OPTIMAL LIFTING FOR THE PROJECTIVE ACTION OF SL$_3(\mathbb{Z})$

AMITAY KAMBER AND HAGAI LAVNER

Abstract. 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 $\gamma \in$ SL$_3(\mathbb{Z})$, with coordinates bounded by $q^{1/3+\epsilon}$, whose projection to SL$_3(\mathbb{F}_q)$ sends $x$ to $y$. The exponent $1/3$ is optimal and the result is a high rank generalization of Sarnak’s optimal strong approximation theorem for SL$_2(\mathbb{Z})$.

1. Introduction

In his letter ([17]), Sarnak proved the following lifting theorem, which he called optimal strong approximation.

Theorem 1.1. Let $\Gamma = \text{SL}_2(\mathbb{Z})$, $q \in \mathbb{Z}_{>0}$, $G_q = \text{SL}_2(\mathbb{Z}/q\mathbb{Z})$ 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 $3/2$ in Theorem 1.1 is optimal, as the the size of $G_q$ is asymptotic to $q^3$, while the number of $\gamma \in \text{SL}_2(\mathbb{Z})$ satisfying $\|\gamma\|_{\infty} \leq T$ grows asymptotically like the Haar measure of the ball $B_T$ of radius $T$ in SL$_2(\mathbb{R})$ ([4, 14]), i.e., $\mu(B_T) \asymp T^2$.

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

We wish to discuss extensions of this theorem to SL$_3$, with a view towards general SL$_N$. If $\Gamma = \text{SL}_N(\mathbb{Z})$, 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 SL$_N(\mathbb{R})$, i.e., $\mu(B_T) \asymp T^{N^2-N}$ ([4, 14]), while the size of $G_q = \text{SL}_N(\mathbb{Z}/q\mathbb{Z})$ is $|G_q| \asymp q^{N^2-1}$. One is therefore led to the following:

Conjecture 1.2. Let $\Gamma = \text{SL}_N(\mathbb{Z})$, $q \in \mathbb{Z}_{>0}$, $G_q = \text{SL}_N(\mathbb{Z}/q\mathbb{Z})$ 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 non-principal congruence subgroup of SL$_3(\mathbb{Z})$. For a prime $q$, let $\mathbb{F}_q$ be the field with $q$ elements, let $P_q = \mathbb{P}^2(\mathbb{F}_q)$ 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 \( \left( \begin{array}{c} a \\ b \\ c \end{array} \right) \sim \left( \begin{array}{c} a \alpha \\ b \alpha \\ c \alpha \end{array} \right) \) for \( \alpha \in \mathbb{F}_q^\times \). The group \( \text{SL}_3(\mathbb{F}_q) \) acts naturally on \( P_q \), and by composing this action with \( \pi_q \) we have an action \( \Phi_q : \text{SL}_3(\mathbb{Z}) \to \text{Sym}(P_q) \).

**Theorem 1.3.** Let \( \Gamma = \text{SL}_3(\mathbb{Z}) \), and for a prime \( q \) let \( P_q = P^2(\mathbb{F}_q) \) and \( \Phi_q : \text{SL}_3(\mathbb{Z}) \to \text{Sym}(P_q) \) as above. Then for every \( \epsilon > 0 \), as \( q \to \infty \), there exists a set \( Y \subset P_q \) of size \(|Y| \geq (1 - o(1)) |P_q|\), such that for every \( x \in Y \), there exists a set \( Z_x \subset P_q \) of size \(|Z_x| \geq (1 - o(1)) |P_q|\), 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 1/3 is optimal, since the size of \( P_q \) is \(|P_q| \asymp q^2 \), while the number of elements \( \gamma \in \text{SL}_3(\mathbb{Z}) \) satisfying \( \|\gamma\|_\infty \leq T \) is \( \asymp T^6 \).

An alternative formulation of Theorem 1.3 is that for all but \( o(q^2) \) 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 \( x = 1 = \left( \begin{array}{c} 0 \\ 1 \end{array} \right) \in P_q \). Elements sending 1 to \( \left( \begin{array}{c} a \\ b \\ c \end{array} \right) \in P_q \) necessarily have the third column modulo \( q \) equivalent to \( \left( \begin{array}{c} a \\ b \\ c \end{array} \right) \) (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 \( x = 1 \) to almost all \( y \in P_q \). As a matter of fact, one may use the explicit property (T) of \( \text{SL}_3(\mathbb{R}) \) from [15] together with ideas from [7] 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 ([18]). To state it, we first define a different gauge of largeness on \( \text{SL}_3(\mathbb{Z}) \) by \( \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \). The number of \( \gamma \in \text{SL}_3(\mathbb{Z}) \) satisfying \( \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T \) grows asymptotically like \( T^2 \log(T) \) ([14]). Note that if \( \|\gamma\|_\infty \leq T \) then \( \|\gamma^{-1}\|_\infty \leq 2T^2 \). In particular, the ball of radius \( 2T \) relatively to \( \|\cdot\|_\infty \) contains the ball of radius \( T^{1/3} \) relatively to \( \|\cdot\|_\infty \), and their volume is asymptotically the same up to \( T^{\epsilon(1)} \). The counting result is as follows:

**Theorem 1.4.** Let \( \Gamma = \text{SL}_3(\mathbb{Z}) \), and for a prime \( q \) let \( P_q = P^2(\mathbb{F}_q) \) and \( \Phi_q : \text{SL}_3(\mathbb{Z}) \to \text{Sym}(P_q) \) 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

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

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

\[
\Gamma_0^\prime(q) = \{ \gamma \in \text{SL}_3(\mathbb{Z}) : \Phi_q(\gamma) (1) = 1 \} = \left\{ \begin{pmatrix} * & * & a \\ * & * & b \\ * & * & * \end{pmatrix} \in \text{SL}_3(\mathbb{Z}) : a = b = 0 \mod q \right\}
\]

is a non-principal congruence subgroup of \( \text{SL}_3(\mathbb{Z}) \). Theorem 1.3 says that Conjecture 1.2 holds “on average” for the non-principal congruence subgroup \( \Gamma_0^\prime(q) \).
Conjecturally, such “optimal lifting on average” should hold for every sequence of congruence subgroups of \( \Gamma = \text{SL}_N(\mathbb{Z}) \), i.e., subgroups of some \( \Gamma(q) \), \( q > 1 \) an integer. We provide a further example of this phenomena for the action of \( \text{SL}_3(\mathbb{Z}) \) 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 [8] for a more detailed discussion.

