EFFECTIVE RATNER THEOREM FOR SL(2, R) ⋊ R^2 AND GAPS IN \sqrt{n} MODULO 1

TIM BROWNING AND ILYA VINOGRA DOV

Abstract. Let G = SL(2, R) ⋊ R^2 and \Gamma = SL(2, Z) ⋊ Z^2. Building on recent work of Strömbergsson we prove a rate of equidistribution for the orbits of a certain 1-dimensional unipotent flow of \Gamma\backslash G, which projects to a closed horocycle in the unit tangent bundle to the modular surface. We use this to answer a question of Elkies and McMullen by making effective the convergence of the gap distribution of \sqrt{n} mod 1.

1. Introduction

Results of Ratner on measure rigidity and equidistribution of orbits [8, 9] play a fundamental role in the study of unipotent flows on homogeneous spaces. They have many applications beyond the world of dynamics, ranging from problems in number theory to mathematical physics. This paper is concerned with the problem of obtaining effective versions of results that build on Ratner’s theorem and is inspired by recent work of Strömbergsson [12].

Let G = ASL(2, R) = SL(2, R) ⋊ R^2 be the group of affine linear transformations of R^2. We define the product on G by 

(M, \mathbf{x})(M', \mathbf{x}') = (MM', \mathbf{x}M' + \mathbf{x}')

and the right action is given by \mathbf{x}(M, \mathbf{x}') = \mathbf{x}M + \mathbf{x}'. We always think of \mathbf{x} ∈ R^2 as a row vector. Put Γ = ASL(2, Z) = SL(2, Z) ⋊ Z^2 and let X = Γ\backslash G be the associated homogeneous space. The group G is unimodular and so the Haar measure \mu on G projects to a right-invariant measure on X. The space X is non-compact, but it has finite volume with respect to the projection of \mu. Following the usual abuse of notation, we denote the projected measure by \mu and normalize it so that \mu(X) = 1.

Let \nabla(x) = \begin{pmatrix} \sqrt{y} & 0 \\ 0 & 1/\sqrt{y} \end{pmatrix},

and write A⁺ = \{a(y) : y > 0\}. In what follows we will use the embedding SL(2, R) ↪ G, given by M ↦ (M, 0), which thereby allows us to think of SL(2, R) as a subgroup of G.

Strömbergsson [12] works with the unipotent flow on X generated by right multiplication by the subgroup

U₀ = \left\{ \left( \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, (0, 0) \right) : x ∈ \mathbb{R} \right\}.

He considers orbits of a point (Id₂, (\xi₁, \xi₂)) subject to a certain Diophantine condition. In [12, Thm. 1.2], effective rates of convergence are obtained for the equidistribution of such orbits under the flow a(y) as y → 0. The goal of the present paper is to extend the methods

2010 Mathematics Subject Classification. 37A25 (11L07, 11J71 37A17).
of Strömbergsson to handle the orbit generated by right multiplication by the subgroup $U = \{u(x) : x \in \mathbb{R}\}$, where

$$u(x) = \left( \begin{array}{cc} 1 & x \\ 0 & 1 \end{array} \right), \left( \frac{x}{2}, \frac{x^2}{4} \right).$$

As noted by Strömbergsson [12, §3.6], any Ad-unipotent 1-parameter subgroup of $G$ with non-trivial image in $\text{SL}(2, \mathbb{R})$ is conjugate to either $U_0$ or $U$.

With this notation we will study the rate of equidistribution of the closed orbit $\Gamma \backslash \Gamma U$ under the action of $a(y)$, as $y \to 0$. Geometrically this orbit is a lift of a closed horocycle in $\text{SL}(2, \mathbb{Z}) \backslash \text{SL}(2, \mathbb{R})$ to $\Gamma \backslash G$, and the behaviour of horocycles under the flow $A^+$ on $\text{SL}(2, \mathbb{Z}) \backslash \text{SL}(2, \mathbb{R})$ is very well understood. The main obstruction to treating the problem of horocycle lifts with the usual techniques of ergodic theory (such as thickening) is the fact that $U$ is not the entire unstable manifold of the flow $a(y)$, but only a codimension 1 submanifold. Elkies and McMullen [6] used Ratner’s measure classification theorem [8] to prove that the horocycle lifts equidistribute, but their method is ineffective. In [6, §3.6] they ask whether explicit error estimates can be obtained. The following result answers this affirmatively.

**Theorem 1.1.** There exists $C > 0$ such that for every $f \in C^8_b(X)$ and $y > 0$ we have

$$\left| \frac{1}{2} \int_{-1}^1 f(u(x)a(y)) \, dx - \int_X f \, d\mu \right| < C\|f\|_{C^8_b} y^{\frac{4}{5}} \log^2(2 + y^{-1}).$$

Here $C^k_b(X)$ denotes the space of $k$ times continuously differentiable functions on $X$ whose left-invariant derivatives up to order $k$ are bounded (see equation (2.6) for the exact definition of the norm).

Our next result shows that we can replace $dx$ by a sufficiently smooth absolutely continuous measure. Let $\rho : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a compactly supported function that has $1+\varepsilon$ derivatives in $L^1$. For simplicity we follow [12] and interpolate between the Sobolev norms $\|\rho\|_{W^{1,1}}$ and $\|\rho\|_{W^{2,1}}$, which give the $L^1$ norms of first and second derivatives, respectively. This interpolation allows us to treat the case of piecewise constant functions with an $\varepsilon$-loss in the rate.

**Theorem 1.2.** Let $\eta \in (0, 1)$. There exists $K > 1$ and $C(\eta) > 0$ such that for every $f \in C^8_b(X)$ and $y > 0$ we have

$$\left| \int_{\mathbb{R}} f(u(x)a(y))\rho(x) \, dx - \int_X f \, d\mu \int_{\mathbb{R}} \rho(x) \, dx \right| < C(\eta)\|\rho\|_{W^{1,1}}^{1-\eta}\|\rho\|_{W^{2,1}}^\eta \|f\|_{C^8_b} y^{\frac{4}{5}} \log^K(2 + y^{-1}).$$

The constant $K$ in this result is absolute and does not depend on $\eta$. The proof of Theorems 1.1 and 1.2 builds on the proof of [12 Thm. 1.2]. It relies on Fourier analysis and estimates for complete exponential sums which are essentially due to Weil. Let us remark that while we strive to obtain the best possible decay in $y$, we take little effort to optimize the norms of $f$ and $\rho$ that appear in the estimates. The exponent $\frac{1}{4}$ in the error term is optimal for our method, but we surmise it can be improved by exploiting additional cancellation in certain two dimensional exponential sums. The natural upper limit is $\frac{1}{2}$, which holds for horocycles on $\text{SL}(2, \mathbb{Z}) \backslash \text{SL}(2, \mathbb{R})$ due to work of Sarnak [10].

We may apply Theorem 1.1 to study gaps between the fractional parts of $\sqrt{n}$. Consider the sequence $\sqrt{n} \mod 1 \subset \mathbb{R}/\mathbb{Z} \cong S^1$. It is easy to see from Weyl’s criterion that this
sequence is uniformly distributed on the circle. This means that for every interval $J \subset S^1$, we have
\[
\lim_{N \to \infty} \frac{|\{\sqrt{n} \mod 1 : 1 \leq n \leq N\} \cap J|}{N} = |J|,
\]
where $| \cdot |$ denotes length. The statistic we focus on is the gap distribution. For each $N \in \mathbb{N}$, we consider the set $\{\sqrt{n} \mod 1 \}_{1 \leq n \leq N}$ and we allow 0 $\in \mathbb{R}/\mathbb{Z}$ to be included for each perfect square. This set of $N$ points divides the circle into $N$ intervals (a few of which could be of zero length) which we refer to as gaps. For $t \geq 0$, we define the gap distribution $\lambda_N(t)$ to be the proportion of gaps whose length is less than $t/N$. This function satisfies $\lambda_N(0) = 0$ and $\lambda_N(\infty) = 1$, and it is left-continuous.

The behaviour of $\lambda_N(t)$, as $N \to \infty$, has been analyzed by Elkies and McMullen [6] and later also by Sinai [11]. It is shown in [6] that there exists a function $\lambda_\infty(t)$ such that $\lambda_N(t) \to \lambda_\infty(t)$ for each $t$. We have
\[
\lambda_\infty(t) = \int_0^t F(\xi) d\xi,
\]
where $F$ is given in [6, Thm. 1.1]. It is defined by analytic functions on three intervals, but it is not analytic at the endpoints joining these intervals. Moreover, it is constant on the interval $[0, 1/2]$.

