduction

Let Λ be a finitely generated subgroup of $SL_2(\mathbf{Z})$. In dimension two being finitely generated is equivalent to being geometrically finite, i.e., the fundamental domain $\mathcal{F} = \Lambda\backslash\mathbf{H}$ has finitely many bounding sides [2]. The limit set of Λ, denoted by $L(\Lambda)$, is a subset of $\mathbf{R}\cup\infty$; it was observed a century ago by Poincaré and Klein that if $\mathcal{F}$ has infinite volume (and Λ is not elementary) $L(\Lambda)$ is a Cantor-like set, we will denote by $\delta(L(\Lambda))$ its Hausdorff dimension. The spectrum of the Laplacian $\boldsymbol{\Delta}$ on $L^2(\mathcal{F})$ will be denoted by $\Omega(\mathcal{F})$.

The spectrum of geometrically finite Fuchsian groups was investigated by Patterson [32, 33]; Sullivan [46], and Lax and Phillips [21] generalized and extended his results in higher dimensions. The main result for $\Omega(\Lambda\backslash\mathbf{H})$ is the following

* The author was supported in part by the NSF graduate fellowship.
Received August 19, 1999

THEOREM (Patterson, Lax and Phillips): *Assume that* $\delta > \frac{1}{2}$. *Then*

(1) *The bottom of the spectrum*[1], $\lambda_0(\mathcal{F}) = \delta(1-\delta)$; *it is an isolated eigenvalue of multiplicity one.*

(2) *There are finitely many discrete eigenvalues in the interval* $[0, 1/4)$.

(3) *If* $\mathrm{vol}(\mathcal{F}) = \infty$ *the spectrum* $\Omega(\mathcal{F})$ *is continuous in* $[1/4, \infty]$.

Now let $\Lambda(p) = \Lambda \cap \Gamma(p)$, where $\Gamma(p)$ is a principal congruence subgroup of level p:

$$\Gamma(p) = \{\gamma \in \mathrm{SL}_2(\mathbf{Z}) \colon \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod p\}. \tag{1.1}$$

As a subgroup of finite index in $\Lambda(1) = \Lambda$, $\Lambda(p)$ has the same bottom of the spectrum, $\lambda_0(\Lambda(p)\backslash\mathbf{H}) = \lambda_0(\Lambda(1)\backslash\mathbf{H})$.

We are interested in estimating $\lambda_1(\Lambda(p)\backslash\mathbf{H})$ as $p \to \infty$. For congruence subgroups Selberg proved the following celebrated result [42].

THEOREM (Selberg, 1965): *Let* $\Gamma'(N)$ *be any congruence subgroup of* $\mathrm{SL}_2(\mathbf{Z})$, *i.e., subgroup containing principal congruence subgroup* $\Gamma(N)$. *Let* $X'(N) = \Gamma'(N)\backslash\mathbf{H}$. *For* $N \geq 1$

$$\lambda_1(X'(N)) \geq 3/16. \tag{1.2}$$

In 1995 Luo, Rudnick and Sarnak [26] established a better bound[2] by using properties of the Rankin–Selberg convolution L-functions on GL_3:

THEOREM (Luo, Rudnick, Sarnak, 1995):

$$\lambda_1(X'(N)) \geq 171/784, \tag{1.3}$$

taking us about halfway beetween Selberg's Theorem and the following remarkable conjecture of his:

CONJECTURE (Selberg, 1965): *Conditions being as above*

$$\lambda_1(X'(N)) \geq 1/4. \tag{1.4}$$

The conjecture is known to be true for groups of small level [11]. It remains one of the fundamental unsolved analytic questions in modular forms (see [38] for a

---

1 To put it sonorously, à la Mark Kac [16], one can hear the fractal shape of the boundary in the bass note.

2 Shortly afterwards, Iwaniec [15] established a sligtly weaker bound $\lambda_1(X'(N)) \geq \frac{10}{49}$, using only GL_2 theory.

tantalizing discussion). It has many applications to classical number theory (see [14, 37], for example); from the point of view of modern representation theory it is a generalisation [40] of the famous Ramanujan conjectures solved by Deligne.

Selberg's approach was to relate this problem to arithmetic questions about Kloosterman sums; the key ingredient in his proof is Weil's bound[3] on Kloosterman sums [49]. He remarked that "a natural question arises [as to] what happens if we give up the assumption that Γ' is a congruence subgroup", but pointed out that the "methods which depend on estimates for Kloosterman sums are not capable of extension to the general case" and constructed examples of surfaces (which correspond to subgroups of $SL_2(\mathbf{Z})$) with arbitrarily small first eigenvalue; examples of compact Riemann surfaces with this property were also given by Randol [35].

In 1991 Sarnak and Xue [39] considered the case of cocompact arithmetic subgroups of $SL_2(\mathbf{R})$. By the classification results of Weil a cocompact arithmetic subgroup is commensurable with a group Φ_A derived from certain types of quaternion algebras A over F, where F is a totally real number field with ring of integers $\mathcal{O}$. The group Φ_A has its family of congruence subgroups, defined similarly to the above by

$$\Delta(\mathcal{P}) = \{\alpha \in \Phi \mid \alpha \equiv 1 \operatorname{mod}(\mathcal{P})\}. \tag{1.5}$$

Sarnak and Xue obtained the following result (cf. also Huxley [12] where the number 5/36 appears for related reasons)

THEOREM (Sarnak and Xue, 1991): *Let $\Gamma \subset \Phi_A$ be a subgroup of finite index. Let $\Gamma(\mathcal{P}) = \Gamma \cap \Delta(\mathcal{P})$. For large enough prime ideals $\mathcal{P}$ of $\mathcal{O}$*

$$\Omega(\Gamma(\mathcal{P})\backslash\mathbf{H}) \cap [0, 5/36) = \Omega(\Gamma(1)\backslash\mathbf{H}) \cap [0, 5/36).$$

COROLLARY (Sarnak and Xue, 1991): *For Γ cocompact as above and $\mathcal{P}$ a prime ideal of sufficiently large norm*

$$\lambda_1(\Gamma(\mathcal{P})\backslash\mathbf{H}) \geq \min(\lambda_1(\Gamma(1)\backslash\mathbf{H}), 5/36).$$

The proof of Sarnak and Xue stems from the observation that if the spectrum $\Omega(\Gamma(\mathcal{P})\backslash\mathbf{H})$ has a new eigenvalue λ in $[0, \frac{1}{4})$, it must be of high multiplicity[4]. This

3 Weil's bound in turn is a consequence of the Riemann hypothesis for curves which he had proven earlier. Iwaniec [13] has given a proof of (1.2), which, while using Kloosterman sums, avoids appealing to Weil's bound. Gelbart and Jacquet [8], using methods very different from Selberg's, showed that $\frac{3}{16}$ is not attained in (1.2).

4 For the origin of this idea see [38].

follows (as we explain in more detail in Section 6) from the result going back to Frobenius, that the smallest dimension of a nontrivial irreducible representation of $\mathrm{SL}_2(\mathbf{F}_p)$ is $\frac{p-1}{2}$, which is large compared to the size of the group (which is of order p^3). The proof then proceeds by estimating the number of lattice points and using the (pre)trace formula to show that for small eigenvalues one cannot accommodate such a high multiplicity.

Proceeding along the lines of Sarnak and Xue we prove the following result for infinite index subgroups of $\mathrm{SL}_2(\mathbf{Z})$, providing in passing a new proof of Selberg's theorem[5] with somewhat weaker bound.

MAIN THEOREM: *Let $\Lambda = \langle A_1, \ldots, A_k\rangle$ be a finitely generated subgroup of $\mathrm{SL}_2(\mathbf{Z})$ with $\delta > \frac{5}{6}$. Let $\mathcal{F}(p) = \Lambda(p)\backslash\mathbf{H}$. For p large enough*

$$\Omega(\mathcal{F}(p)) \cap [\delta(1-\delta), 5/36) = \Omega(\mathcal{F}(1)) \cap [\delta(1-\delta), 5/36).$$

MAIN COROLLARY: *Suppose that $\delta > \frac{5}{6}$. Then $\Omega(\mathcal{F}(p))$ has a spectral gap, that is for p large*

$$\lambda_1(\mathcal{F}(p)) \geq \min(\lambda_1(\mathcal{F}(1)), 5/36).$$

ORGANISATION OF THE PAPER AND STRATEGY OF THE PROOF. The observation that the multiplicity of new eigenvalues is large,

$$m(\lambda, \mathcal{F}(p)) > (p-1)/2,$$

carries through to our context, once we prove that for p large enough $\Lambda(1)/\Lambda(p) \cong \mathrm{SL}_2(\mathbf{F}_p)$; we attend to this matter at once in section 2. The proof of the theorem hinges on bounding the multiplicity from above, more precisely, obtaining the bound

$$m(\lambda, \mathcal{F}(p)) \ll p^{6(1-s)},$$

where $\lambda = s(1-1)$. In order to extend the approach of Sarnak and Xue in obtaining this bound, the main obstacle we have to overcome is the infinitude of the volume of $\mathcal{F}(p)$. After the review of the pertinent geometric facts, we begin the attack by decomposing the fundamental domain into compact and infinite parts in section 3. Following that, in section 4, we prove the key collar lemmas, (Lemmas 4.1 and 4.2) which state, rougly speaking, that for low-lying eigenvalues (below $\frac{1}{4}$) the L^2 norm of the eigenfunction of the Laplacian in a collar of fixed width, contiguous with a cusp or a flare, is of the same order of magnitude as

5 It follows from the main theorem and the well known fact (see, e.g., [37], p. 34) that $\lambda_1(\mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}) \geq \frac{1}{4}$.

its L^2 norm in the whole cusp or flare. In a sense, this lemma could be viewed as a generalization of the following fact about the zero eigenvalue and constant eigenfuction (area) in the cusp: the hyperbolic area of the collar of width $\ln 2$ is the same as the area of contiguous cusp.

Arithmetic of the problem comes into play in the following estimate[6] on the number of lattice points (the implied constant is independent of p):

$$N_1(T,\Gamma(p)) \stackrel{\text{def}}{=} \sum_{\substack{\gamma\in\Gamma(p)\\ \|\gamma\|\le T}} 1 \ll \frac{T^{2+\epsilon}}{p^3} + \frac{T^{1+\epsilon}}{p^2} + 1.$$

This estimate is proved in section 5, where we also set up the trace formalism by proving the pretrace inequality (Proposition 5.2), and computing the Selberg transform and convolution of a point-pair invariant given by characteristic function (Proposition 5.1). In section 6 we put everything together and prove the main theorem.

Then in section 7 we exploit the consequences of the main theorem to address the question of Alex Lubotzky, pertaining to expander graphs.

Expander graphs are widely used in Computer Science, in areas ranging from parallel computation to complexity theory and cryptography. Identifying synapses with the vertices and dendrites with the edges, we can view the network of 10^{12} neurons in the brain as an expander graph. Intuitively, to be an expander graph, a graph has to be sparse and highly connected. Clearly, high connectivity is desirable in any communication network. The necessity of sparsity is, perhaps, best seen in the case of the brain-graph: since the "wires" have finite thickness, their total length cannot exceed the quotient of the average volume of one head and the area of the wires' cross-section. Also note that the thinner the wire, the longer the time of propagation.

There are several ways of making the intuitive notions of connectivity and sparsity precise; the simplest and most widely used is the following.

Definition 1.1: Given an undirected k-regular graph $\mathcal{G}$ and a subset X of V, the **expansion** of X, $c(X)$, is defined to be the ratio $|N(X)|/|X|$, where $N(X)$ is a set of neighbors of X. The expansion coefficient of a graph $\mathcal{G}$

$$c(\mathcal{G}) = \inf\left\{c(X) \mid |X| < \frac{1}{2}|\mathcal{G}|\right\}. \tag{1.7}$$

6 It is only here that we use the fact that the homomorphism $\Lambda \to \mathrm{SL}_2(\mathbf{F}_p)$ is the reduction modulo p and not an arbitrary one; in fact most of the proof goes through with the much weaker assumption $\Lambda(1)/\Lambda(p) \cong \mathrm{SL}_2(\mathbf{F}_p)$. We are indebted to a referee for this remark.

The expansion coefficient is a discrete analogue of Cheeger's constant for Riemannian manifold. The discrete analogue of the Laplace operator is the nearest neighbor averaging operator; there is a discrete analogue of Cheeger–Buser inequalities relating $c(\mathcal{G})$ to $\lambda_1(\mathcal{G})$ (see [24] for a very clear exposition of this and other topics pertaining to expander graphs).

Definition 1.2: A family of k-regular graphs $\mathcal{G}_n$ forms a family of expanders if there is a fixed positive constant C, such that

$$\liminf_{n\to\infty} c(\mathcal{G}_n) \geq C. \tag{1.8}$$

By the discrete Cheeger–Buser inequality $\mathcal{G}_n$ is a family of expanders iff

$$\liminf_{n\to\infty} \lambda_1(\mathcal{G}_n) > 0. \tag{1.9}$$

It is not difficult to see that a random regular graph is a good expander. However the explicit construction of expander graphs is much more difficult and was first achieved by Margulis [29], who used Kazhdan property (T) from representation theory of semi-simple Lie groups [17]. Lubotzky, Phillips and Sarnak [22] constucted expanders based on Selberg's theorem, and later the optimal ones based on the (proven) Ramanujan conjectures. As Lubotzky wrote in [25]:

> What is very frustrating is that all these deep theories give some examples with very special sets of generators. A small change of the construction – which seems to be meaningless from the combinatorial point of view – leaves these tools helpless.

Lubotzky illustrated this by the following example. For a prime $p \geq 5$ let us define

$$S_p^1 = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\},$$
$$S_p^2 = \left\{ \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \right\},$$
$$S_p^3 = \left\{ \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \right\}$$

and for $i = 1, 2, 3$ let $\mathcal{G}_p^i = \mathcal{G}\left(\mathrm{SL}_2(\mathbf{F}_p), S_p^i\right)$, a Cayley graph[7] of $\mathrm{SL}_2(\mathbf{F}_p)$ with respect to S_p^i; these graphs are connected as we will show in the next section.

