Optimal lifting for the projective action of $SL_3(\mathbb{Z})$

Amitay Kamber and Hagai Lavner

Optimal lifting for the projective action of SL₃(ℤ)

Amitay Kamber and Hagai Lavner

Let $\epsilon > 0$ and let $q \to \infty$ be a prime. We prove that with high probability, given x, y in the projective plane over $\mathbb{F}_q$, there exists $\gamma \in \mathrm{SL}_3(\mathbb{Z})$, with coordinates bounded by $q^{1/3+\epsilon}$, whose projection to $\mathrm{SL}_3(\mathbb{F}_q)$ sends x to y. The exponent $\frac{1}{3}$ is optimal and the result is a high rank generalization of Sarnak's optimal strong approximation theorem for $\mathrm{SL}_2(\mathbb{Z})$.

1. Introduction

In a letter to Miller and Talebizadeh, Sarnak [2015] proved the following lifting theorem, which he called optimal strong approximation.

Theorem 1.1. *Let* $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, $q \in \mathbb{Z}_{>0}$, $G_q = \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ *and let* $\pi_q : \Gamma \to G_q$ *be the quotient map. Then for every* $\epsilon > 0$, *as* $q \to \infty$, *there exists a set* $Y \subset G_q$ *of size* $|Y| \geq |G_q|(1 - o_\epsilon(1))$, *such that for every* $y \in Y$ *there exists* $\gamma \in \Gamma$ *of norm* $\|\gamma\|_\infty \leq q^{3/2+\epsilon}$, *with* $\pi_q(\gamma) = y$, *where* $\|\cdot\|_\infty$ *is the infinity norm on the coordinates of the matrix.*

The exponent $\frac{3}{2}$ in Theorem 1.1 is optimal, as the size of G_q is asymptotic to q^3, while the number of $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ satisfying $\|\gamma\|_\infty \leq T$ grows asymptotically like the Haar measure of the ball B_T of radius T in $\mathrm{SL}_2(\mathbb{R})$ [Duke et al. 1993; Maucourant 2007], i.e., $\mu(B_T) \asymp T^2$.

We use the standard notation $x \ll_z y$ to say that there is a constant C depending only on z such that $x \leq Cy$, and $x \asymp_z y$ means that $x \ll_z y$ and $y \ll_z x$.

We wish to discuss extensions of this theorem to SL_3, with a view towards general SL_N. If $\Gamma = \mathrm{SL}_N(\mathbb{Z})$, then the number of $\gamma \in \Gamma$ of satisfying $\|\gamma\|_\infty \leq T$ also grows like the Haar measure of the ball of radius T in $\mathrm{SL}_N(\mathbb{R})$, i.e., $\mu(B_T) \asymp T^{N^2-N}$ [Duke et al. 1993; Maucourant 2007], while the size of $G_q = \mathrm{SL}_N(\mathbb{Z}/q\mathbb{Z})$ is $|G_q| \asymp q^{N^2-1}$. One is therefore led to the following:

Conjecture 1.2. *Let* $\Gamma = \mathrm{SL}_N(\mathbb{Z})$, $q \in \mathbb{Z}_{>0}$, $G_q = \mathrm{SL}_N(\mathbb{Z}/q\mathbb{Z})$ *and let* $\pi_q : \Gamma \to G_q$ *be the quotient map. Then for every* $\epsilon > 0$, *as* $q \to \infty$, *there exists a set* $Y \subset G_q$ *of size* $|Y| \geq |G_q|(1 - o_\epsilon(1))$, *such that for every* $y \in Y$ *there exists* $\gamma \in \Gamma$ *of norm* $\|\gamma\|_\infty \leq q^{(N^2-1)/(N^2-N)+\epsilon}$, *with* $\pi_q(\gamma) = y$, *where* $\|\cdot\|_\infty$ *is the infinity norm on the coordinates of the matrix.*

While we were unable to prove Conjecture 1.2 even for $N = 3$, we prove a similar theorem for a nonprincipal congruence subgroup of $\mathrm{SL}_3(\mathbb{Z})$. For a prime q, let $\mathbb{F}_q$ be the field with q elements, let

$P_q = P^2(\mathbb{F}_q)$ be the 2-dimensional projective space over $\mathbb{F}_q$, i.e., the set of vectors $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$, $a, b, c \in \mathbb{F}_q$ not all 0, modulo the equivalence relation $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \sim \begin{pmatrix} \alpha a \\ \alpha b \\ \alpha c \end{pmatrix}$ for $\alpha \in \mathbb{F}_q^\times$. The group $\mathrm{SL}_3(\mathbb{F}_q)$ acts naturally on P_q, and by composing this action with π_q we have an action $\Phi_q \colon \mathrm{SL}_3(\mathbb{Z}) \to \mathrm{Sym}(P_q)$.

Theorem 1.3. *Let* $\Gamma = \mathrm{SL}_3(\mathbb{Z})$*, and for a prime* q *let* $P_q = P^2(\mathbb{F}_q)$ *and* $\Phi_q : \mathrm{SL}_3(\mathbb{Z}) \to \mathrm{Sym}(P_q)$ *as above. Then for every* $\epsilon > 0$*, as* $q \to \infty$*, there exists a set* $Y \subset P_q$ *of size* $|Y| \geq (1 - o_\epsilon(1))|P_q|$*, such that for every* $x \in Y$*, there exists a set* $Z_x \subset P_q$ *of size* $|Z_x| \geq (1 - o_\epsilon(1))|P_q|$*, such that for every* $y \in Z_x$*, there exists an element* $\gamma \in \Gamma$ *satisfying* $\|\gamma\|_\infty \leq q^{1/3+\epsilon}$*, such that* $\Phi_q(\gamma)x = y$.

The exponent $\frac{1}{3}$ is optimal, since the size of P_q is $|P_q| \asymp q^2$, while the number of elements $\gamma \in \mathrm{SL}_3(\mathbb{Z})$ satisfying $\|\gamma\|_\infty \leq T$ is $\asymp T^6$.

An alternative formulation of Theorem 1.3 is that for all but $o_\epsilon(|P_q|^2)$ of pairs $(x, y) \in P_q \times P_q$, there exists an element $\gamma \in \Gamma$ satisfying $\|\gamma\|_\infty \leq q^{1/3+\epsilon}$ such that $\Phi_q(\gamma)x = y$. However, in this formulation it is a bit harder to see why the exponent $1/3$ is optimal, and our proof actually uses the formulation of Theorem 1.3 as stated.

An important observation is that the premise of Theorem 1.3 actually fails for the point $x = \mathbf{1} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \in P_q$. Elements sending $\mathbf{1}$ to $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \in P_q$ necessarily have the third column modulo q equivalent to $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ (modulo the action of $\mathbb{F}_q^\times$). Since there are only $\asymp T^3$ possibilities for the third column, we need to consider matrices of infinity norm at least $q^{2/3}$ in order to reach from $x = \mathbf{1}$ to almost all of $y \in P_q$. As a matter of fact, one may use the explicit property (T) of $\mathrm{SL}_3(\mathbb{R})$ from [Oh 2002] together with ideas from [Ghosh et al. 2018] to deduce that if we allow the size of the matrices to reach $q^{2/3+\epsilon}$ we may replace the set Y in Theorem 1.3 by the entire set P_q.

We deduce Theorem 1.3 from a lattice point counting argument, in the spirit of the work of Sarnak and Xue [1991]. To state it, we first define a different gauge of largeness on $\mathrm{SL}_3(\mathbb{Z})$ by $\|\gamma\|_\infty \|\gamma^{-1}\|_\infty$. The number of $\gamma \in \mathrm{SL}_3(\mathbb{Z})$ satisfying $\|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T$ grows asymptotically like $T^2 \log T$ [Maucourant 2007]. Note that if $\|\gamma\|_\infty \leq T$ then $\|\gamma^{-1}\|_\infty \leq 2T^2$. In particular, the ball of radius $2T$ relatively to $\| \cdot \|_\infty \| \cdot^{-1} \|_\infty$ contains the ball of radius $T^{1/3}$ relatively to $\| \cdot \|_\infty$, and their volume is asymptotically the same up to $T^{o(1)}$. The counting result is as follows:

Theorem 1.4. *Let* $\Gamma = \mathrm{SL}_3(\mathbb{Z})$*, and for a prime* q *let* $P_q = P^2(\mathbb{F}_q)$ *and* $\Phi_q : \mathrm{SL}_3(\mathbb{Z}) \to \mathrm{Sym}(P_q)$ *as above. Then there exists a constant* $C > 0$ *such that for every prime* q*,* $T \leq Cq^2$ *and* $\epsilon > 0$ *it holds that*

$$|\{(\gamma, x) \in \mathrm{SL}_3(\mathbb{Z}) \times P^2(\mathbb{F}_q) : \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T, \Phi_q(\gamma)(x) = x\}| \ll_\epsilon q^{2+\epsilon} T.$$

Underlying Conjecture 1.2 is the principal congruence subgroup $\Gamma(q) = \ker \pi_q$. Let $\mathbf{1} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \in P_q$. Then the group

$$\Gamma_0'(q) = \left\{ \gamma \in \mathrm{SL}_3(\mathbb{Z}) : \Phi_q(\gamma)(\mathbf{1}) = \mathbf{1} \right\} = \left\{ \begin{pmatrix} * & * & a \\ * & * & b \\ * & * & * \end{pmatrix} \in \mathrm{SL}_3(\mathbb{Z}) : a = b = 0 \bmod q \right\}$$

is a nonprincipal congruence subgroup of $\mathrm{SL}_3(\mathbb{Z})$. Theorem 1.3 says that Conjecture 1.2 holds "on average" for the nonprincipal congruence subgroup $\Gamma_0'(q)$.

Conjecturally, such "optimal lifting on average" should hold for every sequence of congruence subgroups of $\Gamma = SL_N(\mathbb{Z})$, i.e., subgroups of some $\Gamma(q)$, $q > 1$ an integer. We provide a further example of this phenomenon for the action of $SL_3(\mathbb{Z})$ on flags of $\mathbb{F}_q^3$ in Theorem 5.1.

Let us provide a spectral context for our results, namely Sarnak's density conjecture for exceptional eigenvalues. See also [Golubev and Kamber 2020] for a more detailed discussion.

Theorem 1.1 follows from Selberg's conjecture about the smallest nontrivial eigenvalue of the Laplacian of the hyperbolic surfaces $\Gamma(q)\backslash\mathcal{H}$, where $\mathcal{H}$ is the hyperbolic plane and $\Gamma(q)$ is the q-th principal congruence subgroup of $\Gamma = SL_2(\mathbb{Z})$. While Selberg's conjecture remains widely open, Sarnak proved Theorem 1.1 using density estimates on exceptional eigenvalues of the Laplacian, which are due to Huxley [1986]. Similar density results were proved by Sarnak and Xue [1991] using lattice point counting arguments, but only for arithmetic quotients which are compact. The compactness assumption was removed in [Huntley and Katznelson 1993; Gamburd 2002] (and the results were moreover extended to some thin subgroups of $SL_2(\mathbb{Z})$). As a matter of fact, in rank 1 the density property is equivalent to the lattice point counting property [Golubev and Kamber 2020].

In higher rank, Conjecture 1.2 would similarly follow from a naive Ramanujan conjecture for $\Gamma(q)\backslash SL_N(\mathbb{R})$, $\Gamma = SL_N(\mathbb{Z})$, which says (falsely!) that the representation of $SL_N(\mathbb{R})$ on $L^2(\Gamma(q)\backslash SL_N(\mathbb{R}))$ decomposes into a trivial representation and a tempered representation. Burger, Li and Sarnak's explanation of the failure of the naive Ramanujan conjecture [Burger et al. 1992] is closely related to the behavior of the point $x_0 = \mathbf{1} \in P_q$. As in rank 1, Theorem 1.4 should be equivalent to density estimates for $\Gamma_0'(q)$, but there are some technical problems coming from the fact that $SL_3(\mathbb{Z})$ is not cocompact [Golubev and Kamber 2020]. Closely related density results were recently proven by Blomer, Buttcane and Maga for $N = 3$ in [Blomer et al. 2017], and for general N by [Blomer 2019], using the Kuznetsov trace formula, and it is very likely that Theorem 1.3 can also be proven (and generalized to $N > 3$) using those density arguments. There are some technical problems with the implementation of this approach, for example, the results of [Blomer et al. 2017] and [Blomer 2019] concern cusp forms, and one has to deal with the presence of nontempered Eisenstein representations and some other technical issues. Moreover, those results are limited to subgroups similar to $\Gamma_0'(q)$, and are not available in the context of Theorem 5.1. Our counting approach is more elementary, and is easier to generalize to other contexts.