Theorem 1.1 follows from Selberg’s conjecture about the smallest non-trivial 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 \( \Gamma = \text{SL}_2(\mathbb{Z}) \). While Selberg’s conjecture is widely open, Sarnak proved Theorem 1.1 using density estimates on exceptional eigenvalues of the Laplacian, which are due to Huxley ([12]). Similar density results were proved by Sarnak and Xue using lattice point counting arguments in [18], but only for arithmetic quotients which are compact. The compactness assumption was removed in [11, 5] (and the results were moreover extended to some thin subgroups of \( \text{SL}_2(\mathbb{Z}) \)). As a matter of fact, in rank 1 the density property is equivalent to the lattice point counting property ([8]).

In higher rank, Conjecture 1.2 would similarly follow from a naive Ramanujan conjecture for \( \Gamma(q) \backslash \text{SL}_N(\mathbb{R}) \), \( \Gamma = \text{SL}_N(\mathbb{Z}) \), which says (false!) that the representation of \( \text{SL}_N(\mathbb{R}) \) on \( L^2(\Gamma(q) \backslash \text{SL}_N(\mathbb{R})) \) decomposes into a trivial representation and a tempered representation. The Burger-Li-Sarnak explanation of the failure of the naive Ramanujan conjecture ([3]) is closely related to the behavior of the point \( x_0 = 1 \in \mathbb{P}_q \). 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 \( \text{SL}_3(\mathbb{Z}) \) is not cocompact ([8]). Closely related density results were recently proven by Blomer, Buttcane and Maga for \( N = 3 \) in [2], and for general \( N \) by Blomer in [1], 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. However, the results of [2] and [1], concern cusp forms, and one has to deal with the presence of non-tempered Eisenstein representations and some other technical issues. Our counting approach is more elementary, and allows simpler generalizations, such as Theorem 5.1.

**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 \( \text{SL}_3 \). The main difference between our proof and the proof in [17] is that we avoid using spectral decomposition, which is far harder in \( \text{SL}_3 \).

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 pre-trace formula (in a disguised form), and bounds on Harish-Chandra’s \( \Xi \) function. This section is based on a general framework developed by the first author with Konstantin Golubev surrounding similar questions ([8]).

Finally, in Section 5 we prove Theorem 5.1 which is a variant of Theorem 1.3 for the action of \( \text{SL}_3(\mathbb{Z}) \) on flags of \( \mathbb{F}_q^3 \).
Acknowledgments. We are grateful to Amos Nevo and Elon Lindenstrauss for various discussions surrounding this project, and for Peter Sarnak for his continued encouragement. We also thank Amos Nevo for his careful reading of a previous version of this article and for pointing out some inaccuracies.

This work is part of the Ph.D. thesis of the both authors at the Hebrew University of Jerusalem. The first author is under the guidance of Prof. Alex Lubotzky, and is supported by the ERC grant 692854. The second author is under the guidance of Prof. Tamar Ziegler, and is supported by the ERC grant 682150.

2. Proof of Theorem 1.1

The basic input for the proof of Theorem 1.1 is the following counting result, proved in [5, Lemma 5.3] (it also appeared earlier, e.g., in [12]).

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

\[
\{ \gamma \in \text{SL}_2(\mathbb{Z}) : \gamma = I \mod q, \|\gamma\|_\infty \leq T \}
\]

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

**Proof.** Let \( \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z}) \) be in the set. It holds that \( \gamma - I \in qM_n(\mathbb{Z}) \), so \( \det(\gamma - I) = 0 \mod q^2 \), or explicitly

\[
(a - 1)(d - 1) - bc = 0 \mod q^2.
\]

Since \( ad - bc = 1 \), we have \( a + d = 2 \mod q^2 \). 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 \( ad - bc = 1 \), then \( bc = 1 - ad \neq 0 \), and by standard divisor bounds this gives \( \ll_t T^\epsilon \) options for \( (b,c) \). Otherwise, assuming \( q > 2 \), \( a = d = 1 \), and then either \( b = 0 \) or \( c = 0 \). 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

\[
\ll_t (T/q^2 + 1)(T/q + 1) T^\epsilon + T/q + 1 \ll T^\epsilon (T^2/q^3 + T/q + 1).
\]

\( \Box \)

Our proof of Theorem 1.1 proceeds with some spectral analysis of hyperbolic surfaces associated to \( \text{SL}_2(\mathbb{Z}) \) 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 \( d(x + iy, x' + iy') = \text{arccosh} \left( 1 + \frac{(x-x')^2 + (y-y')^2}{2yy'} \right) \) and a measure defined by \( \frac{dx dy}{yy'} \).

It also has a natural \( \text{SL}_2(\mathbb{R}) \) action by Möbius transformation, i.e., \( \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) z = \frac{az + b}{cz + d} \).

This action allows us to identify \( \mathcal{H} \) with \( G/K \), where \( G = \text{SL}_2(\mathbb{R}) \), and \( K = \text{SO}(2) \) 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 $\|g\|_{\mathcal{H}} = e^{d(\gamma, gT)/2}$. Explicitly, by the Cartan decomposition of $G$, $g$ can be written as

$$g = k_1 \left( e^{r/2} \ e^{-r/2} \right) k_2,$$

with $k_1, k_2 \in K = \text{SO}(2)$, and $r \in \mathbb{R}_{\geq 0}$ unique. Then $\|g\|_{\mathcal{H}} = e^{r/2}$. As the $L^2$-norm of the coordinates of $\gamma$ 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 sub-multiplicativity $\|g_1 g_2\|_{\mathcal{H}} \leq \|g_1\|_{\mathcal{H}} \|g_2\|_{\mathcal{H}}$. The sub-multiplicativity 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 $\{g \in G: \|g\|_{\mathcal{H}} \leq T\}$, i.e.,

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

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 non-negative function with integral 1.

We also define $\psi_T \in L^1(K \backslash G/K)$ as the function $\psi_T(g) = \frac{1}{T} \begin{cases} \|g\|_{\mathcal{H}}^{-1} & \|g\|_{\mathcal{H}} \leq T, \\ 0 & \|g\|_{\mathcal{H}} > T. \end{cases}$

There is a convolution of $f \in L^\infty(G/K) \approx L^\infty(\mathcal{H})$ and $\chi \in L^1(K \backslash G/K)$, which we usually think as an action of $\chi$ 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 [18, Lemma 2.1] or [5, 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 $\text{SL}_3(\mathbb{R})$, which also works for $\text{SL}_2(\mathbb{R})$, but adds a factor that is logarithmic in $T$. 


As in the statement Theorem 1.1, let \( q \in \mathbb{Z}_{>0} \), \( \Gamma = \text{SL}_2(\mathbb{Z}) \), \( G_q = \text{SL}_2(\mathbb{Z}/q\mathbb{Z}) \) 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 \( \chi_T \).

