AN EFFECTIVE RATNER EQUIDISTRIBUTION RESULT FOR $\mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$

ANDREAS STRÖMBERGSSON

ABSTRACT. Let $G = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$ be the affine special linear group of the plane, and set $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2$. We prove a polynomially effective asymptotic equidistribution result for the orbits of a 1-dimensional, non-horospherical unipotent flow on $\Gamma\backslash G$.

1. Introduction

In the theory of unipotent flows on homogeneous spaces, a fundamental role is played by the theorems by M. Ratner on measure rigidity, topological rigidity, and orbit equidistribution, [38], [39]; these results also appear as a crucial ingredient in numerous, and surprisingly diverse, applications. See [52] and [24] for expositions and references; some more recent works where important use is made of Ratner's theorems are [10], [13], [43], [44], [32], [33], to just mention a few.

In the last decade there has been an increased interest in obtaining *effective* versions of Ratner's results, that is, to provide an explicit rate of density or equidistribution for the orbits of a unipotent flow. This problem was raised for example in [26, Probl. 7]. There are two general cases where it has been known for a fairly long time that effective results may be proved, namely when the group generating the flow is either horospherical or "large" in an appropriate sense (cf. [5, §1.5.2] for a discussion; compare also p. 3 below). Recently, however, some new important cases have been established: Green and Tao [15] have proved effective equidistribution of polynomial orbits on nilmanifolds; this is an important input in their work on linear equations in primes [14], [16]. Moreover, Einsiedler, Margulis and Venkatesh [5] have proved effective equidistribution for large closed orbits of semisimple groups on homogeneous spaces; see also Mohammadi [34] for a more explicit result in the special case of closed $\mathrm{SO}(2,1)$-orbits in $\mathrm{SL}(3,\mathbb{Z})\backslash\mathrm{SL}(3,\mathbb{R})$. Recently also Lindenstrauss and Margulis [25] have obtained an effective density-type result for *arbitrary* $\mathrm{SO}(2,1)$-orbits in $\mathrm{SL}(3,\mathbb{Z})\backslash\mathrm{SL}(3,\mathbb{R})$, and used this to give an effective proof of a theorem of Dani and Margulis regarding the values of indefinite ternary quadratic forms at primitive integer vectors.

Our purpose in the present paper is to establish effective Ratner equidistribution in a new particular setting: We let G be the semidirect product group $G = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$ with multiplication law

$$(M,\boldsymbol{v})(M',\boldsymbol{v}') = (MM', \boldsymbol{v}M' + \boldsymbol{v}').$$

Let $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2$ and $X = \Gamma\backslash G$, and consider the flow on X which is generated by right multiplication by the (Ad-)unipotent 1-parameter subgroup $U^{\mathbb{R}} = \{U^t \,:\, t \in \mathbb{R}\}$, where

$$U^t = \left(\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, (0,0) \right).$$

The Ratner measure rigidity and equidistribution for this particular flow, and closely related ones, have found several applications in number theory and in mathematical physics; cf. [48,

The work was conducted while Strömbergsson was a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

2010 *Mathematics Subject Classification.* Primary 37A17, 37A45; Secondary 11K60.

Remark 4], [30], [29], [8], [31], [32], [28, Thm. 1.10], [6], [7]; we discuss this further in Section 1.3. Note that $\{U^t\}$ on X is a *1-dimensional, non-horospherical* unipotent flow on a non-solvable homogeneous space. As far as we are aware, there is only one previous setting of this kind where effective equidistribution has been established; namely, the results by Venkatesh [50, §3.1] and Sarnak and Ubis [41, Thm. 4.11] for orbits of the discrete horocycle flow can be viewed as giving effective equidistribution for the flow generated by $U^t = ((\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}), s^{-1}t)$ (any fixed $s > 0$) in $(\Gamma' \times \mathbb{Z})\backslash(\mathrm{SL}(2,\mathbb{R}) \times \mathbb{R})$, with either $\Gamma' = \mathrm{SL}(2,\mathbb{Z})$ or Γ' a cocompact subgroup of $\mathrm{SL}(2,\mathbb{R})$.

The group $G = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$ can be viewed as the group of area and orientation preserving affine maps of the plane $\mathbb{R}^2$, with the action given by

$$\boldsymbol{y}(M,\boldsymbol{v}) := \boldsymbol{y}M + \boldsymbol{v}, \qquad \forall (M,\boldsymbol{v}) \in G,\ \boldsymbol{y} \in \mathbb{R}^2,$$

and a central property of $X = \Gamma\backslash G$ is that it can be naturally identified with the space of translates of unimodular lattices in $\mathbb{R}^2$, through $\Gamma g \mapsto \mathbb{Z}^2 g = \{\boldsymbol{m}g \,:\, \boldsymbol{m} \in \mathbb{Z}^2\}$. Then the subspace of (non-translated) lattices becomes identified with $X' = \Gamma'\backslash G'$, where $G' = \mathrm{SL}(2,\mathbb{R})$, which we always view as a subgroup of G through $M \mapsto (M,\boldsymbol{0})$, and $\Gamma' = \Gamma \cap G' = \mathrm{SL}(2,\mathbb{Z})$. Note that X' is an embedded submanifold of X. Furthermore, $U^{\mathbb{R}}$ is contained in G', and the flow $U^{\mathbb{R}}$ on X' is the standard horocycle flow. There is also a natural projection $D : G \to G'$ sending $(M,\boldsymbol{v})$ to M, which makes X into a torus fiber bundle over X'. We write D also for the projection map $X \to X'$. Note that the embeddings $G' \subset G$ and $X' \subset X$ are sections of D. In the language of lattice translates, the fiber over a lattice $L \in X'$ equals the torus $\mathbb{R}^2/L$ consisting of all translates of L.

Let μ be the (left and right invariant) Haar measure on G, normalized so as to induce a probability measure on X, which we also denote by μ. Then $\mu' := D_*\mu$ is the Haar measure on G' which induces a probability measure on X'.

We will start by discussing the case of $U^{\mathbb{R}}$-orbits in X which project to *closed* orbits in X'; we then turn to the case of general $U^{\mathbb{R}}$-orbits is Section 1.2.

1.1. **Lifts of pieces of closed horocycles.** Set

$$\Phi^t = \begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix} \in G' \subset G \qquad (t \in \mathbb{R}).$$

We also write $1_2 = (\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix})$. For given $\boldsymbol{\xi} \in \mathbb{R}^2$ and $t \in \mathbb{R}$, we consider pieces of the $U^{\mathbb{R}}$-orbit through the point $\Gamma\,(1_2,\boldsymbol{\xi})\,\Phi^t \in X$. These are exactly those $U^{\mathbb{R}}$-orbits in X which project to closed orbits, i.e. closed horocycles, in X'. From the relation