7 A Cayley graph of a group G with respect to a symmetric set of generators S, which we denote by $\mathcal{G}(G, S)$, is a graph whose vertices are elements of G and $a \in G$ is adjacent to σa, $\sigma \in S$.

The graphs $\mathcal{G}_p^i$ can be viewed as a "discrete approximation" of the hyperbolic manifolds $X_p^i = \Lambda^i(p)\backslash\mathbf{H}$, where Λ^i is subgroup of $\mathrm{SL}_2(\mathbf{Z})$ generated by $\langle(\begin{smallmatrix}1 & i\\ 0 & 1\end{smallmatrix}), \langle(\begin{smallmatrix}1 & 0\\ i & 1\end{smallmatrix})\rangle$. By Selberg's theorem the families X_p^1 and X_p^2 have a spectral gap (form a family of "expander surfaces") and from this one deduces (as we do in section 7) that $\mathcal{G}_p^1$ and $\mathcal{G}_p^2$ are families of expander graphs. Given this fact, it is difficult to believe that $\mathcal{G}_p^3$ is not an expander family. Note, however, that the group generated by $\langle(\begin{smallmatrix}1 & 3\\ 0 & 1\end{smallmatrix}), \langle(\begin{smallmatrix}1 & 0\\ 3 & 1\end{smallmatrix})\rangle$ has infinite index and thus does not come under the purview of Selberg's theorem. These considerations led Lubotzky to pose the following question[8]: can a finite set of elements in $\mathrm{SL}_2(\mathbf{Z})$ give rise to a family of expanders even if the subgroup they generate has infinite index?

Recently Shalom [43] proved that such expanders exist and gave an explicit example of an infinite-index subgroup in $\mathrm{PSL}_2(\mathbf{Z}[\omega])$ (where ω is a primitive third root of unity) yielding a family of expanders; his approach (based on the use of invariant means on the profinite completion of the finite groups) is non-constructive and does not yield explicit examples in $\mathrm{SL}_2(\mathbf{Z})$. Restating the corollary of the main theorem in representation-theoretic language and using Fell's continuity of induction we prove the following theorem, complementing results of Shalom in addressing Lubotzky's problem.

THEOREM: *Let $S = \{A_1, \ldots, A_k\}$ be a symmetric set of generators in $\mathrm{SL}_2(\mathbf{Z})$ and let $\Lambda = \langle A_1, \ldots, A_k\rangle$. If the Hausdorff dimension of the limit set $\delta(L(\Lambda)) > 5/6$ then $\mathcal{G}_p = G(\mathrm{SL}_2(\mathbf{F}_p), S)$ is a family of expanders.*

After proving this theorem in section 7 we conclude by giving the examples of generators satisfying its conditions.

For the subgroups with Hausdorff dimension of the limit set less than $\frac{5}{6}$ (in particlular for the group generated by $\langle(\begin{smallmatrix}1 & 3\\ 0 & 1\end{smallmatrix}), \langle(\begin{smallmatrix}1 & 0\\ 3 & 1\end{smallmatrix})\rangle$, whose Hausdorff dimension was found to be 0.753 ± 0.003 by Phillips and Sarnak [34]) the question remains open[9] and demands for its solution a fundamentally new idea. Several remaining problems are more tractable: extending the main theorem to $\mathrm{SL}_2(\mathbf{C})$ (where in the cocompact arithmetic case Sarnak and Xue [39] have a result analogous to the one in $\mathrm{SL}_2(\mathbf{R})$) and finding conditions (perhaps resembling those in [7]) under which a subgroup has a high enough Hausdorff dimension of the limit set.

---

8 The general problem, raised by Lubotzky in [24, 25, 23], is whether being an expander family is the property of groups alone, independent of the choice of generators. Very little is known about this intriguing fundamental problem; see [44] for a discussion.

9 In fact, numerical experiments of Lafferty and Rockmore [18, 19, 20] indicate that a "generic" element in the group ring of $\mathrm{SL}_2(\mathbf{Z})$ has a spectral gap (cf. [7]).

ACKNOWLEDGEMENT: I thank my advisor Peter Sarnak for his guidance, encouragement, and inspiration. I am grateful to Alex Lubotzky and Yehuda Shalom for interest in my work and stimulating conversations.

2. A little algebra

In this section we prove that for p large enough $\Lambda(1)/\Lambda(p) \cong \mathrm{SL}_2(\mathbf{F}_p)$. Along the way we show that the graphs $\mathcal{G}(\mathrm{SL}_2(\mathbf{F}_p), \langle A_1, \ldots, A_k \rangle)$, described at the end of introduction, are connected and establish a lower bound for their girth.

We can assume that $\Lambda(1)$ is torsion free, if necessary passing to a subgroup of finite index and using Selberg's Lemma [41, 1], which states that a finitely generated group of matrices over a field of characterisitc zero has a torsion free subgroup of finite index.

Now consider $\Lambda(2) = \Lambda(1) \cap \Gamma(2)$. The group $\Gamma(2)$ is a free group, as its subgroup $\Lambda(2)$ is free as well. Moreover, since Hausdorff dimension of the limit set of $\Lambda(1)$, and, consequently of $\Lambda(2)$, is greater than $\frac{1}{2}$, $\Lambda(1)$ is nonelementary and hence nonabelian. Now by Stallings's Theorem [45], which states that a finitely generated torsion-free subgroup which contains a free subgroup of finite index is free, we conclude that $\Lambda(1)$ is free as well.

Consider now the Cayley graphs

$$\mathcal{G}_p = \mathcal{G}(\Lambda(1)/\Lambda(p), S), \quad S = \{A_1, \ldots, A_k\}.$$

We estimate their girth $c(\mathcal{G}_p)$ (the length of the shorterst cycle) from below following the method used by Margulis [30]. To this end we estimate the quantity $d(\mathcal{G}_p)$, defined as the largest integer such that any two walks in $\mathcal{G}_p$ of length less than $d(\mathcal{G}_p)$ starting at $E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ end at different vertices. (By a walk of length k we mean a sequence of steps along adjacent vertices $x_0, \ldots, x_l$, such that $x_{j-1} \neq x_{j+1}$.) By the homogeneity of $\mathcal{G}_p$ we have

$$c(\mathcal{G}_p) \geq 2d(\mathcal{G}_p) - 1.$$

Let ϕ_p be a homomorphism of $\mathrm{SL}_2(\mathbf{Z})$ onto $\mathrm{SL}_2(\mathbf{F}_p)$ which associates with each matrix $X \in \mathrm{SL}_2(\mathbf{Z})$ the matrix $\phi_p(X)$, obtained by reducing each element of X modulo p. We have

$$\Lambda(1)/\Lambda(p) \cong \phi_p(\Lambda(1)) \cong Y_p,$$

where Y_p is some subgroup of $\mathrm{SL}_2(F_p)$.

Let us set

$$A_{1,p} = \phi_p(A_1), \ldots, A_{k,p} = \phi_p(A_k); \quad S_p = \phi_p(S).$$

So we have $\mathcal{G}_p = \mathcal{G}(Y_p, S_p)$. Assume we are given two walks in $\mathcal{G}_p$, $p = (p_0, p_1, \ldots, p_r)$ and $s = (s_0, s_1, \ldots, s_t)$, both starting at $E = p_0 = s_0$ and having a common end $p_r = s_t$. By the definition of the graph $\mathcal{G}_p$, we find that $p_i = p_{i-1}v_i$ and $s_j = s_{j-1}w_j$, $1 \leq i \leq r, 1 \leq j \leq t$, where $v_i, w_j \in S_p$. The walks p and s correspond to the words $V = (v_1, \ldots, v_r)$ and $W = (w_1, \ldots, w_t)$ over S_p. Clearly $p_i = v_1 \ldots v_i$ and $s_j = w_1 \ldots w_j$.

Hence, since $p_r = s_t$, we have

$$v_1 \ldots v_r = w_1 \ldots w_t. \tag{2.1}$$

Let us define the word $\tilde{V} = (\tilde{v_1}, \ldots, \tilde{v_r})$ by

$$\tilde{v_i} = \begin{cases} A_l & \text{if } v_i = A_{l,p}; \\ A_l^{-1} & \text{if } v_i = A_{l,p}^{-1} \end{cases}$$

and $\tilde{W} = (\tilde{w_1}, \ldots, \tilde{w_t})$ analogously.

By our definition of a walk, the words V and W and hence the words $\tilde{V}$ and $\tilde{W}$ are reduced. Since the walks p and s are different and $p_0 = s_0 = E$, the words V and W and thus the words $\tilde{V}$ and $\tilde{W}$ are different. As V and W are different reduced words over $A_1, \ldots, A_k$, we infer that

$$\tilde{v_1} \ldots \tilde{v_r} \neq \tilde{w_1} \ldots \tilde{w_t} \tag{2.2}$$

since there are no multiplicative relations between $A_1, \ldots, A_k$. On the other hand, by (2.1) we have

$$\phi_p(\tilde{v_1} \ldots \tilde{v_r}) = \phi_p(\tilde{w_1} \ldots \tilde{w_t}). \tag{2.3}$$

Therefore all the elements of the matrix $\tilde{v_1} \ldots \tilde{v_r} - \tilde{w_1} \ldots \tilde{w_t}$ are divisible by p and at least one of them is different from zero by (2.2).

We conclude that

$$\|\tilde{v_1} \ldots \tilde{v_r} - \tilde{w_1} \ldots \tilde{w_t}\| \geq p, \tag{2.4}$$

where the norm of a matrix L is defined by

$$\|L\| = \sup_{x \neq 0} \frac{\|Lx\|}{\|x\|},$$

and the norm of $x = (x_1, x_2)$ is given by $\|x\| = \sqrt{x_1^2 + x_2^2}$. From (2.4) we infer

$$\max \|\tilde{v_1} \ldots \tilde{v_r}\|, \|\tilde{w_1} \ldots \tilde{w_t}\| \geq \frac{p}{2}. \tag{2.5}$$

Let $\alpha = \max_k \|A_k\|$. By submultiplicativity of the norm of matrices, from (2.5) we deduce

$$\alpha^{\max(r,t)} \geq \frac{p}{2}. \tag{2.6}$$

Keeping in mind the definition of $d(\mathcal{G}_p)$, by (2.6) we obtain the inequaliy

$$d(\mathcal{G}_p) \geq \log_\alpha \left(\frac{p}{2}\right), \tag{2.7}$$

and consequently

$$c(\mathcal{G}_p) \geq 2\log_\alpha \left(\frac{p}{2}\right) - 1. \tag{2.8}$$

Let us recall now the classification of subgroups of $\mathrm{SL}_2(\mathbf{F}_p)$ from [47].

THEOREM 1: *Let p be a prime with $p \geq 5$. Then any subgroup of $\mathrm{SL}_2(\mathbf{F}_p)$ is isomorphic to one of the following subgroups:*

1. *The dihedral groups of order $2(\frac{p\pm1}{2})$ and their subgroups.*
2. *A group H of order $p(\frac{p-1}{2})$ and its subgroups. If $H_1 \equiv N_1$ is a subgroup of H, then its factor group H/H_1 is cyclic.*
3. *A_4, S_4, or A_5.*

Following Davidoff and Sarnak [5] we now prove that $Y_p \cong \mathrm{SL}_2(\mathbf{F}_p)$ for p large enough. Suppose not. Then Y_p is a certain proper subgroup listed in Theorem 1. Certain proper subgroups can be eliminated immediately as possibilities for Y_p since they contain elements of small order which clearly violate the girth bound (2.8). However, we see that the remaining subgroups have trivial second commutator, that is to say for all $x_1, x_2, y_1, y_2 \in Y_p$ we have

$$(x_1y_1x_1^{-1}y_1^{-1})(x_2y_2x_2^{-1}y_2^{-1})(y_1x_1y_1^{-1}x_1^{-1})(y_2x_2y_2^{-1}x_2^{-1}) = 1.$$

If we take x_1, y_1, x_2, y_2 to be any generators in the construction of our graphs, then we see that this condition provides a closed cycle of length 16. However, such a cycle also violates the girth bound whenever

$$2\log_\alpha \left(\frac{p}{2}\right) \geq 17. \tag{2.9}$$

So for $p > 2\alpha^{17/2}$ we get a contradiction.

3. Anatomy of the fundamental domain

We will use alternately the upper half-plane and unit disc models; in this subsection we will summarize the pertinent formulae. The Poincaré upper half-plane model is the following subset of the complex plane $\mathbf{C}$:

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

with the hyperbolic metric

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

The distance function on $\mathbf{H}$ is explicitly given by

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

It will often be more convenient to 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 mapping

$$z \longrightarrow \frac{z - i}{z + i}, \quad z \in \mathbf{H}$$

maps $\mathbf{H}$ biholomorphically onto the unit disc $\mathbf{D}$,

$$\mathbf{D} = \left\{w = x + iy \in \mathbf{C} \mid x^2 + y^2 < 1\right\}.$$

The induced metric is

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - (x^2 + y^2))^2}. \tag{3.5}$$

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

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

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

$$||g||^2 = a^2 + b^2 + c^2 + d^2 \quad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{3.6}$$

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

$$\|g\|^2 = a^2 + b^2 + c^2 + d^2 = 4u(gi, i) + 2. \tag{3.8}$$

The Riemannian measure on $\mathbf{H}$ is expressed in terms of Lebesgue measure by the formula

$$d\mu(z) = \frac{dxdy}{y^2}, \tag{3.8}$$

and on $\mathbf{D}$ by the formula

$$d\mu(z) = \left(\frac{2}{1 - |z|^2}\right)^2 dxdy. \tag{3.9}$$

The area of a hyperbolic disc of radius r is

$$\text{area}(r) = 4\pi \sinh^2\left(\frac{r}{2}\right). \tag{3.10}$$

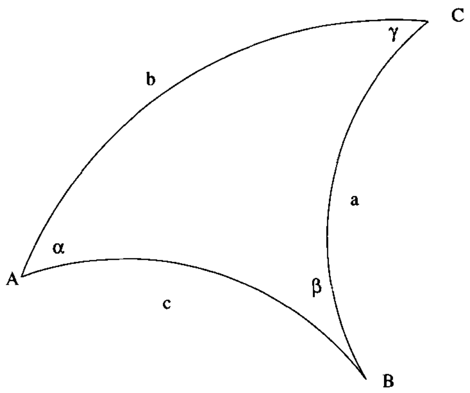

Figure 1. Hyperbolic triangle.