The approach of Elkies and McMullen [6] is to relate $\lambda_N(t)$ to a function on $X$, so that the problem of understanding $\lambda_N(t)$ is translated into studying
\[
\frac{1}{2} \int_{-1}^1 f(u(x)a(1/N)) dx,
\]
as $N \to \infty$, for a certain function $f$ that depends on $t$. The error terms appearing in this step are worked out explicitly in [6]. In fact, $f$ is directly related to
\[
\sigma_N(t) = \int_0^t \xi d\lambda_N(\xi),
\]
which is the total length of gaps whose length is less than $t/N$. The key input in [6] comes from Ratner’s theorem [8], which is used in [6, Thm. 2.2] to find the limiting distribution of $\sigma_N(t)$. Armed with Theorem 1.1, we will refine this approach to get the following result.

**Corollary 1.3.** Let $\lambda_N(t), \lambda_\infty(t)$ be as above and let $t \geq N^{-1/40} \log N$. Then
\[
|\lambda_N(t) - \lambda_\infty(t)| \ll \max\{1, t^{-1/9}\} N^{-1/36} \log^{2/9} N
\]
for any $N \geq 2$.

The sequence $\sqrt{n} \mod 1$ has also been studied from the perspective of its pair correlation function. This is a useful statistic for measuring randomness in sequences and, in this setting, it has been shown to converge to that of a Poisson point process by El-Baz, Marklof, and the second author [5]. In the light of Theorem 1.1, although we will not carry out the details here, by developing effective versions of the results in [5] it would be possible to conclude that the pair correlation function converges effectively. By way of comparison, we remark that Strömbergsson [12, §1.3] indicates how one might make effective the convergence of the pair correlation function in the problem of visible lattice points (see [4]).

The plan of the paper is as follows. In Section 2 we embark on the proof of Theorem 1.1 by developing $f$ into a Fourier series in the torus coordinate. Section 3 is dedicated to
estimating certain complete exponential sums that are required in Section 4 to control the error terms. Corollary 1.3 is proved in Section 5 and, finally, the proof of Theorem 1.2 is sketched in Section 6.

**Notation.** Given functions $f, g : S \to \mathbb{R}$, with $g$ positive, we will write $f \ll g$ if there exists a constant $c$ such that $|f(s)| \leq cg(s)$ for all $s \in S$.

**Acknowledgements.** The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreements 291147 and 306457. The authors are grateful to Andreas Strömbergsson and Jens Marklof for helpful discussions and comments on an earlier draft. Special thanks are due to the referee for numerous useful remarks that have improved the paper considerably.

2. Fourier Decomposition

In this section we develop the tools necessary to prove Theorem 1.1 and decompose $f$ into a Fourier series on the torus. We proceed exactly as in [12]. To begin with we note that

$$f((1, \xi)M) = f((1, \xi + n)M)$$

for $n \in \mathbb{Z}^2$. So for $M$ fixed, $f$ is a well defined function on $\mathbb{R}^2 / \mathbb{Z}^2$ and we can expand it into a Fourier series as

$$f((1, \xi)M) = \sum_{m \in \mathbb{Z}^2} \hat{f}(M, m)e(m.\xi),$$

(2.1)

where

$$\hat{f}(M, m) = \int_{\mathbb{T}^2} f((1, \xi')M)e(-m.\xi')d\xi'.$$

Note that

$$\hat{f}(TM, m) = \hat{f}(M, (m(T^{-1})^t)),$$

(2.2)

for $T \in \text{SL}(2, \mathbb{Z})$. Set $\tilde{f}_n(M) = \hat{f}(M, (n, 0))$. These functions of $M \in \text{SL}(2, \mathbb{R})$ are left-invariant under the group $(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix})$ by (2.2).

Now it follows from (2.2) that

$$\tilde{f}_n \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} M \right) = \hat{f} \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} M, (n, 0) \right) = \hat{f} \left( M, (n, 0) \begin{pmatrix} d & -c \\ -b & a \end{pmatrix} \right)$$

$$= \hat{f}(M, (nd, -nc)).$$

Therefore we can rewrite (2.1) with $\xi = (x/2, -x^2/4)$ as

$$f \left( (1, (x/2, -x^2/4)) M \right) = \tilde{f}_0(M) + \sum_{n \geq 1} \sum_{(c, d) = 1} \tilde{f}_n \left( \begin{pmatrix} * & * \\ c & d \end{pmatrix} M \right) e \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} \right) \right),$$

(2.3)

where $\begin{pmatrix} * & * \\ c & d \end{pmatrix}$ is any matrix in $\text{SL}(2, \mathbb{Z})$ with $c$ and $d$ in the second row as specified. Integrating (2.3) over $x$, we obtain

$$\frac{1}{2} \int_{-1}^1 f(u(x)a(y))dx = M(y) + E(y),$$

where

$$M(y) = \frac{1}{2} \int_{-1}^1 \tilde{f}_0 \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix} dx$$

(2.4)
and
\[ E(y) = \sum_{n \geq 1} \frac{1}{n} \int_{-1}^{1} e \left( n \left( \frac{dx}{2} + \frac{c x^2}{4} \right) \right) \tilde{f}_n \left( \begin{pmatrix} * & * \\ c & d \end{pmatrix} \begin{pmatrix} \sqrt{y} \\ 0 \end{pmatrix} \right) \right) dx. \] (2.5)

The main term in this expression is \( M(y) \) and, as is well-known (cf. [10, 7, 2, 12]), we have
\[ M(y) = \int_X f \, d\mu + O(\|f\|_{C_b^0} y^{1/2-\varepsilon}). \]

This statement is nothing more than effective equidistribution of horocycles under the geodesic flow on \( SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R}) \). We need not seek the best error term for this problem, since there will be larger contributions to the error term in Theorem 1.1.

It remains to estimate \( E(y) \) as \( y \to 0 \), which we do in Section 4.

We end this section with a pair of technical results that will help us to estimate \( E(y) \). First, however, we give a precise definition of \( \| \cdot \|_{C_b^m} \) for functions on \( G \) and hence also on \( X \). Following [12], we let \( \mathfrak{g} = \mathfrak{sl}(2, \mathbb{R}) \oplus \mathbb{R}^2 \) be the Lie algebra of \( G \) and fix
\[ X_1 = ((\frac{1}{0}, 0), 0), \quad X_2 = ((\frac{0}{1}, 0), 0), \quad X_3 = ((\frac{0}{1}, 1), 0), \]
\[ X_4 = ((\frac{0}{1}, 1), (0, 1)), \quad X_5 = ((\frac{0}{1}, 0), (0, 1)) \]
to be a basis of \( \mathfrak{g} \). Every element of the universal enveloping algebra \( U(\mathfrak{g}) \) corresponds to a left-invariant differential operator on functions on \( X \). We define
\[ \|f\|_{C_b^m} = \sum_{\text{deg } D \leq m} \|Df\|_{L^\infty}, \] (2.6)
where the sum runs over monomials in \( X_1, \ldots, X_5 \) of degree at most \( m \).

The following result is [12, Lemma 4.2].

**Lemma 2.1.** Let \( m \geq 0 \) and \( n > 0 \) be integers. Then
\[ \tilde{f}_n \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) \ll_m \frac{\|f\|_{C_b^m}}{n^m(c^2 + d^2)^{m/2}}, \quad \forall \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) \in SL(2, \mathbb{R}). \]

Passing to Iwasawa coordinates in \( SL(2, \mathbb{R}) \), we write
\[ \tilde{f}_n(u, v, \theta) = \tilde{f}_n \left( \begin{pmatrix} u \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} \sqrt{v} & 0 \\ 0 & 1/\sqrt{v} \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \right). \] (2.7)

for \( u \in \mathbb{R}, v > 0 \) and \( \theta \in \mathbb{R}/2\pi \mathbb{R} \). The following is [12, Lemma 4.4].