$$U^x\Phi^t = \Phi^t U^{e^t x} \tag{1}$$

we see that $\Gamma\,(1_2,\boldsymbol{\xi})\,\Phi^t U^{\mathbb{R}} = \Gamma\,(1_2,\boldsymbol{\xi})\,U^{\mathbb{R}}\Phi^t$, that is, the $U^{\mathbb{R}}$-orbit through $\Gamma\,(1_2,\boldsymbol{\xi})\,\Phi^t$ is obtained as the Φ^t-push-forward of the $U^{\mathbb{R}}$-orbit through $\Gamma\,(1_2,\boldsymbol{\xi})$. It also follows from (1) that the projected orbit, $x \mapsto \Gamma'\Phi^t U^x \in X'$, has period e^t with respect to x. It is well-known that these closed horocycles, and more generally the Φ^t-push-forwards of any fixed segment $\{\Gamma' U^x \,:\, x \in [\alpha,\beta]\}$, become asymptotically equidistributed in (X',μ') as $t \to \infty$. These facts are also known with precise rates; cf. [40], [18], [46], [11]. As to the orbits in X, it turns out that the Φ^t-push-forwards of a fixed segment $\{\Gamma(1_2,\boldsymbol{\xi})U^x \,:\, x \in [\alpha,\beta]\}$ become asymptotically equidistributed in (X,μ) as $t \to \infty$ *if and only if $\boldsymbol{\xi}$ is irrational.* We state the non-trivial direction of this implication as Theorem 1.1 below; it is a special case of a theorem of Shah, [42, Thm. 1.4] (cf. [32, proof of Thm. 5.2]), and also a special case of Elkies and McMullen, [8, Thm. 2.2]. Both proofs depend crucially on Ratner's classification of invariant measures. (See [9, §3] for a discussion of the proof of Ratner's theorem in exactly our setting with $G = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$, $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2$.)

Theorem 1.1. *([42] or [8]) Fix any $\boldsymbol{\xi} \in \mathbb{R}^2$ with at least one irrational coordinate, i.e. $\boldsymbol{\xi} \notin \mathbb{Q}^2$. Then the Φ^t-push-forwards of any fixed portion of the orbit $\Gamma\,(1_2,\boldsymbol{\xi})\,U^{\mathbb{R}}$ become asymptotically* *equidistributed in (X, μ) as $t \to \infty$. In other words, for any fixed $\alpha < \beta$ and any bounded continuous function $f : X \to \mathbb{R}$,*

$$\lim_{t\to\infty} \frac{1}{\beta-\alpha} \int_\alpha^\beta f\Big(\Gamma\,(1_2, \boldsymbol{\xi})\, U^x \Phi^t\Big)\, dx = \int_X f\, d\mu. \tag{2}$$

To see that the assumption $\boldsymbol{\xi} \notin \mathbb{Q}^2$ in Theorem 1.1 is a necessary condition, set, for any positive integer q,

$$X_q := \big\{\Gamma(1_2, \boldsymbol{v})M \,:\, M \in G',\ \boldsymbol{v} \in \mathbb{Q}^2,\ \mathbf{d}(\boldsymbol{v}) = q\big\},$$

where for any vector $\boldsymbol{v} \in \mathbb{Q}^2$ we write $\mathbf{d}(\boldsymbol{v})$ for its denominator, i.e. the smallest positive integer d such that $\boldsymbol{v} \in d^{-1}\mathbb{Z}^2$. Then X_q is a closed embedded 3-dimensional submanifold of X; this is an easy consequence of the fact that $\{\boldsymbol{v} \in \mathbb{Q}^2 \,:\, \mathbf{d}(\boldsymbol{v}) = q\}$ is an invariant subset for the action of Γ on $\mathbb{R}^2$. Note in particular that $X_1 = X'$. Now if $\boldsymbol{\xi} \in \mathbb{Q}^2$ then $\Gamma\,(1_2, \boldsymbol{\xi})\, U^{\mathbb{R}} \Phi^t \subset X_{\mathbf{d}(\boldsymbol{\xi})}$ holds for every t; hence the orbit certainly cannot become equidistributed in (X, μ).

The map $G' \ni M \mapsto \Gamma\big(1_2, (0, q^{-1})\big)M \in X_q$ gives an identification of X_q with the homogeneous space $\Gamma_1(q)\backslash G'$, where $\Gamma_1(q)$ is the congruence subgroup

$$\Gamma_1(q) = \big\{\big(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\big) \in \Gamma' \,:\, a \equiv d \equiv 1 \ (\mathrm{mod}\ q),\ c \equiv 0 \ (\mathrm{mod}\ q)\big\}.$$

(To see this, note that Γ acts transitively on $\{\boldsymbol{v} \in \mathbb{Q}^2 \,:\, \mathbf{d}(\boldsymbol{v}) = q\}$.) If $\boldsymbol{\xi} \in \mathbb{Q}^2$ with $\mathbf{d}(\boldsymbol{\xi}) = q$ then the curves studied in Theorem 1.1 correspond to pieces of closed horocycles in $\Gamma_1(q)\backslash G'$, and hence as $t \to \infty$ they go asymptotically equidistributed *in* X_q, i.e. in place of (2) we have

$$\lim_{t\to\infty} \frac{1}{\beta-\alpha} \int_\alpha^\beta f\Big(\Gamma\,(1_2, \boldsymbol{\xi})\, U^x \Phi^t\Big)\, dx = \int_{X_q} f\, d\mu_q,$$

where μ_q is the measure which corresponds to Haar measure on G', normalized to give a probability measure on $\Gamma_1(q)\backslash G'$ (cf., e.g., [11]).

The main result of the present paper is Theorem 1.2 below, which is an *effective* version of Theorem 1.1. It is clear from the preceding discussion that the rate of convergence in (2) is necessarily quite sensitive to the Diophantine properties of the vector $\boldsymbol{\xi}$.

One should note that the flow $\{\Phi^t\}$ on X is Anosov, with unstable directions generated by the flows $U^{\mathbb{R}}$ and $(1_2, (0, \mathbb{R}))$ and stable directions generated by the flows $\big(\big(\begin{smallmatrix} 1 & 0 \\ \mathbb{R} & 1 \end{smallmatrix}\big), (0,0)\big)$ and $(1_2, (\mathbb{R}, 0))$. In fact, for any fixed metric on X coming from a left invariant Riemannian metric on G, the tangent vectors in the direction of $U^{\mathbb{R}}$ are expanded at a rate e^t by the flow Φ^t (cf. (1)), the tangent vectors in the direction of $(1_2, (0, \mathbb{R}))$ are expanded at a rate $e^{t/2}$, while vectors in the direction of $\big(\big(\begin{smallmatrix} 1 & 0 \\ \mathbb{R} & 1 \end{smallmatrix}\big), (0,0)\big)$ are contracted at a rate e^{-t} and those in the direction of $(1_2, (\mathbb{R}, 0))$ are contracted at a rate $e^{-t/2}$. If, in place of *1-dimensional* averages along $U^{\mathbb{R}}$-orbits, we would instead consider *2-dimensional* averages taken over some bounded open subset of the unstable manifold, then there exists a by now standard approach to establishing effective results by using mixing properties of the flow $\Phi^{\mathbb{R}}$; the origin of this technique can be traced back to the thesis of Margulis, [27], where it was used in the context of general Anosov flows. However, it seems that this technique cannot be carried over to the 1-dimensional averages which we consider; instead our proof relies on Fourier analysis and methods from number theory, in particular Weil's bound on Kloosterman sums.