Let T be a hyperbolic triangle (see Figure 1) with vertices labelled A, B, C; the sides opposite these vertices labelled a, b, and c respectively, and the interior angles at the vertices α, β, and γ.

The following hyperbolic sine and cosine laws hold:

$$\frac{\sinh a}{\sin \alpha} = \frac{\sinh b}{\sin \beta} = \frac{\sinh c}{\sin \gamma}, \tag{3.11}$$

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma. \tag{3.12}$$

We will need the notion of **Fermi coordinates** [3]. They are defined as follows. Let η be the geodesic in the hyperbolic plane parametrized with the unit speed in the form

$$t \to \eta(t) \in \mathbf{H}, \quad t \in \mathbf{R}.$$

Then η separates $\mathbf{H}$ into two half-planes: a left hand side and a right hand side of η. For each $p \in \mathbf{H}$ 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

$$ds^2 = d\rho^2 + \cosh^2 \rho dt^2. \tag{3.13}$$

DECOMPOSITION OF THE FUNDAMENTAL DOMAIN OF $\mathcal{F}(p)$. We begin by decomposing the fundamental domain of a given geometrically finite group $\Lambda(1)$. As is well known (and discussed, for example, in [32]) we have a decomposition of the fundamental domain $\mathcal{F}(1) = \Lambda(1)\backslash\mathbf{D}$ of the following form:

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

where

(1) $\overline{\mathcal{K}(1)}$ is relatively compact in $\mathbf{D}$
(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 $\text{flare}_j(\alpha)$ is isometric to a standard hyperbolic fundamental domain $F(\alpha)$ of the form

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

where $\alpha < \pi/2$.

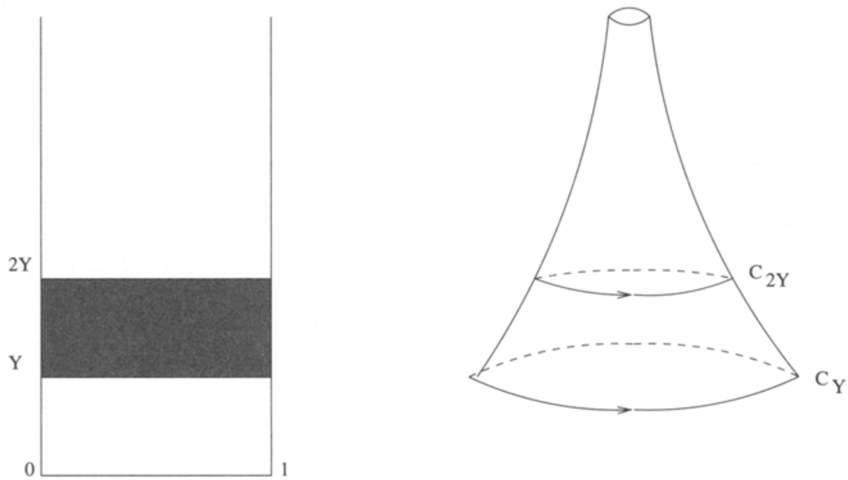

Figure 2. A cusp and its collar.

Now we subdivide each cusp_i and flare_j into a collar of width $\ln 2$ and a contiguous $\overline{\text{cusp}_i}$ and $\overline{\text{flare}_j}$ respectively:

$$\text{cusp}_i = \text{collar}_i \sqcup \overline{\text{cusp}_i} \quad \text{and} \quad \text{flare}_j = \text{collar}_j \sqcup \overline{\text{flare}_j}.$$

In more detail the definitions are as follows. For a cusp (see Figure 2), a collar of width a is a set of points in $P(Y)$ whose distance from h_Y does not exceed a:

$$\text{collar}_a(Y) = \{z \in P(Y) \mid \rho(z, h_Y) \le a\}.$$

Fixing $a = \ln 2$ we have

$$\text{collar}(Y) = \{x + iy \mid 0 < x < 1, Y < y < 2Y\}.$$

In other words, $\text{collar}(Y)$ is a set of points between the horocycles h_Y and h_{2Y}.

Given a hyperbolic fundamental domain $F(\alpha)$ (see Figure 3) let η be the geodesic $\tau \to \eta(\tau) = ie^{\tau}$.

Introducing Fermi coordinates based on η we obtain the description of a flare as a surface $[\rho_0, \infty) \times [0, \kappa)$ with Riemannian metric

$$ds^2 = d\rho^2 + \cosh^2 \rho \, dt^2. \tag{3.14}$$

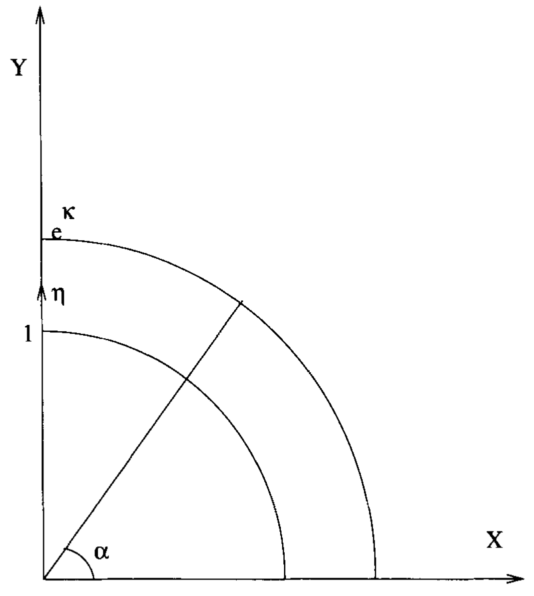

Figure 3. Hyperbolic fundamental domain.

Here ρ is the distance to the geodesic η, $\rho_0 = \rho(\alpha) = 1/\sin\alpha$ (see Figure 4 and formulas below). Using elementary hyperbolic geometry we easily obtain

$$\cosh\rho = 1/\sin\phi, \quad \sinh\rho = \cot\phi, \quad \tanh\rho = \cos\phi.$$

By a collar of width a in a flare (see Figure 5) we mean a set of points in $F(\alpha)$, whose distance from η does not exceed $\rho(\alpha) + a$:

$$\text{collar}_a(\alpha) = \{z \in F(\alpha) \mid \rho(\alpha) \leq \rho(z) \leq \rho(\alpha) + a\};$$

fixing again $a = \ln 2$ we have

$$\text{collar}(\alpha) = \{z \in F(\alpha) \mid \rho(\alpha) \leq \rho(z) \leq \rho(\alpha) + \ln 2\}.$$

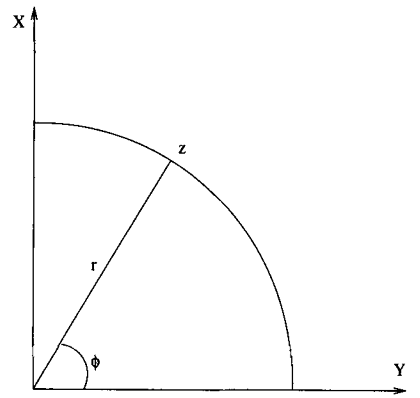

Figure 4. Polar coordinates in a flare.

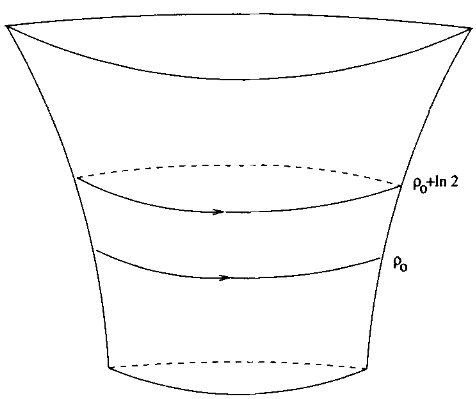

Figure 5. Flare and its collar.

So we have a decomposition

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

We define the compact part of $\mathcal{F}(1)$ as follows:

$$\mathcal{K}(1) = \overline{\mathcal{K}(1)} \cup \bigcup_{i \in Cu(1)} \text{collar}_i \cup \bigcup_{j \in Fl(1)} \text{collar}_j\,.$$

Now consider $\Lambda(p)$ and $\mathcal{F}(p) = \Lambda(p)\backslash\mathbf{H}$. As we have shown in section 2, for p large enough the deck group of covering π_p: $\mathcal{F}(p) \to \mathcal{F}(1)$ is $\mathrm{SL}_2(\mathbf{F}_p)$. We define

$$\overline{\mathcal{K}(p)} = \bigcup_{\gamma\in\Lambda(1)/\Lambda(p)} \gamma\overline{\mathcal{K}(1)}.$$

We proceed to examine the structure of the lifted cusps and flares.

As a surface, cusp is isometric to

$$]-\infty, Y] \times S^1 =]-\infty, Y] \times \mathbf{R}/[t \to t+1]$$

with the Riemanninan metric

$$ds^2 = d\rho^2 + e^{2\rho}dt^2.$$

The horocycles h_Y and h_{2Y} bounding the collar map to closed curves C_Y and C_{2Y} of lengths $1/Y$ and $1/2Y$ respectively. Now consider the lifting of a loop C_Y. Since the covering π_p: $X(p) \to X(1)$ is regular, the points in $X(p)$ with the same projection in $X(1)$ are inditinguishable from each other and each lift of the loop C_Y is conjugate to any other. Consequently

$$\pi_p^{-1}(C_Y) = \bigcup_{j=1}^{m} C^n_{Y_p},$$

where $C^n_{Y_p}$ is a loop of length n/Y_p and $mn = |\mathrm{SL}_2(\mathbf{F}_p)|$. We repeat the same argument for C_{2Y} and observe that while we have no control over the values of Y_p, the width of a collar, defined along the geodesic η, is preserved.

Making the same argument for a flare and repeating it for each flare and each cusp we obtain the following:

$$\pi_p^{-1}[\overline{\mathrm{cusp}_i} \sqcup \mathrm{collar}_i] = \bigcup_{l=1}^{m_i} [\overline{\mathrm{cusp}^{r_i}_{i,p}} \sqcup \mathrm{collar}^{r_i}_{i,p}],$$

$$\pi_p^{-1}[\overline{\mathrm{flare}_j} \sqcup \mathrm{collar}_j] = \bigcup_{k=1}^{n_j} [\overline{\mathrm{flare}^{s_j}_{j,p}} \sqcup \mathrm{collar}^{s_j}_{j,p}],$$

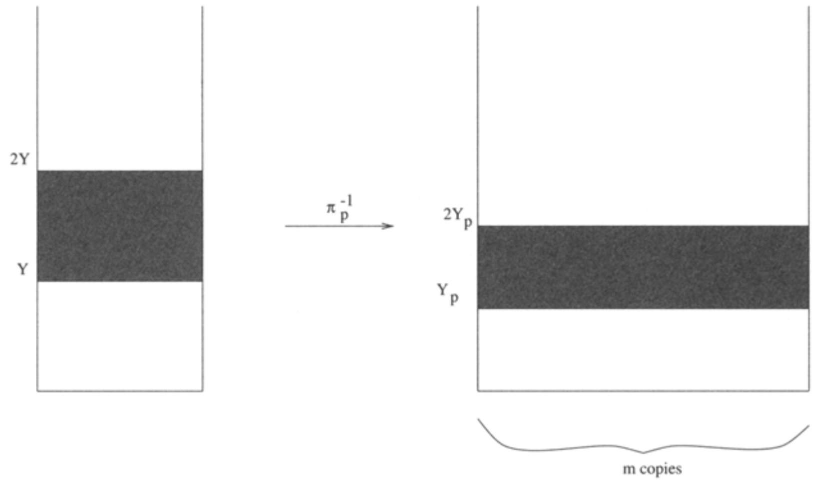

Figure 6. Lifting a cusp and its collar.

where $\overline{\text{cusp}_{i,p}^{r_i}}$ and $\text{collar}_{i,p}^{r_i}$ are contiguous cusp and collar of width $\ln 2$, and $\overline{\text{flare}_{j,p}^{s_j}}$ and $\text{collar}_{j,p}^{s_j}$ are contiguous flare and collar of width $\ln 2$, respectively. Moreover, for each i and each j we have

$$r_i m_i = n_j s_j = |\,\mathrm{SL}_2(\mathbf{F}_p)|.$$

Now collecting the lifting of the collars and lifting of $\overline{\mathcal{K}(1)}$ we obtain the compact part of $\mathcal{F}(p)$:

$$\mathcal{K}(p) = \pi_p^{-1}(\mathcal{K}(1)) = \overline{\mathcal{K}(p)} \cup \bigcup_{i\in Cu(1)} \bigcup_{l=1}^{m_i} \text{collar}_{i,p}^{r_i} \cup \bigcup_{j\in Fl(1)} \bigcup_{k=1}^{n_j} \text{collar}_{j,p}^{s_j}. \tag{3.15}$$

The collar lemma, to be proved in section 4, will be applied to each of the contiguous liftings of collars and cusps

$$\text{collar}_{i,p}^{r_i} \sqcup \overline{\text{cusp}_{i,p}^{r_i}} \tag{3.16}$$

and collars and flares

$$\text{collar}_{j,p}^{s_j} \sqcup \overline{\text{flare}_{j,p}^{s_j}}. \tag{3.17}$$

4. Collar lemma for cusp and flare

COLLAR LEMMA FOR CUSP. Given Γ, a geometrically-finite subgroup[10] of $SL_2(\mathbf{Z})$, let $f \in A_s(\Gamma\backslash\mathbf{H})$ be an automorphic form with respect to Γ for the eigenvalue $\lambda = s(1-s)$:

$$(\Delta + \lambda)f = 0,$$
$$f(\gamma z) = f(z) \quad \forall \gamma \in \Gamma.$$

Assume, moreover, that

$$f \in L^2(\Gamma\backslash\mathbf{H}) = \{f \in A(\Gamma\backslash\mathbf{H}) \mid \|f\| < \infty\}$$

where

$$\langle f, g\rangle \stackrel{\text{def}}{=} \int_F f(z)\overline{g(z)}d\mu$$

and normalized to have norm 1, $\|f\| = 1$.