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,
\[ \langle f, b_{T,x_0} \rangle = f \ast \chi_T(x_0). \]

**Proof.** By unfolding,
\[
\langle f, b_{T,x_0} \rangle = \int_{x \in \Gamma(q) \backslash \mathcal{H}} f(x) \sum_{\gamma \in \Gamma(q)} \chi_T(x_0^{-1}\gamma x) \, dx \\
= \int_{x \in \Gamma(q) \backslash \mathcal{H}} \sum_{\gamma \in \Gamma(q)} f(\gamma x) \chi_T(x_0^{-1}\gamma x) \, dx \\
= \int_{x \in \mathcal{H}} f(x) \chi_T(x_0^{-1}x) \, dx \\
= \int_{x \in \mathcal{H}} f(x) \chi_T(x^{-1}x_0) \, dx \\
= f \ast \chi_T(x_0).
\]

Notice that we used the fact that \( \chi_T(g) = \chi_T(g^{-1}) \), which is a simplification that will not occur in \( \text{SL}_3 \). \( \square \)

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

**Lemma 2.4.** It holds that
\[ \|b_{T,e}\|_2 \ll \epsilon \, T^\epsilon \left( \frac{1}{q^3} + \frac{1}{T^2} \right). \]

In particular, for \( T = q^{3/2} \),
\[ \|b_{T,e}\|_2 \ll \epsilon \, \frac{T^\epsilon}{q^3}. \]
OPTIMAL LIFTING FOR THE PROJECTIVE ACTION OF $\text{SL}_3(\mathbb{Z})$ 

**Proof.** By Lemma 2.3, and Lemma 2.2,

$$\|b_{T,e}\|_2^2 = b_{T,e} \ast \chi_T(e)$$

$$= \sum_{\gamma \in \Gamma(q)} (\chi_T \ast \chi_T)(\gamma)$$

$$\ll \sum_{\gamma \in \Gamma(q)} \psi_{T^2}(\gamma)$$

$$= \frac{1}{T^2} \sum_{\gamma \in \Gamma(q) : \|\gamma\|_{\mathcal{H}} \leq T^2} \|\gamma\|_{\mathcal{H}}^{-1}.$$

We next apply discrete partial summation ([9, Theorem 421]) which says that for $g : \Gamma(q) \rightarrow [1, \infty], f : [1, \infty] \rightarrow \mathbb{R}$ nice enough it holds that

$$(2.1) \sum_{\gamma : 1 \leq g(\gamma) \leq Y} f(g(\gamma)) = f(Y) \left| \{ \gamma : 1 \leq g(\gamma) \leq Y \} \right| - \int_1^Y \left| \{ \gamma : g(\gamma) \leq S \} \right| \frac{df}{dS}(S) dS.$$

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

$$\sum_{\gamma \in \Gamma(q) : \|\gamma\|_{\mathcal{H}} \leq T^2} \|\gamma\|_{\mathcal{H}}^{-1}$$

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

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

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

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

The first inequality follows from Theorem 2.1.

Let $\pi \in L^2(X_q)$ be the constant probability function on $X_q$ (recall that the space has finite volume). Denote by $L^2_0(X_q)$ the set of functions of integral 0, or alternatively the set of functions orthogonal to $\pi$. 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 $f \in L^2_0(X_q)$ and $T > 0$ holds that $\|f \ast \chi_T\|_2 \ll T^{-\tau} \|f\|_2$.

The important part of the theorem is the independence of $\tau$ from $q$. 

\[\square\]
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 (see, e.g., [6, Section 4] for a generalized statement). There are various results improving the value of $\tau$ in Selberg’s theorem (see [16]), but those improvement are inconsequential to our theorem.

From Selberg’s theorem we deduce:

**Lemma 2.6.** For $T = q^{3/2}$,

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

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

Therefore,

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

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. □

The last lemma implies that the function $b_{T,e} \ast \chi_T$ is very close to the constant probability function $\pi$. Let us show how this implies Theorem 1.1.

We have a map $\iota: G_q \cong \Gamma(q) \backslash G \rightarrow X_q \cong \Gamma(q) \backslash G/K$, 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 $\gamma \in SL_2(\mathbb{Z})$ are disjoint. We also notice that $\iota$ has fibers of bounded size, specifically $|SL_2(\mathbb{Z}) \cap K| = 4$. In addition, there is a uniform (in $q$) upper bound on the norm $\|b_{T_0,\iota(y)}\|_2$.

**Lemma 2.7.** Assume that $b_{T_0,\iota(y)} \ast b_{T_0,\iota(y)} > 0$, then there is $\gamma \in \Gamma$ such that $\pi_q(\gamma) = y$, and $\|\gamma\|_H \leq T_0 T^{1+\eta}$.

**Proof.** By Lemma 2.3, the condition implies that

$$
(b_{T,e} \ast \chi_T \ast \pi_{T_0})(\iota(y)) > 0.
$$

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

By the definition of convolution, there are $g_1', g_2, g_3 \in G$, such that $g_1' \in supp(b_{T,e})$, $g_2 \in supp(\chi_T)$, $g_3 \in supp(\chi_{T_0})$, and such that $g_1' g_2 g_3 = \gamma_y$. Looking at the definition of $b_{T,e}$ and $g_1'$, there are $g_1 \in supp(\chi_T)$, $\gamma \in \Gamma(q)$ 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 $e^{-1} \gamma g_1' g_2 g_3 = \gamma_y$.

Write $g = g_1 g_2 g_3$. By the above, $\|g\|_H \leq \|g_1\|_H \|g_2\|_H \|g_3\|_H \leq T_0 T^{1+\eta}$. In addition, $e g = \gamma g_y$, so that $g \in \Gamma(\gamma_g)$. Therefore $g \in \Gamma$ and $\pi_q(g) = y$, as needed.
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 $\|\gamma_y\| \leq T_0 T_1^{1+\eta}$ and $\pi_q(\gamma_y) = y$. It suffices to prove that for a fixed $\eta > 0$ it holds that $|Z| = o(q^3)$.

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

$$\langle b_{T,e} * \chi y, b_{T_0,\epsilon}(y) \rangle = 0.$$

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

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

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-Schwartz,

$$|\langle b_{T,e} * \chi y - \pi, B \rangle| \ll \|B\|_2 \|b_{T,e} * \chi y - \pi\|_2 \ll \epsilon_{\eta} \sqrt{|Z| q^{-2-3\eta+\epsilon}}.$$

Combining the two estimates and taking $\epsilon$ small enough gives

$$|Z| \ll_{\epsilon} q^{3-2\eta+2\epsilon} = o(q^3),$$

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 $T \leq C q^2$, we have

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

If $\gamma \mod q$ has no eigenspace of dimension 2, then it has at most 3 eigenvectors in $P^2(\mathbb{F}_q)$. Call such a $\gamma$ good mod $q$ and otherwise call it bad mod $q$. Therefore for $T \leq q^2$,

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

We therefore need to bound the number of bad $\gamma$-s. Notice that bad elements do exist and may have a lot of fixed points: e.g., the element $I \in \text{SL}_3(\mathbb{Z})$ is bad mod $q$ and $\Phi_q(I)$ fixes all of $P^2(\mathbb{F}_q)$.