We now state Theorem 1.2. Let $\mathrm{C}_{\mathrm{b}}^k(X)$ be the space of k times continuously differentiable functions on X whose all left invariant derivatives up to order k are bounded. Choose, once and for all, a norm $\|\cdot\|_{\mathrm{C}_{\mathrm{b}}^k}$ on $\mathrm{C}_{\mathrm{b}}^k(X)$ involving the supremum norms of all these derivatives. (For definiteness, we fix a precise choice of $\|\cdot\|_{\mathrm{C}_{\mathrm{b}}^k}$; cf. (11) below.) Set

$$a(y) = \Phi^{-\log y} = \begin{pmatrix} \sqrt{y} & 0 \\ 0 & 1/\sqrt{y} \end{pmatrix} \qquad \text{for } y > 0.$$

(As a motivation, note that $U^x a(y)(i) = x + iy$, for the standard action of G' on the Poincaré upper half plane model of the hyperbolic plane.) For $x \in \mathbb{R}$ we write $\langle x \rangle$ for the distance to the nearest integer; $\langle x\rangle = \min_{n\in\mathbb{Z}} |x-n|$. For any $\boldsymbol{\xi} = (\xi_1, \xi_2) \in \mathbb{R}^2$, $L > 0$ and $y > 0$ we set

$$b_{\boldsymbol{\xi},L}(y) := \max_{q\in\mathbb{Z}^+} \min\Big(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\langle q\xi_2\rangle}\Big). \tag{3}$$

(Convention: if $\langle q\xi_1\rangle = 0$ or $\langle q\xi_2\rangle = 0$ then the corresponding entry is removed from the minimum; in particular if both $\langle q\xi_1\rangle = \langle q\xi_2\rangle = 0$ then the minimum equals $1/q^2$.) Note that the entry $1/q^2$ ensures that the maximum is attained, and $0 < b_{\boldsymbol{\xi},L}(y) \le 1$ for all $y, \boldsymbol{\xi}, L$; furthermore, $b_{\boldsymbol{\xi},L}(y)$ depends continuously on $y, \boldsymbol{\xi}, L$.

Theorem 1.2. *Given any $\varepsilon > 0$, there exists a constant $C > 0$ such that, for any $f \in \mathrm{C}^8_{\mathrm{b}}(X)$ and any $\alpha < \beta$, $\boldsymbol{\xi} \in \mathbb{R}^2$ and $0 < y < 1$,*

$$\left|\frac{1}{\beta-\alpha}\int_\alpha^\beta f\Big(\Gamma\,(1_2,\boldsymbol{\xi})\,U^x a(y)\Big)\,dx - \int_{\Gamma\backslash G} f\,d\mu\right| \le C\|f\|_{\mathrm{C}^8_{\mathrm{b}}} \frac{L}{\beta-\alpha}\big(b_{\boldsymbol{\xi},L}(y) + y^{\frac14}\big)^{1-\varepsilon}, \tag{4}$$

where $L = \max(1, |\alpha|, |\beta|)$.

Let us make some comments on this result. First of all, note that for any fixed $\boldsymbol{\xi} \in \mathbb{R}^2$ and $L > 0$, we have $\lim_{y\to 0} b_{\boldsymbol{\xi},L}(y) = 0$ if (and only if) $\boldsymbol{\xi} \notin \mathbb{Q}^2$. Hence Theorem 1.2 is indeed an effective version of Theorem 1.1.

In order to discuss the rate of decay of our bound as $y \to 0$, we recall the following definition: We say that a vector $\boldsymbol{\xi} \in \mathbb{R}^2$ is of *(Diophantine) type* K if there is some constant $c > 0$ such that $\|\boldsymbol{\xi} - q^{-1}\boldsymbol{m}\| > cq^{-K}$ for all $q \in \mathbb{Z}^+$ and $\boldsymbol{m} \in \mathbb{Z}^2$. The smallest possible value for K is $K = \frac32$, and it is known that Lebesgue-almost all $\boldsymbol{\xi} \in \mathbb{R}^2$ are of type $K = \frac32 + \varepsilon$ for any $\varepsilon > 0$. In fact, by a result of Jarnik [22], for any $K \ge \frac32$, the set of those $\boldsymbol{\xi} \in \mathbb{R}^2$ which are not of type K has Hausdorff dimension $3/K$. Now from the definition (3) one easily verifies that, for any fixed $\boldsymbol{\xi}$ and L and any given $\delta > 0$, we have $b_{\boldsymbol{\xi},L}(y) \ll y^\delta$ as $y \to 0$ if and only if $\delta \le \frac12$ and $\boldsymbol{\xi}$ is of type $K = \delta^{-1}$. Hence we get:

Corollary 1.3. *For any $\varepsilon > 0$, $f \in \mathrm{C}^8_{\mathrm{b}}(X)$, $\alpha < \beta$ and any $\boldsymbol{\xi} \in \mathbb{R}^2$ of Diophantine type $K \ge \frac32$, there is a constant $C = C(\varepsilon, f, \alpha, \beta, \boldsymbol{\xi}) > 0$ such that*

$$\left|\frac{1}{\beta-\alpha}\int_\alpha^\beta f\Big(\Gamma\,(1_2,\boldsymbol{\xi})\,U^x a(y)\Big)\,dx - \int_{\Gamma\backslash G} f\,d\mu\right| < Cy^{\min(\frac14,\frac1K)-\varepsilon}, \qquad \forall 0 < y < 1. \tag{5}$$

In particular, in view of Jarnik's result, we obtain the rate $y^{\frac14-\varepsilon}$ for any fixed $\boldsymbol{\xi} \in \mathbb{R}^2$ away from a set of Hausdorff dimension $\frac34$. It seems that the exponent $\frac14$ in (5) is not the best possible, and that optimally one might hope to prove that the left hand side of (4) decays with a rate $y^{\frac12-\varepsilon}$ as $y \to 0$, for any $\boldsymbol{\xi}$ satisfying an appropriate Diophantine condition; cf. Remark 8.1 below.

Regarding the dependence of our bound on α, β, we remark that we could have chosen to state Theorem 1.2 with the extra restriction $-1 \le \alpha < \beta \le 1$ (viz., $L = 1$); the general case can be deduced aposteriori from that case by using invariance under $U^n \in \Gamma$, $n \in \mathbb{Z}$, and splitting $[\alpha, \beta]$ into subintervals of length ≤ 1; this will be seen in Section 10 where we discuss basic properties of the majorant function $b_{\boldsymbol{\xi},L}(y)$. We have not given any special attention to the case of $\beta - \alpha$ becoming small in our proof of Theorem 1.2, and there seems to be room for improvement in this direction. (Cf. [46], where the case of both $\beta - \alpha$ and y being small is considered for the case of pieces of closed horocycles in X' and other homogeneous spaces of $\mathrm{SL}(2,\mathbb{R})$.) Also we have made no effort to optimize the dependence on f in Theorem 1.2.

A point to note is that the orbit $\Gamma(1_2,\boldsymbol{\xi})U^{\mathbb{R}}$ is *closed* in X if and only if $\xi_1 \in \mathbb{Q}$, and in this case its period equals the denominator of ξ_1; a corresponding fact also holds for any Φ^t-push-forward of that orbit. This is to some extent reflected in the bound (4): for fixed $y, \boldsymbol{\xi}$, we have $\lim_{L\to\infty} b_{\boldsymbol{\xi},L}(y) = 0$ if and only if $\xi_1 \notin \mathbb{Q}$.

1.2. **General orbits.** We now turn to the case of arbitrary $U^{\mathbb{R}}$-orbits. According to Ratner's equidistribution theorem [39], *every* $U^{\mathbb{R}}$-orbit in X has a closure which is homogeneous. Stated in more detail, for any given $x = \Gamma g \in X$ ($g \in G$) there exists a closed connected subgroup

$H \subset G$ such that $U^{\mathbb{R}} \subset H$, $\Gamma \cap gHg^{-1}$ is a lattice in gHg^{-1}, and the closure of $xU^{\mathbb{R}}$ in X equals $xH = \Gamma\backslash\Gamma gH$. Furthermore the orbit $xU^{\mathbb{R}}$ is then asymptotically equidistributed in xH with respect to ν_H, the H-invariant Borel probability measure on X supported on xH [39, Thm. B].