The approach of this article can be carried far more generally. We refer to [Golubev and Kamber 2020] where the approach is studied in detail and in great generality. In particular, counting results such as Theorem 1.4 (called the *weak injective radius property* by Golubev and Kamber), imply optimal lifting results such as Theorem 1.3; see [loc. cit., Theorem 1.5]. In particular, the right counting theorem will imply Conjecture 1.2. Our restriction to $N = 3$ and the cases considered in Theorems 1.3 and 5.1 follows from the fact that in those cases we can prove the relevant counting theorems, namely Theorems 1.4 and 5.2. Such counting results are available in rank 1 following [Sarnak and Xue 1991], but as far as we know are new in rank greater than 1. As explained in [Golubev and Kamber 2020], the counting theorems are closely related to some spectral questions as in the rank 1 case discussed above, but the fact that the space is not compact significantly complicates matters. We refer again to [Sarnak and Xue 1991] for a more complete discussion.

Structure of the article. We provide a proof of Theorem 1.1 in Section 2, which serves as a guideline for the harder case of SL_3. The main difference between our proof and the proof in [Sarnak 2015] is that we avoid using spectral decomposition, which is far harder in SL_3.

In Section 3 we prove Theorem 1.4. The proof uses basic number theory and linear algebra.

In Section 4 we deduce Theorem 1.3 from Theorem 1.4. The argument is analytic, and uses various tools from spectral analysis and representation theory, which include property (T), the pretrace formula (in a disguised form), and bounds on Harish-Chandra's Ξ function. This section is based on a general framework developed by the first author with Konstantin Golubev surrounding similar questions [Golubev and Kamber 2020].

Finally, in Section 5 we prove Theorem 5.1 which is a variant of Theorem 1.3 for the action of $SL_3(\mathbb{Z})$ on flags of $\mathbb{F}_q^3$.

2. Proof of Theorem 1.1

The basic input for the proof of Theorem 1.1 is the following counting result, proved in [Gamburd 2002, Lemma 5.3]; it also appeared earlier, e.g., in [Huxley 1986].

Lemma 2.1. *Let $\epsilon > 0$. Then for every $q \in \mathbb{N}$, the size of the set*

$$\{\gamma \in SL_2(\mathbb{Z}) : \gamma = I \bmod q, \|\gamma\|_\infty \leq T\}$$

is bounded by $\ll_\epsilon T^\epsilon (T^2/q^3 + T/q + 1)$.

Proof. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$ be in the set. It holds that $\gamma - I \in qM_n(\mathbb{Z})$, so $\det(\gamma - I) = 0 \bmod q^2$, or explicitly

$$(a-1)(d-1) - bc = 0 \bmod q^2.$$

Since $ad - bc = 1$, we have $a + d = 2 \bmod q^2$. Since both a and d are bounded in absolute value by T, the number of options for $a + d$ is at most $4T/q^2 + 1$. Similarly, the number of options for a is at most $2T/q + 1$. Therefore, the number of options for (a, d) is $\ll (T/q^2 + 1)(T/q + 1)$.

To determine b, c, note that if $ad \neq 1$, then $bc = 1 - ad \neq 0$, and by standard divisor bounds this gives $\ll_\epsilon T^\epsilon$ options for (b, c). Otherwise, assuming $q > 2$, $a = d = 1$, and then $b = 0$ or $c = 0$ (or both). If $b = 0$ then c has at most $2T/q + 1$ options, while if $c = 0$, then b has at most $2T/q + 1$ options.

All in all, the number of solution is bounded by

$$\ll_\epsilon (T/q^2 + 1)(T/q + 1)T^\epsilon + T/q + 1 \ll T^\epsilon (T^2/q^3 + T/q + 1). \qquad \square$$

Our proof of Theorem 1.1 proceeds with some spectral analysis of hyperbolic surfaces associated to $SL_2(\mathbb{Z})$ and its congruence subgroups, which will require some preliminaries. Let $\mathcal{H}$ be the hyperbolic plane, with the model $\mathcal{H} = \{z = x + iy \in \mathbb{C} : y > 0\}$. The space $\mathcal{H}$ is equipped with the metric defined by $d(x + iy, x' + iy') = \text{arccosh}(1 + ((x - x')^2 + (y - y')^2)/(2yy'))$ and a measure defined by $dx\,dy/y^2$. It also has a natural $SL_2(\mathbb{R})$ action by Möbius transformation, i.e., $\begin{pmatrix} a & b \\ c & d \end{pmatrix}z = (az + b)/(cz + d)$.

This action allows us to identify $\mathcal{H}$ with G/K, where $G = SL_2(\mathbb{R})$, and $K = SO(2)$ is the stabilizer of the point $i \in \mathcal{H}$. We also assume that the Haar measure on G is normalized to agree with the measure on $\mathcal{H}$ on right K-invariant measurable sets.

When using spectral arguments, it will be useful to use a bi-K-invariant (i.e., left and right K-invariant) gauge of largeness of an element. We therefore define $\|g\|_{\mathcal{H}} = e^{d(i,gi)/2}$. Explicitly, by the Cartan decomposition of G, g can be written as

$$g = k_1 \begin{pmatrix} e^{r/2} & \\ & e^{-r/2} \end{pmatrix} k_2,$$

with $k_1, k_2 \in K = SO(2)$, and $r \in \mathbb{R}_{\geq 0}$ unique. Then $\|g\|_{\mathcal{H}} = e^{r/2}$. As the L^2-norm of the coordinates of γ is $\sqrt{e^r + e^{-r}}$, $\|g\|_{\mathcal{H}}$ is closely related to the infinity norm on the coordinates, namely, there exists a constant $C > 0$ such that $C^{-1}\|g\|_\infty \leq \|g\|_{\mathcal{H}} \leq C\|g\|_\infty$. We may therefore prove Theorem 1.1 using the gauge $\|\cdot\|_{\mathcal{H}}$ instead of $\|\cdot\|_\infty$. Two important properties of $\|\cdot\|_{\mathcal{H}}$ are symmetry $\|g\|_{\mathcal{H}} = \|g^{-1}\|_{\mathcal{H}}$, and submultiplicativity $\|g_1 g_2\|_{\mathcal{H}} \leq \|g_1\|_{\mathcal{H}}\|g_2\|_{\mathcal{H}}$. The submultiplicativity follows from the fact that d is a G-invariant metric on $\mathcal{H}$.

We define the function $\chi_T \in L^1(K\backslash G/K)$ as the normalized probability characteristic function of the set $\{g \in G \colon \|g\|_{\mathcal{H}} \leq T\}$, i.e.,

$$\chi_T(g) = \frac{1}{2\pi(\cosh(2\log T) - 1)} \begin{cases} 1 & \text{if } \|g\|_{\mathcal{H}} \leq T, \\ 0 & \text{if } \|g\|_{\mathcal{H}} > T. \end{cases}$$

Notice that $2\pi(\cosh r - 1)$ is the volume of the hyperbolic ball of radius r. Here and later by a probability function we mean a nonnegative function with integral 1.

We also define $\psi_T \in L^1(K\backslash G/K)$ as the function

$$\psi_T(g) = \frac{1}{T} \begin{cases} \|g\|_{\mathcal{H}}^{-1} & \text{if } \|g\|_{\mathcal{H}} \leq T, \\ 0 & \text{if } \|g\|_{\mathcal{H}} > T. \end{cases}$$

There is a convolution of $f \in L^\infty(G/K) \cong L^\infty(\mathcal{H})$ and $\chi \in L^1(K\backslash G/K)$, which we usually think as an action of χ on f. It is simply the convolution of the two functions, when both are considered as invariant functions on G:

$$f * \chi(x) = \int_{g \in G} f(xg^{-1})\chi(g)\, dg = \int_{g \in G} f(g^{-1})\chi(gx)\, dg.$$

It holds that $f * \chi \in L^\infty(\mathcal{H})$. For example, the value of $f * \chi_T$ at g_0, is the average of f over the ball $\{g_0 g \in G : \|g\|_{\mathcal{H}} \leq T\}$.

Lemma 2.2 (convolution lemma). *For every $g \in G$, $(\chi_T * \chi_T)(g) \ll \psi_{T^2}(g)$.*

We refer to [Sarnak and Xue 1991, Lemma 2.1] or [Gamburd 2002, Proposition 5.1] for a proof. Geometrically, the proof calculates the volume of an intersection of two hyperbolic balls. In Lemma 4.2 we give a spectral proof of a similar statement for $SL_3(\mathbb{R})$, which also works for $SL_2(\mathbb{R})$, but adds a factor that is logarithmic in T.

As in the statement Theorem 1.1, let $q \in \mathbb{Z}_{>0}$, $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, $G_q = \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ and let $\pi_q : \Gamma \to G_q$ be the quotient map. Let $\Gamma(q) = \ker \pi_q$.

We look at the locally symmetric space $X_q := \Gamma(q)\backslash\mathcal{H} \cong \Gamma\backslash G/K$. This space is a hyperbolic orbifold of finite volume. By $L^2(X_q)$ we mean the Hilbert space of measurable functions on X_q with bounded L^2-norm relative to the finite measure on X_q, with the obvious inner-product. We still consider a function on $X_q = \Gamma(q)\backslash\mathcal{H} = \Gamma(q)\backslash G/K$ as a left $\Gamma(q)$-invariant function on $\mathcal{H}$ or on G. Right convolution by functions from $L^1(K\backslash G/K)$ is defined for bounded functions on X_q, and extends to functions in $L^2(X_q)$ as the convolution defines a bounded operator. In particular, we will consider right convolution of $f \in L^2(X_q)$ with χ_T.

For $x_0 \in X_q$, denote $b_{T,x_0}(x) := \sum_{\gamma \in \Gamma(q)} \chi_T(\tilde{x}_0^{-1}\gamma x)$, when $\tilde{x}_0$ is any lift of x_0 to G. It holds that $b_{T,x_0} \in L^2(X_q)$, and $\int_{X_q} b_{T,x_0}(x)\,dx = 1$.

In particular $b_{T,e}$ corresponds to the point $\Gamma(q)eK \in \Gamma(q)\backslash\mathcal{H}$, where e is the identity matrix in G.

Lemma 2.3. *For $f \in L^2(X_q)$ bounded,*

$$\langle f, b_{T,x_0}\rangle = f * \chi_T(x_0).$$

Proof. By unfolding,

$$
\begin{aligned}
\langle f, b_{T,x_0}\rangle &= \int_{x\in\Gamma(q)\backslash\mathcal{H}} f(x) \sum_{\gamma\in\Gamma(q)} \chi_T(x_0^{-1}\gamma x)\,dx \\
&= \int_{x\in\Gamma(q)\backslash\mathcal{H}} \sum_{\gamma\in\Gamma(q)} f(\gamma x)\chi_T(x_0^{-1}\gamma x)\,dx \\
&= \int_{x\in\mathcal{H}} f(x)\chi_T(x_0^{-1}x)\,dx \\
&= \int_{x\in\mathcal{H}} f(x)\chi_T(x^{-1}x_0)\,dx \\
&= f * \chi_T(x_0).
\end{aligned}
$$

Notice that we used the fact that $\chi_T(g) = \chi_T(g^{-1})$, which is a simplification that will not occur in SL_3. $\quad\square$

The following lemma uses the combinatorial Lemma 2.1 to get analytic information:

Lemma 2.4. *It holds that*

$$\|b_{T,e}\|_2^2 \ll_\epsilon T^\epsilon \left(\frac{1}{q^3} + \frac{1}{T^2}\right).$$

In particular, for $T = q^{3/2}$,

$$\|b_{T,e}\|_2^2 \ll_\epsilon \frac{T^\epsilon}{q^3}.$$

Proof. By Lemmas 2.3 and 2.2,

$$\|b_{T,e}\|_2^2 = b_{T,e} * \chi_T(e) = \sum_{\gamma\in\Gamma(q)} (\chi_T * \chi_T)(\gamma) \ll \sum_{\gamma\in\Gamma(q)} \psi_{T^2}(\gamma) = \frac{1}{T^2} \sum_{\gamma\in\Gamma(q):\|\gamma\|_{\mathcal{H}}\leq T^2} \|\gamma\|_{\mathcal{H}}^{-1}.$$