We want to examine the behavior of f in cusps and flares of Γ. We begin with a cusp.

Let $P(Y)$ be a cuspidal fundamental domain, the semistrip

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

with a collar

$$C(Y) = \{x + iy \mid 0 < x < q, Y < y < 2Y\}.$$

LEMMA 4.1 (Collar Lemma for Cusp): *In the notation as above for any $Y \in (0,\infty)$ and $s \in (s_0, 1)$, where $s_0 > \frac{1}{2}$, there is a constant c_{s_0} (independent of Y) such that we have*

$$\frac{\int_Y^{2Y}\int_0^q |f(z)|^2 d\mu(z)}{\int_{2Y}^{\infty}\int_0^q |f(z)|^2 d\mu(z)} \geq c_{s_0} > 0.$$

Proof: We can assume that $q = 1$: our domain can be conjugated to such a one with Y getting mapped to Y/q. As discussed for example in [14], we have the following expansion:

$$f(z) = \hat{f}(0)y^{1-s} + \sum_{n\neq 0}\hat{f}(n)W_s(nz)$$

where

$$W_s(z) = 2y^{1/2}K_{s-1/2}(2\pi y)e(x).$$

---

10 We are interested in $\Lambda(p)$, but the results to be proved in this section hold for cusps and flares of any geometrically finite subgroup of $SL_2(\mathbf{R})$.

We get by Parseval

$$|\hat{f}(0)y^{1-s}|^2 + \sum_{n\neq 0} |\hat{f}(n)W_s(iny)|^2 = \int_0^1 |f(x+iy)|^2 dx.$$

Hence

$$\int_Y^\infty \int_0^1 |f(z)|^2 d\mu(z) = |\hat{f}(0)|^2 \int_Y^\infty y^{-2s} dy + \sum_{n\neq 0} |\hat{f}(n)|^2 \int_Y^\infty W_s^2(iny) y^{-2} dy. \tag{4.1}$$

Let $\nu = s - \frac{1}{2}$. For the zeroeth term we get

$$\frac{\int_Y^{2Y} y^{-2s} dy}{\int_{2Y}^\infty y^{-2s} dy} = \frac{Y^{-2\nu} - (2Y)^{-2\nu}}{(2Y)^{-2\nu}} = 2^{2\nu} - 1 > 0. \tag{4.2}$$

The lemma would thus follow from the following claim:

CLAIM 4.1: *For any $Y \in (0,\infty)$, $\nu \in (\nu_0, 1/2)$ we have*

$$\frac{\int_Y^{2Y} \frac{K_\nu^2(x)}{x} dx}{\int_{2Y}^\infty \frac{K_\nu^2(x)}{x} dx} \geq c_{\nu_0} > 0. \tag{4.3}$$

Remark: 1. The K-Bessel function K_ν is real for ν real.

2. When we apply this claim to the n-th term of equation (4.1) the n-dependence scales away, so this suffices.

Proof: We recall the following facts regarding the Bessel function $K_\nu(x)$ [48]. One way of defining it is via the series

$$K_\nu(x) \stackrel{\text{def}}{=} \frac{\pi}{2\sin(\pi\nu)} (I_{-\nu}(x) - I_\nu(x)), \tag{4.4}$$

where

$$I_\nu(x) \stackrel{\text{def}}{=} \sum_{k=0}^\infty \frac{1}{k!\Gamma(k+1+\nu)} \left(\frac{x}{2}\right)^{\nu+2k} = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^\infty \frac{(\frac{x^2}{4})^{2k}}{k!\Gamma(k+1+\nu)}. \tag{4.5}$$

The following asymptotic estimates hold:

$$K_\nu(x) \sim \frac{\Gamma(\nu)}{2} \left(\frac{x}{2}\right)^{-\nu} \quad \text{as } x \to 0. \tag{4.6}$$

For $0 < x < 1$

$$K_\nu(x) = \frac{2^{\nu-1}\Gamma(\nu)}{x^\nu} + O(x^{2-\nu}), \tag{4.7}$$

where the implied constant is uniform for $\nu \in (\nu_0, \frac{1}{2})$ and $x \in (0, \frac{1}{2})$.

For $x > 1 + \nu^2$

$$K_\nu(x) = \left(\frac{\pi}{2x}\right)^{\frac{1}{2}} e^{-x}\left(1 + O\left(\frac{1+\nu^2}{x}\right)\right), \tag{4.8}$$

where the implied constant is uniform for $\nu \in (\nu_0, \frac{1}{2})$ and $x > 2$.

More precisely, with $\mu = 4\nu^2$

$$K_\nu(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x}\left\{1 + \frac{\nu - 1}{8x} + \frac{(\nu-1)(\nu-9)}{2!(8x)^2} + \cdots\right\}.$$

Occasionally it would be more convenient to work with the integral representation of $K_\nu(x)$:

$$K_\nu(x) = \int_0^\infty e^{-x\cosh t}\cosh(\nu t)dt. \tag{4.9}$$

From the integral representation the following two properties easily follow:

(1) For fixed ν throughout the x interval $(0, \infty)$ the function $K_\nu(x)$ is positive and decreasing.

(2) If $x > 0$ is fixed, then throughout the ν interval $(0, \infty)$ the function $K_\nu(x)$ is increasing.

We divide the analysis of (4.3) into three parts depending on the value of Y. We begin with large values of Y, $Y > A$, where the constant A is fixed below. Equation (4.8) and the comment following it imply that there is a constant C_2 such that

$$\left|K_\nu(x) - \left(\frac{\pi}{2x}\right)^{\frac{1}{2}}\right| \le C_2\frac{e^{-x}}{x^{\frac{3}{2}}} \tag{4.10}$$

uniformly for all $x \ge 2$ and $\nu \in (\nu_0, \frac{1}{2})$. Let

$$A = \max(4C_2, 2) + 1. \tag{4.11}$$

For $Y > A$ we have

$$\left|\int_Y^{2Y} \frac{K_\nu^2(x)}{x}dx - -\frac{\pi}{2}\int_Y^{2Y}\frac{e^{-2x}}{x^2}dx\right| \le C_2^2\int_Y^{2Y}\frac{e^{-2x}}{x^4}dx,$$

from which we obtain

$$\begin{aligned}\int_Y^{2Y}\frac{K_\nu^2(x)}{x}dx &\ge \frac{\pi}{2}\int_Y^{2Y}\frac{e^{-2x}}{x^2}dx - C_2^2\int_Y^{2Y}\frac{e^{-2x}}{x^4}dx \\ &\ge \frac{\pi}{2}\frac{1}{(2Y)^2}\int_Y^{2Y} e^{-2x}dx - C_2^2\frac{1}{Y^4}\int_Y^{2Y} e^{-2x}dx \\ &\ge \frac{\pi}{2}\frac{e^{-2Y}}{16Y^2} - C_2^2\frac{e^{-2Y}}{2Y^4}.\end{aligned}$$

Similarly we have

$$\left|\int_{2Y}^{\infty} \frac{K_\nu^2(x)}{x} dx - -\frac{\pi}{2}\int_{2Y}^{\infty} \frac{e^{-2x}}{x^2} dx\right| \le C_2^2 \int_{2Y}^{\infty} \frac{e^{-2x}}{x^4} dx,$$

from which we obtain

$$\begin{aligned}\int_{2Y}^{\infty} \frac{K_\nu^2(x)}{x} dx &\le \frac{\pi}{2}\int_{2Y}^{\infty} \frac{e^{-2x}}{x^2} dx + C_2^2 \int_{2Y}^{\infty} \frac{e^{-2x}}{x^4} dx \\ &\ge \frac{\pi}{2}\frac{e^{-4Y}}{8Y^2} + C_2^2 \frac{e^{-4Y}}{8Y^4}.\end{aligned}$$

Hence for $Y > A$ we have

$$\frac{\int_{Y}^{2Y} \frac{K_\nu^2(x)}{x} dx}{\int_{2Y}^{\infty} \frac{K_\nu^2(x)}{x} dx} \ge \frac{\frac{\pi}{2}\frac{e^{-2Y}}{16Y^2} - C_2^2 \frac{e^{-2Y}}{2Y^4}}{\frac{\pi}{2}\frac{e^{-4Y}}{8Y^2} + C_2^2 \frac{e^{-4Y}}{8Y^4}} \ge e^{2Y} \frac{1 - \frac{16C_2^2}{\pi Y^2}}{1 + \frac{2C_2^2}{\pi Y^2}};$$

the last expression, since $Y > A > 4C_2$, is no less than

$$e^{4C_2} \frac{1 - \frac{16C_2^2}{\pi 16 C_2^2}}{1 + \frac{2}{16\pi}} \ge \frac{2}{3} e^{4C_2} > 0.$$

Next we consider the case of small Y, $0 < Y < a/2$, where the the value of a is fixed below. Equation (4.7) and the comment following it imply that there is a constant C_1 such that

$$\left|K_\nu(x) - \frac{2^{\nu-1}\Gamma(\nu)}{x^\nu}\right| \le C_1 x^{2-\nu} \tag{4.12}$$

uniformly for all $x \in (0, \frac{1}{2})$ and $\nu \in (\nu_0, \frac{1}{2})$. Let $b_\nu = 2^{\nu-1}\Gamma(\nu)$. The function

$$h(\nu) = \left[\frac{b_\nu^2}{10\nu}\left(1 - \frac{1}{2^{2\nu}}\right)\right]^{\frac{1}{4}} \tag{4.13}$$

is positive and continuous on $[\nu_0, \frac{1}{2}]$, hence its minimum on this interval is positive. We now fix the value of a:

$$a = \frac{\min_{\nu \in [\nu_0, \frac{1}{2}]} h(\nu)}{\sqrt{C_1}}. \tag{4.14}$$

Equation (4.12) yields

$$\left|\int_{Y}^{2Y} \frac{K_\nu^2(x)}{x} dx - b_\nu^2 \int_{Y}^{2Y} x^{-2\nu-1} dx\right| \le C_1^2 \int_{Y}^{2Y} x^{3-2\nu} dx,$$

therefore we have

$$\int_Y^{2Y} \frac{K_\nu^2(x)}{x} dx \geq \frac{b_\nu^2}{2\nu} Y^{-2\nu}(1 - \frac{1}{2^{2\nu}}) - 5C_1^2 Y^{4-2\nu}.$$

Similarly we obtain

$$\left| \int_{2Y}^{a} \frac{K_\nu^2(x)}{x} dx - b_\nu^2 \int_{2Y}^{a} x^{-2\nu-1} dx \right| \leq C_1^2 \int_{2Y}^{a} x^{3-2\nu} dx,$$

and therefore we have

$$\int_{2Y}^{a} \frac{K_\nu^2(x)}{x} dx \leq \frac{b_\nu^2}{2\nu} \frac{Y^{-2\nu}}{2^{2\nu}} + \frac{C_1^2 a^{4-2\nu}}{3}.$$

On the other hand, for $\nu \in (\nu_0, \frac{1}{2})$ we have, using property (2) (following (4.9)),

$$\int_a^\infty \frac{K_\nu^2(x)}{x} dx \leq \int_a^\infty \frac{K_{\frac{1}{2}}^2(x)}{x} dx = C_a.$$

Hence we obtain

$$\frac{\int_Y^{2Y} \frac{K_\nu^2(x)}{x} dx}{\int_{2Y}^\infty \frac{K_\nu^2(x)}{x} dx} \geq \frac{\frac{b_\nu^2 Y^{-2\nu}}{2\nu 2^{2\nu}}(2^{2\nu} - 1) - 5C_1^2 Y^{4-2\nu}}{\frac{b_\nu^2 Y^{-2\nu}}{2\nu 2^{2\nu}} + \frac{C_1^2 a^{4-2\nu}}{3} + C_a}.$$

Now since $2Y$ is less than a the latter expression is no less than

$$\frac{b_\nu^2(2^{2\nu} - 1) - 10\nu 2^{2\nu} C_1^2 a^4}{b_\nu^2 + C_1^2 a^4 + C_a a^{2\nu}} > 0,$$

which is positive due to our choice of a in (4.14).