For our specific space X it is fairly easy to list explicitly those subgroups H which can occur, and in particular to give a precise criterion for when $xU^{\mathbb{R}} = \Gamma gU^{\mathbb{R}}$ is asymptotically equidistributed in (X, μ). Clearly a necessary condition for the latter is that the projected orbit $D(xU^{\mathbb{R}})$ should be equidistributed in X'. By a theorem of Dani [3] (a very special case of Ratner's [39]), $D(xU^{\mathbb{R}})$ is equidistributed in X' unless $D(xU^{\mathbb{R}})$ is a closed horocycle, viz., unless the lattice $\mathbb{Z}^2 D(g)$ contains some point along the line $(0,\mathbb{R}) := \{0\} \times \mathbb{R}$ other than the origin. Assuming that $D(xU^{\mathbb{R}})$ is equidistributed in X', one finds (cf. the discussion in [8, §2.6] applied to the measure ν_H; see in particular [8, Cor. 2.11 and Cor. 2.12, corrected]) that either $H = G$ and $xH = X$, or else there is some $\beta \in \mathbb{R}$ such that $(0,\beta)g^{-1} \in \mathbb{Q}^2$, and then $H = \big(1_2, -(0,\beta)\big)G'\big(1_2,(0,\beta)\big)$ and $xH = X_q\big(1_2,(0,\beta)\big)$, where $q = \mathbf{d}\big((0,\beta)g^{-1}\big)$. (For clarity, note that in the second case, β is uniquely determined. Indeed, if the point set $\mathbb{Q}^2 g$ intersects the line $(0,\mathbb{R})$ in more than one point then by subtraction $\mathbb{Q}^2 D(g)$ contains a non-zero point on $(0,\mathbb{R})$; hence so does the lattice $\mathbb{Z}^2 D(g)$, contradicting our assumption that $D(xU^{\mathbb{R}})$ is equidistributed in X'.)

In particular we have:

Theorem 1.4. *(Special case of Ratner,* [39]*.) Fix any $g \in G$ satisfying $\mathbb{Z}^2 D(g) \cap (0,\mathbb{R}) = \{\mathbf{0}\}$ and $(0,\beta)g^{-1} \notin \mathbb{Q}^2$ for all $\beta \in \mathbb{R}$. Then the orbit $\Gamma gU^{\mathbb{R}}$ is asymptotically equidistributed in (X,μ). In other words, for any bounded continuous function f on X, $\frac{1}{T}\int_0^T f(\Gamma gU^t)\,dt \to \int_X f\,d\mu$ as $T \to \infty$.*

As an application of our main result, Theorem 1.2, and using the technique of approximating nonclosed horocycles by pieces of closed horocycles (cf. [41]), we will prove an effective version of Theorem 1.4; see Theorem 1.6 below. Before stating it, it is useful to recall the effective equidistribution result for horocycles in X' proved in [47] (viz., an effective version of Dani's theorem [3]); cf. also [2], [11], [41]. For $g \in G'$ we write $\boldsymbol{\ell}(g) > 0$ for the Euclidean length of the shortest non-zero vector in the lattice $\mathbb{Z}^2 g$. Note that $\boldsymbol{\ell}(\gamma g) = \boldsymbol{\ell}(g)$ for all $\gamma \in \Gamma'$, i.e. $\boldsymbol{\ell}$ is a function on X'; in fact $\boldsymbol{\ell}(g)$ equals the inverse square root of the invariant height function $\mathcal{Y}_{\Gamma'}(g)$ used in [47]. More generally for $g \in G$ we set $\boldsymbol{\ell}(g) = \boldsymbol{\ell}(D(g))$. Finally for $g \in G$ and $T > 0$ we set

$$y_g(T) := T^{-1}\boldsymbol{\ell}(g\,a(T))^{-2}. \tag{6}$$

Theorem 1.5. *(*[47, Thm. 1]*; cf. also* [41]*) There exists an absolute constant $C > 0$ such that, for any $g \in G'$, $T \geq 1$, and any $f \in \mathrm{C}_{\mathrm{b}}^4(X')$:*

$$\left|\frac{1}{T}\int_0^T f\big(\Gamma' gU^t\big)\,dt - \int_{X'} f\,d\mu'\right| \leq C\|f\|_{\mathrm{C}_{\mathrm{b}}^4}\, y_g(T)^{\frac{1}{2}} \log^3(2 + y_g(T)^{-1}). \tag{7}$$

Note that for given $g \in G'$, $\lim_{T\to\infty} y_g(T) = 0$ holds if and only if the horocycle $\Gamma' gU^{\mathbb{R}}$ is not closed; hence Theorem 1.5 is indeed an effective version of Dani's equidistribution result. For given $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G'$, the rate of decay of $y_g(T)$ as $T \to \infty$ is directly related to the Diophantine properties of the number $\frac{a}{c}$ (assuming $c \neq 0$): If $\frac{a}{c}$ is of Diophantine type $K \geq 2$ (viz., $\inf_{q\in\mathbb{Z}^+} q^{K-1}\langle q\frac{a}{c}\rangle > 0$), then there is $C = C(g,K) > 0$ such that $y_g(T) \leq CT^{-2/K}$ for all $T \geq 1$. In particular, for (Haar-)almost all $g \in G'$, the right hand side of (7) decays more rapidly than $T^{\varepsilon-\frac{1}{2}}$ as $T \to \infty$ ($\forall \varepsilon > 0$). The rate of decay of the right hand side in (7) is in fact essentially optimal, for any given $g \in G'$; cf. [47, Thm. 2 and §§4-5]. We also remark that [47, Thm. 1] is more general in that it holds for an arbitrary cofinite subgroup of $\mathrm{PSL}(2,\mathbb{R})$ in place of Γ' (the bound then depends on the small eigenvalues of the Laplace-Beltrami operator on the corresponding hyperbolic surface); also the bound holds with a weaker function space norm than the $\|\cdot\|_{\mathrm{C}_{\mathrm{b}}^4}$ used above.

We are now ready to state our effective version of Theorem 1.4. For $T > 0$, let $\mathfrak{R}_T$ be the closed rectangle $\mathfrak{R}_T := [-T^{-1}, T^{-1}] \times [-1, 1] \subset \mathbb{R}^2$. We also use the shorthand notation $\mathbb{Z}^+_{\leq a} = (0, a] \cap \mathbb{Z}$. Set, for $g \in G$ and $T > 0$,

$$b_g(T) = \inf\Big\{\delta > 0 \;:\; \Big[\forall q \in \mathbb{Z}^+_{\leq \delta^{-1/2}} : \; (q^{-1}\mathbb{Z}^2)g \cap \frac{1}{\delta q^2}\mathfrak{R}_T = \emptyset\Big]\Big\}. \tag{8}$$

(This can be viewed as a generalization of the notation $b_{\boldsymbol{\xi},L}(y)$ introduced previously; cf. equation (100) on p. 35.)