We next apply discrete partial summation [Hardy and Wright 1979, Theorem 421] which says that for $g : \Gamma(q) \to [1, \infty]$, $f : [1, \infty] \to \mathbb{R}$ nice enough it holds that

$$\sum_{\gamma : 1 \le g(\gamma) \le Y} f(g(\gamma)) = f(Y) \left| \{ \gamma : 1 \le g(\gamma) \le Y \} \right| - \int_1^Y |\{ \gamma : g(\gamma) \le S \}| \frac{df}{dS}(S) \, dS. \qquad (2\text{-}1)$$

Apply this to $g(\gamma) = \|\gamma\|_{\mathcal{H}}$, $f(x) = x^{-1}$ and $Y = T^2$,

$$\frac{1}{T^2} \sum_{\gamma \in \Gamma(q) : \|\gamma\|_{\mathcal{H}} \le T^2} \|\gamma\|_{\mathcal{H}}^{-1} = \frac{1}{T^2} \left(\frac{1}{T^2} |\{ \gamma \in \Gamma(q) : \|\gamma\|_{\mathcal{H}} \le T^2 \}| + \int_1^{T^2} |\{ \gamma \in \Gamma(q) : \|\gamma\|_{\mathcal{H}} \le S \}| S^{-2} \, dS \right)$$

$$\ll_\epsilon T^\epsilon \frac{1}{T^2} \left(\frac{1}{T^2} \left(\frac{T^4}{q^3} + \frac{T^2}{q} + 1 \right) + \int_1^{T^2} \frac{1}{S^2} \left(\frac{S^2}{q^3} + \frac{S}{q} + 1 \right) dS \right)$$

$$\ll_\epsilon T^\epsilon \frac{1}{T^2} \left(\frac{T^2}{q^3} + \frac{1}{q} + \frac{1}{T^2} + 1 \right)$$

$$\ll T^\epsilon \left(\frac{1}{q^3} + \frac{1}{T^2} \right).$$

The first inequality follows from Lemma 2.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Let $\pi \in L^2(X_q)$ be the constant probability function on X_q (recall that the space has finite volume). Denote by $L_0^2(X_q)$ the set of functions of integral 0, or alternatively the set of functions orthogonal to π. The deepest input to the proof is the following celebrated theorem of Selberg:

Theorem 2.5 (Selberg's spectral gap theorem). *There is an explicit $\tau > 0$ such that for every $f \in L_0^2(X_q)$ and $T > 0$ is holds that $\|f * \chi_{T^\eta}\|_2 \ll T^{-\eta\tau} \|f\|_2$.*

Selberg's theorem is usually stated as a lower bound on the spectrum of the Laplacian. However, it is well known that it can be translated to a spectral gap of the convolution operators by large balls. The statement is true in great generality (see, e.g., [Ghosh et al. 2013, Section 4]), but for the benefit of the reader we give a sketch of the proof, based on [Golubev and Kamber 2019]. For $r \ge 0$ we define an operator

$$A_r : L^2(X_q) \to L^2(X_q),$$

by

$$A_r f(x) = \int_K \int_K f \left(x k_1 \begin{pmatrix} e^{r/2} & 0 \\ 0 & e^{-r/2} \end{pmatrix} k_2 \right) dk_1 \, dk_2.$$

By [Golubev and Kamber 2019, Proposition 7.2] (or alternatively, by bounds on Harish-Chandra's function for $\mathrm{SL}_2(\mathbb{R})$), if the smallest nonzero eigenvalue of the Laplacian Δ on $L^2(X_q)$ is larger than $\frac{1}{4} - \left(\frac{1}{2} - p^{-1} \right)^2$, then for every $f \in L_0^2(X_q)$ it holds that

$$\|A_r f\|_2 \le (r+1) e^{-r/p} \|f\|_2.$$

Selberg's spectral gap theorem says that the smallest nontrivial eigenvalue of the Laplacian is at least $\frac{3}{16}$, so the above holds with $p = 4$. There are various results improving the value of p in Selberg's

theorem (see [Sarnak 2005]), and the best one is due to Kim and Sarnak, giving $p = \frac{64}{25}$. However, those improvements are inconsequential for us. In any case, by comparing the definition, we see that

$$f * \chi_T(x) = \frac{1}{2\pi \cosh(2\log T) - 1} \int_0^{2\log T} 2\pi \sinh r (A_r f)(x) \, dr,$$

and therefore

$$\|f * \chi_T\|_2 \leq \frac{1}{2\pi \cosh(2\log T) - 1} \int_0^{2\log T} 2\pi \sinh r \|A_r f\|_2 \, dr$$

$$\leq \frac{\|f\|_2}{\cosh(2\log T) - 1} \int_0^{2\log T} (r+1) e^{r(1-1/4)} \, dr$$

$$\leq \frac{2(\log T + 1)^2 T^{2(1-1/4)} \|f\|_2}{\cosh(2\log T) - 1}.$$

For T large enough the above is

$$\ll T^{-2/5} \|f\|_2,$$

which give us the needed result, with $\tau = \frac{2}{5}$. The Kim–Sarnak bounds allows us to take $\tau = \frac{32}{25} + \epsilon$ for any $\epsilon > 0$.

The important part of the theorem is the independence of τ from q. We fix this $\tau > 0$ for the rest of this section.

From Selberg's theorem we deduce:

Lemma 2.6. *For $T = q^{3/2}$, and every $\epsilon > 0$*

$$\|b_{T,e} * \chi_{T^\eta} - \pi\|_2 \ll_\epsilon q^{-3/2 - \eta\tau + \epsilon}.$$

Proof. We have $b_{T,e} - \pi \in L_0^2(X_q)$ and $\pi * \chi_T = \pi$ (as an average of the constant function is the constant function).

Therefore,

$$\|b_{T,e} * \chi_{T^\eta} - \pi\|_2 = \|(b_{T,e} - \pi) * \chi_{T^\eta}\|_2 \ll T^{-\eta\tau} \|b_{T,e} - \pi\|_2 \ll_\epsilon q^{-3/2 - \eta\tau + \epsilon},$$

where in the first inequality we applied Theorem 2.5, and in the second inequality we applied $\|b_{T,e} - \pi\|_2 \leq \|b_{T,e}\|_2$ ($b_{T,e} - \pi$ is the orthogonal projection of $b_{T,e}$ onto $L_0^2(X_q)$) and Lemma 2.4. $\square$

The last lemma implies that the function $b_{T,e} * \chi_{T^\eta}$ is very close to the constant probability function π. Let us show how this implies Theorem 1.1.

We have a map $\iota : G_q \cong \Gamma(q) \backslash \Gamma \to X_q \cong \Gamma(q) \backslash G / K$, defined as $\iota(\Gamma(q)\gamma) = \Gamma(q)\gamma K$. For $y \in G_q$, we may consider the function $b_{T_0, \iota(y)}$. We choose T_0 small enough (independently of q), so that the functions $b_{T_0, \iota(y)}$ will have disjoint supports for $\iota(y) \neq \iota(y')$. Specifically, it is enough to choose T_0 such that the ball of radius $2\log T_0$ around i and around $\gamma i \neq i$ for $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ are disjoint. We also notice that ι has fibers of bounded size, specifically $|\mathrm{SL}_2(\mathbb{Z}) \cap K| = 4$. This implies that for every $\iota(y), x \in X_q$,

identified with some lifts $\tilde{y}, \tilde{x}$ to $\mathrm{SL}_2(\mathbb{R})$, there are at most 4 element $\gamma \in \Gamma(q)$ such that $\chi_{T_0}(\tilde{y}^{-1}\gamma\tilde{x}) \neq 0$. Therefore,

$$\|b_{T_0,\iota(y)}\|_2^2 \leq 16\|\chi_{T_0}\|_2^2.$$

In particular, $\|b_{T_0,\iota(y)}\|_2$ is bounded uniformly in q.

Lemma 2.7. *Assume that $\langle b_{T,e} * \chi_{T^\eta}, b_{T_0,\iota(y)}\rangle > 0$. Then there exists $\gamma \in \Gamma$ such that $\pi_q(\gamma) = y$, and $\|\gamma\|_{\mathscr{H}} \leq T_0 T^{1+\eta}$.*

Proof. By Lemma 2.3, the condition implies that

$$(b_{T,e} * \chi_{T^\eta} * \chi_{T_0})(\iota(y)) > 0.$$

Treat the function as a left $\Gamma(q)$-invariant and right K-invariant function on G. Let γ_y be a lift of y to Γ, i.e., $\pi_q(\gamma_y) = y$. Therefore, $b_{T,e} * \chi_{T^\eta} * \chi_{T_0}(\gamma_y) > 0$.

By the definition of convolution, there are $g_1', g_2, g_3 \in G$, such that $g_1' \in \mathrm{supp}(b_{T,e})$, $g_2 \in \mathrm{supp}(\chi_{T^\eta})$, $g_3 \in \mathrm{supp}(\chi_{T_0})$, and such that $g_1' g_2 g_3 = \gamma_y$. Looking at the definition of $b_{T,e}$ and g_1', there are $g_1 \in \mathrm{supp}(\chi_T)$, $\gamma \in \Gamma(q)$ such that $e^{-1}\gamma g_1' = g_1$ (we write e for the identity element instead of discarding it, anticipating the case of SL_3 below). Therefore $\gamma^{-1}eg_1 g_2 g_3 = \gamma_y$.

Write $g = g_1 g_2 g_3$. By the above, $\|g\|_{\mathscr{H}} \leq \|g_1\|_{\mathscr{H}}\|g_2\|_{\mathscr{H}}\|g_3\|_{\mathscr{H}} \leq T_0 T^{1+\eta}$. In addition, $eg = \gamma\gamma_y$, so that $g \in \Gamma(q)\gamma_y$. Therefore $g \in \Gamma$ and $\pi_q(g) = y$, as needed. $\square$

We may now finish the proof of Theorem 1.1. Let $\eta > 0$ and write $T = q^{3/2}$. Assume that $Z \subset G_q$ is the set of $y \in G_q$ such that there is no $\gamma_y \in \Gamma$ with $\|\gamma_y\|_{\mathscr{H}} \leq T_0 T^{1+\eta}$ and $\pi_q(\gamma_y) = y$. It suffices to prove that for a fixed $\eta > 0$ it holds that $|Z| = o_\eta(q^3)$.

By Lemma 2.7, for every $y \in Z$,

$$\langle b_{T,e} * \chi_{T^\eta}, b_{T_0,\iota(y)}\rangle = 0.$$

Let $B = \sum_{y \in Z} b_{T_0,\iota(y)}$. Then by the above and the fact that $\langle \pi, b_{T_0,\iota(y)}\rangle = 1/\mathrm{Vol}(\Gamma(q)\backslash\mathscr{H}) \gg 1/q^3$,

$$|\langle b_{T,e} * \chi_{T^\eta} - \pi, B\rangle| \gg \frac{|Z|}{q^3}.$$

On the other hand, by the choice of T_0 and the remarks following it, $\|B\|_2^2 \ll |Z|$. Therefore, using Lemma 2.6 and Cauchy-Schwarz,

$$|\langle b_{T,e} * \chi_{T^\eta} - \pi, B\rangle| \ll \|B\|_2\|b_{T,e} * \chi_{T^\eta} - \pi\|_2 \ll_\epsilon \sqrt{|Z|}q^{-3/2-\eta\tau+\epsilon}.$$

Combining the two estimates and taking $\epsilon = \eta\tau/2$ gives

$$|Z| \ll_\eta q^{3-\eta\tau} = o_\eta(q^3),$$

as needed.