Assuming that we choose $C < 1/4$, it will hold that either $\|\gamma\|_\infty < q/2$ or $\|\gamma^{-1}\|_\infty < q/2$. Therefore if $\gamma \neq I$ then $\gamma \mod q \neq I_{\text{SL}_3(\mathbb{F}_q)}$, and thus $\Phi_q(\gamma)$ fixes at most $q + 1$ elements in $P^2(\mathbb{F}_q)$. It thus suffices to prove that for some $C > 0$, and $T \leq C q^2$,

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

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

$$|\{\gamma \in SL_3(Z) : \|\gamma\|_{\infty} \leq S, \|\gamma^{-1}\|_{\infty} \leq R, \gamma \text{ bad mod } q\}| \ll Cq^{1+\epsilon}.$$

It will be useful to understand the behavior of bad $\gamma$. Let $\alpha \in F_q \setminus \{0\}$ 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 $(\gamma - \alpha I)(\gamma - \alpha^{-2} I) = 0 \mod q$, or,

$$\gamma + \alpha^{-1} \gamma^{-1} = \alpha + \alpha^{-2} \mod q. \tag{3.1}$$

By considering the trace of $\gamma$ and $\gamma^{-1}$ we have that

$$\text{tr } \gamma = \alpha + 2\alpha^{-2} \mod q, \quad \text{tr } \gamma^{-1} = \alpha^{-1} + 2\alpha^{2} \mod q. \tag{3.2}$$

Finally, identify $\alpha$ with some lift of it in $Z$. Then $\gamma - \alpha I \mod q$ is of rank 1, which means that $\det(\gamma - \alpha I) = 0 \mod q^2$. Since $\det \gamma = 1$, it holds that $\det(\gamma - x I) = 1 - \text{tr } \gamma^{-1} x + \text{tr } \gamma x^2 - x^3$, and we get

$$\alpha^2 \text{tr } \gamma - \alpha \text{tr } \gamma^{-1} = \alpha^3 - 1 \mod q^2. \tag{3.3}$$

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

There are $\leq (2S+1)^3$ options for choosing the diagonal $a_{11}, a_{22}, a_{33}$ of $\gamma$, and once we know them, we know $\text{tr } \gamma$. By Equation (3.2) $\alpha$ is a root of a known third degree polynomial, so there are at most 3 options for $\alpha$. By Equation (3.3) we know $\text{tr } \gamma^{-1} \mod q^2$. Since $R \leq RS \leq Cq^2 < q^2/4$, we may assume that $|\text{tr } \gamma^{-1}| < q^2/2$, so now we know $\text{tr } \gamma^{-1}$.

By Equation (3.1) we now know the diagonal $b_{11}, b_{22}, b_{33} \mod q$ of $\gamma^{-1} \mod q$. 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 $\text{tr } \gamma^{-1}$.

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 non-exceptional 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 non-exceptional case, the right hand side is not 0. By the divisor bound there are at most $\ll q^3$ options for $a_{12}, a_{21}$. Similarly, all the other entries $a_{13}, a_{31}, a_{23}, a_{32}$ have at most $\ll q^3$ options.

In total, we counted $\ll q^3 S^3(R/q + 1)^2$ bad $\gamma$-s in the non-exceptional 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 \( S \leq R, RS \leq Cq^2 \).

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

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 \( RS \leq Cq^2 \), this reduces to showing that \( S \ll q \), which is obvious since \( S^2 \leq RS \leq Cq^2 \).

**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 \( 2 \times 2 \) submatrix of \( \gamma \) 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 \( \alpha \) 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:
\[
(a_{31} - \alpha)(a_{22} - \alpha) = a_{21}a_{32} = 0 \mod q
\]
\[
a_{31}a_{23} = a_{21}(a_{33} - \alpha) = 0 \mod q.
\]

We now divide into two cases according to whether \( a_{31} = 0 \) or not:

1. **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 \( A = \begin{pmatrix} a_{22} & a_{23} \\
a_{32} & a_{33} \end{pmatrix} \). 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 \( \pm 1 \) (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 \( a_{22} = \alpha \).

(a) **Subcase 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 \( \ll S^e \) options for \( a_{23}, a_{32} \) and both are non-zero. 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.

(b) **Subcase 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.

(2) **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 $\gamma$ 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 $\gamma$.

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 $\text{SL}_3(\mathbb{R})$ by $\text{SL}_3(\mathbb{Z})$ from the left, we apply a transpose to the question as stated in Theorem 1.3.

Let 

$$\Gamma_0(q) = \left\{ \left( \begin{array}{ccc} * & * & * \\ * & * & * \\ a & b & * \end{array} \right) : a = b = 0 \mod q \right\}.$$ 

We have a right action of $\Gamma = \text{SL}_3(\mathbb{Z})$ on $\Gamma_0(q)$. We let $P^\text{tr}_q = \Gamma_0(q) \setminus \Gamma$ (it is obviously isomorphic to $P_q$ as a set with a $\Gamma$ 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 $Y \subset \Gamma_0(q) \setminus \Gamma = P^\text{tr}_q$ of size $|Y| \geq (1 - o_{\epsilon}(1)) |P^\text{tr}_q|$, such that for every $x_0 \in Y$, there exists a set $Z_{x_0} \subset P^\text{tr}_q$ of size $|Z_{x_0}| \geq (1 - o_{\epsilon}(1)) |P^\text{tr}_q|$, 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 $K = \text{SO}(3)$ be the maximal compact subgroup of $G = \text{SL}_3(\mathbb{R})$. 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 \( k_1, k_2 \in \text{SO}(3) \), and unique \( a_1, a_2, a_3 \in \mathbb{R}_{>0} \), satisfying \( a_1 \geq a_2 \geq a_3 > 0 \) and \( a_1a_2a_3 = 1 \).

Define \( \|g\|_K = a_1 \). Since \( K = \text{SO}(3) \) 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 \( \|\cdot\|_H \) in the SL\(_2\) case. Let us note some of its properties. There is a constant \( C > 0 \) such that \( \|g_1g_2\|_K \leq C \|g_1\|_K \|g_2\|_K \) (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 \( \|g\|_\delta = a_1a_3^{-1} \), where \( a_1, a_3 \) are as in the Cartan decomposition. It holds that there is a constant \( C > 0 \) such that

\[
\text{(4.1)} \quad C^{-1} \|g\|_\infty \|g^{-1}\|_\infty \leq \|g\|_\delta \leq C \|g\|_\infty \|g^{-1}\|_\infty.
\]

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 \( \|g_1g_2\|_\delta \leq C \|g_1\|_\delta \|g_2\|_\delta \).