Theorem 1.6. *Given any $\varepsilon > 0$, there exists a constant $C > 0$ such that, for any $g \in G$, $T \geq 2$ and $f \in \mathrm{C}^8_{\mathrm{b}}(\Gamma\backslash G)$, we have*

$$\left|\frac{1}{T}\int_0^T f(\Gamma g U^t)\,dt - \int_{\Gamma\backslash G} f\,d\mu\right| \leq C\|f\|_{\mathrm{C}^8_{\mathrm{b}}}\big(y_g(T)^{\frac{1}{4}} + b_g(T)\big)^{\frac{1}{2}-\varepsilon}. \tag{9}$$

Note that for any given $g \in G$ we have $\lim_{T\to\infty}(y_g(T)^{\frac{1}{4}} + b_g(T)) = 0$ if (and only if) $D(\Gamma g U^{\mathbb{R}})$ is not a closed horocycle in X' and $\mathbb{Q}^2 g \cap (0, \mathbb{R}) = \emptyset$, viz. $(0, \beta)g^{-1} \notin \mathbb{Q}^2$ for all $\beta \in \mathbb{R}$. Hence Theorem 1.6 is indeed an effective version of Theorem 1.4. We will also see that for μ-almost all $g \in G$, we have $\lim_{T\to\infty} b_g(T)T^{\delta} = 0$ for all $\delta < \frac{1}{2}$ (cf. Proposition 11.4); hence, recalling the earlier discussion about $y_g(T)$, we see that for μ-almost all $g \in G$, the right hand side in (9) decays more rapidly than $T^{\varepsilon-\frac{1}{8}}$ as $T \to \infty$ ($\forall \varepsilon > 0$). As we discuss in Remark 11.1 below, optimally one might hope to improve Theorem 1.6 so as to yield a rate of decay $T^{\varepsilon-\frac{1}{2}}$ for any g satisfying appropriate Diophantine conditions.

1.3. **Applications and extensions.** As we have mentioned, cases of Ratner equidistribution in settings closely related to that of the present paper have played a crucial role in the solution of several problems in number theory and in mathematical physics. We discuss some of these here.

In [30], [29], Marklof proved that the limit local pair correlation density of the sequence $\|\boldsymbol{m} - \boldsymbol{\alpha}\|^k$, $\boldsymbol{m} \in \mathbb{Z}^k$ ($k \geq 2$) is that of a Poisson process, under Diophantine conditions on the fixed vector $\boldsymbol{\alpha} \in \mathbb{R}^k$. In particular for $k = 2$ this gives a quantitative Oppenheim type statement for the inhomogeneous quadratic form $(x_1 - \alpha)^2 + (x_2 - \beta)^2 - (x_3 - \alpha)^2 - (x_4 - \beta)^2$. The proof makes use of an analogue of Theorem 1.1 for $G = \mathrm{SL}(2, \mathbb{R}) \ltimes (\mathbb{R}^2)^{\oplus k}$ and Γ a congruence subgroup of $\mathrm{SL}(2, \mathbb{Z}) \ltimes (\mathbb{Z}^2)^{\oplus k}$. In joint work with Pankaj Vishe, [49], we generalize the methods of the present paper to that case, and apply this to obtain an effective rate of convergence for the pair correlation density of $\|\boldsymbol{m} - \boldsymbol{\alpha}\|^k$.

In particular it is noted in [49] that the methods of the present paper can without serious difficulty be extended to the case of Γ being an arbitrary congruence subgroup of $\mathrm{SL}(2, \mathbb{Z}) \ltimes \mathbb{Z}^2$. However, already in a case such as $\Gamma = \widetilde{\Gamma}' \ltimes \mathbb{Z}^2$, with $\widetilde{\Gamma}'$ a noncongruence subgroup of finite index of $\mathrm{SL}(2, \mathbb{Z})$, new ideas would be needed to extend the results of the present paper. (We remark that *every* lattice Γ in $G = \mathrm{SL}(2, \mathbb{R}) \ltimes \mathbb{R}^2$ can be conjugated within $\mathrm{GL}(2, \mathbb{R}) \ltimes \mathbb{R}^2$ into a position where $D(\Gamma)$ is a finite index subgroup of $\mathrm{SL}(2, \mathbb{Z})$ and $(\{1_2\} \ltimes \mathbb{R}^2) \cap \Gamma = \{1_2\} \ltimes \mathbb{Z}^2$; cf. [36, Cor. 8.28]. However it is not always possible to conjugate into a situation where Γ contains $\widetilde{\Gamma}' \ltimes L$ for some subgroup $\widetilde{\Gamma}'$ of finite index in $\mathrm{SL}(2, \mathbb{Z})$ and a lattice $L \subset \mathbb{Z}^2$. Indeed, consider for example the lattice Γ generated by $((\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}), \boldsymbol{v})$, $((\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}), \boldsymbol{v}')$, $(1_2, (1, 0))$, $(1_2, (0, 1))$, for some fixed $\boldsymbol{v}, \boldsymbol{v}' \in \mathbb{R}^2$ such that the first coordinate of $\boldsymbol{v}$ is irrational. Recall in this connection that $(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix})$ and $(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix})$ are free generators of the principal congruence subgroup $\Gamma(2)$ in $\mathrm{SL}(2, \mathbb{Z})$.)

Quantitative Oppenheim type results for more general inhomogeneous quadratic forms have recently been obtained by Margulis and Mohammadi [28], using a method different from Marklof's. For the special case of forms of signature (2,1) whose homogeneous part is a split rational form (see [28, Thm. 1.10]), the proof depends on equidistribution of unipotent orbits in homogeneous spaces of the group $\mathrm{SL}(2, \mathbb{R}) \ltimes \mathrm{Sym}_2(\mathbb{R})$. It seems that it should be possible to extend the methods of the present paper to these homogeneous spaces, and also to more general groups of the form $\mathrm{SL}(2,\mathbb{R}) \ltimes V$ where V is the vector space of a finite dimensional linear representation of $\mathrm{SL}(2,\mathbb{R})$.

Elkies and McMullen [8] have shown that the gaps between the fractional parts of $\sqrt{n}$ for $n = 1,\ldots,N$, have a limit distribution as N tends to infinity, and they compute this limit distribution explicitly. In a recent paper, El-Baz, Marklof and Vinogradov [7] also prove convergence of the local pair-correlation and more general mixed moments. The proofs make crucial use of an analogue of Theorem 1.1 for the flow $U_1^{\mathbb{R}}$, with $U_1^x = \big(\big(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\big), -(x/2, x^2/4)\big)$. Since $U_1^{\mathbb{R}}$ is not conjugate to $U^{\mathbb{R}}$, Theorem 1.2 does not apply to this setting. In fact *any* Ad-unipotent 1-parameter subgroup in G with nontrivial image in G' is conjugate to either $U^{\mathbb{R}}$ or $U_1^{\mathbb{R}}$. Recently, Browning and Vinogradov [1] have extended the methods of the present paper so as to yield an effective equidistribution result for certain orbits of the flow $U_1^{\mathbb{R}}$, and applied this to establish an effective rate for the convergence of the gap distribution of $\sqrt{n} \bmod 1$. (Note also that Sinai [45] has proposed an alternative approach to the statistics of $\sqrt{n} \bmod 1$.)