**Lemma 2.2.** Let \( m, k_1, k_2, k_3 \geq 0 \) and \( n > 0 \) be integers, and let \( k = k_1 + k_2 + k_3 \). Then
\[ \partial_u^{k_1} \partial_v^{k_2} \partial_\theta^{k_3} \tilde{f}_n(u, v, \theta) \ll_{m, k} \|f\|_{C_b^m} v^{-m/2-k_1-k_2}. \]

3. **Complete exponential sums**

In this section we make a detailed examination of the exponential sum
\[ T_q(A, B) = \sum_{n \mod q} e_q \left( A n^2 + B \bar{n} \right). \] (3.1)

for \( A, B \in \mathbb{Z} \) and \( q \in \mathbb{N} \). Here \( e_q(\cdot) = e(\frac{\cdot}{q}) \) and \( \bar{n} \) is the multiplicative inverse of \( n \) modulo \( q \). Our main tool is Weil’s resolution of the Riemann hypothesis for function fields in one variable (see Bombieri [11]), together with some general results due to Cochrane and Zheng.
Suppose first that \( p > r = 3 \) we have

\[
\sum_{n \mod p^m} e_{p^m} \left( \frac{f_1(n)}{f_2(n)} \right),
\]

where \( f_1(x) = Ax^3 + B \) and \( f_2(x) = x \). The symbol \( \sum^* \) means that \( n \) is only taken over values for which \( p \nmid f_2(n) \), in which scenario \( f_1(n)/f_2(n) \) means \( f_1(n)f_2(n) \). We proceed by establishing the following result, which deals with the odd prime powers.

**Lemma 3.1.** Let \( p > 2 \) and \( m \in \mathbb{N} \). Then we have

\[
|T_{p^m}(A, B)| \leq \begin{cases} 
3p^{m/2}(p^m, A, B)^{1/2}, & \text{if } p > 3, \\
3^{1+3m/4}(3^m, A, B)^{1/4}, & \text{if } p = 3.
\end{cases}
\]

**Proof.** When \( m = 1 \) the sum in which we are interested is a classical exponential sum over a finite field and we may use the Weil bound, in the form developed by Bombieri \([1]\) for rational functions. This leads to the satisfactory estimate

\[
|T_p(A, B)| \leq 2p^{1/2}(p, A, B)^{1/2}.
\]

Our investigation of the case \( m \geq 2 \) is founded on work of Cochrane and Zheng \([3]\, \text{§3}], \) with \( f(x) = f_1(x)/f_2(x) \). Note that

\[
f'(x) = \frac{2Ax^3 - B}{x^2}.
\]

Following \([3\, \text{Eq. (1.8)]} \) and recalling that \( p \) is odd, we put

\[
t = \text{ord}_p(f') = \text{ord}_p(2Ax^3 - B) - \text{ord}_p(x^2) = \nu_p((A, B)).
\]

Here, if \( \text{ord}_p(h) \) is the largest power of \( p \) dividing all of the coefficients of a polynomial \( h \in \mathbb{Z}[x] \), then \( \text{ord}_p(f_1/f_2) = \text{ord}_p(f_1) - \text{ord}_p(f_2) \). Next, we put

\[
\mathcal{A} = \{ \alpha \in \mathbb{F}_p^* : 2A'\alpha^3 \equiv B' \mod p \},
\]

where \( A' = p^{-t}A \) and \( B' = p^{-t}B \). In particular \( (p, A', B') = 1 \) and \#\( \mathcal{A} \leq 3 \). The elements of \( \mathcal{A} \) are called the critical points. If \( p \mid A' \) or \( p \mid B' \) then \( \mathcal{A} \) is empty since \( (p, A', B') = 1 \). We therefore suppose that \( p \nmid A'B' \).

The strength of our estimate for \( T_{p^m}(A, B) \) depends on the multiplicity \( \nu_\alpha \) of each \( \alpha \in \mathcal{A} \). Suppose first that \( p > 3 \) and write \( r(x) = 2A'x^3 - B' \). Any root of multiplicity exceeding 1 must also be a root of \( r'(x) = 6A'x^2 \). Hence any \( \alpha \in \mathcal{A} \) satisfies \( \nu_\alpha = 1 \) if \( p > 3 \). When \( p = 3 \) we have \( 2A'\alpha^3 - B' = (2A'\alpha - B')^3 \) in \( \mathbb{F}_3 \) and so \( \mathcal{A} \) contains a single element \( \alpha \) of multiplicity \( \nu_\alpha = 3 \).

Next, as in \([3\, \text{§1}]\), one writes

\[
T_{p^m}(A, B) = \sum_{\alpha \in \mathbb{F}_p^*} S_\alpha,
\]
with

\[ S_\alpha = \sum_{n \equiv \alpha \mod p}^* e_{p^n} \left( \frac{f_1(n)}{f_2(n)} \right). \]

We are now ready to establish Lemma 3.1. When \( m \leq t \) we have \( T_{p^m}(A, B) = \varphi(p^m) \), which is satisfactory. When \( m = t + 1 \) we have

\[ T_{p^m}(A, B) = p^{m-1} T_p(A', B'), \]

which has absolute value at most \( 2p^{m-1/2} \leq 2p^{m/2+t/2} \), by (3.4). It remains to deal with the case \( m \geq t + 2 \). Then [3 Thm. 3.1(a)] implies that \( S_\alpha = 0 \) unless \( \alpha \in \mathcal{A} \). If \( \alpha \in \mathcal{A} \) then this same result yields

\[ |S_\alpha| \leq \nu_\alpha p^{t/(\nu_\alpha + 1)} p^{m(1-1/(\nu_\alpha + 1))}. \]

We recall that \( \#\mathcal{A} \leq 3 \) and \( \nu_\alpha = 1 \) for each \( \alpha \in \mathcal{A} \) if \( p > 3 \), while \( \#\mathcal{A} = 1 \) and \( \nu_\alpha = 3 \) if \( p = 3 \). Substituting this into our expression for \( T_{p^m}(A, B) \), this therefore concludes the proof of the lemma. \( \square \)

We complement our analysis of \( T_{p^m}(A, B) \) for odd \( p \) by studying the exponential sum

\[ T_{2^m}(A, B; \delta) = \sum_{n \equiv 0 \mod 2^m} e_{2^m+\delta} \left( An^2 + 2^\delta Bn \right), \quad (3.5) \]

for \( A, B \in \mathbb{Z} \) and \( \delta \in \{0, 1\} \). When \( \delta = 0 \) we have \( T_{2^m}(A, B; 0) = T_{2^m}(A, B) \), in our earlier notation. Furthermore, on writing \( x = u + 2^m v \) for \( u \in (\mathbb{Z}/2^m \mathbb{Z})^* \) and \( v \in \mathbb{Z}/2^m \mathbb{Z} \), it is easy to check that

\[ \sum_{x \equiv 0 \mod 2^{m+1}}^* e_{2^{m+1}} \left( \frac{Ax^3 + 2B}{x} \right) = 2T_{2^m}(A, B; 1). \]

Hence we have

\[ T_{2^m}(A, B; \delta) = \frac{1}{2^\delta} \sum_{x \equiv 0 \mod 2^{m+\delta}}^* e_{2^{m+\delta}} \left( \frac{Ax^3 + 2^\delta B}{x} \right), \]

for \( \delta \in \{0, 1\} \), which brings our sum in line with the exponential sums considered by Cochrane and Zheng [3]. We proceed to establish the following result.

**Lemma 3.2.** Let \( \delta \in \{0, 1\} \) and \( m \in \mathbb{N} \). Then we have

\[ |T_{2^m}(A, B; \delta)| \leq 6 \cdot 2^{3m/4} (2^m, A, B)^{1/4}. \]

**Proof.** Let us put \( t = \nu_2 ((2A, 2^\delta B)) \). Then \( u \leq t \leq 1 + u \), with \( u = \nu_2 ((A, B)) \). Suppose first that \( m \leq t + 2 \). Then the trivial bound gives

\[ |T_{2^m}(A, B; \delta)| \leq \varphi(2^m) = 2^{m-1} \leq 2^{(m+\min\{m, u\}+1)/2}, \]

which is satisfactory for the lemma. We henceforth assume that \( m \geq t + 3 \).

We are interested in a complete exponential sum modulo \( 2^{m+\delta} \). Arguing as in the proof Lemma 3.1, we have

\[ f'(x) = \frac{2Ax^3 - 2^\delta B}{x^2}. \]

and \( \text{ord}_2(f') = t \). Next, we put

\[ \mathcal{A} = \{ \alpha \in \mathbb{F}_2^*: A'\alpha^3 \equiv B' \mod 2 \}. \]
where $A' = 2^{1-t}A$ and $B' = 2^{3-t}B$. In particular, $A', B'$ are integers which cannot both be even and $A$ consists of at most 1 element and it has multiplicity at most 3. It therefore follows from [3] Thm. 3.1(b) that $S_\alpha = 0$ unless $\alpha \in A$, in which case $|S_\alpha| \leq 3 \cdot 2^{t/4 + 3(m+\delta)/4}$. But then

$$|T_{2m}(A, B; \delta)| \leq 3 \cdot 2^{(1+u)/4 + 3(m+\delta)/4} = 6 \cdot 2^{3m/4 + u/4}.$$ 

This too is satisfactory for the lemma and so completes its proof.