3. Proof of Theorem 1.4

Our goal is to prove that there exists a constant $C > 0$ such that for every prime q, $\epsilon > 0$ and $T \leq Cq^2$, we have

$$\left|\left\{(\gamma, x) \in \mathrm{SL}_3(\mathbb{Z}) \times P^2(\mathbb{F}_q) : \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T, \Phi_q(\gamma)x = x\right\}\right| \ll_\epsilon Tq^{2+\epsilon}.$$

If $\gamma \mod q$ has no eigenspace of dimension 2 or 3, then it has at most 3 eigenvectors in $P^2(\mathbb{F}_q)$. Call such a γ *good* mod q and otherwise call it *bad* mod q. Recall that the number of $\gamma \in \mathrm{SL}_3(\mathbb{Z})$ such that $\|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T$ is bounded up to a constant by the measure of the corresponding set in $\mathrm{SL}_3(\mathbb{R})$ [Duke et al. 1993; Maucourant 2007], which is bounded for every $\epsilon > 0$ by $T^{2+\epsilon}$ [Maucourant 2007]; see also (4-3). Therefore, for $T \leq q^2$,

$$\left|\left\{(\gamma, x) \in \mathrm{SL}_3(\mathbb{Z}) \times P^2(\mathbb{F}_q) : \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T, \Phi_q(\gamma)x = x, \gamma \text{ good} \mod q\right\}\right|$$
$$\ll \left|\left\{\gamma \in \mathrm{SL}_3(\mathbb{Z}) : \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T\right\}\right| \ll T^{2+\epsilon} \ll Tq^{2+\epsilon}.$$

We therefore restrict to the case of bad γ-s. Notice that bad elements do exist and may have a lot of fixed points: e.g., the element $I \in \mathrm{SL}_3(\mathbb{Z})$ is bad $\mod q$ and $\Phi_q(I)$ fixes all of $P^2(\mathbb{F}_q)$. There are two types of bad elements:

- Elements $\gamma \in \mathrm{SL}_3(\mathbb{Z})$ such that $\Phi_q(\gamma) = \alpha I_{\mathrm{SL}_3(\mathbb{F}_q)}$, for $\alpha \in \mathbb{F}_q$, $\alpha^3 = 1$. Such elements will fix the entire space $P^2(\mathbb{F}_q)$.

- In any other case, $\Phi_q(\gamma)$ will have one eigenspace of dimension 2, and possibly another eigenspace of dimension 1. Thus $\Phi_q(\gamma)$ fixes at most $q + 1 + 1$ elements in $P^2(\mathbb{F}_q)$.

Assuming that we choose $C < \frac{1}{4}$, it will hold that either $\|\gamma\|_\infty < q/2$ or $\|\gamma^{-1}\|_\infty < q/2$. On the other hand, if $\gamma \neq I$ and $\Phi_q(\gamma) = \alpha I_{\mathrm{SL}_3(\mathbb{F}_q)}$, then γ and γ^{-1} will have a nonzero entry divisible by q, which is a contradiction. Therefore, we may assume that for each bad γ, $\Phi_q(\gamma)$ will fix at most $q + 1$ elements in $P^2(\mathbb{F}_q)$.

It thus suffices to prove that for some $C > 0$, and $T \leq Cq^2$,

$$\left|\left\{\gamma \in \mathrm{SL}_3(\mathbb{Z}) : \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T, \gamma \text{ bad} \mod q\right\}\right| \ll_\epsilon Tq^{1+\epsilon}.$$

Assume that γ is bad $\mod q$ and $\|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T$. Without loss of generality assume that $\|\gamma\|_\infty \leq \|\gamma^{-1}\|_\infty \leq T^{1/2} < q/2$. We identify elements of $\mathbb{F}_q$ with integers of absolute value at most $q/2$. Thus, once we know the value of an entry of $\gamma \mod q$ we know the same entry in γ.

We divide the range of $\|\gamma\|_\infty$ into $O(\log T)$ dyadic subintervals. Denote by S the bound on $\|\gamma\|_\infty$ and by R the bound on $\|\gamma^{-1}\|_\infty$. Then it is enough to prove that there exists $C > 0$ such that for every $RS \leq Cq^2$ and $S \leq R$ it holds that

$$\left|\left\{\gamma \in \mathrm{SL}_3(\mathbb{Z}) : \|\gamma\|_\infty \leq S, \|\gamma^{-1}\|_\infty \leq R, \gamma \text{ bad} \mod q\right\}\right| \ll_\epsilon RSq^{1+\epsilon}.$$

It will be useful to understand the behavior of bad γ. Let $\alpha \in \mathbb{F}_q \setminus \{0\}$ be the eigenvalue of $\gamma \mod q$ with an eigenspace of dimension 2. Then the third eigenvalue is $\alpha^{-2} \mod q$.

From this it follows that $(\gamma - \alpha I)(\gamma - \alpha^{-2} I) = 0 \bmod q$, or,

$$\gamma + \alpha^{-1} \gamma^{-1} = \alpha + \alpha^{-2} \bmod q. \tag{3-1}$$

By considering the trace of γ and γ^{-1} we have that

$$\operatorname{tr} \gamma = \alpha + 2\alpha^{-2} \bmod q, \quad \operatorname{tr} \gamma^{-1} = \alpha^{-1} + 2\alpha^2 \bmod q. \tag{3-2}$$

Finally, identify α with some lift of it in $\mathbb{Z}$. Then $\gamma - \alpha I \bmod q$ is of rank 1, which means that $\det(\gamma - \alpha I) = 0 \bmod q^2$. Since $\det \gamma = 1$, we have

$$\det(\gamma - xI) = 1 - \operatorname{tr} \gamma^{-1} x + \operatorname{tr} \gamma x^2 - x^3,$$

and hence

$$\alpha^2 \operatorname{tr} \gamma - \alpha \operatorname{tr} \gamma^{-1} = \alpha^3 - 1 \bmod q^2. \tag{3-3}$$

Denote the entries of γ by a_{ij}, $1 \le i, j \le 3$ and the entries of γ^{-1} by b_{ij}, $1 \le i, j \le 3$.

There are $\le (2S+1)^3$ options for choosing the diagonal a_{11}, a_{22}, a_{33} of γ, and once we know them, we know $\operatorname{tr} \gamma$. By (3-2) α (when considered as an element of $\mathbb{F}_q$) is a root of a known third degree polynomial, so there are at most 3 options for α. By (3-3) we know $\operatorname{tr} \gamma^{-1} \bmod q^2$. Since $R \le RS \le Cq^2 < q^2/4$, we may assume that $|\operatorname{tr} \gamma^{-1}| < q^2/2$, so now we know $\operatorname{tr} \gamma^{-1}$.

By (3-1) we now know the diagonal $b_{11}, b_{22}, b_{33} \bmod q$ of $\gamma^{-1} \bmod q$. Since the entries b_{11}, b_{22}, b_{33} are bounded in absolute value by R, we have at most $2R/q + 1$ options for each of them. We may guess b_{11}, b_{22} and get b_{33} since we know $\operatorname{tr} \gamma^{-1}$.

In total, we had $\ll S^3 (R/q + 1)^2$ options so far. We call the case where $a_{ii} a_{jj} = b_{kk}$ for some $\{i, j, k\} = \{1, 2, 3\}$ exceptional. We will deal with it later and assume for now that we are in the nonexceptional case.

Notice that $a_{11} a_{22} - a_{12} a_{21} = b_{33}$, or

$$a_{12} a_{21} = a_{11} a_{22} - b_{33}.$$

Since we are in the nonexceptional case, the right hand side is not 0. By the divisor bound there are at most $\ll_\epsilon q^\epsilon$ options for a_{12}, a_{21}. Similarly, all the other entries $a_{13}, a_{31}, a_{23}, a_{32}$ have at most $\ll_\epsilon q^\epsilon$ options.

In total, we counted $\ll_\epsilon q^\epsilon S^3 (R/q + 1)^2$ bad γ-s in the nonexceptional case. We postpone the exceptional case to the end of the proof. The same (and better) bounds hold for it as well.

It remains to show that

$$S^3 (R/q + 1)^2 \ll RSq,$$

assuming $S \le R$, $RS \le Cq^2$.

If $R \le q$, then we need to show that $S^3 \ll RSq$, or $S^2 \ll Rq$, which is obvious since $S \le R \le q$.

If $R > q$ then we need to show that $S^3 R^2/q^2 \ll RSq$, or $S^2 R \ll q^3$. Since $RS \le Cq^2$, this reduces to showing that $S \ll q$, which is obvious since $S^2 \le RS \le Cq^2$.

Exceptional cases. Recall that the exceptional case is when $a_{ii}a_{jj} = b_{kk}$ for some $\{i, j, k\} = \{1, 2, 3\}$. Assume without loss of generality that $a_{11}a_{22} = b_{33}$. Therefore $a_{12}a_{21} = a_{11}a_{22} - b_{33} = 0$.

We know that $\gamma - \alpha I \mod q$ is of rank 1, so each determinant of a 2×2 submatrix of γ equals $0 \mod q$. Therefore

$$(a_{11} - \alpha)(a_{22} - \alpha) - a_{12}a_{21} = 0 \mod q,$$

so

$$(a_{11} - \alpha)(a_{22} - \alpha) = 0 \mod q.$$

Without loss of generality again, we may assume that $a_{11} = \alpha \mod q$. By our assumptions on the size of the matrix, we may lift α to some fixed element in $\mathbb{Z}$ of absolute value $\leq q/2$ and let $a_{11} = \alpha$. By the above, $a_{12}a_{21} = 0$, and by symmetry again, we may assume that $a_{21} = 0$. Some more minors give:

$$a_{31}(a_{22} - \alpha) = a_{21}a_{32} = 0 \mod q. \tag{3-4}$$

$$a_{31}a_{23} = a_{21}(a_{33} - \alpha) = 0 \mod q. \tag{3-5}$$

We now divide into two cases according to whether $a_{31} = 0$:

Case 1 $a_{11} = \alpha$, $a_{21} = 0$, $a_{31} = 0$. In this case, the matrix is of the form

$$\gamma = \begin{pmatrix} \alpha & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix}.$$

Denote $A = \left(\begin{smallmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{smallmatrix}\right)$. It holds that $\alpha \det A = 1$. Therefore $\alpha = \pm 1$ and $\det A = \pm 1$. We also know that the eigenvalues of $A \mod q$ are either ± 1 (if $\alpha = -1$) or 1 with multiplicity 2 (if $\alpha = 1$). Therefore the trace of A is either 0 or 2. We now separate into two further cases. In the first case $a_{22} \neq \alpha$ and $a_{33} \neq \alpha$, or equivalently $a_{22}a_{33} \neq \det A$. In the second case we may assume without loss of generality that $a_{22} = \alpha$.

    Case 1a $a_{11} = \alpha$, $a_{21} = 0$, $a_{31} = 0$, $a_{22} \neq \alpha$, $a_{33} \neq \alpha$. The entry a_{22} has $2S + 1$ options, and it determines the value of a_{33} since we know the trace of A. In this subcase it holds that $a_{23}a_{32} = \det A - a_{22}a_{33} \neq 0$. By the divisor bound there are $\ll_\epsilon S^\epsilon$ options for a_{23}, a_{32} and both are nonzero. We also know that the third column of $\gamma - \alpha I \mod q$ is a multiple of the second column, and now we know the ratio. This means that after we choose a_{12} in $2S + 1$ ways it sets a_{13} uniquely. Therefore there are $\ll_\epsilon S^{2+\epsilon} \leq RSq^\epsilon$ options in this case.

    Case 1b $a_{11} = \alpha$, $a_{21} = 0$, $a_{31} = 0$, $a_{22} = \alpha$, $a_{33} = 1$. In this case $a_{23}a_{32} = \det A - a_{22}a_{33} = 0$. If $a_{23} \neq 0$ then $a_{32} = a_{12} = 0$ and there are $\leq (2S + 1)^2$ options for a_{23}, a_{13}. Similarly, if $a_{32} \neq 0$ then $a_{23} = 0$ and once we know a_{12} we also know a_{13}. Therefore there are $\ll S^2 \leq RS$ option in this case.

Case 2 $a_{11} = \alpha$, $a_{21} = 0$, $a_{31} \neq 0$. By (3-4), (3-5) we have $a_{22} = \alpha$, $a_{23} = 0$, and hence

$$\gamma - \alpha I = \begin{pmatrix} 0 & a_{12} & a_{13} \\ 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} - \alpha \end{pmatrix}$$

Since its rank mod q is 1 and $a_{31} \neq 0$ the second and third columns are scalar multiples of the first, thus $a_{12} = a_{13} = 0$. Therefore γ is of the form

$$\gamma = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Since $\det \gamma = 1$ it holds that $\alpha = \pm 1$, $a_{33} = 1$ and there are $\ll S^2 \leq RS$ options for γ.