Another application concerns the local statistics of directions to lattice points: Consider a fixed lattice translate L in $\mathbb{R}^2$ and record the directions of all lattice vectors of length at most T. In joint work with Marklof we proved in [32, Thm. 1.3; see also Thm. 2.1] that the distribution of gaps between the lattice directions has a limit as T tends to infinity; see also El-Baz, Marklof and Vinogradov [6] regarding convergence of the local pair-correlation and more general mixed moments. Assuming that L is an 'irrational' translate, the limit distribution is universal and in fact coincides with the limiting gap distribution for $\sqrt{n} \bmod 1$ found by Elkies and McMullen. The proofs of these facts make use of equidistribution of expanding translates of $\mathrm{SO}(2)$-orbits in the same space $X = \Gamma\backslash G$ as we consider here. By a standard approximation argument this is reduced to the equidistribution of pieces of $U^{\mathbb{R}}$-orbits (cf. the proof of Cor. 5.4 in [32]), and thus using our Theorem 1.6 it should be possible to prove an effective rate of convergence in [32, Thm. 1.3], for 'irrational' lattice translates. However several technicalities remain to be worked out to carry this through.

As a final example, in [48, Remark 4 ($n = 2$)] it is noted that the number of values modulo one of a random linear form ωn for $n = 1,\ldots,N$ which fall inside a given small interval of length c/N centered at a fixed irrational point $\xi \in \mathbb{R}/\mathbb{Z}$, has a limit distribution as $N \to \infty$, which is independent of ξ. The proof is an application of Theorem 1.1 in the special case $\boldsymbol{\xi} = (0,\xi)$, and thus using our Theorem 1.2 it would be possible to prove an effective rate for the convergence to the limit distribution, depending on the Diophantine properties of ξ.

We hope to return to several of the above-mentioned questions in later work.

1.4. **Outline of the paper.** Sections 2–5 lay down the setup of our approach: In Section 2 we set some basic notation; in Section 3 we smooth the (α,β)-integral appearing in Theorem 1.2; in Section 4 we discuss the Fourier decomposition of the given test function on $X = \Gamma\backslash G$ with respect to the torus fiber variable; and in Section 5 we handle the contribution from the zeroth Fourier term; this reduces to a known result on the effective equidistribution of horocycle orbits in X'.

The basic idea of our approach appears in Sections 6–7; we first rewrite the remaining terms of the Fourier decomposition in an appropriate format, and then prove a lemma (Lemma 7.1) which can be used to establish cancellation in the sum; this lemma is nothing but a standard application of the classical Weil's bound on Kloosterman sums.

The proof of Theorem 1.2 is given in Sections 8–9: In Section 8 we carry out those steps which utilize only the irrationality properties of ξ_1 and not those of ξ_2; the outcome is a weaker version of the theorem, Proposition 8.3, which is strong enough to imply the equidistribution in Theorem 1.1 whenever ξ_1 is irrational, with the error bound decaying as a power of y whenever ξ_1 is of Diophantine type; however for ξ_1 rational it does not imply any equidistribution whatsoever. To complete the proof of Theorem 1.2, in Section 9 (the longest section of the paper), we consider more carefully those terms in the Fourier decomposition which give the largest contribution in the treatment of Section 8; these correspond to good rational approximations of ξ_1; we collect these terms in a way which allows us to utilize also the irrationality properties of ξ_2 to establish cancellation. The error bound which we finally arrive at in Theorem 1.2 incorporates the Diophantine properties of both ξ_1 and ξ_2, the bound being far from zero only if ξ_1 and ξ_2 are well approximable by rational numbers with a *common* small denominator q; cf. the definition of the error majorant $b_{\boldsymbol{\xi},L}(y)$ in (3).

The precise format of this bound plays a crucial role when we apply Theorem 1.2 to deduce the effective equidistribution of general $U^{\mathbb{R}}$-orbits, Theorem 1.6. To illustrate this point, note that to establish a result which could be called "an effective version of Theorem 1.1", it would suffice to complement Proposition 8.3 with an effective equidistribution result for ξ_1 *rational* and ξ_2 irrational. This would be quite a bit easier than what we do in Section 9; however it would not be sufficient for our goal of deriving a satisfactory effective equidistribution for general $U^{\mathbb{R}}$-orbits, basically since our proof of Theorem 1.6 for a given $g = (1_2, \boldsymbol{\xi})M$ generally involves applying Theorem 1.2 with $\boldsymbol{\xi}\gamma$ in place of $\boldsymbol{\xi}$, where γ varies through more and more elements of Γ' as $T \to \infty$.

In Section 10 we establish some important basic properties of the error majorant $b_{\boldsymbol{\xi},L}(y)$. Finally in Section 11 we prove Theorem 1.6, by approximating the given $U^{\mathbb{R}}$-orbit by one or several lifts of pieces of closed horocycles in X' and applying Theorem 1.2 to each of these.

1.5. **Acknowledgments.** I am grateful to Livio Flaminio, Giovanni Forni, Han Li, Jens Marklof, Amir Mohammadi, Hee Oh, Wolfgang Staubach, Akshay Venkatesh and Pankaj Vishe for helpful and inspiring discussions. I would also like to thank the referees for their valuable comments; in particular Remark 6.2 below is based on a suggestion by one of the referees.

2. Some notation

We shall use the standard notation $A = O(B)$ or $A \ll B$ meaning $|A| \leq CB$ for some constant $C > 0$. We shall also write $A \asymp B$ as a substitute for $A \ll B \ll A$. To indicate that the implicit constant C may depend on some quantities or functions f, g, h we will use the notation $A \ll_{f,g,h} B$ or $A = O_{f,g,h}(B)$. The constant C will not depend on any other variable, except in a statement that contains an implication of the kind "if $A_1 = O(B_1)$ then $A_2 = O(B_2)$"; in that case the constant implicit in $O(B_2)$ may also depend on the one in $O(B_1)$. (We will use the last convention only in Remarks 10.1 and 10.2.)

Recall from Section 1 that $G = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$, $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2$, $G' = \mathrm{SL}(2,\mathbb{R})$ and $\Gamma' = \mathrm{SL}(2,\mathbb{Z})$. We will also write $\Gamma'_\infty := \{\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) : x \in \mathbb{Z}\}$.

Let $\mathfrak{g}$ be the Lie algebra of G. We may identify $\mathfrak{g}$ in a natural way with the space $\mathfrak{sl}(2,\mathbb{R}) \oplus \mathbb{R}^2$, with Lie bracket $[(X, \boldsymbol{v}), (Y, \boldsymbol{w})] = (XY - YX, \boldsymbol{v}Y - \boldsymbol{w}X)$ (cf., e.g., [23, Prop. 1.124]). Using this notation, we fix the following basis of $\mathfrak{g}$:

(10)
$$X_1 = \left(\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right), \mathbf{0}\right); \quad X_2 = \left(\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right), \mathbf{0}\right); \quad X_3 = \left(\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right), \mathbf{0}\right); \quad X_4 = \big(0_2, (1,0)\big); \quad X_5 = \big(0_2, (0,1)\big).$$

To each $Y \in \mathfrak{g}$ corresponds a left invariant differential operator on functions on G, and thus also a differential operator on $\Gamma\backslash G$, which we will also denote by Y. We let $\mathrm{C}_\mathrm{b}^k(\Gamma\backslash G)$ be the space of k times continuously differentiable functions on $\Gamma\backslash G$ such that $\|Df\|_{\mathrm{L}^\infty} < \infty$ for every left invariant differential operator D on G of order $\leq k$. For $f \in \mathrm{C}_\mathrm{b}^k(\Gamma\backslash G)$ we set

$$\|f\|_{\mathrm{C}_\mathrm{b}^k} := \sum_{\mathrm{ord}(D) \leq k} \|Df\|_{\mathrm{L}^\infty}, \tag{11}$$

the sum being over all monomials in $X_1, \ldots, X_5$ of degree $\leq k$. Note in particular that $\mathrm{C}_\mathrm{b}^0(\Gamma\backslash G)$ is the space of bounded continuous functions on $\Gamma\backslash G$, and $\|\cdot\|_{\mathrm{C}_\mathrm{b}^0}$ is the supremum norm.