For any $q \in \mathbb{N}$ we will henceforth write $q = q_0q_1$, where

$$q_1 = \prod_{p \mid q} p^j.$$ \hspace{1cm} (3.6)

That is, $q_0$ is not divisible by primes other than 2 and 3, while $q_1$ is coprime to 6. Using the multiplicativity property (3.2), we may combine Lemma 3.1 and Lemma 3.2 with $\delta = 0$ to arrive at the following result.

**Lemma 3.3.** Let $q \in \mathbb{N}$ and let $A, B \in \mathbb{Z}$. Then we have

$$|T_q(A, B)| \leq 18 \cdot 3^{\omega(q_1)}q_0^{3/4}q_1^{1/2}(q_0, A, B)^{1/4}(q_1, A, B)^{1/2},$$

where $\omega(q_1)$ is the number of distinct prime factors of $q_1$.

## 4. Error terms

The purpose of this section is to estimate $E(y)$ in (2.5). We begin with the case $c = 0$. Then $d = \pm 1$ by coprimality, and [12] Eq. (25)] yields

$$\frac{1}{2} \int_{-1}^{1} \tilde{f}_n \left( \pm \left( \begin{array}{c} \sqrt{y} \\ 0 \\ 1/\sqrt{y} \end{array} \right) \frac{x}{\sqrt{y}} \right) dx \ll \|f\|c^2 n^{-2} y/n^2.$$

After summing over $n$, the contribution from this term is clearly much smaller than that claimed in Theorem 1.1.

Next we consider the effect of shifting the interval of integration by 2 in (2.5). For this it will be convenient to note that

$$\tilde{f}_n \left( \left( \begin{array}{c} a \\ b \\ c \\ d \end{array} \right) \left( \sqrt{y} \\ 0 \\ 1/\sqrt{y} \right) \right) = \tilde{f}_n \left( \left( \begin{array}{c} a \\ b \\ c \\ d \end{array} \right) \left( \begin{array}{c} 1 \\ -2 \\ 0 \\ 1 \end{array} \right) \left( \begin{array}{c} \sqrt{y} \\ x/\sqrt{y} \\ 0 \\ 1/\sqrt{y} \end{array} \right) \right)

= \tilde{f}_n \left( \left( \begin{array}{c} a \\ b - 2a \\ c \\ d - 2c \end{array} \right) \left( \sqrt{y} \\ x/\sqrt{y} \\ 0 \\ 1/\sqrt{y} \right) \right)$$

and

$$e \left( n \left( \frac{d(x-2)}{2} + \frac{c(x-2)^2}{4} \right) \right) = e \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} - cx \right) \right)

= e \left( n \left( \frac{(d-2c)x}{2} + \frac{cx^2}{4} \right) \right).$$
Bearing these in mind it follows that for any $D \in \mathbb{Z}$ and $s \in \mathbb{R}$ we have

\[
\sum_{(c,d)=1 \atop d|D,D+2c} \frac{1}{2} \int_{s}^{s+2} e \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} \right) \right) \tilde{f}_n \left( \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \left( \begin{array}{c} \sqrt{y} \\ 0 \\ 1/\sqrt{y} \end{array} \right) \right) \, dx
\]

\[
= \sum_{(c,d)=1 \atop d|D,D+2c} \frac{1}{2} \int_{s}^{s+4} e \left( n \left( \frac{(d-2c)x}{2} + \frac{cx^2}{4} \right) \right) \tilde{f}_n \left( \left( \begin{array}{cc} a & b-2a \\ c & d-2c \end{array} \right) \left( \begin{array}{c} \sqrt{y} \\ 0 \\ 1/\sqrt{y} \end{array} \right) \right) \, dx
\]

\[
= \sum_{(c,d)=1 \atop d|D-2c,D} \frac{1}{2} \int_{s}^{s+4} e \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} \right) \right) \tilde{f}_n \left( \left( \begin{array}{cc} a & b-2a \\ c & d \end{array} \right) \left( \begin{array}{c} \sqrt{y} \\ 0 \\ 1/\sqrt{y} \end{array} \right) \right) \, dx.
\]

But the values of $a$ and $b$ are immaterial and so the contribution to \([2.5]\) from terms with $c \neq 0$ is

\[
\sum_{n \geq 1 \atop c \geq 1} \frac{1}{2} \sum_{(c,d)=1 \atop d \equiv 2c \mod 2c} \int_{\mathbb{R}} e \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} \right) \right) \tilde{f}_n \left( \left( \begin{array}{cc} * & * \\ c & d \end{array} \right) \left( \begin{array}{c} \sqrt{y} \\ 0 \\ 1/\sqrt{y} \end{array} \right) \right) \, dx.
\]

Next, we change to Iwasawa coordinates as in \([2.7]\) (cf. \[12, Lemma 6.1\]). This leads to the expression

\[
\int_{\mathbb{R}} e \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} \right) \right) \tilde{f}_n \left( \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \left( \begin{array}{c} \sqrt{y} \\ 0 \\ 1/\sqrt{y} \end{array} \right) \right) \, dx
\]

\[
= \int_{0}^{\pi} \tilde{f}_n \left( \frac{a}{c} - \frac{\sin 2\theta}{2c^2 y} \right) e \left( -\frac{nd^2}{4c} + \frac{ncy^2 \cot^2 \theta}{4} \right) \, y \, d\theta.
\] (4.1)

for positive $c$. For negative $c$, the limits on the integral are $-\pi$ and 0. Since $ad - bc = 1$ and $a$ and $b$ are otherwise arbitrary, we write $a = \tilde{d}$ for any integer such that $\tilde{d} \equiv 1 \mod c$. Combining the integrals for positive and negative $c$ we get the contribution

\[
\sum_{n \geq 1 \atop c \geq 1} \int_{0}^{\pi} \sum_{(c,d)=1 \atop d \equiv 2c \mod 2c} \tilde{f}_n \left( \frac{\tilde{d}}{c} - \frac{\sin 2\theta}{2c^2 y} \right) e \left( -\frac{nd^2}{4c} + \frac{ncy^2 \cot^2 \theta}{4} \right) \, y \, d\theta.
\] (4.2)

Recall that $\tilde{f}_n$ is left-invariant under $\left( \begin{array}{cc} 1 & \frac{y}{z} \\ 0 & 1 \end{array} \right)$, which in Iwasawa coordinates translates into having period 1 in the first coordinate. Therefore we can expand $\tilde{f}_n$ as a Fourier series to get

\[
\tilde{f}_n \left( \frac{\tilde{d}}{c} - \frac{\sin 2\theta}{2c^2 y} \right) = \sum_{l \in \mathbb{Z}} b_{l}^{(n,c)}(\theta) e \left( \frac{ld}{c} \right) e \left( -\frac{l \sin 2\theta}{2c^2 y} \right),
\] (4.3)

whence the expression in \([4.2]\) is at most

\[
\ll \sum_{n,c,l} \int_{-\pi}^{\pi} \sum_{d \equiv 2c \mod 2c} e \left( -\frac{nd^2}{4c} + \frac{ld}{c} \right) \left| b_{l}^{(n,c)}(\theta) \right| \, y \, d\theta.
\] (4.4)

We need bounds for the Fourier coefficients and the exponential sum in \([4.4]\). Beginning with the former we have the following result.
Lemma 4.1. We have
\[ b_t^{(n,c)}(\theta) \ll \begin{cases} \|f\|_{C^m_b} \min \left\{ 1, \left( \frac{|\sin \theta|}{nc\sqrt{y}} \right)^m \right\} & \text{for any } m \geq 0, \\ l^{-2}\|f\|_{C^{m+2}b} n^{-4} \min \left\{ 1, \left( \frac{|\sin \theta|}{nc\sqrt{y}} \right)^{m-4} \right\} & \text{for any } m \geq 4. \end{cases} \]

Proof. The first inequality follows from Lemma 2.1 by taking the smaller of the estimate for general \( m \) and \( m = 0 \). To obtain the second inequality we observe that
\[ b_t^{(n,c)}(\theta) = \int_0^1 \tilde{f}_n \left( u, \frac{\sin^2 \theta}{c^2 y}, \theta \right) e(-lu) du. \]

We then apply integration by parts twice, followed by two applications of Lemma 2.2 one with \( k_1 = 2, k_2 = k_3 = 0, m = 4 \), and the other with \( k_1 = 2, k_2 = k_3 = 0, \) and \( m \) general. Taking the small of the two outcomes yields the result. \( \square \)

Next we turn to the exponential sum, with the following outcome.