The relation between the two sizes is that \( \|g\|_\delta \leq \|g\|_K^3 \), which follows from the fact that in the Cartan decomposition \( a_3^{-1} = a_1a_2 \leq a_1^3 \), so \( a_1a_3^{-1} \leq a_1^3 \).

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 \( \mu \) of \( G \) ([13, Proposition 5.28]), which holds up to multiplication by a scalar:

\[
\int \int \int f(g) d\mu = \int K K K \alpha_+ f(k \exp(a)k') S(a) dk dk' da,
\]

where

\[
\alpha_+ = \left\{ \alpha = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \in M_3(\mathbb{R}) : \alpha_1 \geq \alpha_2 \geq \alpha_3, \alpha_1 + \alpha_2 + \alpha_3 = 0 \right\},
\]

and

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

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

\[
\mu (\{ g \in G : \|g\|_K \leq T \}) \asymp T^6,
\]

and

\[
\mu (\{ g \in G : \|g\|_\delta \leq T \}) \asymp \log(T)T^2.
\]

See also [14] for more accurate similar statements.
Let $\chi_T, \chi_{T,\delta} \in L^1(K \backslash G/K)$ be

$$
\chi_T(g) = \frac{1}{\mu(\{g \in G : \|g\|_K \leq T\})} \begin{cases} 1 & \|g\|_K \leq T, \\ 0 & \text{else} \end{cases},
$$

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

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

Notice that for every $g \in G$,

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

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

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

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

Now we have the following version of Lemma 2.2:

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

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

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

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

**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

$$
\delta \begin{pmatrix} a_1 & * & * \\ 0 & a_2 & * \\ 0 & 0 & a_3 \end{pmatrix} = a_1^2 a_3^{-2}.
$$

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

There are standard bounds on $\Xi$, given by (see, e.g., [19, 2.1])

$$
(4.2) \quad \|g\|_\delta^{-1} \leq \Xi(g) \ll (\log\|g\| + 1)^{C_0} \|g\|_\delta^{-1}
$$

for some $C_0 > 0$. Using these upper bounds, we find that for some $C_2 > 0$,
\[
\int_G \chi_{T,\delta} \Xi (g) \, dg = \frac{1}{\mu (\{ g \in G : \| g \|_{\delta} \leq T \})} \int_{g : \| g \|_{\delta} \leq T} \Xi (g) \, dg \ll (\log (T) + 1)^{C_2} T^{-1}.
\]

Harish-Chandra’s function $\Xi$ arises as follows (see, e.g., [6, Section 3]). Let $(\pi, V)$ 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
\[
\pi (f) v = \int_G f (g) \pi (g) vdg = \left( \int_G f (g) \Xi (g) \, dg \right) v.
\]

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 $C^{-1} \| g \|_{\delta} \leq \| g' \|_{\delta} \leq C \| g \|_{\delta}$ 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 Equation (4.2).

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

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 $L^2_0 (X_q)$ the functions in $L^2 (X_q)$ of integral 0. Since $\chi_{T}$ is bi-$K$-invariant and sufficiently nice, the function $\chi_{T}$ 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 $\chi_{T}$ satisfies for every $f \in L^2_0 (X_q)$,
\[
\| f * \chi_{T} \|_2 \ll T^{-\tau} \| f \|_2.
\]

The theorem follows from explicit versions of property (T), or explicit versions of the mean ergodic theorem (e.g., [6, Section 4]) which are actually true for all lattices in $G = \text{SL}_3(\mathbb{R})$ 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 $\iota(\Gamma_0(q)x_0) = \Gamma_0x_0K \in X_q$. By a slight abuse of notation we write $\iota(\Gamma_0(q)x_0) = \iota(x_0)$.

The map $\iota$ 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 $T' \leq Cq^2$, \[ |\{ \gamma \in \Gamma : \|\gamma\|_\delta \leq T', x_0 \gamma = x_0 \}| \ll \epsilon_0 q^\alpha T'. \]

Then there exists $C' > 0$ depending only on $C$ such that for $T = C'q^{1/3}$ it holds that for every $\epsilon > 0$, \[ \|b_{T,\iota(x_0)}\|_2 \ll \epsilon_0 \epsilon q^{-1+\alpha+\epsilon}. \]

**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 $\Gamma$). Therefore we may rewrite the assumption in the following manner: For every $T' \leq Cq^2$,

$$\|b_{T,\iota(x_0)}\|_2 \ll \epsilon_0 q^\alpha T'.$$

(4.3)

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

Write using Lemma 4.4,

$$\|b_{T,\iota(x_0)}\|_2^2 = \langle b_{T,\iota(x_0)}, b_{T,\iota(x_0)} \rangle = b_{T,\iota(x_0)} * \chi_T^*(\iota(x_0)) = \sum_{\gamma \in \Gamma_0(q)} (\chi_T * \chi_T^*) (x_0^{-1} \gamma x_0) \ll T^* \psi_{C_1 T_0} (x_0^{-1} \gamma x_0),$$

where in the last inequality we used Lemma 4.2.

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

$$\sum_{\gamma \in \Gamma_0} \psi_{C_1 T_0} (x_0^{-1} \gamma x_0) = T^{-6} \sum_{\gamma \in \Gamma_0(q) : \|x_0^{-1} \gamma x_0\|_\delta \leq C_1 T_0} \|x_0^{-1} \gamma x_0\|_\delta^{-1} \ll q^{-2} \sum_{\gamma \in \Gamma_0(q) : \|x_0^{-1} \gamma x_0\|_\delta \leq C_2 q^2} \|x_0^{-1} \gamma x_0\|_\delta^{-1} \ll q^{-2+\epsilon_0+\epsilon},$$

$$!$$

$$\ll \epsilon q^{-2+\epsilon_0+\epsilon},$$
where $C_2 = C_1 C^6$.