We will also have occasion to use Sobolev L^1-norms on functions on $\mathbb{R}$: For $1 \leq p < \infty$, k a positive integer and $\nu \in \mathrm{C}^k(\mathbb{R})$ we set

$$\|\nu\|_{\mathrm{W}^{k,p}} = \sum_{j=0}^{k} \|\nu\|_{\mathrm{L}^p} = \sum_{j=0}^{k} \Big(\int_{\mathbb{R}} |\nu^{(j)}(x)|^p \, dx\Big)^{1/p}.$$

We will only use these for $p = 1$.

We will use the standard notation $e(x) = e^{2\pi i x}$. We write $\gcd(c,d)$, or just (c,d), for the greatest common divisor of two integers c, d. For n a positive integer, we write $\sigma(n)$ for the number of (positive) divisors of n, and $\sigma_1(n)$ for their sum: $\sigma(n) = \sum_{d|n} 1$ and $\sigma_1(n) = \sum_{d|n} d$.

3. Smoothed ergodic averages

As a first step in our proof of Theorem 1.2 we replace the sharp cutoff "$\int_\alpha^\beta$" by a compactly supported cutoff function $\nu(x)$ satisfying a mild regularity assumption. Basically we need control on the L^1-norm of "$1+\varepsilon$" derivatives of ν; in order to avoid a technical overhead we formulate the bound using a crude interpolation between the Sobolev norms $\|\nu\|_{\mathrm{W}^{1,1}}$ and $\|\nu\|_{\mathrm{W}^{2,1}}$ (cf., e.g., [48, Sec. 2]). We will prove the following theorem.

Theorem 3.1. *Let $0 < \eta < 1$ and $\varepsilon > 0$ be fixed. Then for any $f \in \mathrm{C}_\mathrm{b}^8(\Gamma\backslash G)$, any $\nu \in \mathrm{C}^2(\mathbb{R})$ with compact support, and any $\boldsymbol{\xi} \in \mathbb{R}^2$, $0 < y < 1$,*

$$\int_{\mathbb{R}} f\Big(\Gamma\,(1_2, \boldsymbol{\xi})\, U^x a(y)\Big)\, \nu(x)\, dx = \int_{\Gamma\backslash G} f\, d\mu \int_{\mathbb{R}} \nu\, dx$$
$$+O_{\eta,\varepsilon}\bigg\{\|f\|_{\mathrm{C}_\mathrm{b}^8} \|\nu\|_{\mathrm{W}^{1,1}}^{1-\eta} \|\nu\|_{\mathrm{W}^{2,1}}^{\eta}\, y^{\frac14}\log(1+y^{-1}) + \|f\|_{\mathrm{C}_\mathrm{b}^4} L \|\nu\|_{\mathrm{L}^\infty}\big(b_{\boldsymbol{\xi},L}(y) + y^{\frac14}\big)^{1-\varepsilon}\bigg\}, \tag{12}$$

where L is the smallest real number ≥ 1 such that $\mathrm{supp}(\nu) \subset [-L, L]$.

Proof that Theorem 3.1 implies Theorem 1.2. This is a standard approximation argument. Fix $g \in \mathrm{C}_c^\infty(\mathbb{R})$ satisfying $g \geq 0$, $\int_{\mathbb{R}} g = 1$ and $\mathrm{supp}(g) \subset [-1,1]$. Set $g_\delta(x) = \delta^{-1} g(\delta^{-1}x)$ for $0 < \delta \leq 1$; then $\mathrm{supp}(g_\delta) \subset [-\delta, \delta]$ and $\int_{\mathbb{R}} g_\delta = 1$. Let $\alpha < \beta$ be given, and set $L' = \beta - \alpha$. We apply Theorem 3.1 with $\nu = \chi_{[\alpha,\beta]} * g_\delta$. Then $\|\nu\|_{\mathrm{W}^{1,1}} \ll L' + 1$ and $\|\nu\|_{\mathrm{W}^{2,1}} \ll L' + \delta^{-1}$; thus $\|\nu\|_{\mathrm{W}^{1,1}}^{1-\eta}\|\nu\|_{\mathrm{W}^{2,1}}^{\eta} \ll (L'+1)\delta^{-\eta}$, and so the error term in Theorem 3.1 is

$$O_{\eta,\varepsilon}\bigg\{\|f\|_{\mathrm{C}_\mathrm{b}^8}(L'+1)\delta^{-\eta}\, y^{\frac14(1-\varepsilon)} + \|f\|_{\mathrm{C}_\mathrm{b}^4} L\big(b_{\boldsymbol{\xi},L}(y) + y^{\frac14}\big)^{1-\varepsilon}\bigg\},$$

with $L = \max(1, |\alpha|+\delta, |\beta|+\delta)$. Furthermore, using $0 \leq \nu \leq 1$ and $\nu(x) = \chi_{[\alpha,\beta]}(x)$ whenever $|x-\alpha| \geq \delta$ and $|x - \beta| \geq \delta$, we see that the difference between the left hand side of (12) and $\int_\alpha^\beta f(\Gamma\,(1_2,\boldsymbol{\xi})\,U^x a(y))\,dx$ is $\ll \|f\|_{\mathrm{C}_\mathrm{b}^0}\,\delta$. Hence, choosing $\delta = y^{\frac14}$, we obtain

$$\int_\alpha^\beta f\Big(\Gamma\,(1_2,\boldsymbol{\xi})\,U^x a(y)\Big)\,dx = (\beta-\alpha)\int_{\Gamma\backslash G} f\,d\mu$$
$$+O_{\eta,\varepsilon}\bigg\{\|f\|_{\mathrm{C}_\mathrm{b}^8}(L'+1)y^{\frac14(1-\varepsilon-\eta)} + \|f\|_{\mathrm{C}_\mathrm{b}^4} L\big(b_{\boldsymbol{\xi},L}(y) + y^{\frac14}\big)^{1-\varepsilon}\bigg\}.$$

This implies Theorem 1.2 with $\varepsilon + \eta$ in place of ε (cf. also Lemma 10.1 below). □

Remark 3.1. The proof shows that the bound in Theorem 1.2 may be improved to

$$C\bigg(\|f\|_{\mathrm{C}_\mathrm{b}^8}\Big(1 + \frac{1}{\beta-\alpha}\Big)y^{\frac14(1-\varepsilon)} + \|f\|_{\mathrm{C}_\mathrm{b}^4}\frac{L}{\beta-\alpha}\big(b_{\boldsymbol{\xi},L}(y) + y^{\frac14}\big)^{1-\varepsilon}\bigg).$$

4. Fourier decomposition in the torus variable

We now start with the proof of Theorem 3.1. In this section we consider the Fourier decomposition of the given test function with respect to the torus variable, and prove bounds on the Fourier coefficients appearing in this decomposition.