Lemma 4.2. We have
\[ \left| \sum_{d \mod 2c \atop (d,c)=1} e \left( -\frac{nd^2}{4c} + \frac{ld\bar{d}}{c} \right) \right| \ll 3w(c_1)^{3/4}c_0^{1/4}c_1^{1/4}(c_0, n, l)^{1/4}(c_1, n, l)^{1/2}, \]
where \( c = c_0c_1 \) and \( c_1 \) is given by (3.6).

Proof. Let \( S(l, n; c) \) denote the exponential sum in the statement of the lemma. We need to relate \( S(l, n; c) \) to the complete exponential sums considered in Section 3.

The sum over \( d \) runs modulo \( 2c \) in \( S(l, n; c) \). Let us write
\[ S(l, n; c) = S_1 + S_2, \]
where \( S_1 \) is the contribution to the sum from even \( d \) and \( S_2 \) is the remaining contribution. Writing \( d = 2d' \), we see that
\[ S_1 = \sum_{d' \mod c \atop (2d', c)=1} e \left( -\frac{nd'^2}{c} + \frac{ld'}{c} \right) = \begin{cases} T_c(-n, 2l), & \text{if } 2 \nmid c, \\ 0, & \text{if } 2 \mid c. \end{cases} \]
in the notation of (3.1). The desired estimate for \( S_1 \) is now a direct consequence of Lemma 3.3. Next, we note that
\[ S_2 = \sum_{d \mod 2c \atop (d,2c)=1} e_{2c} \left( -\frac{nd^2}{2} + 2ld\bar{d} \right), \]
where \( \bar{d} \) is now the multiplicative inverse of \( d \) modulo \( 2c \). If \( n \) is even then \( S_2 = T_{2c}(-n/2, 2l) \), which can again be estimated using Lemma 3.3. If, on the other hand, \( n \) is odd we write \( c = 2^{m-1}c' \) for odd \( c' \in \mathbb{N} \). Then \( S_2 \) factorises as the product of an exponential sum modulo \( 2^m \) and an exponential sum modulo \( c' \). A satisfactory estimate for the latter follows from Lemma 3.3. Using (3.2), the former is equal to \( T_{2m}(-n\bar{c'}, 4lc'; 1) \), in the notation of (3.5), where \( \bar{c'} \in \mathbb{Z} \) is chosen to satisfy \( c'\bar{c'} \equiv 1 \mod 2^{m+1} \). In this case \( (2^m, -n, 4l) = 1 \) since \( n \) is odd. This can be estimated using Lemma 3.2 which ultimately leads to a satisfactory estimate for \( |S_2| \). This concludes the proof of the lemma. \( \square \)
We learn from [12, Eq. (35)] that
\[
\frac{1}{a(1 + a)} \ll \int_{-\pi}^{\pi} \min \left\{ 1, \left( \frac{\sin \theta}{a} \right)^2 \right\} \frac{d\theta}{\sin^2 \theta} \ll \frac{1}{a(1 + a)},
\]
for \(a > 0\). We will apply this with \(a = nc\sqrt{y}\).

Returning to (4.4), we recall the factorisation \(c = c_0 c_1\), where \(c_0\) is not divisible by primes greater than 3, and where \(c_1\) given by (3.6). It will be useful to note that
\[
\sum_{c_0} c_0^{-\gamma} = \sum_{a, \beta \geq 0} 2^{-\alpha \gamma} 3^{-\beta \gamma} = O(1),
\]
for any \(\gamma > 0\). We first consider the case \(l = 0\). Combining the first line of (4.5) with \(m = 2\) and Lemma 4.2, we obtain the contribution
\[
\ll \|f\|_{b^2} \sum_{n \geq 1} \frac{y}{nc\sqrt{y}(1 + nc\sqrt{y})} 3^{\omega(c_1)} c_0^{3/4} c_1^{1/4} (c_0, n)^{1/4} (c_1, n)^{1/2}
\]
\[
\ll \|f\|_{b^2} \sqrt{y} \sum_{n, c_0, c_1} 3^{\omega(c_1)} (c_0, n)^{1/4} (c_1, n)^{1/2}
\]
\[
\ll \|f\|_{b^2} \sqrt{y} \sum_{n, c_1} \frac{3^{\omega(c_1)} (c_1, n)^{1/2}}{n^{3/4} c_1^{1/2} (1 + nc_1 \sqrt{y})},
\]
by (4.8). The resulting sum is only made larger by summing over all positive integers \(c\) and \(c_1\), and so we freely replace \(c_1\) by \(c\). Let us denote the right hand side by \(J\). Writing \(h = (c, n)\) and \(c = hc'\) and \(n = h n'\), we see that
\[
J \ll \|f\|_{b^2} \sqrt{y} \sum_{h, n', c'} \frac{3^{\omega(hc')}}{h^{3/4} n'^{3/4} c'^{1/2} (1 + h^2 n' c' \sqrt{y})}.
\]
To proceed further we recall (see Tenenbaum [13, Ex. I.3.4], for example) that there is an absolute constant \(C > 0\) such that
\[
\sum_{n \leq x} 3^{\omega(n)} = C x \log^2 x + O(x \log x),
\]
for any \(x \geq 2\). The bounds
\[
\sum_{c > x} \frac{3^{\omega(c)}}{c^{3/2}} \ll \frac{\log^2(2 + x)}{x^{1/2}} \quad \text{and} \quad \sum_{c \leq x} \frac{3^{\omega(c)}}{c^{1/2}} \ll x^{1/2} \log^2(2 + x)
\]
now follow from this using partial summation and are valid for any \(x > 0\). We therefore obtain
\[
\sum_{c' > (h^2 n' \sqrt{y})^{-1}} \frac{3^{\omega(hc')}}{h^{3/4} n'^{3/4} c'^{1/2} (1 + h^2 n' c' \sqrt{y})} \ll \frac{3^{\omega(h)} \log^2(2 + (h^2 n' \sqrt{y})^{-1})}{h^{7/4} n'^{5/4} y^{1/4}},
\]
and similarly,
\[
\sum_{c' \leq (h^2 n' \sqrt{y})^{-1}} \frac{3^{\omega(hc')}}{h^{3/4} n'^{3/4} c'^{1/2} (1 + h^2 n' c' \sqrt{y})} \ll \frac{3^{\omega(h)} \log^2(2 + (h^2 n' \sqrt{y})^{-1})}{h^{7/4} n'^{5/4} y^{1/4}}.
\]
It therefore follows that
\[ J \ll \|f\|c_8^2 \sqrt{y} \sum_{h,n'} 3^{\omega(h)} \frac{\log^2(2 + (h^2 n' \sqrt{y})^{-1})}{h^{7/4} n'^{5/4} y^{1/4}} \ll \|f\|c_8^2 y^{1/4} \log^2(2 + y^{-1}). \]

In the case \( l \neq 0 \), we apply the second inequality from (4.5) with \( m = 6 \), together with (4.7). The contribution of these terms is therefore found to be
\[ \ll \|f\|c_8^4 \sum_{n,c,l} 3^{\omega(c_1)} c_0^{3/4} c_1^{1/2} (c_0, n, l)^{1/4} (c_1, n, l)^{1/4} y l^{-2} n^{-4} \frac{1}{nc \sqrt{y} (1 + nc \sqrt{y})}. \]

Using (4.8) and taking \((c_0, n, l)^{1/4} \leq 1/4 \) and \((c_1, n, l)^{1/2} \leq 1/2 \), we see that this is
\[ \ll \|f\|c_8^4 \sum_{n,c} 3^{\omega(c)} c_0^{3/4} c_1^{1/2} \frac{y^{1/2}}{n^4 c^{1/2} (1 + nc \sqrt{y})}. \]
\[ \ll \|f\|c_8^4 \left\{ \sum_n \sum_{c_{\leq 1/4}(n \sqrt{y})} 3^{\omega(c)} y^{1/2} \frac{1}{n^5 c^{3/2} \sqrt{y}} + \sum_n \sum_{c_{\leq 1/4}(n \sqrt{y})} 3^{\omega(c)} y^{1/2} \frac{1}{n^4 c^{1/2}} \right\}, \]
as before. Now we apply formulas (4.9), which shows that the latter expression is at most
\[ \ll \|f\|c_8^4 y^{1/4} \log^2(2 + y^{-1}). \]

This therefore concludes the proof of Theorem 1.1.