4. Proof of Theorem 1.3

As in the proof of Theorem 1.1, the proof of Theorem 1.3 is analytic, and employs the combinatorial Theorem 1.4 as an input. Since we wish to use the usual notations of dividing $SL_3(\mathbb{R})$ by $SL_3(\mathbb{Z})$ from the left, we apply a transpose to the question as stated in Theorem 1.3.

Let

$$\Gamma_0(q) = \left\{ \begin{pmatrix} * & * & * \\ * & * & * \\ a & b & * \end{pmatrix} \in SL_3(\mathbb{Z}) : a = b = 0 \bmod q \right\}.$$

We have a right action of $\Gamma = SL_3(\mathbb{Z})$ on $\Gamma_0(q)$. We let $P_q^{tr} = \Gamma_0(q) \backslash \Gamma$ (it is obviously isomorphic to P_q as a set with a Γ action). Then Theorem 1.3 can be stated in the following equivalent formulation:

Theorem 4.1. *As $q \to \infty$ among primes, for every $\epsilon > 0$ there exists a set $Y \subset \Gamma_0(q) \backslash \Gamma = P_q^{tr}$ of size $|Y| \geq (1 - o_\epsilon(1))|P_q^{tr}|$, such that for every $x_0 \in Y$, there exists a set $Z_{x_0} \subset P_q^{tr}$ of size $|Z_{x_0}| \geq (1 - o_\epsilon(1))|P_q^{tr}|$, such that for every $y \in Z_{x_0}$, there exists an element $\gamma \in \Gamma$ satisfying $\|\gamma\|_\infty \leq q^{1/3+\epsilon}$, such that $x_0 \gamma = y$.*

Let $K = SO(3)$ be the maximal compact subgroup of $G = SL_3(\mathbb{R})$. By the Cartan decomposition each element $g \in G$ can be written as

$$g = k_1 \begin{pmatrix} a_1 & & \\ & a_2 & \\ & & a_3 \end{pmatrix} k_2,$$

with $k_1, k_2 \in SO(3)$, and unique $a_1, a_2, a_3 \in \mathbb{R}_{>0}$, satisfying $a_1 \geq a_2 \geq a_3 > 0$ and $a_1 a_2 a_3 = 1$. Define $\|g\|_K = a_1$. Since $K = SO(3)$ is compact there exists a constant $C > 0$ such that

$$C^{-1}\|g\|_\infty \leq \|g\|_K \leq C\|g\|_\infty.$$

We may therefore prove Theorem 4.1 using $\| \cdot \|_K$ instead of $\| \cdot \|_\infty$.

The size $\| \cdot \|_K$ will play the same role as $\| \cdot \|_{\mathcal{H}}$ in the SL_2 case. Let us note some of its properties. There is a constant $C > 0$ such that $\|g_1 g_2\|_K \leq C\|g_1\|_K \|g_2\|_K$ (actually, one may take $C = 1$, but this detail will not influence us). A big difference from the SL_2 case comes from the fact that $\|\gamma\|_K$ and $\|\gamma^{-1}\|_K$ can be quite different. However, it does hold that $\|\gamma\|_K \ll \|\gamma^{-1}\|_K^2$.

It will also be useful to define another bi-K invariant gauge of largeness, by $\|g\|_\delta = a_1 a_3^{-1}$, where a_1, a_3 are as in the Cartan decomposition. It holds that there is a constant $C > 0$ such that

$$C^{-1}\|g\|_\infty \|g^{-1}\|_\infty \leq \|g\|_\delta \leq C\|g\|_\infty \|g^{-1}\|_\infty. \tag{4-1}$$

Now we have $\|g\|_\delta = \|g^{-1}\|_\delta$, and there is $C > 0$ (which may be chosen to be $C = 1$ by extra analysis) such that $\|g_1 g_2\|_\delta \leq C\|g_1\|_\delta \|g_2\|_\delta$.

The relation between the two sizes is that $\|g\|_\delta \leq \|g\|_K^3$, which follows from the fact that in the Cartan decomposition $a_3^{-1} = a_1 a_2 \leq a_1^2$, so $a_1 a_3^{-1} \leq a_1^3$.

We will want to estimate the size of balls relative to $\|\cdot\|_K$ and $\|\cdot\|_\delta$. For this, we use the following formula for the Haar measure μ of G [Knapp 1986, Proposition 5.28], which holds up to multiplication by a scalar $C > 0$:

$$\int_G f(g)\, d\mu = C \int_K \int_K \int_{\mathfrak{a}_+} f(k \exp(a) k') S(a)\, dk\, dk'\, da,$$

where

$$\mathfrak{a}_+ = \left\{ a = \begin{pmatrix} \alpha_1 & & \\ & \alpha_2 & \\ & & \alpha_3 \end{pmatrix} \in M_3(\mathbb{R}) : \alpha_1 \geq \alpha_2 \geq \alpha_3, \alpha_1 + \alpha_2 + \alpha_3 = 0 \right\},$$

and

$$S(a) = \sinh(\alpha_1 - \alpha_2) \sinh(\alpha_2 - \alpha_3) \sinh(\alpha_3 - \alpha_1).$$

Notice that for $\alpha_1 - \alpha_2 \geq 1$, $\alpha_2 - \alpha_3 \geq 1$, $S(a)$ behaves like $\|a\|_\delta^2$. This implies that

$$\mu(\{g \in G : \|g\|_K \leq T\}) \asymp T^6, \tag{4-2}$$

and

$$\mu(\{g \in G : \|g\|_\delta \leq T\}) \asymp \log(T)T^2. \tag{4-3}$$

For completeness let us explain the calculation of (4-3), the calculation for (4-2) is similar; see [Maucourant 2007; Gorodnik and Weiss 2007] for more accurate and general statements. To simplify notations we identify $a \in \mathfrak{a}_+$ with $a = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3$. The condition $a_1 a_3^{-1} \leq T$ translates under the inverse of the exponential map and the Cartan decomposition to $\alpha_1 - \alpha_3 \leq \log T$. Denote

$$B(T) = \{a \in \mathfrak{a}_+ : \alpha_1 - \alpha_3 \leq \log T\}$$

Since $\{g \in G : \|g\|_\delta \leq T\} = K \exp(B(T))K$,

$$\mu(\{g \in G : \|g\|_\delta \leq T\}) = \int_{a \in B(T)} S(a)\, da.$$

Let us parametrize the set $B(T)$ by looking at the vectors $v_1 = (1, -1, 0)$, $v_2 = (-1, 2, -1)$. Then $B(T) = \left\{ sv_1 + tv_2 : 0 \leq s \leq \frac{1}{2}\log T, 0 \leq t \leq s/2 \right\}$. We therefore get

$$\mu(\{g \in G : \|g\|_\delta \leq T\}) = \int_0^{\frac{1}{2}\log T} \int_0^{s/2} \sinh(2s - 3t) \sinh(3t) \sinh(2s)\, dt\, ds.$$

Using the upper bound $\sinh x \le e^x$, the above is upper bounded by

$$\le \int_0^{\frac{1}{2}\log T} \int_0^{s/2} e^{4s}\, dt\, ds \le \int_0^{\frac{1}{2}\log T} s e^{4s}/2\, ds \ll \log(T)T^2.$$

Using the lower bound $\sinh x \ge e^x/4$ for $x \ge 1$, we get the lower bound

$$\ge \int_{\frac{1}{2}\log T - 1}^{\frac{1}{2}\log T} \int_1^{s/2-1} e^{4s}\, dt\, ds \gg \log(T)T^2.$$

Let $\chi_T, \chi_{T,\delta} \in L^1(K\backslash G/K)$ be

$$\chi_T(g) = \frac{1}{\mu(\{g \in G : \|g\|_K \le T\})} \begin{cases} 1 & \|g\|_K \le T, \\ 0 & \text{else,} \end{cases} \quad \chi_{T,\delta}(g) = \frac{1}{\mu(\{g \in G : \|g\|_\delta \le T\})} \begin{cases} 1 & \|g\|_\delta \le T, \\ 0 & \text{else.} \end{cases}$$

The functions χ_T, $\chi_{T,\delta}$ are simply the probability characteristic functions of the balls according to $\|\cdot\|_K$ and $\|\cdot\|_\delta$.

By (4-2), (4-3) and the definition of $\|\cdot\|_K$, $\|\cdot\|_\delta$, for every $g \in G$,

$$\chi_T(g) \gg \log T \chi_{T^3,\delta}(g).$$

Let $\psi_T : G \to \mathbb{R}$ be

$$\psi_T(g) = \frac{1}{T} \begin{cases} \|g\|_\delta^{-1} & \|g\|_\delta \le T, \\ 0 & \text{else.} \end{cases}$$

For $f : G \to \mathbb{C}$, we let $f^* : G \to \mathbb{C}$ be the function $f^*(g) = \overline{f(g^{-1})}$.

Now we have the following version of Lemma 2.2:

Lemma 4.2 (convolution lemma). *There exists a constant $C > 0$ such that for $T \ge 1$*

$$\chi_{T,\delta} * \chi_{T,\delta}(g) \le (\log T + 2)^C \psi_{CT^2}(g).$$

As a result, there exist a constant $C' > 0$ such that for $T \ge 1$

$$\chi_T * \chi_T^* \le (\log T + 2)^{C'} \psi_{C'T^6}(g).$$

Proof. Normalize K to have measure 1. Let $\Xi : G \to \mathbb{R}_+$ be Harish-Chandra's function, defined as

$$\Xi(g) = \int_K \delta^{-1/2}(gk)\, dk,$$

where $\delta : G \to \mathbb{R}_{>0}$ is defined, using the Iwasawa decomposition $G = KP$, as

$$\delta\left(k \begin{pmatrix} a_1 & * & * \\ 0 & a_2 & * \\ 0 & 0 & a_3 \end{pmatrix}\right) = a_1^2 a_3^{-2}.$$

(When restricted to P, δ is the modular function of P. Notice the similarity between $\delta(g)$ and $\|g\|_\delta^2$, hence the notation.)

There are standard bounds on Ξ, given by (see, e.g., [Trombi and Varadarajan 1972, 2.1])

$$\|g\|_\delta^{-1} \le \Xi(g) \ll (\log\|g\| + 1)^{C_0} \|g\|_\delta^{-1} \tag{4-4}$$

for some $C_0 > 0$. Using these upper bounds, we find that for some $C_2 > 0$,

$$\int_G \chi_{T,\delta} \Xi(g)\, dg = \frac{1}{\mu(\{g \in G : \|g\|_\delta \le T\})} \int_{g:\|g\|_\delta \le T} \Xi(g)\, dg \ll (\log T + 1)^{C_2} T^{-1}.$$

Harish-Chandra's function Ξ arises as follows; see, e.g., [Ghosh et al. 2013, Section 3]. Let (π, V) be the spherical representation of G unitarily induced from the trivial character of P. It holds that if $f \in L^1(K\backslash G/K)$ and $v \in V$ is K-invariant, then

$$\pi(f)v = \int_G f(g)\pi(g)v\, dg = \left(\int_G f(g)\Xi(g)\, dg\right)v.$$

Since $\pi(f_1 * f_2)v = \pi(f_1)\pi(f_2)v$,

$$\int_G (\chi_{T,\delta} * \chi_{T,\delta})(g)\Xi(g)\, dg = \left(\int_G \chi_{T,\delta}(g)\Xi(g)\, dg\right)\left(\int_G \chi_{T,\delta}(g)\Xi(g)\, dg\right) \ll (\log T + 1)^{2C_2} T^{-2}.$$

To show pointwise bounds, we notice that if $\chi_{T,\delta} * \chi_{T,\delta}(g) = R$, then $\chi_{T+1,\delta} * \chi_{T+1,\delta}(g') \gg R$, for g' in an annulus of size similar to that of g, i.e., for $C^{-1}\|g\|_\delta \le \|g'\|_\delta \le C\|g\|_\delta$ for some $C > 1$. This annulus is of measure $\asymp \|g\|_\delta^2$. Therefore,

$$\chi_{T,\delta} * \chi_{T,\delta}(g)\|g\|_\delta^2 \Xi(g) \ll \int_G (\chi_{T+1,\delta} * \chi_{T+1,\delta})(g')\Xi(g')\, dg' \ll (\log T + 1)^{2C_2} T^{-2},$$

and the first bound follows by applying the lower bound of (4-4).