So it suffices to show that
\[
\sum_{\gamma \in \Gamma_0(q) : \|x_0^{-1} \gamma x_0\|_\delta \leq C_2 q^2} \|x_0^{-1} \gamma x_0\|_\delta^{-1} \ll_{\epsilon, \epsilon_0} q^{\epsilon + \epsilon_0}.
\]

We now apply Equation (2.1) (discrete partial summation), with $g(\gamma) = \|\gamma\|$, $f(x) = x^{-1}$ and $Y = C_2 q^2$ we have
\[
\sum_{\gamma \in \Gamma_0(q) : \|x_0^{-1} \gamma x_0\|_\delta \leq C_2 q^2} \|x_0^{-1} \gamma x_0\|_\delta^{-1} \ll \{ \gamma : \|x_0^{-1} \gamma x_0\|_\delta \leq C_2 q^2 \} |q^{-2} + \int_1^C \{ \gamma : \|x_0^{-1} \gamma x_0\|_\delta \leq S \} |S^{-2} dS.
\]

Choosing $C'$ small enough so that $C_2 = C_1 C^6 \leq C$ and applying Equation (4.3) we have that
\[
\ll_{\epsilon, \epsilon_0} q^{\epsilon + \epsilon_0} + q^{\epsilon + \epsilon_0} \int_1^C S^{-1} dS \ll_{\epsilon} q^{2 \epsilon + \epsilon_0},
\]
as needed.

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 $x_0 \in \Gamma_0(q) \setminus \Gamma$ 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 $Y \subset \Gamma_0(q) \setminus \Gamma = P^\tau_q$ of size $|Y| \geq (1 - o_{\epsilon_0}(1)) |\Gamma_0(q) \setminus \Gamma|$, such that for every $\Gamma_0 x_0 \in Y$, it holds for $T = C q^{1/3}$ that
\[
\|b_{T,\gamma(x_0)} \chi_{T^\gamma} - \pi\|_2 \ll_{\epsilon_0} q^{-1 - \eta \tau + \epsilon_0}.
\]

**Proof.** By Theorem 1.4 and Equation (4.1) it holds that for some $C > 0$, for all $T \leq C q^2$ and $\epsilon > 0$
\[
\sum_{x_0 \in \Gamma_0(q) \setminus \Gamma} | \{ \gamma \in \Gamma : \|\gamma\|_\delta \leq T, x_0 \gamma = x_0 \} | \ll_\epsilon q^{2 + \epsilon} T.
\]

Since $|\Gamma_0(q) \setminus \Gamma| = (1 + o(1)) q^2$, we may choose a subset $Y \subset \Gamma_0(q) \setminus \Gamma$ of size
\[
|Y| \geq (1 - o_{\epsilon_0}(1)) |\Gamma_0(q) \setminus \Gamma|,
\]
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
\[
\left\| b_{T,\iota(x_0)} \right\|_2 \ll \epsilon_0 q^{-1+\epsilon_0}.
\]

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

We may now finish the proof of Theorem 4.1, similar to the \(SL_2\) case. First,

**Lemma 4.7.** There is \(C' > 0\) such that for \(x_0, y \in \Gamma_0(q) \setminus \Gamma\), if \(\langle b_{T,\iota(x_0)} \ast \chi_T^q, b_{T,\iota(y)} \rangle > 0\), 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
\[
b_{T,\iota(x_0)} \ast \chi_T^q \ast \chi_{T_0}^\iota(y) > 0.
\]

Denote by \(\tilde{x}_0, \tilde{y}\) as some lifts of \(x_0, y\) to \(\Gamma\). We get \(g_1, g_2, g_3 \in G\), \(\gamma \in \Gamma_0(q)\) such that \(\gamma^{-1}\tilde{x}_0g_1g_2g_3 = \tilde{y}\), with \(g_1 \in \text{supp}(\chi_T^q)\), \(g_2 \in \text{supp}(\chi_T^q)\), \(g_3 \in \text{supp}\left(\chi_{T_0}^\iota\right)\). Writing \(g = g_1g_2g_3\), we have that
\[
\|g\|_K \ll \|g_1\|_K \|g_2\|_K \|g_3\|_K \ll T^{1+\eta}.
\]

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

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

Choose \(T = Cq^{1/3}\), and \(\eta\) small enough so that \(C'T^{1+\eta} \leq q^{1/3+\epsilon}\), 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,
\[
\langle b_{T,x_0} \ast \chi_T^q - \pi, B \rangle = \frac{\left| \tilde{Z}_{x_0} \right|}{\text{Vol}(X_q)} \gg \frac{\left| \tilde{Z}_{x_0} \right|}{q^2}.
\]

On the other hand, by the choice of \(x_0\) and Lemma 4.6,
\[
\langle b_{T,x_0} \ast \chi_T^q - \pi, B \rangle \ll \|B\|_2 \left\| b_{T,x_0} \ast \chi_T^q - \pi \right\|_2 \\
\ll \epsilon_{\epsilon_0} \sqrt{\left| \tilde{Z}_{x_0} \right| q^{-1-\eta\tau+\epsilon_0}}.
\]

By combining the two estimates and choosing \(\epsilon_{\epsilon_0}\) small enough, we get the desired result
\[
\left| \tilde{Z}_{x_0} \right| \ll \epsilon_{\epsilon_0} q^{2-2\eta\tau-2\epsilon_0} = o\left(q^2\right).
\]

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 \(\mathbb{F}^3_q\), i.e.,
\[
B_q = \{(V_1, V_2) : 0 < V_1 < V_2 < \mathbb{F}^3_q\},
\]
i.e., \(V_1 \subset V_2\) are subspaces of \(\mathbb{F}^3_q\), such that \(\dim V_1 = 1, \dim V_2 = 2\).
There is a natural action action $\Phi_q : SL_3(\mathbb{Z}) \to Sym(B_q)$. It gives rise to a non-principal congruence subgroup

$$\Gamma'_2(q) = \left\{ \left( \begin{array}{ccc} a & b \\ c & d & e \\ f & g & h \end{array} \right) \in SL_3(\mathbb{Z}) : a = b = c = 0 \mod q \right\}.$$

Concretely,

$$\Gamma'_2(q) = \{ \gamma \in SL_3(\mathbb{Z}) : \Phi_q(\gamma)(1) = 1 \},$$

where

$$1 = (\text{span}\{\left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right)\}, \text{span}\{\left( \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right)\}).$$

The result reads as follows:

**Theorem 5.1.** Let $\Gamma = SL_3(\mathbb{Z})$, and for a prime $q$ let $B_q$ and $\Phi_q : SL_3(\mathbb{Z}) \to Sym(B_q)$ as above. Then for every $\epsilon > 0$, as $q \to \infty$, there exists a set $Y \subset B_q$ of size $|Y| \geq (1 - o(1))|B_q|$, such that for every $x \in Y$, there exists a set $Z_x \subset B_q$ of size $|Z_x| \geq (1 - o(1))|B_q|$, 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 $1/2$ is optimal, since the size of $B_q$ is $|B_q| \asymp q^3$, while the number of elements $\gamma \in SL_3(\mathbb{Z})$ 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

$$|\{(\gamma, x) \in SL_3(\mathbb{Z}) \times B_q : \|\gamma\|_\infty \|\gamma^{-1}\|_\infty \leq T, \Phi_q(\gamma)(x) = x\}| \ll \epsilon q^{3+\epsilon}T.$$

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 $RS \leq Cq^3$:

$$|\{(\gamma, x) \in SL_3(\mathbb{Z}) \times B_q : \|\gamma\|_\infty \leq S, \|\gamma^{-1}\|_\infty \leq R, \Phi_q(\gamma)(x) = x\}| \ll \epsilon q^{3+\epsilon}RS$$

We divide into several cases according to the Jordan form of $\gamma \mod q$. We leave the verification of the following to the reader:

(1) If $\gamma = I \mod q$ then there are $|B_q| \asymp q^3$ different $x \in B_q$ such that $\Phi_q(\gamma)x = x$. 