Finally we consider the case of $Y \in [\frac{a}{2}, A]$. Let $R(\nu, Y)$ be the following function:

$$R(\nu, Y) = \frac{\int_Y^{2Y} \frac{K_\nu^2(x)}{x} dx}{\int_{2Y}^\infty \frac{K_\nu^2(x)}{x} dx}.$$

It is a positive, continuous function of ν and Y, therefore the minimum it achieves on a compact set $\nu \in [\nu_0, \frac{1}{2}]$, $Y \in [\frac{a}{2}, A]$ is positive. This observation completes the proof of Claim 4.1. ∎

Collar Lemma for Flare. As detailed in Section 3, flare can be viewed as a surface $[\rho_0, \infty) \times [0, \kappa)$ with Riemannian metric

$$ds^2 = d\rho^2 + \cosh^2 \rho \, dt^2;$$

the Riemannian measure is easily seen to be

$$d\mu = \cosh\rho \, d\rho \, dt. \tag{4.15}$$

Since the Laplacian in coordinates

$$ds^2 = ed\xi^2 + 2fd\xi d\eta + gd\eta^2$$

is given by the formula (cf. (64) on p. 225 of [4])

$$\Delta u = \frac{1}{\sqrt{eg - f^2}}\left(\frac{\partial}{\partial\xi}\left(\frac{gu_\xi - fu_\eta}{\sqrt{eg-f^2}}\right) + \frac{\partial}{\partial\eta}\left(\frac{eu_\eta - fu_\xi}{\sqrt{eg-f^2}}\right)\right),$$

we obtain the following expression for the Laplacian in our coordinates:

$$\Delta F = \frac{\partial^2 F}{\partial\rho^2} + \frac{1}{\cosh^2\rho}\frac{\partial^2 F}{\partial t^2} + \frac{\sinh\rho}{\cosh\rho}\frac{\partial F}{\partial\rho}. \tag{4.16}$$

Let F again be an L^2 automorphic form with eigenvalue $\lambda = s(1-s)$. In this subsection we will prove the following

LEMMA 4.2 (Collar Lemma for Flare): *For any $\rho_0 \in (0,\infty)$, $\frac{1}{2} < s_0 < s < 1$ there is a constant c_{s_0} (independent of ρ_0) such that we have*

$$\frac{\int_{\rho_0}^{\rho_0+\ln 2}\int_0^\kappa |F(z)|^2 d\mu}{\int_{\rho_0+\ln 2}^{\infty}\int_0^\kappa |F(z)|^2 d\mu} \geq c_s \geq c_{s_0} > 0.$$

In the flare we have

$$(\Delta + \lambda)F = 0, \quad F(e^\kappa z) = F(z).$$

Separating variables we obtain

$$F(z) = f(\rho)g(t), \quad g(t+\kappa) = g(t).$$

Expanding g in a Fourier series yields

$$F(z) = \sum f_n(\rho)e^{2\pi int/\kappa}c_n.$$

Let $\bar{n} = 2\pi in/\kappa$. Using (4.16) we obtain the following ordinary differential equation for $f_n(\rho)$:

$$f_n''(\rho) + \frac{\sinh\rho}{\cosh\rho}f_n'(\rho) + \left(s(1-s) + \frac{1}{\cosh^2\rho}\left(\frac{2\pi in}{\kappa}\right)^2\right)f_n = 0. \tag{4.17}$$

Introducing the substitution

$$y = \tanh\rho = \cos\phi \tag{4.18}$$

we obtain the equivalent ODE:

$$(1-y^2)\frac{d^2 f_n}{dy^2} - y\frac{df_n}{dy} + \Big(\frac{s(1-s)}{1-y^2} + (\tilde{n})^2\Big) f_n = 0.$$

Let us make another substitution

$$f_n = (1-y^2)^{\frac{1}{4}} v_n = \sqrt{\sin\phi}\, v_n. \tag{4.19}$$

The differential equation for v_n takes the form

$$(1-y^2)\frac{d^2 v_n}{dy^2} - 2y\frac{dv_n}{dy} + \Big(-\frac{1}{4} + (\tilde{n})^2 - (\frac{1}{4} - \lambda)\frac{1}{1-y^2}\Big) v_n = 0.$$

Now we recognize it as a Legendre differential equation

$$(1-x^2)\frac{d^2 v}{dx^2} - 2x\frac{dv}{dx} + (\nu(\nu+1) - \mu^2(1-x^2)^{-1})v = 0 \tag{4.20}$$

with the following values of ν and μ:

$$\mu = \frac{1}{2} - s, \tag{4.21}$$

$$\nu_n = -\frac{1}{2} + \frac{2\pi i n}{\kappa}. \tag{4.22}$$

The variable x in our case is real and lies in $(0,1)$. This equation has two linearly independent solutions, called associated Legendre functions of the first and second kind and denoted by $P_\nu^\mu(x)$ and $Q_\nu^\mu(x)$ [36]. The behavior of these functions at the singular point 1, expressed in terms of the leading term of their asymptotic expansions, is as follows:

$$P_\nu^\mu(x) \sim \frac{2^{\mu/2}(1-x)^{-\mu/2}}{\Gamma(1-\mu)},$$

$$Q_{nu}^\mu(x) \sim 2^{\mu/2}\Gamma(\mu)\cos(\pi\mu)(1-x)^{-\mu/2} \quad \text{if } \Re(\mu) > 0,$$

$$Q_\nu^\mu(x) \sim \frac{\Gamma(-\mu)\Gamma(\nu+\mu+1)}{2^{1+\mu/2}\Gamma(\nu-\mu+1)}(1-x)^{\mu/2} \quad \text{if } \Re(\mu) < 0.$$

Since in our case $\mu = \frac{1}{2} - s < 0$, $F(z)$ lying in L^2 forces us to choose the decaying solution $P_\nu^\mu(x)$.

Thus finally we obtain for f_n

$$f_n(y) = (1-y^2)^{\frac{1}{4}} P_\nu^\mu(y) = \sqrt{\sin\phi}\, P_\nu^\mu(\cos\phi),$$

where μ and ν are as in (4.21) and (4.22). Returning to the variable ρ we have

$$f_n(\rho) = \frac{1}{\sqrt{\cosh\rho}} P_\nu^\mu(\tanh\rho).$$

By the Parseval identity we get

$$\int_{\rho_0}^\infty \int_0^\kappa |F(z)|^2 d\mu = \sum_n |c_n|^2 \int_{\rho_0}^\infty (P_{-\frac{1}{2}+\bar{n}}^{\frac{1}{2}-s}(\tanh\rho))^2 d\rho.$$

The collar lemma for flare would thus follow from the following

LEMMA 4.3: *For any $\rho_0 \in (0,\infty)$, $\frac{1}{2} < s_0 < s < 1$ we have (with $c_{s_0} = 1 - (\frac{1}{2})^{2s_0-1}$)*

$$\frac{\int_{\rho_0}^{\rho_0+\ln 2} (P_{-\frac{1}{2}+\bar{n}}^{\frac{1}{2}-s}(\tanh\rho))^2 d\rho}{\int_{\rho_0+\ln 2}^\infty (P_{-\frac{1}{2}+\bar{n}}^{\frac{1}{2}-s}(\tanh\rho))^2 d\rho} \geq c_s \geq c_{s_0} > 0. \tag{4.23}$$

The proof of Lemma 4.3 in turn hinges on the following claim:

CLAIM 4.2: *For any $\rho > 0$*

$$(P_{-\frac{1}{2}+\bar{n}}^{\frac{1}{2}-s}(\tanh\rho))^2 = \frac{1}{(\cosh\rho)^{2s-1}} \sum_{m=0}^\infty \frac{a_m(n)}{(\cosh\rho)^{2m}}, \tag{4.24}$$

where $a_m(n) > 0$.

Proof: For $\Re(\mu+\nu) < 0$, $\Re(\nu-\mu) > -1$ and $0 < x < 1$ (the case we are in) we have the following remarkable expression for the product of two Legendre functions, involving our old friends I_μ and K_ν ([28], p. 191; [10], p. 667):

$$(P_\nu^\mu(x))^2 = \frac{2\int_0^\infty (I_{-\mu}(-\frac{1}{2}t(1-x^2)^{\frac{1}{2}}))^2 K_{2\nu+1}(t)dt}{\Gamma(\nu+1-\mu)\Gamma(-\nu-\mu)}.$$

In our case with $\mu = \frac{1}{2} - s$, $\nu = -\frac{1}{2} + \bar{n}$, and $x = \tanh\rho$ this yields

$$(P_{-\frac{1}{2}+\bar{n}}^{\frac{1}{2}-s}(\tanh\rho))^2 = 2(\Gamma(s+\bar{n})\Gamma(s-\bar{n}))^{-1} \int_0^\infty \left(I_{s-\frac{1}{2}}\left(\frac{t}{2\cosh\rho}\right)\right)^2 K_{2\bar{n}}(t)dt. \tag{4.25}$$

Now for $I_\nu(x)$ we have

$$I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^\infty \frac{(x^2/4)^k}{k!\Gamma(\nu+k+1)}.$$

Computing the coefficient of $(\frac{x}{2})^{2\nu+2m}$ in the product of two convergent series (cf. [31])

$$\sum_{k=0}^{\infty} \frac{(\frac{x}{2})^{\nu+2k}}{k!\Gamma(\nu+k+1)} \times \sum_{j=0}^{\infty} \frac{(\frac{x}{2})^{\nu+2j}}{j!\Gamma(\nu+j+1)}$$

we obtain the following expression for $I_\nu^2(x)$:

$$I_\nu^2(x) = \left(\frac{x}{2}\right)^{2\nu} \sum_{m=0}^{\infty} \frac{x^{2m}\Gamma(2\nu+2m+1)}{2^{2m}m!\Gamma(2\nu+m+1)(\Gamma(\nu+m+1))^2}. \tag{4.26}$$

On the other hand, using the integral representation for $K_\nu(x)$ when $\Re(\nu) > -\frac{1}{2}$

$$K_\nu(x) = \frac{\pi^{\frac{1}{2}}(\frac{x}{2})^\nu}{\Gamma(\nu+\frac{1}{2})} \int_1^\infty e^{-xt}(t^2-1)^{\nu-\frac{1}{2}}dt$$

and the Beta function integral, we obtain for $\alpha > -1$, $\Re(\alpha+1\pm\nu) > 0$

$$\int_0^\infty t^\alpha K_\nu(t)dt = 2^{\alpha-1}\Gamma\Big(\frac{1+\alpha+\nu}{2}\Big)\Gamma\Big(\frac{1+\alpha-\nu}{2}\Big). \tag{4.27}$$

We now expand $(I_{s-\frac{1}{2}}(\frac{t}{2\cosh\rho}))^2$, using (4.26), and substitute it in (4.25) to obtain

$$\int_0^\infty \Big(I_{s-\frac{1}{2}}\Big(\frac{t}{2\cosh\rho}\Big)\Big)^2 K_{2\bar{n}}(t)dt =$$

$$\int_0^\infty \frac{1}{(\cosh\rho)^{2s-1}}\Big(\frac{t}{4}\Big)^{2s-1} \sum_{m=0}^{\infty} \frac{t^{2m}\Gamma(2s+2m)K_{2\bar{n}}(t)}{(\cosh\rho)^{2m}m!4^{2m}\Gamma(2s+m)(\Gamma(s+m+\frac{1}{2}))^2}$$

$$= \frac{1}{(\cosh\rho)^{2s-1}} \sum_{m=0}^{\infty} \frac{b_m(n)}{(\cosh\rho)^{2m}},$$

where

$$b_m(n) = \frac{\Gamma(2s+2m)\int_0^\infty t^{2m+2s+1}K_{2\bar{n}}(t)dt}{4^{2m+2s-1}m!\Gamma(2s+m)(\Gamma(s+m+\frac{1}{2}))^2}.$$

Using (4.27) we have

$$\int_0^\infty t^{2m+2s+1}K_{2\bar{n}}(t)dt = 2^{2m+2s-2}\Gamma(m+s+\bar{n})\Gamma(m+s-\bar{n}).$$

So finally we obtain

$$(P^{\frac{1}{2}-s}_{-\frac{1}{2}+\bar{n}}(\tanh\rho))^2 = \frac{1}{(\cosh\rho)^{2s-1}} \sum_{m=0}^{\infty} \frac{a_m(n)}{(\cosh\rho)^{2m}}, \tag{4.28}$$

where

$$a_m(n) = \frac{\Gamma(2s+2m)\Gamma(m+s+\bar{n})\Gamma(m+s-\bar{n})}{2^{2m+2s-1}m!\Gamma(s+\bar{n})\Gamma(s-\bar{n})\Gamma(2s+m)(\Gamma(s+m+\frac{1}{2}))^2}. \tag{4.29}$$

This proves[11] Claim 4.2.

Returning now to the proof of Lemma 4.3 we have, using (4.28),

$$\begin{aligned}\frac{\int_{\rho_0}^{\rho_0+\ln 2}\left(P^{\frac{1}{2}-s}_{-\frac{1}{2}+\bar{n}}(\tanh\rho)\right)^2 d\rho}{\int_{\rho_0+\ln 2}^{\infty}(P^{\frac{1}{2}-s}_{-\frac{1}{2}+\bar{n}}(\tanh\rho))^2 d\rho} &= \frac{\int_{\rho_0}^{\rho_0+\ln 2}\frac{d\rho}{(\cosh\rho)^{2s-1}}\sum_{m=0}^{\infty}\frac{a_m(n)}{(\cosh\rho)^{2m}}}{\int_{\rho_0+\ln 2}^{\infty}\frac{d\rho}{(\cosh\rho)^{2s-1}}\sum_{m=0}^{\infty}\frac{a_m(n)}{(\cosh\rho)^{2m}}}\\ &= \frac{\sum_{m=0}^{\infty}a_m(n)\int_{\rho_0}^{\rho_0+\ln 2}\frac{d\rho}{(\cosh\rho)^{2m+2s-1}}}{\sum_{m=0}^{\infty}a_m(n)\int_{\rho_0+\ln 2}^{\infty}\frac{d\rho}{(\cosh\rho)^{2m+2s-1}}}\\ &= \frac{\sum_{m=0}^{\infty}a_m(n)b_m(\rho_0,s)}{\sum_{m=0}^{\infty}a_m(n)c_m(\rho_0,s)}\end{aligned}$$

with

$$b_m(\rho_0,s) = \int_{\rho_0}^{\rho_0+\ln 2}\frac{d\rho}{(\cosh\rho)^{2m+2s-1}} \text{ and } c_m(\rho_0,s) = \int_{\rho_0+\ln 2}^{\infty}\frac{d\rho}{(\cosh\rho)^{2m+2s-1}}.$$

Now $\cosh\rho \geq \frac{1}{2}e^{\rho}$ and for $\rho \geq \rho_0 > 0$ we have

$$\cosh\rho \leq \frac{1}{2}(1+e^{-2\rho_0})e^{\rho} \leq e^{\rho}.$$

Hence we have

$$b_m(\rho_0,s) \geq \int_{\rho_0}^{\rho_0+\ln 2} e^{-\rho(2m+2s-1)}d\rho = \frac{1-2^{-(2m+2s-1)}}{2m+2s-1}e^{-\rho_0(2m+2s-1)},$$

and similarly for $c_m(\rho_0,s)$ we obtain

$$c_m(\rho_0,s) \leq \int_{\rho_0+\ln 2}^{\infty}\frac{d\rho}{(\frac{1}{2}e^{\rho})^{2m+2s-1}} = \frac{1}{2m+2s-1}e^{-\rho_0(2m+2s-1)}.$$

Therefore we have

$$\frac{b_m(\rho_0,s)}{c_m(\rho_0,s)} \geq 1-\left(\frac{1}{2}\right)^{2m+2s-1} \geq 1-\left(\frac{1}{2}\right)^{2s-1} \geq 1-\left(\frac{1}{2}\right)^{2s_0-1} > 0.$$

---

11 One can give a shorter proof using the following expansion, valid for $0 < x < 1$:

$$\Gamma(1-\mu)P^{\mu}_{\nu}(x) = 2^{\mu}(1-x^2)^{-\frac{1}{2}\mu}F\left(\frac{1}{2}+\frac{1}{2}\nu-\frac{1}{2}\mu, -\frac{1}{2}\nu-\frac{1}{2}\mu; 1-\mu; 1-x^2\right).$$

We felt, however, that the longer road was sufficiently more scenic to justify taking it.