The bound on χ_T follows from the bound on $\chi_{T,\delta}$ and the relation between them. $\square$

Now consider the locally symmetric space $X_q = \Gamma_0(q)\backslash G/K$. As in the SL_2 case, it has finite measure, and we will consider the space $L^2(X_q)$, with the natural L^2-norm.

We first discuss the spectral gap. We denote by $L_0^2(X_q)$ the functions in $L^2(X_q)$ of integral 0. Since χ_T is bi-K-invariant and sufficiently nice, the function χ_T acts by convolution from the right on $f \in L^2(X_q)$, and the resulting function is well defined pointwise if f is bounded. The operation sends $L_0(X_q)$ to itself.

Theorem 4.3 (spectral gap). *There exists $\tau > 0$ such that for $T > 0$ the operator χ_T satisfies for every $f \in L_0^2(X_q)$,*

$$\|f * \chi_T\|_2 \ll T^{-\tau}\|f\|_2.$$

The theorem follows from explicit versions of property (T), or explicit versions of the mean ergodic theorem (e.g., [Ghosh et al. 2013, Section 4]) which are actually true for all lattices in $G = SL_3(\mathbb{R})$ uniformly in T and the lattice. It is remarkable that the proof of Theorem 4.3 is much simpler than the proof of Theorem 2.5.

As in the SL_2 case, we define for $x_0 \in X_q$ the function $b_{T,x_0}(x) = \sum_{\gamma \in \Gamma_0(q)} \chi_T(\tilde{x}_0^{-1}\gamma x)$, where $\tilde{x}_0$ is any lift of x_0 to G.

We have a map $\iota : \Gamma_0(q)\backslash\Gamma \to X_q$ defined by $\iota(\Gamma_0(q)x_0) = \Gamma_0 x_0 K \in X_q$. By a slight abuse of notation we write $\iota(\Gamma_0(q)x_0) = \iota(x_0)$.

The map ι has fibers of bounded size (independently of q), and we may choose T_0 small enough so that $\iota(y) \neq \iota(y')$ implies that $b_{T_0,\iota(y)}$ and $b_{T_0,\iota(y')}$ have disjoint supports. In addition, $b_{T_0,\iota(y)}$ will have a bounded L^2-norm as a function in $L^2(X_q)$.

Lemma 4.4. *For $f \in L^2(X_q)$ bounded,*

$$\langle f, b_{T,x_0}\rangle = (f * \chi_T^*)(x_0).$$

The proof is the same as the proof of Lemma 2.3.

Lemma 4.5. *Let $C > 0$, $\epsilon_0 > 0$ fixed. Let $x_0 \in \Gamma_0(q)\backslash\Gamma$ and assume for $T' \leq Cq^2$,*

$$|\{\gamma \in \Gamma : \|\gamma\|_\delta \leq T', x_0\gamma = x_0\}| \ll_{\epsilon_0} q^{\epsilon_0}T'.$$

Then there exists $C' > 0$ depending only on C such that for $T = C'q^{1/3}$ it holds that for every $\epsilon > 0$,

$$\|b_{T,\iota(x_0)}\|_2 \ll_{\epsilon_0,\epsilon} q^{-1+\epsilon_0+\epsilon}.$$

Proof. Notice that $\gamma \in \Gamma$ satisfies $\Gamma_0(q)x_0\gamma = \Gamma_0(q)x_0$ if and only if $\gamma \in x_0^{-1}\Gamma_0(q)x_0$ (the last group is a well defined subgroup of Γ). Therefore we may rewrite the assumption in the following manner: For every $T' \leq Cq^2$,

$$|\{\gamma \in \Gamma_0(q) : \|x_0^{-1}\gamma x_0\|_\delta \leq T'\}| \ll_{\epsilon_0} q^{\epsilon_0}T', \tag{4-5}$$

where we identify x_0 with a fixed element of $\Gamma \leq G$.

Write using Lemma 4.4,

$$\|b_{T,\iota(x_0)}\|_2^2 = \langle b_{T,\iota(x_0)}, b_{T,\iota(x_0)}\rangle = b_{T,\iota(x_0)} * \chi_T^*(\iota(x_0)) = \sum_{\gamma \in \Gamma_0(q)} (\chi_T * \chi_T^*)(x_0^{-1}\gamma x_0) \ll_\epsilon T^\epsilon \psi_{C_1 T^6}(x_0^{-1}\gamma x_0),$$

where in the last inequality we used Lemma 4.2.

Therefore, the lemma will follow if we will prove that for $T = C'q^{1/3}$,

$$\sum_{\gamma \in \Gamma_0} \psi_{C_1 T^6}(x_0^{-1}\gamma x_0) = T^{-6} \sum_{\gamma \in \Gamma_0(q):\|x_0^{-1}\gamma x_0\|_\delta \leq C_1 T^6} \|x_0^{-1}\gamma x_0\|_\delta^{-1}$$

$$\ll q^{-2} \sum_{\gamma \in \Gamma_0(q):\|x_0^{-1}\gamma x_0\|_\delta \leq C_2 q^2} \|x_0^{-1}\gamma x_0\|_\delta^{-1}$$

$$\overset{!}{\ll_\epsilon} q^{-2+\epsilon_0+\epsilon},$$

where $C_2 = C_1 C'^6$.

So it suffices to show that

$$\sum_{\gamma \in \Gamma_0(q):\|x_0^{-1}\gamma x_0\|_\delta^{-1} \leq C_2 q^2} \|x_0^{-1}\gamma x_0\|_\delta \overset{!}{\ll_{\epsilon,\epsilon_0}} q^{\epsilon+\epsilon_0}.$$

Applying (2-1) (discrete partial summation), with $g(\gamma) = \|\gamma\|_\delta$, $f(x) = x^{-1}$ and $Y = C_2 q^2$, we have

$$\sum_{\substack{\gamma \in \Gamma_0(q) \\ \|x_0^{-1}\gamma x_0\|_\delta \leq C_2 q^2}} \|x_0^{-1}\gamma x_0\|_\delta^{-1} \ll \left| \{\gamma : \|x_0^{-1}\gamma x_0\|_\delta \leq C_2 q^2\} \right| q^{-2} + \int_1^{C_2 q^2} \left| \{\gamma : \|x_0^{-1}\gamma x_0\|_\delta \leq S\} \right| S^{-2}\, dS.$$

Choosing C' small enough so that $C_2 = C_1 C'^6 \leq C$ and applying (4-5) we obtain the desired bound for the last value:

$$\ll_{\epsilon, \epsilon_0} q^{\epsilon + \epsilon_0} + q^{\epsilon + \epsilon_0} \int_1^{C_3 q^2} S^{-1}\, dS \ll_\epsilon q^{2\epsilon + \epsilon_0}. \qquad \square$$

We denote by $\pi \in L^2(X_q)$ the constant probability function on X_q.

Using the counting result Theorem 1.4 we will now show that for many points $x_0 \in \Gamma_0(q) \backslash \Gamma$ the condition of Lemma 4.5 holds, and thus obtain:

Lemma 4.6. *There exists $C > 0$, $\tau > 0$, such that for every $\epsilon_0 > 0$, as $q \to \infty$ among primes, there exists a set $Y \subset \Gamma_0(q) \backslash \Gamma = P_q^{tr}$ of size $|Y| \geq (1 - o_{\epsilon_0}(1))|\Gamma_0(q) \backslash \Gamma|$, such that for every $\Gamma_0 x_0 \in Y$, it holds for $T = Cq^{1/3}$ that*

$$\|b_{T, \iota(x_0)} * \chi_{T^\eta} - \pi\|_2 \ll_{\epsilon_0} q^{-1 - \eta\tau + \epsilon_0}.$$

Proof. By Theorem 1.4 and (4-1) it holds that for some $C > 0$, for all $T \leq Cq^2$ and $\epsilon > 0$

$$\sum_{x_0 \in \Gamma_0(q) \backslash \Gamma} |\{\gamma \in \Gamma : \|\gamma\|_\delta \leq T, x_0 \gamma = x_0\}| \ll_\epsilon q^{2+\epsilon} T.$$

Since $|\Gamma_0(q) \backslash \Gamma| = (1 + o(1))q^2$, we may choose a subset $Y \subset \Gamma_0(q) \backslash \Gamma$ of size

$$|Y| \geq (1 - o_{\epsilon_0}(1))|\Gamma_0(q) \backslash \Gamma|,$$

such that for every $x_0 \in Y$,

$$|\{\gamma \in \Gamma : \|\gamma\|_\delta \leq T, x_0 \gamma = x_0\}| \ll_{\epsilon_0} q^{\epsilon_0} T.$$

We now apply Lemma 4.5 to every $x_0 \in Y$ to obtain

$$\|b_{T, \iota(x_0)}\|_2 \ll_{\epsilon_0} q^{-1 + \epsilon_0}.$$

Next, we apply Theorem 4.3 as in Lemma 2.6 to deduce the final result. $\qquad \square$

We may now finish the proof of Theorem 4.1, similar to the SL_2 case.

Lemma 4.7. *There is $C' > 0$ such that for $x_0, y \in \Gamma_0(q) \backslash \Gamma$, if $\langle b_{T, \iota(x_0)} * \chi_{T^\eta}, b_{T_0, \iota(y)} \rangle > 0$, then there is $\gamma \in \Gamma$ such that $x_0 \gamma = y$, and $\|\gamma\|_K \leq C' T^{1+\eta}$.*

Proof. The proof is essentially the same as Lemma 2.7. We have by Lemma 4.4

$$b_{T, \iota(x_0)} * \chi_{T^\eta} * \chi_{T_0}^*(\iota(y_0)) > 0.$$

Denote by $\tilde{x}_0$, $\tilde{y}$ as some lifts of x_0, y to Γ. We get g_1, g_2, $g_3 \in G$, $\gamma \in \Gamma_0(q)$ such that $\gamma^{-1}\tilde{x}_0 g_1 g_2 g_3 = \tilde{y}$, with $g_1 \in \text{supp}(\chi_T)$, $g_2 \in \text{supp}(\chi_{T^\eta})$, $g_3 \in \text{supp}(\chi_{T_0}^*)$. Writing $g = g_1 g_2 g_3$, we have that

$$\|g\|_K \ll \|g_1\|_K \|g_2\|_K \|g_3\|_K \ll T^{1+\eta}.$$

In addition $g = \tilde{x}_0^{-1} \gamma \tilde{y} \in x_0^{-1} \Gamma_0(q) y \subset \Gamma$, which says that $x_0 \gamma = y$, as needed. $\qquad\square$

To complete the proof, fix $\epsilon > 0$. Let $x_0 \in \Gamma_0(q) \backslash \Gamma$ be in the set Y of Lemma 4.6. Denote by $\tilde{Z}_{x_0}$ the set of elements $y \in \Gamma_0(q) \backslash \Gamma$ for which there is no $\gamma \in \Gamma$ with $\|\gamma\|_K \le q_K^{1/3+\epsilon}$ such that $x_0 \gamma = y$. It is enough to prove that $\tilde{Z}_{x_0} = o(|\Gamma_0(q) \backslash \Gamma|) = o(q^2)$.

Choose $T = Cq^{1/3}$, and η small enough so that $C'T^{1+\eta} < q^{1/3+\epsilon}$, with C as in Lemma 4.6 and C' as in Lemma 4.7.

We denote $B = \sum_{y \in \tilde{Z}_{x_0}} b_{T, \iota(y)} \in L^2(X_q)$. Then by Lemma 4.7

$$\langle b_{T,x_0} * \chi_{T^\eta} - \pi, B \rangle = \frac{|\tilde{Z}_{x_0}|}{\text{Vol}(X_q)} \gg \frac{|\tilde{Z}_{x_0}|}{q^2}.$$

On the other hand, by the choice of x_0 and Lemma 4.6,

$$\langle b_{T,x_0} * \chi_{T^\eta} - \pi, B \rangle \ll \|B\|_2 \|b_{T,x_0} * \chi_{T^\eta} - \pi\|_2 \ll_{\epsilon_0} \sqrt{|\tilde{Z}_{x_0}|} q^{-1-\eta\tau+\epsilon_0}.$$

By combining the two estimates and choosing ϵ_0 small enough, we get the desired result

$$|\tilde{Z}_{x_0}| \ll_{\epsilon_0} q^{2-2\eta\tau-2\epsilon_0} = o(q^2).$$

5. Optimal lifting for the action on flags

In this section we prove optimal lifting for another action of $SL_3(\mathbb{Z})$. Let B_q be the set of complete flags in $\mathbb{F}_q^3$, i.e.,

$$B_q = \{(V_1, V_2) : 0 < V_1 < V_2 < \mathbb{F}_q^3\},$$

i.e., $V_1 \subset V_2$ are subspaces of $\mathbb{F}_q^3$, such that $\dim V_1 = 1$, $\dim V_2 = 2$.