5. Proof of Corollary 1.3

We adopt the notation of [6]. Let \( N \geq 2 \) and let \( L_N(\alpha): \mathbb{R}/\mathbb{Z} \to [0, \infty) \) be \( N \) times the length of the gap containing \( \alpha \) (and 0 if \( \alpha \equiv \sqrt{n} \mod 1 \) for some positive integer \( n \in [1, N] \)). Let \( I_N(t) \) denote the union of gaps that are less than \( t/N \) in length. Equivalently, we put \( I_N(t) = \{ \alpha \in \mathbb{R}/\mathbb{Z}: L_N(\alpha) < t \} \). This set is related to \( \lambda_N(t) \) since its length is
\[ \sigma_N(t) = |I_N(t)| = \int_0^t \xi d\lambda_N(\xi), \]
as in (1.2). It is this function that is realized on the homogeneous space \( X \) in [6]. For \( c_+ > 0 > c_- \) let \( \Delta_{c_-,c_+} \subset \mathbb{R}^2 \) be the open triangle bounded by the lines \( w_2 = 1, w_1 = 2c_+ w_2, w_1 = 2c_- w_2 \) in the \((w_1, w_2)\)-plane. Its area is clearly \( c_+ - c_- \). For a lattice translate \( g \in \Gamma \setminus G \), let
\[ L(g) = \sup \{ c_+ - c_- : \Delta_{c_-,c_+} \cap \mathbb{Z}^2 g = \emptyset \}, \]
with the convention that \( L(g) = 0 \) if the set in the definition is empty and that \( L(g) = \infty \) if it is all of \( \mathbb{R}^+ \). \( L(g) \) is the area of the largest triangle in the family \( \Delta_{c_-,c_+} \) that is disjoint
from $\mathbb{Z}^2g$. Elkies and McMullen \cite{6} show that
\[ \sigma_N(t) \approx \frac{1}{2}\{x \in [-1, 1]: L(u(x)a(y)) \leq t\}, \]  
where $N = 1/y$. In fact, they prove this result for a suitable conjugate of $U$. Closeness in \cite{5.1} means that the difference between the two sides tends to zero as $N$ tends to infinity along perfect squares, and that (although not explicitly stated in \cite{6}) the error term can be easily calculated.

We seek to apply Theorem 1.1 to the indicator of the set \{\(g \in X: L(g) \leq t\)\} which we cannot do directly as it is not smooth. Let \(\sigma_\infty(t) = \mu\{g \in X: L(g) \leq t\}\), let \(\Delta_{c-,c+} \subset \mathbb{R}^2\) be the closure of the triangle \(\Delta_{c-,c+}\), and let
\[ \overline{L}(g) = \sup\{c_+ - c_-: \Delta_{c-,c+} \cap \mathbb{Z}^2g = \emptyset\}, \]
with conventions as before. Defining \(\overline{\sigma_N}(t)\) and \(\overline{\sigma_\infty}(t)\) accordingly, one checks that \(\sigma_N(t) = \overline{\sigma_N}(t)\) and \(\sigma_\infty(t) = \overline{\sigma_\infty}(t)\) for all $t$. The advantage of this modification is that the boundary of the set
\[ A_t = \{g \in X: \overline{L}(g) \leq t\} \]
is precisely \(\{g \in X: \overline{L}(g) = t\}\).

We describe the geometry of this set to explain the smoothing procedure following the discussion preceding \cite{6} Prop 3.11. Every $g$ with \(\overline{L}(g) = t\) is constructed as follows. Pick \(v_-, v_+ \in (0, 1]\) and let \(\mathbb{Z}^2g\) contain the points \(P_1 = (xv_+, v_+)\) and \(P_2 = (0, v_-)\). Assume the interior of the segment \(P_1P_2\) contains no lattice points. Then, $g$ will be determined by a third point $P_3$ if the area of the triangle $P_1P_2P_3$ is $\frac{1}{2}$ and $P_3 \not\in \overline{\Delta_0t}$. The possible choices for $P_3$ parametrize a piece of a closed horocycle in $X$; the length of this horocycle is denoted $q_t(v_+, v_-)$. Finally, we can apply the matrix \(\begin{pmatrix} 1 & c_- \\ 0 & 1 \end{pmatrix}\) for some $c_- \in [-t, 0)$ to the lattice, which determines $c_+ = t + c_-$. The boundary is thus a piecewise smooth 4-manifold in $X$ realized as the image of $v_+ \in (0, 1]$, $g \in [0, q_t(v_+, v_-))$ and $c_- \in [-t, 0)$. Its measure is \(tF(t)\) (with $F$ as in \cite{1.1}) as per \cite{6} Prop. 3.11.

Let $d$ be a left-invariant metric on $G$ and let $\delta$ be positive; we will choose it later depending on $t$ and $N$. Define
\[ \partial_{t,\delta} = \{g \in G: \min_{g' \in \partial A_t} d(g, g') < \delta\} \]
and
\[ A_{t,\delta}^+ = A_t \cup \partial_{t,\delta} \]
\[ A_{t,\delta}^- = A_t \backslash \partial_{t,\delta}. \]
As we intend to use smoothed indicators of these sets to bound the indicator of $A_t$ above and below, it is essential that $\mu(\partial_{t,\delta}) \ll \mu(A_{t,\delta}^-)$. To this end we need to understand the “thinnest” part of $A_t$. When $t \ll 1$, the thinnest direction is along the variable $c_-$, which is in an interval of length $t$. Therefore we assume that $\delta \ll t$ when $t \ll 1$. In the complementary case $t \gg 1$, we turn to \cite{6} Lemma 3.12, where $q_t(v_+, v_-)$ is computed. We see that it is of order $1/t$ whenever it does not vanish. Hence $\delta \ll 1/t$ for $t \gg 1$.

Let $\eta_\delta: G \to \mathbb{R}$ be nonnegative, smooth, of unit integral, and supported in a ball of radius $\delta$ around the identity with respect to $d$. Let
\[ f_{t,\delta}^\pm = \eta_\delta * \chi_{A_{t,\delta}^\pm}, \]
which are smooth functions satisfying \( f_{t,\delta}^- \leq \chi_{A_t} \leq f_{t,\delta}^+ \). Applying Theorem 1.1 with \( y = 1/N \) gives
\[
\frac{1}{2} \int_{-1}^1 f_{t,\delta}^\pm(u(t)a(1/N)) \, dt = \mu(f_{t,\delta}^\pm) + O(\|f_{t,\delta}^\pm\|C^8N^{-1/4}\log^2 N)
= \mu(A_t) + O(\mu(\partial_t)) + O(\delta^{-8}N^{-1/4}\log^2 N),
\]
since \( f_{t,\delta}^\pm \) can be chosen to have eighth derivative of order \( \delta^{-8} \). Thus
\[
\sigma_N(t) - \sigma_{\infty}(t) \ll \delta t F(t) + \delta^{-8}N^{-1/4}\log^2 N
\]
for every \( \delta \ll \min\{t,1/t\} \). We recall from [3] that \( F(t) = \frac{1}{\zeta(2)} \) when \( t < \frac{1}{2} \) and \( F(t) \sim \frac{1}{2\zeta(2)t^\gamma} \) as \( t \to \infty \). The optimal choices for \( \delta \) are given by
\[
\delta = \begin{cases} 
  t, & t \ll N^{-1/40}\log^{1/5} N \\
  t^{-1/9}N^{-1/36}\log^{2/9} N, & N^{-1/40}\log^{1/5} N \ll t \ll 1 \\
  t^{2/9}N^{-1/36}\log^{2/9} N, & 1 \ll t \ll N^{1/44}\log^{-2/11} N \\
  1/t, & t \gg N^{1/44}\log^{-2/11} N.
\end{cases}
\]
It now follows that
\[
\sigma_N(t) - \sigma_{\infty}(t) \ll \begin{cases} 
  t^2 + t^{-8}N^{-1/4}\log^2 N, & t \ll N^{-1/40}\log^{1/5} N \\
  t^{8/9}N^{-1/36}\log^{2/9} N, & N^{-1/40}\log^{1/5} N \ll t \ll 1 \\
  t^{-16/9}N^{-1/36}\log^{2/9} N, & 1 \ll t \ll N^{1/44}\log^{-2/11} N \\
  t^{-2} + t^{8}N^{-1/4}\log^2 N, & t \gg N^{1/44}\log^{-2/11} N.
\end{cases}
\]  
In addition to this we have
\[
\sigma_N(t) - \sigma_{\infty}(t) \ll 1
\]  
for all \( t \) and \( N \), following from the trivial bounds \( \sigma_{\infty}(t) \leq 1 \) and \( \sigma_N(t) \leq 1 \).