Consequently we obtain

$$\frac{\int_{\rho_0}^{\rho_0+\ln 2}\left(P^{\frac{1}{2}-s}_{-\frac{1}{2}+\tilde{n}}(\tanh\rho)\right)^2 d\rho}{\int_{\rho_0+\ln 2}^{\infty}\left(P^{\frac{1}{2}-s}_{-\frac{1}{2}+\tilde{n}}(\tanh\rho)\right)^2 d\rho} \geq 1-\left(\frac{1}{2}\right)^{2s_0-1} > 0,$$

concluding the proof. ∎

5. Further analytic preparations

Computation of Convolution. A continuous compactly supported function $k(z,w)$ on $\mathbf{H}\times\mathbf{H}$ is called a **point pair invariant** if

$$k(gz,gw)=k(z,w) \quad \forall g\in \mathrm{SL}_2(\mathbf{R}).$$

Such a function depends only on hyperbolic distance, consequently we can set

$$k(z,w)=k(u(z,w))$$

where $k(u)$ is a function in one variable $u\geq 0$ and u is given by equation (3.4). Given a geometrically finite Fuchsian group Γ we define an automorphic kernel

$$K(z,w)=\sum_{\gamma\in\Gamma}k(z,\gamma w). \tag{5.1}$$

Given two pont-pair invariants k_1, k_2 we define their convolution product k_1*k_2 as follows:

$$k_1*k_2(z,w)=\int_{\mathbf{H}}k_1(z,x)k_2(x,w)\,d\mu(x). \tag{5.2}$$

Let us now consider the following point-pair invariant:

$$k_1(z,w)=\begin{cases}1 & \text{if } u(z,w)\leq (X-2)/4,\\ 0 & \text{if } u(z,w)>(X-2)/4.\end{cases} \tag{5.3}$$

The corresponding automorphic kernel counts the number of lattice points in hyperbolic circle problem:

$$K_1(z,w)=\sum_{\gamma\in\Gamma}k_1(z,\gamma w)=\#\{\gamma\in\Gamma: 4u(z,\gamma w)+2\leq X\}.$$

In particular, by taking $z=w=i$ and using (3.7) we have

$$\begin{aligned}P_1(X,\Gamma)=&\#\{\gamma\in\Gamma:4u(i,\gamma i)+2\leq X\}\\ =&\#\{\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\Gamma: a^2+b^2+c^2+d^2\leq X\}\\ =&\#\{\gamma\in\Gamma:\|\gamma\|^2\leq X\}.\end{aligned}$$

Let us now compute the convolution $k = k_1 * \overline{k_1}$ (cf. Sarnak and Xue [39]). It will be more convenient to work in $\mathbf{D}$. We have

$$k(z,w) = \int_{\mathbf{D}} k_1(z,x) k_1(x,w)\, d\mu(x).$$

In terms of hyperbolic distance ρ

$$k_1(z,w) = \begin{cases} 1 & \text{if } \rho(z,w) \le T, \\ 0 & \text{if } \rho(z,w) > T, \end{cases}$$

where $e^T + e^{-T} = X$ or $\cosh T = X/2$.

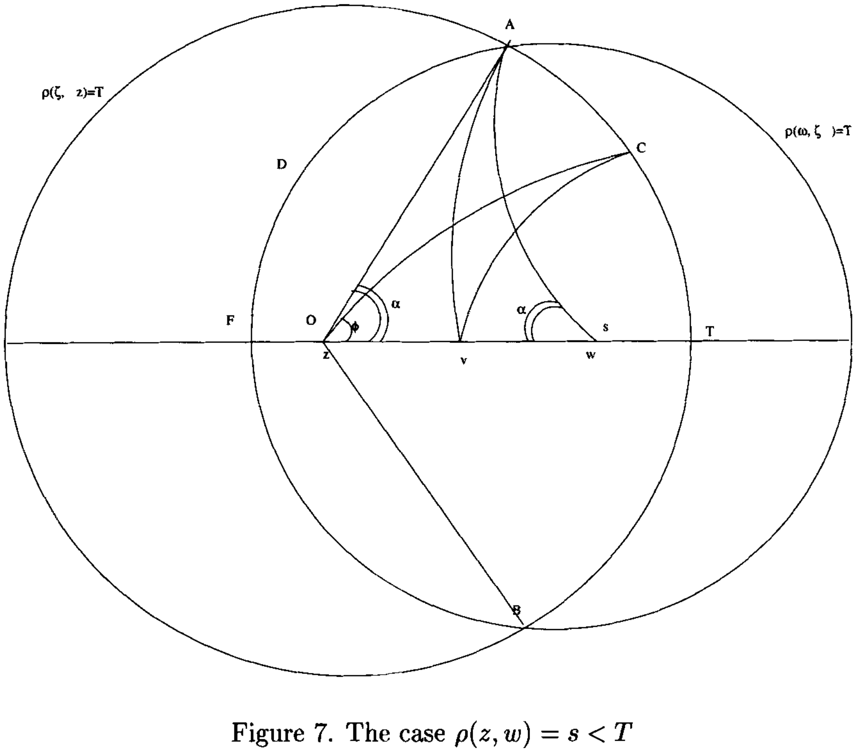

Figure 7. The case $\rho(z,w) = s < T$

Now we observe that

$$k(z,w) = \text{h-area}\,[B(z,T) \cap B(w,T)],$$

where $B(z,T)$ is a hyperbolic disc with center z and hyperbolic radius T. Let $E = B(z,T) \cap B(w,T)$. It is clear that $k(z,w) = 0$ if $\rho(z,w) > 2T$. Suppose

$\rho(z,w) < 2T$. Without loss of generality we can assume that $z = 0$. The set of points ζ satisfying $\rho(0,\zeta) = T$ is a Euclidean circle at 0 with center at 0 and radius R, where

$$\sinh(T/2) = \frac{R}{\sqrt{1-R^2}},$$

or equivalently

$$\tanh(T/2) = R.$$

If $\rho(z,w) = s$, w has Euclidean coordinates $(\tanh(s/2), 0)$. The set of points $\rho(w,\zeta) = T$ is a Euclidean circle with center $(a_0, 0)$ and radius R_0, where (with $d = \tanh(s/2)$)

$$a_0 = d\frac{1-R^2}{1-R^2d^2} \quad \text{and} \quad R_0 = R\frac{1-d^2}{1-R^2d^2}.$$

Using hyperbolic cosine law (3.12) we obtain

$$\begin{aligned}\cosh T &= \cosh s \cosh T - \sinh s \sinh T \cos\alpha,\\ \cos\alpha &= \frac{\cosh s\cosh T - \cosh T}{\sinh s \sinh T}\\ &= \frac{\cosh T(\cosh s - 1)}{\sinh s\sinh T} = \frac{\tanh(s/2)}{\tanh T}.\end{aligned}$$

Let $v = (s/2, 0)$ and suppose c lies on the arc ATB (see Figure 7). The angle $\phi = \widehat{CZV} < \alpha$. Let $r(\phi) = \rho(z, C)$; we have

$$\begin{aligned}\cosh r(\phi) &= \cosh(s/2)\cosh T - \sinh(s/2)\sinh T\cos\phi\\ &\le \cosh(s/2)\cosh T - \sinh(s/2)\sinh T\cos\alpha\\ &= \cosh(s/2)\cosh T - \sinh(s/2)\sinh T\Big(\frac{\tanh(s/2)}{\tanh T}\Big)\\ &= \frac{\cosh T}{\cosh(s/2)}.\end{aligned}$$

Similarly, if $D \in AFB$, the left arc, we get the same estimate by considering the triangle Dvw to obtain

$$\cosh\rho(v,D) < \frac{\cosh T}{\cosh(s/2)}.$$

Thus $E \subset B(v,r)$ with

$$\cosh r = \frac{\cosh T}{\cosh(s/2)},$$

a hyperbolic disc whose area, by (3.10), is

$$4\pi\sinh^2(r/2) = 2\pi(\cosh r - 1) = 2\pi\Big(\frac{\cosh T}{\cosh(s/2)} - 1\Big) \ll 2\pi e^{T-s/2}.$$

Thus we have proved the following

PROPOSITION 5.1: *For k defined as above we have*

$$k(z,w) \ll e^{T-s/2} \quad \text{if } \rho(z,w) \le 2T, \tag{5.4}$$

and is 0 for $\rho(z,w) > 2T$.

PRETRACE INEQUALITY. Now we recall the following crucial fact about point-pair invariants.

THEOREM (Selberg): *Suppose that ϕ is an eigenfunction of the Laplacian with $\Delta\phi = \lambda\phi$. Then ϕ is also an eigenfunction of the integral operator corresponding to any point-pair invariant, and its eigenvalue in the latter context depends only on λ, and not on ϕ; that is there exists a function h, called spherical function, or Selberg/Harish-Chandra transform, defined on the set of all eigenvalues (which in our case is $\mathbf{C}$), such that*

$$\int_{\mathbf{H}} k(x,y)\phi(y)dy = h(\lambda)\phi(x). \tag{5.5}$$

The Selberg/Harish-Chandra transform is computed in the following three steps:

$$q(v) = \int_v^{+\infty} k(u)(u-v)^{-\frac{1}{2}}\,du,$$
$$g(r) = 2q(\sinh(r/2)^2),$$
$$h(t) = \int_{-\infty}^{+\infty} e^{irt} g(r)\,dr.$$

In our case of $k_1 = \chi_{[0,(X-2)/4)}$ such computation yields (see, for example, [14])

$$h_1(s) = \sqrt{\pi}\frac{\Gamma(s-1/2)}{\Gamma(s+1)}X^s + O(X^{1/2}) \quad \text{for } 1/2 < s \le 1, \tag{5.6}$$

where $\lambda = s(s-1)$.

Given two point-pair invariants, their convolution product $k_1 * k_2$ was defined in (5.2) as

$$k_1 * k_2(z,w) = \int_{\mathbf{H}} k_1(z,x)k_2(x,w)\,d\mu(x).$$

For the automorphic kernel we have

$$K(z,w) = \int_{\Gamma\backslash\mathbf{H}} K_1(z,x)K_2(x,w)\,d\mu(x); \tag{5.7}$$

and the Selberg/Harish-Chandra transform of $k_1 * \overline{k_2}$ is $h_1(\lambda)\overline{h_2(\lambda)}$.

In particular, if we take $k_2 = k_1$ we obtain

$$\int_{\mathbf{H}} k_1 * \overline{k_1}(x,y)\phi(y)\,d\mu(y) = |h_1(\lambda)|^2\phi(x). \tag{5.8}$$

We are now ready to prove the following proposition.

PROPOSITION 5.2 (Pretrace Inequality): *Suppose Γ is a geometrically finite Fuchsian group. Let $k = k_1 * \overline{k_1}$ as above and $K(z,w) = \sum_{\gamma\in\Gamma} k(z,\gamma w)$. Let $\lambda_0, \lambda_1, \ldots, \lambda_j$ be the discrete eigenvalues below $\frac{1}{4}$; $\phi_0, \phi_1, \ldots, \phi_j$ be the corresponding eigenfunctions. Then*

$$K(z,z) \geq \sum_{\lambda_i<1/4} |h_1(\lambda_i)|^2|\phi_i(z)|^2.$$

Proof: As was shown by Lax and Phillips in [21], the spectrum of the Laplace operator on $F = \Gamma\backslash\mathbf{H}$ in $[0,1/4)$ consists of a finite number of eigenvalues. (This follows from the energy form being positive definite on a subspace of finite codimension.) Let

$$V_1 = L^2_d(F) = \bigoplus_{\lambda_i<1/4} \phi_i(z) \quad \text{and} \quad V_2 = L^2(F) \ominus V_1.$$

Consider a direct sum decomposition

$$L^2(F) = V_1 \oplus V_2 = L^2_d(F) \oplus L^2_c(F).$$

The operator $K\colon L^2(F) \to L^2(F)$ is nonnegative, i.e.,

$$\langle Kf, f\rangle \geq 0 \quad \forall f \in L^2(F).$$

Its projection onto V_2 is an integral operator $P_{V_2}K$ with a kernel

$$B(z,w) = K(z,w) - \sum_{\lambda_j<1/4} h(\lambda_j)\phi_j(z)\overline{\phi_j(w)}.$$

The operator $B = P_{V_2}K$ is a product of two nonnegative commuting selfadjoint operators; by the standard theorem in functional analysis (see, for example, [27, page 318]) it is also nonnegative, $\langle Bf, f\rangle \geq 0$.

Taking now the sequence of functions (for z fixed)

$$f_n(w) = \begin{cases} 1 & \text{if } d(z,w) \le 1/n, \\ 0 & \text{if } d(z,w) > 1/n, \end{cases}$$

we have

$$B(z,z) = \lim_{n\to\infty} \langle Bf_n, f_n\rangle \ge 0,$$

proving the desired result. ∎

ESTIMATE ON THE NUMBER OF LATTICE POINTS. In this subsection, following the method of Sarnak and Xue ([39], Theorem 3.1) we will prove the following estimate on the number of lattice points of $\Gamma(p)$.

PROPOSITION 5.3: *Let $\Gamma(p)$ be a principal congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$. Then*

$$N_1(T,\Gamma(p)) \overset{\text{def}}{=} \sum_{\substack{\gamma\in\Gamma(p)\\ \|\gamma\|\le T}} 1 \ll \frac{T^{2+\epsilon}}{p^3} + \frac{T^{1+\epsilon}}{p^2} + 1. \tag{5.9}$$

The implied constant is independent of p.