There is a natural action $\Phi_q \colon SL_3(\mathbb{Z}) \to \text{Sym}(B_q)$. It gives rise to a nonprincipal congruence subgroup

$$\Gamma_2'(q) = \left\{ \begin{pmatrix} * & a & b \\ * & * & c \\ * & * & * \end{pmatrix} \in SL_3(\mathbb{Z}) : a = b = c = 0 \bmod q \right\}.$$

Concretely,

$$\Gamma_2'(q) = \{\gamma \in SL_3(\mathbb{Z}) : \Phi_q(\gamma)(\mathbf{1}) = \mathbf{1}\},$$

where

$$\mathbf{1} = \left(\text{span}\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}, \text{span}\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \right)$$

The result reads as follows:

Theorem 5.1. *Let* $\Gamma = \mathrm{SL}_3(\mathbb{Z})$, *and for a prime q let B_q and $\Phi_q : \mathrm{SL}_3(\mathbb{Z}) \to \mathrm{Sym}(B_q)$ as above. Then for every $\epsilon > 0$, as $q \to \infty$, there exists a set $Y \subset B_q$ of size $|Y| \geq (1 - o_\epsilon(1))|B_q|$, such that for every $x \in Y$, there exists a set $Z_x \subset B_q$ of size $|Z_x| \geq (1 - o_\epsilon(1))|B_q|$, such that for every $y \in Z_x$, there exists an element $\gamma \in \Gamma$ satisfying $\|\gamma\|_\infty \leq q^{1/2+\epsilon}$, such that $\Phi_q(\gamma)x = y$.*

The exponent $\frac{1}{2}$ is optimal, since the size of B_q is $|B_q| \asymp q^3$, while the number of elements $\gamma \in \mathrm{SL}_3(\mathbb{Z})$ satisfying $\|\gamma\|_\infty \leq T$ is $\asymp T^6$. This also hints why handling flags is harder than handling the projective plane: The volume of the homogenous space is larger (q^3 instead of q^2). In comparison, the principal congruence subgroup gives the much larger volume q^8, and optimal lifting for it is still open.

The proof of Theorem 5.1 is very similar to the proof of Theorem 1.3. The analytic part is essentially identical to Section 4, with some minor modifications coming from the fact that the size $|P_q| \asymp q^2$ is replaced by $|B_q| \asymp q^3$. We therefore leave it to the reader.

The counting part needs a slightly more delicate argument. The needed result is an analog of Theorem 1.4, as follows:

Theorem 5.2. *There exists a constant $C > 0$ such that for every prime q, $T \leq Cq^3$ and $\epsilon > 0$ it holds that*

$$|\{(\gamma, x) \in \mathrm{SL}_3(\mathbb{Z}) \times B_q : \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T, \Phi_q(\gamma)(x) = x\}| \ll_\epsilon q^{3+\epsilon} T.$$

We prove Theorem 5.2 in the rest of this section.

By dyadically dividing the range of $\|\gamma\|_\infty$ into $O(\log T)$ subintervals, it is enough to prove that there exists $C > 0$ such that for every $S \leq R$ and $RS \leq Cq^3$:

$$|\{(\gamma, x) \in \mathrm{SL}_3(\mathbb{Z}) \times B_q : \|\gamma\|_\infty \leq S, \|\gamma^{-1}\|_\infty \leq R, \Phi_q(\gamma)(x) = x\}| \ll_\epsilon q^{3+\epsilon} RS.$$

We identify $\Phi_q(\gamma) \in \mathrm{SL}_3(\mathbb{F}_q)$, and let $P(t) \in \mathbb{F}_q[t]$ be the characteristic polynomial of $\Phi_q(\gamma)$.

We first notice that if $x = (V_1, V_2) \in B_q$ is a fixed point of $\Phi_q(\gamma)$, then V_1 defines a projective eigenvector of $\Phi_q(\gamma)$, so $\Phi_q(\gamma)$ has an eigenvector and $P(t)$ has a root. Similarly, V_2 is a two-dimensional invariant subspace containing an eigenvector, so $P(t)$ has at least two roots. We deduce that if $\Phi_q(\gamma)$ has a fixed point $x \in B_q$, then the polynomial $P(t)$ splits. Assuming $P(t)$ splits, we divide into several subcases:

(1) Assume that $P(t)$ has three different roots, $\Phi_q(\gamma)$ is diagonalizable, with eigenvectors v_1, v_2, v_3. Then $\Phi_q(\gamma)$ fixes the points of the form (V_1, V_2), $V_1 = \mathrm{span}\{v_i\}$, $V_2 = \mathrm{span}\{v_i, v_j\}$, for $1 \leq i \neq j \leq 3$. So $\Phi_q(\gamma)$ has 6 fixed points in B_q.

(2) Assume that $P(t)$ has the roots $(\alpha, \alpha, \alpha^{-2})$, $\alpha^3 \neq 1$, and the eigenspace $\ker(\Phi_q(\gamma) - \alpha I)$ of eigenvalue α eigenvalue α is of dimension 1. Then let v_1 be an eigenvector of eigenvalue α, v_2 an eigenvector of eigenvalue α^{-2}, and $U = \ker(\Phi_q(\gamma) - \alpha I)^2$ the two-dimensional generalized eigenspace of eigenvalue α. Then the fixed points of $\Phi_q(\gamma)$ are of the form $V_1 = \mathrm{span}\{v_i\}$, $V_2 = \mathrm{span}\{v_1, v_2\}$ for $1 \leq i \neq j \leq 2$, or of the form $V_1 = \mathrm{span}\{v_1\}$, $V_2 = U$. So $\Phi_q(\gamma)$ has 3 fixed points in B_q.

(3) Assume that $P(t)$ has a triple root $\alpha \in \mathbb{F}_q$ (with $\alpha^3 = 1$), and the eigenspace $\ker(\Phi_q(\gamma) - \alpha I)$ is one dimensional, then the only fixed point is of the form $V_1 = \ker(\Phi_q(\gamma) - \alpha I)$, $V_2 = \ker(\Phi_q(\gamma) - \alpha I)^2$. So $\Phi_q(\gamma)$ has a unique fixed point in B_q.

(4) Assume that $P(t)$ has roots $(\alpha, \alpha, \alpha^{-2})$, $\alpha^3 \neq 1$, and the eigenspace $U = \ker(\Phi_q(\gamma) - \alpha I)$ of eigenvalue α is of dimension 2, i.e., the Jordan form of $\Phi_q(\gamma)$ is

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha^{-2} \end{pmatrix}, \quad \alpha^3 \neq 1.$$

Let $v_1 \in U$ denote an eigenvector of eigenvalue α, and let v_2 be an eigenvector of eigenvalue α^{-2}. Then all fixed points of $\Phi_q(\gamma)$ are of the form $V_1 = \text{span}\{v_i\}$, $v_2 = \text{span}\{v_i, v_j\}$, $1 \leq i \neq j \leq 2$, or $V_1 = \text{span}\{v_1\}$, $V_2 = U$ (for different choices of v_1). There are $(q+1)$ options for $\text{span}\{v_1\}$, so in total $\Phi_q(\gamma)$ has $3(q+1)$ fixed points in B_q.

(5) If $P(t)$ has a unique root $\alpha \in \mathbb{F}_q$, $\alpha^3 = 1$, and the eigenspace $U = \ker(\Phi_q(\gamma) - \alpha I)$ of eigenvalue α is of dimension 2, then the Jordan form of $\Phi_q(\gamma)$ is of the form

$$\begin{pmatrix} \alpha & 1 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix} \quad \alpha^3 = 1.$$

In this case, the operator $\Phi_q(\gamma) - \alpha$ is nilpotent, with $\dim \text{Im}(\Phi_q(\gamma) - \alpha) = 1$, $\dim \ker(\Phi_q(\gamma) - \alpha) = 2$. If $x = (V_1, V_2)$ is a fixed point of $\Phi_q(\gamma)$, then $U = V_2 \cap \ker(\Phi_q(\gamma) - \alpha)$ satisfies either $\dim U = 2$ or $\dim U = 1$:

- If $\dim U = 1$, we must choose $V_1 = U$, and $V_1 = (\Phi_q(\gamma) - \alpha)V_2$, and by dimension counting $V_1 = \text{Im}(\Phi_q(\gamma) - \alpha)$. Therefore, V_1 is uniquely defined and V_2 can be chosen as any subspace containing V_1, in $q + 1$ ways.

- If $\dim U = 2$, then $V_2 = \ker(\Phi_q(\gamma) - \alpha I)$ is uniquely defined, and V_1 can be chosen in $q + 1$ ways as a subspace of V_2.

We conclude that $\Phi_q(\gamma)$ has $2(q + 1)$ fixed points in B_q.

(6) If $P(t)$ has a unique root $\alpha \in \mathbb{F}_q$, $\alpha^3 = 1$ and $\Phi_q(\gamma) = \alpha I_{SL_3(\mathbb{F}_q)}$, then every $x \in B_q$ is a fixed point of $\Phi_q(\gamma)$.

As in Section 3, we call $\gamma \in SL_3(\mathbb{Z})$ bad mod q if $\Phi_q(\gamma)$ has an eigenspace of dimension at least 2, i.e., corresponds to one of the last three cases above.

Theorem 5.2 will therefore follow from the following two lemmas:

Lemma 5.3. *There exists $C > 0$ such that for every $\alpha \in \mathbb{F}_q$, $\alpha^3 = 1$, $S \leq R$ and $RS \leq Cq^3$*

$$|\{\gamma \in SL_3(\mathbb{Z}) : \|\gamma\|_\infty \leq S, \|\gamma^{-1}\|_\infty \leq R, \gamma = \alpha I \mod q\}| \ll_\epsilon q^\epsilon RS.$$

Lemma 5.4. *There exists $C > 0$ such that for every $S \leq R$ and $RS \leq Cq^3$*

$$|\{\gamma \in \mathrm{SL}_3(\mathbb{Z}) : \|\gamma\|_\infty \leq S, \|\gamma^{-1}\|_\infty \leq R, \gamma \text{ bad mod } q\}| \ll_\epsilon q^{2+\epsilon} RS.$$

Proof of Lemma 5.3. We will give a short and nonefficient estimate, which may be improved significantly, at least when $\alpha = 1 \in \mathbb{F}_q$.

Fix some lift of α to $\mathbb{Z}$, such that $\alpha^3 = 1 \bmod q^2$. Then $(\gamma - \alpha I)^2 = 0 \bmod q^2$, so $\gamma^2 - 2\alpha\gamma + \alpha^2 I = 0 \bmod q^2$. Multiply by $\alpha\gamma^{-1}$ and we get that

$$\gamma^{-1} = 2\alpha^2 - \alpha\gamma \bmod q^2. \tag{5-1}$$

By the inversion formula, it holds that $\|\gamma^{-1}\|_\infty \leq 2\|\gamma\|_\infty^2$. We may therefore assume that $R \leq 2S^2$, so $R^3 \leq 2R^2 S^2 \leq 2C^2 q^6$. By adjusting the constant C we may assume that $\|\gamma^{-1}\|_\infty \leq R \leq q^2/4$, and (5-1) then implies that given an entry of γ, we know the corresponding entry of γ^{-1}.

As in Section 3, we denote the entries of γ by a_{ij} and the entries of γ^{-1} by b_{ij}.

Since $\gamma = \alpha I \bmod q$, we may choose the diagonal of γ using $\ll (S/q + 1)^3$ options. By (5-1) we know the diagonal of γ^{-1}. We can write $a_{12}a_{21} = a_{11}a_{22} - b_{33}$, and the right hand side is known. If $a_{11}a_{22} - b_{33} \neq 0$ then there are $\ll S^\epsilon$ options for a_{12}, a_{21}. If $a_{11}a_{22} - b_{33} = 0$ then there are $\ll (S/q + 1)$ options for a_{12}, a_{21}. The same is true for the other nondiagonal elements.