(2) If $\gamma \mod q$ has an eigenspace of dimension 2, but is not the identity, i.e., the Jordan form of $\gamma \mod q$ is
\[
\begin{pmatrix}
1 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\text{ or }
\begin{pmatrix}
\alpha & 0 & 0 \\
0 & \alpha & 0 \\
0 & 0 & \alpha^{-2}
\end{pmatrix},
\]
then there are $\asymp q$ different $x \in B_q$ such that $\Phi_q(\gamma)x = x$. As in Section 3, we call such $\gamma$ bad mod $q$.

(3) For all other cases, there are $O(1)$ different $x \in B_q$ such that $\Phi_q(\gamma)x = x$.

Theorem 5.2 will therefore follow from the following two lemmas:

**Lemma 5.3.** There exists $C > 0$ such that for every $S \leq R$ and $RS \leq Cq^3$:
\[
\left| \left\{ \gamma \in \text{SL}_3(\mathbb{Z}) : \|\gamma\|_\infty \leq S, \|\gamma^{-1}\|_\infty \leq R, \gamma = I \mod q \right\} \right| \ll q^2 R S.
\]

**Lemma 5.4.** There exists $C > 0$ such that for every $S \leq R$ and $RS \leq Cq^3$:
\[
\left| \left\{ \gamma \in \text{SL}_3(\mathbb{Z}) : \|\gamma\|_\infty \leq S, \|\gamma^{-1}\|_\infty \leq R, \gamma \text{ bad mod } q \right\} \right| \ll q^{2+\epsilon} R S.
\]

**Proof of Lemma 5.3.** We will actually work a little harder than necessary to show that the count is at most $\ll q^2(S/q + 1)^2$, which is tight up to $q^\epsilon$.

Using ideas of [10, Chapter 6], since for all $x$ it holds that $\gamma - (1 + xq)I = 0 \mod q$, it follows that
\[
\det(\gamma - (1 + xq)I) = 1 - (1 + xq)\tr \gamma^{-1} + (1 + xq)^2 \tr \gamma - (1 + xq)^3 = 0 \mod q^3.
\]

By equating coefficients, one obtains:
\[
\tr \gamma = \tr \gamma^{-1} = 3 \mod q^2
\]
\[
\tr \gamma = \tr \gamma^{-1} \mod q^3
\]
In addition, $(\gamma - I)^2 = 0 \mod q^2$, so $\gamma^{-1} = 2I - \gamma \mod q^2$. We may assume that $\|\gamma^{-1}\|_\infty < q^2/2$, thus $\gamma^{-1} = 2I - \gamma$ and $(\gamma - I)^2 = 0$.

Let us use the above information in order to parametrize the diagonals of $\gamma, \gamma^{-1}$. As in Section 3, we denote the entries of $\gamma$ by $a_{ij}$ and the entries of $\gamma^{-1}$ by $b_{ij}$.

Write $a_{11} = 1 + qa, a_{22} = 1 + qb, a_{33} = 1 - q(a + b), b_{11} = 1 - qa, b_{22} = 1 - qb, b_{33} = 1 + q(a + b)$.

Then:
\[
\gamma = \begin{pmatrix}
1 + qa & * & * \\
* & 1 + qb & * \\
* & * & 1 - (a + b)q
\end{pmatrix}
\]
\[
\gamma^{-1} = \begin{pmatrix}
1 - qa & * & * \\
* & 1 - qb & * \\
* & * & 1 + (a + b)q
\end{pmatrix}
\]
The non-exceptional case happens if \(a, b, a+b\) are all non-zero. Then we shall recover all of \(\gamma\), up to \(q^6\) options as follows. It holds that:

\[
a_{12}a_{21} = a_{11}a_{22} - b_{33} = 1 + (a + b)q + q^2ab - (1 + (a + b)q) = q^2ab \neq 0
\]

so using the divisor bound, we recover \(a_{12}, a_{21}\) up to \(q^6\) options, and similarly for the other entries.

If all three \(a, b, a+b\) vanish, then among any pair \(1 \leq i < j \leq 3\), \(a_{ij}a_{ji} = 0\). There are 3 such pairs, so we recover \(\gamma\) up to \(\ll (S/q+1)^3\) options. As a matter of fact, one can improve this estimate: Either up to permutations \(\gamma\) is upper triangular, or it has a single non-diagonal non-zero contribution to the determinant. The diagonal contributes 1 to the determinat, so there cannot be a non-diagonal contribution. Therefore, we may assume that \(\gamma\) is upper triangular. We know that \((\gamma - I)^2 = 0\), so \(a_{12}a_{23} = 0\). Thus there actually only \(\ll (S/q+1)^2\) options for \(\gamma\).

For the remaining case, we may thus assume \(a + b = 0\), so \(a = -b \neq 0\). Hence, \(\gamma\) and \(\gamma^{-1}\) are of the form:

\[
\gamma = \begin{pmatrix}
1 + qa & cq & * \\
dq & 1 - qa & * \\
* & * & 1
\end{pmatrix}
\]

\[
\gamma^{-1} = \begin{pmatrix}
1 - qa & -cq & * \\
-dq & 1 + qa & * \\
* & * & 1
\end{pmatrix}
\]

There are \(\ll (S/q+1)\) options for \(a\).

It holds that \(cdq^2 = a_{12}a_{21} = a_{11}a_{22} - b_{33} = 1 - q^2a^2 - 1\), hence \(cd = -a^2\). Since \(a \neq 0\) we obtain \(c, d\) from \(a\) up to \(q^4\).

Next, let us note that: \(a_{31}a_{13} = a_{11}a_{33} - b_{22} = 0\), and similarly \(a_{23}a_{32} = 0\).

If both \(a_{13}, a_{31} = 0\), there are \(\ll (S/q+1)\) options for \(a_{23}, a_{32}\).

Otherwise, we may assume that \(a_{13} \neq 0\), and then \(a_{31} = 0\). Multiply the first row of \(\gamma\) with the second column of \(\gamma^{-1}\) to obtain

\[
a_{11}b_{12} + a_{12}b_{22} + a_{31}b_{32} = 0 = (1 + qa)(-cq) + cq(1 + qa) + a_{13}b_{32},
\]

that is \(a_{13}b_{32} = 0\), so \(b_{32} = -a_{32} = 0\).