Proof: We have to estimate the number of integers a, b, c, d satisfying the following conditions:

$$|a|, |b|, |c|, |d| \le T, \tag{5.10}$$

$$ad - bc = 1, \tag{5.11}$$

$$a \equiv d \equiv 1 \pmod p, \tag{5.12}$$

$$b \equiv c \equiv 0 \pmod p. \tag{5.13}$$

There are $O_\epsilon(\frac{T^{1+\epsilon}}{p}) + 1$ choices of a satisfying (5.10) and (5.12). We now observe that

$$a + d \equiv 2 \pmod{p^2}. \tag{5.14}$$

Indeed, rewriting the conditions (5.12) and (5.13) in the form

$$a = \lambda_1 p + 1, \quad b = \lambda_3 p, \quad d = \lambda_2 p + 1, \quad c = \lambda_4 p,$$

and substituting them in (5.11), we obtain

$$(\lambda_1 p + 1)(\lambda_2 p + 1) - \lambda_3 p \lambda_4 p = 1.$$

Opening parentheses and rearranging the terms we easily have

$$(\lambda_1+\lambda_2)p = (\lambda_3\lambda_4-\lambda_1\lambda_2)p^2,$$

so $\lambda_1+\lambda_2\equiv 0 \pmod p$ and

$$a+d = 2+(\lambda_1+\lambda_2)p = 2+(\lambda_3\lambda_4-\lambda_1\lambda_2)p^2,$$

proving (5.14).

There are $O_\epsilon(\frac{T^{1+\epsilon}}{p^2})$ choices of a,d satisfying (5.14) and (5.10). For each satisfying choice of a, d let $\xi = ad-1$. From (5.10), $|\xi|<T^2$, so for each choice of a,d as above the number of b,c, satisfying $bc=\xi$ is at most $O_\epsilon(T^{2\epsilon})$. Hence

$$\begin{aligned} N_1(T,\Gamma(p)) &\ll_\epsilon \frac{T^{1+\epsilon}}{p^2}\left(\frac{T^{1+\epsilon}}{p}+1\right)+1\\ &\ll_\epsilon \frac{T^{2+\epsilon}}{p^3}+\frac{T^{1+\epsilon}}{p^2}+1,\end{aligned}$$

which proves the proposition. ∎

6. Proof of the Main Theorem

Utilizing the results obtained in the previous sections we will now prove the main theorem. As we have shown in section 2 for p large enough, we have $\Lambda(1)/\Lambda(p)\cong \mathrm{SL}_2(\mathbf{F}_p)$. Now suppose that

$$\Omega(\mathcal{F}(p))\cap(\delta(1-\delta),5/36)\neq\Omega(\mathcal{F}(1))\cap(\delta(1-\delta),5/36),$$

i.e., that there is a new discrete eigenvalue λ. Let V_λ be the corresponding eigenspace. The Laplacian on $\mathcal{F}(p)$ commutes with the deck transformations and consequently $\mathrm{SL}_2(\mathbf{F}_p)$ acts on V_λ. Since by assumption this action is nontrivial, V_λ must contain a nontrivial irreducible representaion of $\mathrm{SL}_2(\mathbf{F}_p)$. A result going back to Frobenius asserts that any nontrivial irreducible representaion of $\mathrm{SL}_2(\mathbf{F}_p)$ has dimension at least $\frac{p-1}{2}$, thus we conclude that

$$m(\lambda,\mathcal{F}(p))\geq(p-1)/2, \tag{6.1}$$

where $m(\lambda,\mathcal{F}(p))$ is a multiplicity of a new eigenvalue λ.

To bound the multiplicity from above we evaluate the automorphic kernel $K_X(z,w)$ corresponding to point-pair invariant $k=k_1*\bar{k_1}$ (with k_1 defined by equation (5.3)) on $\mathcal{K}_p$, the compact part of $\mathcal{F}(p)$, as defined in equation (3.15):

$$\begin{aligned}\int_{\mathcal{K}_p} K(z,z)d\mu(z) &= \sum_{\gamma\in\Lambda(p)} \int_{\mathcal{K}_p} k(z,\gamma z)d\mu(z)\\ &= \sum_{\gamma\in\Lambda(p)} \sum_{\delta\in\Lambda(1)/\Lambda(p)} \int_{\mathcal{K}_1} k(\delta^{-1}z,\delta^{-1}\gamma z)d\mu(z)\\ &\ll p^3 \sum_{\gamma\in\Lambda(p)} \int_{\mathcal{K}_1} k(z,\gamma z)d\mu(z) \ll p^3 \sum_{\gamma\in\Gamma(p)} \int_{\mathcal{K}_1} k(z,\gamma z)d\mu(z),\end{aligned}$$

where in the last line, not being in the position to exploit the sparsity of the set of lattice points of $\Lambda(p)$, we simply recorded the observation that $\Lambda(p) \subset \Gamma(p)$. As we will see below, this is a sacrifice we can get away with.

Now recall that for $z = i$ we have

$$\begin{aligned}K_{1,X}(i,i) &= \sum_{\gamma\in\Gamma(p)} k_{1,X}(i,\gamma i)\\ &= \#\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \Gamma(p) : a^2+b^2+c^2+d^2 \le \frac{X-2}{4}\right\}.\end{aligned}$$

For $z \neq i$ we have

$$K_{1,X}(z,z) = \#\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \Gamma(p) : f_z(a,b,c,d) \le \frac{X-2}{4}\right\},$$

where f_z is a homogeneous positive-definite quadratic form in a, b, c, d, continuously depending on z. Since $\mathcal{K}_1$ is compact, there is a constant σ (independent of p, X, z) such that

$$\sigma^{-1} \le \frac{f_z(a,b,c,d)}{a^2+b^2+c^2+d^2} \le \sigma$$

and therefore

$$K_{1,X/\sigma}(i,i) \le K_{1,X}(z,z) \le K_{1,\sigma X}(i,i),$$

and a similar inequality holds for the convolved kernel $K = K_1 * K_1$.

Using the estimate on the number of lattice points given in Proposition 5.3 and the estimate for convolved point-pair invariant k recorded in Proposition 5.1, we obtain

$$\int_{\mathcal{K}_p} K(z,z)d\mu(z) \ll p^3\left(\frac{X^{2+\epsilon}}{p^3} + X^{1+\epsilon}\right) \tag{6.2}$$

for arbitrary $\epsilon > 0$.

On the other hand, using pretrace inequality (Proposition 5.2) and recalling that the Selberg's transform of k_1 is given by equation (5.6), we have

$$\int_{\mathcal{K}_p} K(z,z)d\mu(z) \geq \int_{\mathcal{K}_p} \sum_{\lambda_{j,p}<\frac{1}{4}} X^{2s_{j,p}} |\phi_{j,p}(z)|^2 d\mu(z).$$

Interchanging the order of summation and applying the collar lemma for cusp (Lemma 4.1) to each of the contiguous cusps and collars in $\mathcal{F}(p)$ (see discussion in section 3 leading to (3.16)) and the collar lemma for flare (Lemma 4.2) to each of the contiguous cusps and collars in $\mathcal{F}(p)$, we obtain

$$\begin{aligned}\sum_{\lambda_{j,p}<\frac{1}{4}} X^{2s_{j,p}} \int_{\mathcal{K}_p} |\phi_{j,p}(z)|^2 d\mu(z) &\geq C \sum_{\lambda_{j,p}<\frac{1}{4}} X^{2s_{j,p}} \int_{\mathcal{F}_p} |\phi_{j,p}(z)|^2 d\mu(z) \\ &\gg \sum_{\lambda_{j,p}<\frac{1}{4}} X^{2s_{j,p}} m(\lambda_{j,p}).\end{aligned}$$

So we have

$$\sum_{\lambda_{j,p}<\frac{1}{4}} X^{2s_{j,p}} m(\lambda_{j,p}) \ll p^3 \left(\frac{X^{2+\epsilon}}{p^3} + X^{1+\epsilon}\right)$$

for any $\epsilon > 0$.

Rewriting it as

$$m(\lambda, p) \ll X^{2(1-s)} + p^3 X^{1-2s},$$

and taking X of order p^3, we obtain

$$m(\lambda, p) \ll p^{6(1-s)}.$$

Combining this with the lower bound (6.1), we get that for p large enough

$$6(1-s) > 1 \Leftrightarrow s < \frac{5}{6} \Leftrightarrow \lambda > \frac{5}{36}.$$

Hence we have proved the following

Main Theorem: *Let $\Lambda = \langle A_1, \ldots, A_k \rangle$ be a finitely generated subgroup of $\mathrm{SL}_2(\mathbf{Z})$ with $\delta > \frac{5}{6}$. Let $\mathcal{F}(p) = \Lambda(p)\backslash\mathbf{H}$. For p large enough*

$$\Omega(\mathcal{F}(p)) \cap \left[\delta(1-\delta), \frac{5}{36}\right) = \Omega(\mathcal{F}(1)) \cap \left[\delta(1-\delta), \frac{5}{36}\right).$$

Main Corollary: *Suppose that $\delta > \frac{5}{6}$. Then $\Omega(\mathcal{F}(p))$ has a spectral gap, that is for p large*

$$\lambda_1(\mathcal{F}(p)) \geq \min\left(\lambda_1(\mathcal{F}(1)), \frac{5}{36}\right).$$

7. Lubotzky's problem

In this section we address Lubotzky's problem described at the end of the introduction.

First, we translate the main corollary into representation theoretic language. By the Duality Theorem [9], $\pi_s \in \widehat{G}$ (where $G = \mathrm{SL}_2(\mathbf{R})$) in the spherical complimentary series representation of $\mathrm{SL}_2(\mathbf{R})$ appears as a subrepresentation of $L^2(\Gamma\backslash G)$ iff $\lambda(s) \in \Omega(\Gamma\backslash G/K) = \Omega(\Gamma\backslash\mathbf{H})$. Here $\lambda(s) = 1/4 - s^2$, $s(\lambda) = \sqrt{1/4 - \lambda^2}$. Hence we can restate the corollary as follows:

THEOREM 2: *Notation being as above, let*

$$\lambda_0 = \lambda_0(\Lambda(1)\backslash\mathbf{H}), \qquad s_0 = s(\lambda_0);$$
$$\lambda_1 = \min(\lambda_1(\Lambda(1)\backslash\mathbf{H}), 5/36), \qquad s_1 = s(\lambda_1).$$

For p large enough no complimentary series representation π_s with $s \in (s_1, s_0)$ occurs as a subrepresentation of $L^2(\Lambda(p)\backslash \mathrm{SL}_2(\mathbf{R}))$.

Let us write

$$L^2(\Lambda(p)\backslash \mathrm{SL}_2(\mathbf{R})) = L^2_0(\Lambda(p)\backslash \mathrm{SL}_2(\mathbf{R})) \bigoplus V_{s_0},$$

where V_{s_0} is invariant subspace corresponding to representation π_{s_o}; it occurs with multiplicity one by the Patterson theorem (quoted in the Introduction) and Duality Theorem. Another way of stating Theorem 2 is the following.

THEOREM 2′: *Let R be the following subset of $\widetilde{\mathrm{SL}_2(\mathbf{R})}$:*

$$R = \bigcup_p L^2_0(\Lambda(p)\backslash \mathrm{SL}_2(\mathbf{R})). \tag{7.1}$$

Then π_{s_0} is isolated with respect to R, i.e., $\pi_{s_0} \notin \overline{R}$.

Recalling that we have shown in section 2 that for p large enough we have $\Lambda(1)/\Lambda(p) \cong \mathrm{SL}_2(\mathbf{F}_p)$, let us turn to the family of Cayley graphs

$$\mathcal{G}_p = \mathcal{G}(\mathrm{SL}_2(\mathbf{F}_p), \{A_1, \ldots, A_k\}).$$

Remark 7.1: We are ignoring a finite set of primes p which are not large enough to satisfy the conditions of (2.9). But a finite set of graphs trivially forms a (finite) family of expanders.

The passage from Theorem 2′ to the expansion property of graphs $\mathcal{G}_p$ is a straightforward application of Fell's continuity of induction; we give the details below.

Let $H(p) = L^2(\Lambda(1)/\Lambda(p)) = L^2(\mathrm{SL}_2(\mathbf{F}_p))$ be the vector space of functions on a finite set $V_p = \mathrm{SL}_2(\mathbf{F}_p)$ with the norm $\|f\|^2 = \sum_{x\in V_p} |f(x)|^2$. Let

$$H_0(p) = \Big\{ f \in H(p) \mid \sum_{x\in V_p} f(x) = 0 \Big\}.$$

Then $\Lambda(1)$ acts on $H(p)$ by

$$(\gamma f)(x) = f(x\gamma),$$

and $H(p) = H_0(p) \oplus \mathbf{C}_{\chi_V}$ as $\Lambda(1)$-module.

The action of $\Lambda(1)$ on V_p is transitive, so the only $\Lambda(1)$-invariant functions on V_p are the constants $\mathbf{C}_{\chi_V}$. Thus $H_0(p)$ does not contain ρ_0, the trivial representation of $\Lambda(1)$.

Now inducing to $\mathrm{SL}_2(\mathbf{R})$ we have (by the induction in stages theorem)

$$\mathrm{Ind}_{\Lambda(p)}^{\mathrm{SL}_2(\mathbf{R})} 1 = \mathrm{Ind}_{\Lambda(1)}^{\mathrm{SL}_2(\mathbf{R})} \mathrm{Ind}_{\Lambda(p)}^{\Lambda(1)} 1 = \mathrm{Ind}_{\Lambda(1)}^{\mathrm{SL}_2(\mathbf{R})} \rho_0 \oplus \mathrm{Ind}_{\Lambda(1)}^{\mathrm{SL}_2(\mathbf{R})} H_0(p).$$

We have

$$\mathrm{Ind}_{\Lambda(1)}^{\mathrm{SL}_2(\mathbf{R})} \rho_0 \succ \pi_{s_0}.$$

Now let T be the subset of $\widetilde{\Lambda(1)}$ consisisting of representations $H_0(p)$, $T = \bigcup_p H_0(p)$. Then ρ_0 is isolated with respect to T. Indeed, assume the contrary, i.e., that there is a sequence of $\tau_j \in T$ such that $\tau_j \nearrow \rho_0$. Using Fell's continuity of induction [6] we would then have

$$\mathrm{Ind}_{\Lambda(1)}^{\mathrm{SL}_2(\mathbf{R})} \tau_j \nearrow \mathrm{Ind}_{\Lambda(1)}^{\mathrm{SL}_2(\mathbf{R})} \rho_0 \succ \pi_{s_0}$$

so that

$$\pi_{s_0} \in \overline{\mathrm{Ind}_{\Lambda(1)}^{\mathrm{SL}_2(\mathbf{R})} \Big(\mathrm{Ind}_{\Lambda(p)}^{\Lambda(1)} 1 \ominus \rho_0\Big)} \subset \overline{R},$$

contradicting Theorem $2'$.