Equipped with bounds (5.2) and (5.3), we are ready to transfer savings to \( \lambda_N \). We write
\[
\lambda_N(t) - \lambda_{\infty}(t) = \int_t^\infty \frac{d(\sigma_N(\xi) - \sigma_{\infty}(\xi))}{\xi}
= \frac{\sigma_{\infty}(t) - \sigma_N(t)}{t} + \int_t^\infty \frac{\sigma_N(\xi) - \sigma_{\infty}(\xi)}{\xi^2} \, d\xi. 
\]  
In order to prove Corollary 1.3 we need to show that
\[
\lambda_N(t) - \lambda_{\infty}(t) \ll \max\{1,t^{-1/9}\}N^{-1/36}\log^{2/9} N, 
\]  
for any \( t \gg N^{-1/40}\log N \) and any \( N \geq 2 \). Suppose first that \( t \gg N^{1/36}\log^{-2/9} N \). Then (5.5) follows on using the trivial bound (5.3) in (5.4). Next, when \( t \) lies in the interval \( N^{1/36}\log^{-2/9} N \gg t \gg N^{1/44}\log^{-2/11} N \), we split the integral at \( \xi = N^{1/36}\log^{-2/9} N \). The tail of the \( \xi \)-integral is treated using the bound (5.3), while the remaining portion of the integral and the first term are controlled using the fourth case of (5.2). The outcome of this is the bound
\[
\lambda_N(t) = \lambda_{\infty}(t) \ll t^{-3} + t^{7}N^{-1/4}\log^2 N + N^{-1/36}\log^{2/9} N \ll N^{-1/36}\log^{2/9} N,
\]  
which is satisfactory for (5.5). We complete the proof of (5.5) when \( 1 \ll t \ll N^{1/44}\log^{-2/11} N \) (resp. \( N^{-1/40}\log N \leq t \ll 1 \)) by applying the third (resp. second) bound from (5.2).
6. Sketch of the proof of Theorem 1.2

The proof of Theorem 1.2 proceeds analogously to that of Theorem 1.1. The main term in equation (2.4) becomes
\[
\int_{\mathbb{R}} \tilde{f}_0 \left( \frac{\sqrt{y}}{0} x/\sqrt{y} \right) \rho(x) \, dx.
\]
From [12, Eqs. (23), (24)] this equals
\[
\int_{X} f \, d\mu \int_{\mathbb{R}} \rho(x) \, dx + O(\|f\|_{C^4} \|\rho\|_{W^{1,1}} y^{1/2} \log^3(2 + 1/y)),
\]
which is satisfactory for the theorem.

Turning to the new error terms, we treat the term \( c = 0 \) following [12, Eq. (25)]. Our analogue of the remaining contribution to the error term \( E(y) \) in (2.5) is
\[
E_{\rho}(y) = \sum_{c,n \geq 1} \int_{\mathbb{R}} c \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} \right) \right) \tilde{f}_n \left( \left( \begin{array}{cc} * & \ast \\ c & d \end{array} \right) \left( \frac{\sqrt{y}}{0} x/\sqrt{y} \right) \right) \rho(x) \, dx.
\]
Now we proceed directly to the change of variables (4.1), following [12, Lemma 6.1]. This gives
\[
\int_{\mathbb{R}} e \left( n \left( \frac{dx}{2} + \frac{cx^2}{4} \right) \right) \tilde{f}_n \left( \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \left( \frac{\sqrt{y}}{0} x/\sqrt{y} \right) \right) \rho(x) \, dx = \int_{0}^{\pi} g(\theta) d\theta,
\]
for \( c > 0 \), where
\[
g(\theta) = \tilde{f}_n \left( \frac{\theta}{c} - \frac{-\sin 2\theta}{2c^2 y}, \frac{\sin^2 \theta}{c^2 y}, \theta \right) e \left( -\frac{nd^2}{4c} + \frac{ncy^2 \ctg^2 \theta}{4} \right) \rho \left( -\frac{d}{c} + y \ctg \theta \right) \frac{y}{\sin^2 \theta}.
\]
We have the same integral with limits \(-\pi\) and 0 if \( c < 0 \). Combining terms with positive and negative \( c \) gives
\[
E_{\rho}(y) = \sum_{c,n \geq 1} \int_{-\pi}^{\pi} \sum_{d \mod 2c} g(\theta) d\theta.
\]
We periodize \( \rho \) with period 2, by setting \( P(z) = \sum_{m \in \mathbb{Z}} \rho(z + 2m) \). Now the latter expression can be rewritten using only periodic functions as
\[
E_{\rho}(y) = \sum_{c,n \geq 1} \int_{-\pi}^{\pi} \sum_{d \mod 2c} \tilde{f}_n \left( \frac{d}{c} - \frac{-\sin 2\theta}{2c^2 y}, \frac{\sin^2 \theta}{c^2 y}, \theta \right) e \left( -\frac{nd^2}{4c} + \frac{ncy^2 \ctg^2 \theta}{4} \right) \times P \left( \frac{-d}{c} + y \ctg \theta \right) \frac{y d\theta}{\sin^2 \theta}.
\]
Exploiting periodicity of \( \tilde{f}_n \) and \( P \), we replace them by their Fourier series leaving an exponential sum and two Fourier coefficients to control. Thus
\[
E_{\rho}(y) \ll \sum_{c,n \geq 1} \int_{-\pi}^{\pi} \sum_{k,l \in \mathbb{Z}} \left| \sum_{d \mod 2c} e \left( -\frac{nd^2}{4c} + \frac{l\tilde{d}}{c} - \frac{kd}{2c} \right) \right| |b_{l}^{(n,c)}(\theta) a_k| \frac{y d\theta}{\sin^2 \theta}, \tag{6.1}
\]
where \( a_k \) are the Fourier coefficients of \( P \) and \( b_{l}^{(n,c)}(\theta) \) are as in (4.3) and (4.6). In particular, we have
\[
a_k \ll \eta \left( 1 + |k| \right)^{-1-\eta} \|\rho\|_{W^{1,1}} \|\rho\|_{W^{2,1}} \eta, \tag{6.2}
\]
for all \( \eta \in (0, 1) \), as can be seen using integration by parts.

The exponential sum in (6.1) can be estimated using the tools developed in Section 3. Note that the case \( k = 0 \) corresponds precisely to the sum considered in Lemma 4.2. We say \( u \in \mathbb{N} \) is square-free if \( p \mid u \) implies \( p^2 \nmid u \) for every prime \( p \), and similarly \( v \in \mathbb{N} \) is square-full if \( p \mid v \) implies \( p^2 \mid v \).

**Lemma 6.1.** Write \( c = c_0c_1 = c_0uv \) with \( u \) square-free, \( v \) square-full, and furthermore \((uv, 6) = (u, v) = 1\). Then, there exists an absolute constant \( K \in \mathbb{N} \) such that

\[
\left| \sum_{d \mod 2c \atop (c, d) = 1} e\left(-\frac{nd^2}{4c} + \frac{ld}{c} - \frac{kd}{2c}\right) \right| \leq K^{\omega(c)} c_0^{3/4} u^{1/2} v^{3/4} (c_0, k, n, l)^{1/4} (u, k, n, l)^{1/2} (v, k, n, l)^{1/3}.
\]

**Proof.** We will sketch the proof of this result based on the methods of Section 3. In doing so we will not pay heed to the particular value of \( K \). Arguing as in the proof of Lemma 4.2 the main task is to estimate the exponential sum

\[
T_{p^m}(A, B, C) = \sum_{n \mod p^m} \ast c_{p^m} \left( An^2 + B\bar{n} + Cn \right),
\]

for \( A, B, C \in \mathbb{Z} \) and a prime power \( p^m \). Note that \( T_{p^m}(A, B, 0) = T_{p^m}(A, B) \) in the notation of (3.1). We may now write \( T_{p^m}(A, B, C) \) in the form (3.3), with \( f_1(x) = Ax^3 + B + Cx^2 \) and \( f_2(x) = x \). When \( m = 1 \) it follows from Bombieri [1] that

\[
|T_p(A, B, C)| \leq 2p^{1/2}(p, A, B, C)^{1/2}.
\]

When \( m \geq 2 \), we apply Cochrane and and Zheng [3], as before. We see that

\[
f'(x) = \frac{2Ax^3 - B + Cx^2}{x^2},
\]

whence \( t = \text{ord}_p(f') = v_p((2A, B, C)) \), in the notation of [3, Eq. (1.8)]. This time we have

\[
A = \{ \alpha \in \mathbb{F}_p^* : A'\alpha^3 - B' + C'\alpha^2 \equiv 0 \mod p \},
\]

where \( A' = p^{-t}2A, B' = p^{-t}B \) and \( C' = p^{-t}C \). In particular \((p, A', B', C') = 1\). We may henceforth assume that \( m \geq t + 3 \) since the desired conclusion follows from the trivial bound otherwise. One finds that any critical point \( \alpha \in A \) has multiplicity \( \nu_\alpha \leq 2 \) when \( p \neq 3 \) and multiplicity \( \nu_\alpha \leq 3 \) when \( p = 3 \). Next, one applies [3, Thm. 3.1] to deduce that

\[
T_{p^m}(A, B, C) \ll \begin{cases} p^{2m/3 + \min\{m, t\}/3}, & \text{if } p \neq 3, \\ 3^{3m/4 + \min\{m, t\}/4}, & \text{if } p = 3, \end{cases}
\]

for \( m \geq 2 \).