All in all, there are

$$\ll_\epsilon (S/q + 1)^3 (S/q + 1 + S^\epsilon)^3$$

options for γ. If $S \leq q$ this is obviously smaller than $q^\epsilon RS$. If $S \geq q$, then we need to show that

$$(S/q)^6 \ll q^\epsilon RS$$

or $S^5/R \ll q^{6+\epsilon}$, which is true since

$$S^5/R \leq S^4 \leq (RS)^2 \leq q^6.$$

$\square$

For the proof of Lemma 5.4 we will need the following:

Lemma 5.5. *The number of solutions for* (3-2), (3-3) *in* $\mathrm{tr}\,\gamma, \mathrm{tr}\,\gamma^{-1} \in \mathbb{Z}, \alpha \in \mathbb{F}_q, |\mathrm{tr}\,\gamma| \leq S, |\mathrm{tr}\,\gamma^{-1}| \leq R$ *is bounded by* $\ll (S/q + 1)(R/q + 1) + q$.

Proof. Assume that $(x_1, y_1, \alpha), (x_2, y_2, \alpha)$ are solutions. Then by (3-2), $x_1 - x_2 = y_1 - y_2 = 0 \bmod q$. Denote $z = (x_1 - y_1)/q, w = (x_2 - y_2)/q$. Notice that $|z| \leq 2S/q, |w| \leq 2R/q$. By (3-3) (z, w, α) is a solution to $\alpha qz - qw = 0 \bmod q^2$, or

$$\alpha z - w = 0 \bmod q. \tag{5-2}$$

Therefore, A solutions with the same $\alpha \in \mathbb{F}_q$ for (3-2),(3-3) give A solutions to (5-2) with the same $\alpha \in \mathbb{F}_q$. So the total number of solutions is bounded by the number of solutions of (5-2) with $|z| \leq 2S/q, |w| \leq 2R/q$,

$\alpha \in \mathbb{F}_q$. The last number is bounded by $\ll (S/q+1)(R/q+1)+q$, since every choice of z, w sets α uniquely, unless $z = w = 0$. $\qquad\square$

Proof of Lemma 5.4. Since our definition of a bad element mod q agrees with the definition in Section 4, by Lemma 5.5 there are $\ll (S/q+1)(R/q+1)+q$ options for $\mathrm{tr}\,\gamma$, $\mathrm{tr}\,\gamma^{-1}$, α. In our range of parameters it holds that $RS \leq Cq^3$ and since $\|\gamma^{-1}\|_\infty \leq 2\|\gamma\|_\infty^2$, we may assume that $R \leq 2S^2$, so $R \ll q^2$, and therefore $(S/q+1)(R/q+1)+q \ll q$.

There are at most S^2 options for a_{11}, a_{22}, and knowing $\mathrm{tr}\,\gamma$, we have now all of the diagonal of γ. By (3-1), the diagonal of γ determines the diagonal of $\gamma^{-1} \bmod q$. Lifting, the first two entries b_{11}, b_{22} have just $(R/q+1)^2$ options, giving b_{33} for free. Thus there are at most $\ll qS^2(R/q+1)^2$ options.

In the nonexceptional case when the nondiagonal entries are nonzero, the rest of the matrix has $\ll_\epsilon q^\epsilon$ options. So we should show that

$$qS^2(R/q+1)^2 \ll RSq^2,$$

or $S(R/q+1)^2 \ll Rq$. For $R < q$, this reduces to $S \ll Rq$, which is obvious. For $R > q$, this reduces to $RS \ll q^3$, which is again true.

Let us deal with the exceptional case. Without loss of generality we may assume that $a_{11}a_{22} = b_{33}$ and $a_{21} = 0$. We further separate into cases:

(1) If all other nondiagonal entries besides a_{21} and a_{12} are nonzero, then we may guess the diagonal of γ and γ^{-1} as before, and get the other nondiagonal entries using divisor bounds. The matrix γ is then of the form

$$\begin{pmatrix} * & ? & \times \\ 0 & * & \times \\ \times & \times & * \end{pmatrix},$$

with a_{12} the only unknown and where $\times$ denotes a nonzero value. Then we get that $\det \gamma = Ea_{12} + F$, with $E = a_{23}a_{31} \neq 0$, F known, so a_{12} is determined uniquely from $\det \gamma = 1$.

(2) If $a_{31} = 0$, then $a_{11} = \alpha = \pm 1$, and the matrix is of the form

$$\begin{pmatrix} \pm 1 & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix}.$$

As in the first exceptional case of Section 3, denote $A = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}$. We know that $\det A = \alpha = \pm 1$, and either $\mathrm{tr}\,A = 0 \bmod q$ or $\mathrm{tr}\,A = 2 \bmod q$. Therefore, a_{22}, a_{33} have at most $\ll S(S/q+1)$ options. If $a_{22}a_{33} \neq \det A = \pm 1$ then we get q^ϵ options for a_{23}, a_{32} by the divisor bound. If $a_{22}a_{33} = \det A = \pm 1$, then they are both ± 1, and $a_{23}a_{32} = 0$, so there are $\ll S$ options for A. So in any case A has at most $q^\epsilon S(S/q+1)$ options. The remaining two entries have at most S^2 options, so all in all there are $S^3(S/q+1)$ options. It remains to prove that

$$S^3(S/q+1) \ll RSq^2,$$

which is a simple verification.

(3) If $a_{23} = 0$ then $a_{22} = \alpha = \pm 1$, and the matrix is of the form

$$\begin{pmatrix} * & * & * \\ 0 & \pm 1 & 0 \\ * & * & * \end{pmatrix}.$$

We reduce to the previous case (after permuting indices and transposing).

(4) We may now assume $a_{31} \neq 0$, $a_{23} \neq 0$. If $a_{13} = 0$, we may assume $a_{12} \neq 0$, otherwise we reduce to a previous case. We now guess the diagonals as before, and further diverge into subcases:

(a) If $a_{32} \neq 0$: Then since $a_{23} \neq 0$ we have $a_{23}a_{32} = a_{22}a_{33} - b_{11}$, so we have $\ll_\epsilon q^\epsilon$ options for a_{23}, a_{32} by the divisor bound. Then the matrix is of the form

$$\begin{pmatrix} * & ? & 0 \\ 0 & * & \times \\ ? & \times & * \end{pmatrix}.$$

From $\det \gamma = 1$ we get $a_{12}a_{31}$, which is nonzero. By the divisor bound we are done.

(b) If $a_{32} = 0$, the matrix is of the form

$$\begin{pmatrix} * & ? & 0 \\ 0 & * & ? \\ ? & 0 & * \end{pmatrix}.$$

From $\det \gamma = 1$ we get $a_{12}a_{23}a_{31}$, which is again nonzero, and by the divisor bound we are done.

(5) If $a_{13} \neq 0$, $a_{23} \neq 0$, $a_{31} \neq 0$, $a_{32} = 0$. We may assume that $a_{12} \neq 0$ otherwise we reduce to a previous case. Then we guess the diagonals as usual, and since $a_{31}a_{13} \neq 0$ we know them in $\ll_\epsilon q^\epsilon$ ways by the divisor bound. Then the matrix is of the form

$$\begin{pmatrix} * & ? & \times \\ 0 & * & ? \\ \times & 0 & \alpha^{-2} \end{pmatrix}.$$

From $\det \gamma = 1$ we get $a_{12}a_{23}$ which is nonzero, and by the divisor bound we are done. $\qquad\square$

Acknowledgments

Added in proof

In a recent preprint Jana and Kamber [2022, Theorem 6], following a breakthrough of Assing and Blomer [2022], proved Conjecture 1.2 for q square-free.

References

[Assing and Blomer 2022] E. Assing and V. Blomer, "The density conjecture for principal congruence subgroups", preprint, 2022. arXiv 2204.08868

[Blomer 2019] V. Blomer, "Density theorems for GL(n)", preprint, 2019. arXiv 1906.07459

[Blomer et al. 2017] V. Blomer, J. Buttcane, and P. Maga, "Applications of the Kuznetsov formula on GL(3) II: the level aspect", *Math. Ann.* **369**:1-2 (2017), 723–759. MR

[Burger et al. 1992] M. Burger, J.-S. Li, and P. Sarnak, "Ramanujan duals and automorphic spectrum", *Bull. Amer. Math. Soc. (N.S.)* **26**:2 (1992), 253–257. MR Zbl

[Duke et al. 1993] W. Duke, Z. Rudnick, and P. Sarnak, "Density of integer points on affine homogeneous varieties", *Duke Math. J.* **71**:1 (1993), 143–179. MR Zbl

[Gamburd 2002] A. Gamburd, "On the spectral gap for infinite index "congruence" subgroups of $SL_2(\mathbb{Z})$", *Israel J. Math.* **127** (2002), 157–200. MR Zbl

[Ghosh et al. 2013] A. Ghosh, A. Gorodnik, and A. Nevo, "Diophantine approximation and automorphic spectrum", *Int. Math. Res. Not.* **2013**:21 (2013), 5002–5058. MR Zbl

[Ghosh et al. 2018] A. Ghosh, A. Gorodnik, and A. Nevo, "Best possible rates of distribution of dense lattice orbits in homogeneous spaces", *J. Reine Angew. Math.* **745** (2018), 155–188. MR Zbl

[Golubev and Kamber 2019] K. Golubev and A. Kamber, "Cutoff on hyperbolic surfaces", *Geom. Dedicata* **203** (2019), 225–255. MR Zbl

[Golubev and Kamber 2020] K. Golubev and A. Kamber, "On Sarnak's density conjecture and its applications", preprint, 2020. arXiv 2004.00373

[Gorodnik and Weiss 2007] A. Gorodnik and B. Weiss, "Distribution of lattice orbits on homogeneous varieties", *Geom. Funct. Anal.* **17**:1 (2007), 58–115. MR Zbl

[Hardy and Wright 1979] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th ed., Oxford University Press, New York, 1979. MR Zbl

[Huntley and Katznelson 1993] J. Huntley and Y. R. Katznelson, "Density theorems for congruence groups in real rank 1", *Duke Math. J.* **71**:2 (1993), 463–473. MR Zbl

[Huxley 1986] M. N. Huxley, "Exceptional eigenvalues and congruence subgroups", pp. 341–349 in *The Selberg trace formula and related topics* (Brunswick, Maine, 1984), edited by D. A. Hejhal et al., Contemp. Math. **53**, Amer. Math. Soc., Providence, RI, 1986. MR Zbl

[Jana and Kamber 2022] S. Jana and A. Kamber, "On the local L^2-bound of the Eisenstein series", preprint, 2022. arXiv 2210.16291

[Knapp 1986] A. W. Knapp, *Representation theory of semisimple groups*, Princeton Mathematical Series **36**, Princeton University Press, 1986. MR Zbl

[Maucourant 2007] F. Maucourant, "Homogeneous asymptotic limits of Haar measures of semisimple linear groups and their lattices", *Duke Math. J.* **136**:2 (2007), 357–399. MR Zbl

[Oh 2002] H. Oh, "Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants", *Duke Math. J.* **113**:1 (2002), 133–192. MR Zbl

[Sarnak 2005] P. Sarnak, "Notes on the generalized Ramanujan conjectures", pp. 659–685 in *Harmonic analysis, the trace formula, and Shimura varieties*, edited by J. Arthur et al., Clay Math. Proc. **4**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[Sarnak 2015] P. Sarnak, letter to Stephen D. Miller and Naser Talebizadeh Sardari on optimal strong approximation by integral points on quadrics, 2015.

[Sarnak and Xue 1991] P. Sarnak and X. X. Xue, "Bounds for multiplicities of automorphic representations", *Duke Math. J.* **64**:1 (1991), 207–227. MR Zbl

[Trombi and Varadarajan 1972] P. C. Trombi and V. S. Varadarajan, "Asymptotic behaviour of eigen functions on a semisimple Lie group: the discrete spectrum", *Acta Math.* **129**:3-4 (1972), 237–280. MR Zbl

ak2356@dpmms.cam.ac.uk   *Centre for Mathematical Sciences, University of Cambridge, Cambridge, United Kingdom*

hagai.lavner@mail.huji.ac.il   *Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel*

Algebra & Number Theory

msp.org/ant

Algebra & Number Theory

Volume 17   No. 3   2023