Recalling the definition of Fell's toplogy [6], this implies that there is $\epsilon > 0$, depending only on $\Lambda(1)$ and $S = \{A_1, \ldots, A_k\}$ but not on p, such that $\forall f \in H_0(p)$

$$\|\gamma f - f\| > \epsilon \|f\| \quad \text{for some } \gamma \in S.$$

Let A be a subset of V of size a and B its compliment of size $b = n - a$, where $n = |V|$. Let

$$f(x) = \begin{cases} b, & \text{if } x \in A; \\ -a, & \text{if } x \in B. \end{cases}$$

Then $f \in H_0$ and

$$\|f\|^2 = ab^2 + ba^2 = nab$$

while for every $\gamma \in S$

$$\|\gamma f - f\|^2 = (b+a)^2 |E_\gamma(A,B)|,$$

where

$$E_\gamma(A,B) = \{x \in V \mid x \in A \text{ and } x\gamma \in B \text{ or } x \in B \text{ and } x\gamma \in A\}.$$

To summarize, there exists $\gamma \in S$ such that

$$|\partial A| \geq \frac{1}{2}|E_\gamma(A,B)| = \frac{\|\gamma f - f\|^2}{2n^2} \geq \frac{\epsilon^2 \|f\|^2}{2n^2} = \epsilon^2 \frac{ab}{2n} = \frac{\epsilon^2}{2}\Big(1 - \frac{|A|}{n}\Big)|A|.$$

Hence $\mathcal{G}(\mathrm{SL}_2(\mathbf{F}_p), S)$ are expanders with $c \geq \epsilon^2/2$. We have therefore proved the following theorem:

THEOREM 3: *Let $S = \{A_1, \ldots, A_k\}$ be a symmetric set of generators in $\mathrm{SL}_2(\mathbf{Z})$ and let $\Lambda = \langle A_1, \ldots, A_k \rangle$. If the Hausdorff dimension of the limit set $\delta(L(\Lambda)) > 5/6$, then $\mathcal{G}_p = G(\mathrm{SL}_2(\mathbf{F}_p), S)$ is a family of expanders.*

Now, as promised at the end of the introduction, we give examples of groups satisfying $\delta(L(\Lambda)) > \frac{5}{6}$. Let

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; \quad S_j = T^j S T^{-j}.$$

Let $\Lambda_n = \langle S, S_1, \ldots, S_n \rangle$. We claim that for n sufficiently large $\delta(L(\Lambda)) > \frac{5}{6}$, or equivalently that $\lambda_0(\Lambda_n \backslash \mathbf{H}) < \frac{5}{36}$.

To this end, we recall that by the variational characterisation of the bottom of the spectrum

$$\lambda_0 = \inf_{\substack{u \in L^2(\mathcal{F}) \\ \nabla u \in L^2(\mathcal{F})}} \frac{\int_{\mathcal{F}} |\nabla u|^2 d\mu}{\int_{\mathcal{F}} u^2 d\mu}. \tag{7.1}$$

The fundamental domain of Λ_n is an exterior of $n+1$ circles $C_0, C_1, \ldots, C_n$ (see Figure 8 for $n = 3$).

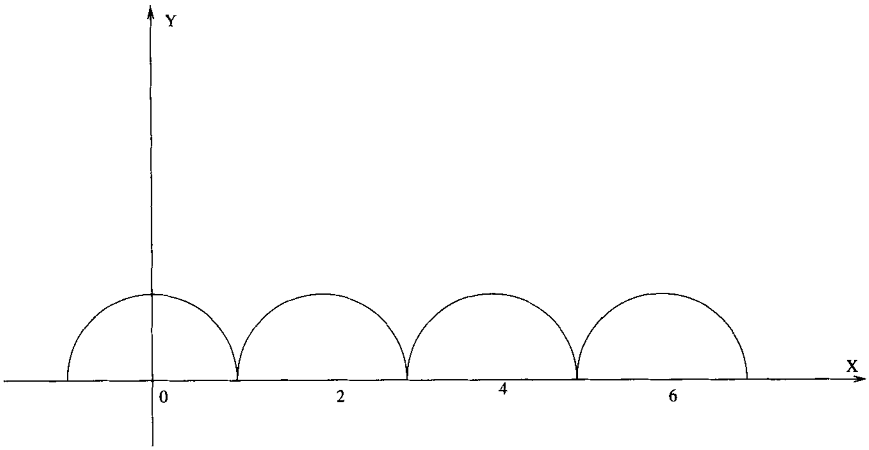

Figure 8. Fundamental domain of Λ_3.

Let us take the following test function u:

$$u_{n,A,\epsilon}(x+iy) = f_n(x)h_{\epsilon,A}(y),$$

where

$$f_n(x) = \begin{cases} 0 & \text{for } x \le -2, \\ 2+x & \text{for } -2 < x \le -1, \\ 1 & \text{for } -1 < x \le 2n+1, \\ 2n+2-x & \text{for } 2n+1 < x \le 2n+2, \\ 0 & \text{for } x \ge 2n+2; \end{cases}$$

and

$$h_{\epsilon,A}(y) = \begin{cases} \frac{y}{\epsilon} & \text{for } 0 \le y \le \epsilon, \\ 1 & \text{for } \epsilon < y \le A, \\ 2-\frac{y}{A} & \text{for } A < y \le 2A, \\ 0 & \text{for } y > 2A; \end{cases}$$

the constants ϵ and A are fixed below. Computation yields

$$\begin{aligned} \lambda_0(\mathcal{F}_n) &\le \frac{\int_{\mathcal{F}_n} |\nabla u_{n,\epsilon,A}|^2 d\mu}{\int_{\mathcal{F}_n} u_{n,\epsilon,A}^2 d\mu} \\ &\le \frac{1}{n}\left(\frac{1}{\epsilon} + \frac{4}{3}A + \frac{1}{A} + n\Big(\frac{1}{A} + \frac{2}{3}\epsilon\Big)\right) \\ &= \Big(\frac{1}{A} + \frac{2}{3}\epsilon\Big) + \frac{1}{n}\Big(\frac{1}{\epsilon} + \frac{4}{3}A + \frac{1}{A}\Big). \end{aligned}$$

If we take $A = 9$ and $\epsilon = \frac{1}{48}$, we obtain that for n greater than 4392 we have $\lambda_0(\mathcal{F}_n) < \frac{5}{36}$, hence for these n the groups Λ_n come under the purview of Theorem 3.

References

[1] R. Alperin, *An Elementary Account of Selberg's Lemma,* L'Enseignement Mathematique **33** (1987), 269–273.

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

[3] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces,* Progress in Mathematics, **106**, Birkhäuser, Boston, 1992.

[4] R. Courant and D. Hilbert, *Methods of Mathematical Physics,* Interscience Publishers, Inc, New York, 1953.

[5] G. Davidoff, P. Sarnak and A. Valette, *Ramanujan Graphs, Elementary Number Theory and Group Theory,* Unpublished Manuscript.

[6] J. M. G. Fell, *Weak containment and induced representations of groups,* Canadian Journal of Mathematics **14** (1962), 237–268.

[7] A. Gamburd, D. Jakobson and P. Sarnak, *Spectra of elements in the group ring of* SU(2), Journal of European Mathematical Society **1** (1999), 51–85.

[8] S. Gelbart and H. Jacquet, *A relation between automorphic representations of* GL(2) *and* GL(3), Annales Scientifiques de l'École Normale Supérieure **11** (1978), 471–552.

[9] I. M. Gelfand, M. I. Graev and I. I. Pyatetskii-Shapiro, *Representation Theory and Automorphic Functions*, W. B. Saunders Company, Philadelphia, Pa, 1969.

[10] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products,* Academic Press, New York–London–Toronto–Sydney–San Francisco, 1980.

[11] M. Huxley, *Introduction to Kloostermania,* in *Elementary and Analytic Theory of Numbers*, Banach Center Publ. **17**, Warsaw, 1985, pp. 217–306.

[12] M. Huxley, *Eigenvalues of congruence subgroups,* Contemporary Mathematics **53** (1986), 341–349.

[13] H. Iwaniec, *Selberg's lower bound of the first eigenvalue of congruence subgroups,* in *Number Theory, Trace Formulas, Discrete Groups* (E. J. Avers, ed.), Academic Press, New York, 1989, pp. 371–375.

[14] H. Iwaniec, *Introduction to the spectral theory of automorphic forms,* Biblioteca de la Revista Mathematica Iberoamericana, Revista Mathematica Iberoamericana, Madrid, 1995.

[15] H. Iwaniec, *The lowest eigenvalue for congruence groups,* Progress in Nonlinear Differential Equations and Applications **20**, Birkhäuser, Boston, 1996, pp. 203–212.

[16] M. Kac, *Can one hear the shape of a drum?*, The American Mathematical Monthly **73** (1966), 1–23.

[17] D. A. Kazhdan, *Connection of the dual space of a group with the structure of its closed subgroups*, Functional Analysis and its Applications **1** (1967), 63–65.

[18] J. Lafferty and D. Rockmore, *Fast Fourier analysis for* SL_2 *over a finite field and related numerical experiments*, Experimental Mathematics **1** (1992), 115–139.

[19] J. Lafferty and D. Rockmore, *Numerical investigation of the spectrum for certain families of Cayley graphs*, in DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 10 (J. Friedman, ed.), 1993, pp. 63–73.

[20] J. Lafferty and D. Rockmore, *Level spacings for Cayley graphs*, The IMA Volumes in Mathematics and its Applications **109** (1999), 373–387.

[21] 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.

[22] A. Lubotzky, R. Phillips and P. Sarnak, *Ramanujan Graphs*, Combinatorica **8** (1988), 261–277.

[23] A. Lubotzky and B. Weiss, *Groups and Expanders*, in DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 10 (J. Friedman, ed.), 1993, pp. 95–109.

[24] A. Lubotzky, *Discrete Groups, Expanding Graphs and Invariant Measures*, Progress in Mathematics Vol. 195, Birkhäuser, Boston, 1994.

[25] A. Lubotzky, *Cayley graphs: eigenvalues, expanders and random walks*, in *Surveys in Combinatorics* (P. Rowbinson, ed.), London Mathematical Society Lecture Note Series 218, Cambridge University Press, 1995, pp. 155–189.

[26] W. Luo, Z. Rudnick and P. Sarnak, *On Selberg's eigenvalue conjecture*, Geometric and Functional Analysis **5** (1995), 387–401.

[27] L. A. Lusternick and V. I. Sobolev, *Elements of Functional Analysis*, Nauka, Moscow, 1965.

[28] W. Magnus, F. Oberhettinger and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physiscs*, Springer-Verlag, New York, 1966.

[29] G. A. Margulis, *Explicit constructions of concentrators*, Problems of Information Transmission **10** (1975), 325–332.

[30] G. A. Margulis, *Explicit construction of graphs without short cycles and low density codes*, Combinatorica **2** (1982), 71–78.

[31] N. Nielsen, *Sur les produits de deux fonctions cylindriques*, Mathematische Annalen **LII** (1899), 228–242.

[32] S. J. Patterson, *The limit set of a Fuchsian group*, Acta Mathematica **136** (1975), 241–273.

[33] S. J. Patterson, *The Laplacian operator on the Riemann surface I, II, III,* Compositio Mathematica **31** (1975), 83–107; **32** (1976), 71–112; **33** (1976), 227–259.

[34] P. S. Phillips and P. Sarnak, *On the spectrum of the Hecke groups,* Duke Mathematical Journal **52** (1985), 211–221.

[35] B. Randol, *Small eigenvalues of the Laplace operator on compact Riemann surfaces,* Bulletin of the American Mathematical Society **80** (1974), 996–1008.

[36] L. Robin, *Fonctions Spheriques de Legendre et fonctions spheroidales,* Tome I, II, III, Collection Technique et Scientifique du C.N.E.T., Gauthier-Villars, Paris, 1959.

[37] P. Sarnak, *Some Applications of Modular Forms,* Vol. 99, Cambridge University Press, 1990.

[38] P. Sarnak, *Selberg's eigenvalue conjecture,* Notices of the American Mathematical Society **42** (1995), 1272–1277.

[39] P. Sarnak and X. Xue, *Bounds for multiplicities of automorphic representations,* Duke Mathematical Journal **64** (1991), 207–227.

[40] I. Satake, *Spherical functions and Ramanujan conjectures,* in *Algebraic Groups and Discontinuous Subgroups,* Proceedings of Symposia in Pure Mathematics, American Mathematical Society, Providence, RI, 1966, pp. 258–264.

[41] A. Selberg, *On Discontinous Groups in Higher Dimensional Symmetric Spaces,* in *Contributions to Function Theory,* Tata Institute of Fundamental Research, 1960, pp. 147–164.

[42] A. Selberg, *On the estimation of Fourier coefficients of modular forms,* Proceedings of Symposia in Pure Mathematics **VII**, American Mathematical Society, 1965, pp. 1–15.

[43] Y. Shalom, *Expanding graphs and invariant means,* Combinatorica **17** (1997), 555–575.

[44] Y. Shalom, *Expander graphs and amenable quotients,* The IMA Volumes in Mathematics and its Applications **109** (1999), 571–581.

[45] J. R. Stallings, *Group Theory and Three-Dimensional Manifolds,* Yale Mathematical Monographs, Vol. 4, Yale University Press, New Haven, 1971.

[46] D. Sullivan, *Discrete conformal groups and measurable dynamics,* Bulletin of the American Mathematical Society **6** (1982), 57–73.

[47] M. Suzuki, *Group Theory I,* Springer-Verlag, Berlin–Heidelberg–New York, 1982.

[48] G. N. Watson, *A Treatise on the Theory of Bessel Functions,* Cambridge University Press, London, 1962.

[49] A. Weil, *On some exponential sums,* Proceedings of the National Academy of Sciences of the United States of America **34** (1948), 204–207.