Once combined with our treatment of the case \( m = 1 \), one arrives at the statement of the lemma on invoking multiplicativity. \( \square \)

We have four regimes in (6.1) to consider, according to whether or not \( k \) or \( l \) vanish. The cases with \( k = 0 \) are identical to those already dealt with in Section 4. We proceed to present the argument needed to handle the case \( l = 0 \) and \( k \neq 0 \). After using (4.5) with \( m = 2 \) in
\[ (6.1) \text{, followed by } (4.7), \text{ Lemma } 6.1 \text{ and } (6.2), \text{ we arrive at the bound} \]
\[
E_{\rho}(y) \ll \| f \| c_{b}^{2} \| \rho \|_{W_{1,1}}^{1-\eta} \| \rho \|_{W_{2,1}}^{\eta} \\
\times \sum_{c = c_{0}uv, n \geq 1} \sum_{k \neq 0} \frac{y}{|k|^{1+\eta}} \frac{K_{\omega(c)}^{3/4} c_{0}^{1/2} u^{1/2} v^{2/3} (c_{0}, k, n)^{1/4} (u, k, n)^{1/2} (v, k, n)^{1/3}}{nc_{v} \sqrt[3]{3} (1 + nc_{v} \sqrt[3]{3})} \]  
(6.3)

\[ = \| f \| c_{b}^{2} \| \rho \|_{W_{1,1}}^{1-\eta} \| \rho \|_{W_{2,1}}^{\eta} S(y), \]

say. We will apply the upper bounds \((c_{0}, k, n)^{1/4} \leq c_{0}^{1/4-\eta/4} |k|^{\eta/4}, (u, k, n)^{1/2} \leq (u, k)^{1/2}, \text{ and } (v, k, n)^{1/3} \leq v^{1/3-\eta/4} |k|^{\eta/4} \) in order to simplify this expression. In particular the resulting sum over \( c_{0} \) is absolutely convergent by \((1.8)\). Next we divide the sum so that \( uv \) belongs to the dyadic intervals \([2^{j-1}, 2^{j})\) for \( j \in \mathbb{N} \). In this way we deduce that

\[
S(y) \ll \sum_{n \geq 1} \sum_{k \neq 0} \frac{y^{1/2}}{|k|^{1+\eta/2}} \sum_{j \geq 1} \sum_{v \sqrt{2} \text{ full}} \frac{K_{\omega(uv)}^{1/2} v^{1/2-\eta/4}}{n \sqrt{uv} (1 + n uv \sqrt{3})} \sum_{u \leq 2^{j}/v} K_{\omega(u)}^{1/2-\eta/4} \sum_{v \sqrt{2} \text{ full}} K_{\omega(v)}^{1/2-\eta/4} (u, k)^{1/2}.
\]

Now, we have \( \sum_{n \leq x} K_{\omega(n)}^{(n, k)^{1/2}} \ll x \log K^{-1} x \) for any \( K > 1 \) and \( x \geq 2 \) (see \((13, \text{ Thm. II.6.1})\), for example). Hence it follows that

\[
\sum_{n \leq x} K_{\omega(n)}^{(n, k)^{1/2}} \ll \sum_{h \geq k} h^{1/2} \sum_{n \leq x} K_{\omega(h)}^{(h, k)^{1/2}} \ll x \log K^{-1} x \sum_{h \geq k} h^{-1/2} K_{\omega(h)},
\]

for \( x \geq 2 \). The remaining sum over \( h \) is at most \( \ll K \tau(k) \), where \( \tau \) denotes the divisor function, whence

\[
S(y) \ll y^{1/2} \sum_{n \geq 1} \frac{1}{n} \sum_{k \neq 0} \frac{\tau(k)}{|k|^{1+\eta/2}} \sum_{j \geq 1} \sum_{v \sqrt{2} \text{ full}} \frac{2^{j/2} \log K^{-1} 2^{j}}{1 + n \sqrt{3} 2^{j}} \sum_{v \sqrt{2} \text{ full}} K_{\omega(v)}^{v^{1/2+\eta/4}}.
\]

The sum over \( k \) is convergent. Furthermore, since square-full integers have square root density, the sum over \( v \) is also convergent. Hence

\[
S(y) \ll y^{1/2} \sum_{n \geq 1} \frac{1}{n} \sum_{j \geq 1} \frac{2^{j/2} \log K^{-1} 2^{j}}{1 + n \sqrt{3} 2^{j}}
\]

Once substituted into \((6.3)\), this leads to the satisfactory contribution

\[
\ll \| f \| c_{b}^{2} \| \rho \|_{W_{1,1}}^{1-\eta} \| \rho \|_{W_{2,1}}^{\eta} y^{1/4} \log K^{-1} (2 + y^{-1}),
\]

to \( E_{\rho}(y) \), for any \( \eta \in (0, 1) \).

In a similar manner, using instead the second estimate from \((4.5)\) with \( m = 6 \), the contribution from \( kl \neq 0 \) is found to be

\[
\ll \| f \| c_{b}^{2} \| \rho \|_{W_{1,1}}^{1-\eta} \| \rho \|_{W_{2,1}}^{\eta} y^{1/4} \log K^{-1} (2 + y^{-1}),
\]

for any \( \eta \in (0, 1) \). This completes the sketch of the proof of Theorem 1.2.
References

[1] Enrico Bombieri. On exponential sums in finite fields. *Amer. J. Math.*, 88:71–105, 1966.
[2] Marc Burger. Horocycle flow on geometrically finite surfaces. *Duke Math. J.*, 61(3):779–803, 1990.
[3] Todd Cochrane and Zhiyong Zheng. Exponential sums with rational function entries. *Acta Arithmetica*, 95:67–95, 2000.
[4] Daniel El-Baz, Jens Marklof, and Ilya Vinogradov. The distribution of directions in an affine lattice: two-point correlations and mixed moments. arXiv e-print 1306.0028, IMRN, doi:10.1093/imrn/rnt258, May 2013.
[5] Daniel El-Baz, Jens Marklof, and Ilya Vinogradov. The two-point correlation function of the fractional parts of $\sqrt{n}$ is poisson. arXiv 1306.6543, accepted to Proc. of AMS, June 2013.
[6] Noam D. Elkies and Curtis T. McMullen. Gaps in $\sqrt{n}$ mod 1 and ergodic theory. *Duke Math. J.*, 123(1):95–139, 2004.
[7] Livio Flaminio and Giovanni Forni. Invariant distributions and time averages for horocycle flows. *Duke Math. J.*, 119(3):465–526, 2003.
[8] Marina Ratner. On Raghunathan’s measure conjecture. *Annals of Mathematics. Second Series*, 134(3):545–607, 1991.
[9] Marina Ratner. Raghunathan’s topological conjecture and distributions of unipotent flows. *Duke Mathematical Journal*, 63(1):235–280, 1991.
[10] Peter Sarnak. Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series. *Comm. Pure Appl. Math.*, 34(6):719–739, 1981.
[11] Ya. G. Sinai. Statistics of gaps in the sequence $\{\sqrt{n}\}$. In *Dynamical systems and group actions*, volume 567 of *Contemp. Math.*, pages 185–189. Amer. Math. Soc., Providence, RI, 2012.
[12] Andreas Strömbergsson. An effective Ratner equidistribution result for ASL(2,R). *arXiv:1309.6103 [math]*, September 2013.
[13] Gérard Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 46 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Translated from the second French edition (1995) by C. B. Thomas.

School of Mathematics, University of Bristol, Bristol, BS8 1TW

E-mail address: t.d.browning@bristol.ac.uk

School of Mathematics, University of Bristol, Bristol, BS8 1TW

E-mail address: ilya.vinogradov@bristol.ac.uk