Multiply the first row of \(\gamma\) with the third column of \(\gamma^{-1}\) to show that \(a_{13}\) determines \(a_{23}\).

All in all, this brings us to \(\ll q^6(S/q+1)^2\) options for \(\gamma\), as needed. \(\square\)

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

**Lemma 5.5.** The number of solutions for Equations (3.2), (3.3) in \(\text{tr} \gamma, \text{tr} \gamma^{-1} \in \mathbb{Z}, \alpha \in \mathbb{F}_q, |\text{tr} \gamma| \leq S, |\text{tr} \gamma^{-1}| \leq R \) 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 Equation (3.2), \(x_1 - x_2 = y_1 - y_2 = 0 \mod q\). Denote \(z = (x_1 - y_1)/q, w = (x_2 - y_2)/q\). Notice that \(|z| \leq 2S/q, |w| \leq 2R/q\). By
Equation (3.3) \((z, w, \alpha)\) is a solution to \(\alpha qz - qw = 0 \mod q^2\), or
\[
(5.1)\quad \alpha z - w = 0 \mod q.
\]

Therefore, \(A\) solutions with the same \(\alpha \in \mathbb{F}_q\) for (3.2),(3.3) give \(A\) solutions to (5.1) with the same \(\alpha \in \mathbb{F}_q\). So the total number of solutions is bounded by the number of solutions of Equation (5.1) 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 \(\alpha\) uniquely, unless \(z = w = 0\).

\[\square\]

**Proof of Lemma 5.4.** By Lemma 5.5 there are \(\ll (S/q + 1)(R/q + 1) + q\) options for \(\text{tr} \gamma, \text{tr} \gamma^{-1}, \alpha\). In our range of parameters it holds that \(RS \leq Cq^3\) and since \(\|\gamma^{-1}\|_\infty \leq 2\|\gamma\|\), 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 \(\text{tr} \gamma\), we have now all of the diagonal of \(\gamma\). By Equation (3.1), the diagonal of \(\gamma\) determines the diagonal of \(\gamma^{-1} \mod q\). 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 non-exceptional case when the non-diagonal entries are non-zero, the rest of the matrix has \(\ll q^e\) 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 non-diagonal entries besides \(a_{21}\) and \(a_{12}\) are non-zero, then we may guess the diagonal of \(\gamma\) and \(\gamma^{-1}\) as before, and get the other non-diagonal entries using divisor bounds. The matrix \(\gamma\) 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 non-zero 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 \( \text{tr} A = 0 \mod q \) or \( \text{tr} A = 2 \mod q \). 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 \( q^e \) options for \( a_{23}, a_{32} \) by the divisor bound. If \( a_{22}a_{33} = \det A = \pm 1 \), then they are both \( \pm 1 \), and \( a_{23}a_{32} = 0 \), so there are \( \ll S \) options for \( A \). So in any case \( A \) has at most \( q^e S \) 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:

\[
S^3(S/q + 1) \ll R^2q^2,
\]

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 \( \ll q^e \) options for \( a_{23}, a_{32} \) by the divisor bound. Then the matrix is of the form:

\[
\begin{pmatrix}
* & ? & 0 \\
0 & * & ? \\
? & ? & *
\end{pmatrix}
\]

From \( \det \gamma = 1 \) we get \( a_{12}a_{31} \), which is non-zero. 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 non-zero, 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 \( \ll q^e \) 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 non-zero, and by the divisor bound we are done.

□

References

[1] Blomer, V. Density theorems for GL(n). arXiv preprint arXiv:1906.07459 (2019).
[2] Blomer, V., Buttcane, J., and Maga, P. Applications of the Kuznetsov formula on GL(3) II: the level aspect. Mathematische Annalen 369, 1-2 (2017), 723–759.
[3] Burger, M., Li, J.-S., and Sarnak, P. Ramanujan duals and automorphic spectrum. Bulletin of the American Mathematical Society 26, 2 (1992), 253–257.
[4] Duke, W., Rudnick, Z., and Sarnak, P. Density of integer points on affine homogeneous varieties. Duke mathematical journal 71, 1 (1993), 143–179.
[5] Gamburd, A. On the spectral gap for infinite index "congruence" subgroups of SL$_2$(Z). Israel Journal of Mathematics 127, 1 (2002), 157–200.
[6] Ghosh, A., Gorodnik, A., and Nevo, A. Diophantine approximation and automorphic spectrum. International Mathematics Research Notices 2013, 21 (2013), 5002–5058.
[7] Ghosh, A., Gorodnik, A., and Nevo, A. Best possible rates of distribution of dense lattice orbits in homogeneous spaces. Journal für die reine und angewandte Mathematik (Crelles Journal) (2014).
[8] Golubev, K., and Kamber, A. On Sarnak’s density conjecture and its applications. arXiv preprint arXiv:2004.00373 (2020).
[9] Hardy, G. H., and Wright, E. M. An introduction to the theory of numbers. Oxford university press, 1979.
[10] Heath, T. C. On a Spectral Bound for Congruence Subgroup Families in SL(3, Z). PhD thesis, Columbia University, 2015.
[11] Huntley, J., and Katznelson, Y. Density theorems for congruence groups in real rank 1. Duke Math. J 71 (1993), 463–473.
[12] Huxley, M. N. Exceptional eigenvalues and congruence subgroups. The Selberg Trace Formula and Related Topics, Contemp. Math 53 (1986), 341–349.
[13] Knapp, A. W. Representation Theory of Semisimple Groups: An Overview Based on Examples (PMS-36). Princeton university press, 2016.
[14] Maucourant, F. Homogeneous asymptotic limits of haar measures of semisimple linear groups and their lattices. Duke Mathematical Journal 136, 2 (2007), 357–399.
[15] Oh, H. Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants. Duke mathematical journal 113, 1 (2002), 133–192.
[16] Sarnak, P. Notes on the generalized Ramanujan conjectures. Harmonic analysis, the trace formula, and Shimura varieties 4 (2005), 659–685.
[17] Sarnak, P. Letter to Stephen D. Miller and Naser Talebizadeh Sardari on optimal strong approximation by integral points on quadrics. 2015.
[18] Sarnak, P., and Xue, X. X. Bounds for multiplicities of automorphic representations. Duke Math. J 64, 1 (1991), 207–227.
[19] Trombi, P. C., and Varadarajan, V. S. Asymptotic behaviour of eigen functions on a semisimple Lie group: the discrete spectrum. Acta Mathematica 129, 1 (1972), 237–280.