Assume that $f \in \mathrm{C}_\mathrm{b}^2(\Gamma\backslash G)$. We view f as a function on G which is Γ-left invariant. In particular we have $f((1_2,\boldsymbol{\xi})M) = f((1_2,\boldsymbol{\xi}+\boldsymbol{n})M)$ for all $\boldsymbol{n} \in \mathbb{Z}^2$, and hence for any fixed $M \in G'$, the function $\boldsymbol{\xi} \mapsto f((1_2, \boldsymbol{\xi})M)$ is a C^2-function on the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. Decomposing this function as a Fourier series we have

$$f((1_2,\boldsymbol{\xi})M) = \sum_{\boldsymbol{m}\in\mathbb{Z}^2} \widehat{f}(M,\boldsymbol{m})e(\boldsymbol{m}\cdot\boldsymbol{\xi}), \tag{13}$$

where the Fourier coefficients $\widehat{f}(M,\boldsymbol{m})$ are given by

$$\widehat{f}(M,\boldsymbol{m}) = \int_{\mathbb{T}^2} f((1_2,\boldsymbol{\xi})M)e(-\boldsymbol{m}\cdot\boldsymbol{\xi})\,d\boldsymbol{\xi}. \tag{14}$$

Here $d\boldsymbol{\xi}$ denotes Lebesgue measure on $\mathbb{R}^2$. Note that the sum in (13) is absolutely convergent, uniformly[1] over $(M,\boldsymbol{\xi})$ in any compact subset of G, since $f \in \mathrm{C}^2_{\mathrm{b}}(\Gamma\backslash G)$ implies that the function $\boldsymbol{\xi} \mapsto f(M,\boldsymbol{\xi})$ is in $\mathrm{C}^2_{\mathrm{b}}(\mathbb{T}^2)$, with $\|f(M,\cdot)\|_{\mathrm{C}^2_{\mathrm{b}}(\mathbb{T}^2)}$ depending continuously on $M \in G'$.

Now the fact that f is also Γ'-left invariant leads to an invariance relation for $\widehat{f}(M,\boldsymbol{m})$, which allows us to group together terms in (13) in a convenient way. Let us write $\widehat{\mathbb{Z}}^2$ for the set of primitive lattice points in $\mathbb{Z}^2$, i.e. the set of integer vectors (c,d) with $\gcd(c,d)=1$. Recall that $\Gamma'_\infty := \{\left(\begin{smallmatrix}1 & x\\ 0 & 1\end{smallmatrix}\right) : x \in \mathbb{Z}\}$.

Lemma 4.1. *In the above situation we have*

$$\widehat{f}(TM,\boldsymbol{m}) = \widehat{f}(M,\boldsymbol{m}\,{}^{\mathrm{t}}T^{-1}), \qquad \forall T\in\Gamma',\ M\in G',\ \boldsymbol{m}\in\mathbb{Z}^2. \tag{15}$$

In particular, for each $n \in \mathbb{Z}_{\geq 0}$, the function

$$\widetilde{f}_n(M) := \widehat{f}(M,(n,0)) \qquad (M \in G') \tag{16}$$

is left Γ'_∞-invariant, and $\widetilde{f}_0(M)$ is even left Γ'-invariant. We have

$$f((1_2,\boldsymbol{\xi})M) = \widetilde{f}_0(M) + \sum_{n=1}^{\infty}\sum_{(c,d)\in\widehat{\mathbb{Z}}^2} \widetilde{f}_n\left(\begin{pmatrix} * & * \\ c & d\end{pmatrix}M\right)\cdot e(n(d\xi_1 - c\xi_2)), \qquad \forall (M,\boldsymbol{\xi})\in G, \tag{17}$$

where $\left(\begin{smallmatrix}* & *\\ c & d\end{smallmatrix}\right)$ denotes any matrix in Γ' having lower entries c and d. The sum in (17) is absolutely convergent, uniformly over $(M,\boldsymbol{\xi})$ in any compact subset of G.

(To see that the sum in (17) is well-defined, note that for any $(c,d)\in\widehat{\mathbb{Z}}^2$, the set of matrices $\left(\begin{smallmatrix}* & *\\ c & d\end{smallmatrix}\right) \in \Gamma' = \mathrm{SL}(2,\mathbb{Z})$ is a coset of the form $\Gamma'_\infty\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$, and since $\widetilde{f}_n$ is left Γ'_∞-invariant, $\widetilde{f}_n(TM)$ takes the same value for every matrix T in this coset.)

Proof. For any $T\in\Gamma'$ we have,

$$\begin{aligned}\widehat{f}(TM,\boldsymbol{m}) &= \int_{\mathbb{Z}^2\backslash\mathbb{R}^2} f((1_2,\boldsymbol{\xi})TM)e(-\boldsymbol{m}\cdot\boldsymbol{\xi})\,d\boldsymbol{\xi} = \int_{\mathbb{Z}^2\backslash\mathbb{R}^2} f(T(1_2,\boldsymbol{\xi}T)M)e(-\boldsymbol{m}\cdot\boldsymbol{\xi})\,d\boldsymbol{\xi}\\ &= \int_{\mathbb{Z}^2\backslash\mathbb{R}^2} f(T(1_2,\boldsymbol{\xi})M)e(-\boldsymbol{m}\cdot\boldsymbol{\xi}T^{-1})\,d\boldsymbol{\xi} = \int_{\mathbb{Z}^2\backslash\mathbb{R}^2} f((1_2,\boldsymbol{\xi})M)e(-\boldsymbol{m}\cdot\boldsymbol{\xi}T^{-1})\,d\boldsymbol{\xi},\end{aligned}$$

where in the third identity we used the fact that $\boldsymbol{\xi}\mapsto\boldsymbol{\xi}T$ is a diffeomorphism of $\mathbb{Z}^2\backslash\mathbb{R}^2$ preserving the area measure $d\boldsymbol{\xi}$, and in the last identity we used the fact that f is left Γ-invariant. Using now $\boldsymbol{m}\cdot\boldsymbol{\xi}T^{-1} = \boldsymbol{m}\,{}^{\mathrm{t}}T^{-1}\cdot\boldsymbol{\xi}$ we obtain (15).

Next, note that every non-zero vector $\boldsymbol{m}\in\mathbb{Z}^2$ can be uniquely expressed as $n(d,-c)$ with $n\in\mathbb{Z}^+$ and $(c,d)\in\widehat{\mathbb{Z}}^2$. Hence the Fourier series (13) can be expressed as

$$f((1_2,\boldsymbol{\xi})M) = \widehat{f}(M,\mathbf{0}) + \sum_{n=1}^{\infty}\sum_{(c,d)\in\widehat{\mathbb{Z}}^2}\widehat{f}\big(M,n(d,-c)\big)e(n(d,-c)\cdot\boldsymbol{\xi}).$$

[1]This is for any fixed exhaustion of $\mathbb{Z}^2$ by an increasing sequence of finite subsets.

However if T is any matrix of the form $\left(\begin{smallmatrix} * & * \\ c & d\end{smallmatrix}\right) \in \Gamma'$ then $n(d,-c) = (n,0)\,{}^{\mathrm{t}}T^{-1}$, and now by using (15) we obtain (17). The uniform absolute convergence on compacta holds since it holds in (13). ☐

Note that the functions $\widetilde{f}_n$ are well-defined for any $f \in \mathrm{C}(\Gamma\backslash G)$, through (16), (14).

Lemma 4.2. *For any $m \in \mathbb{Z}_{\geq 0}$, $n \in \mathbb{Z}^+$ and $f \in \mathrm{C}_{\mathrm{b}}^m(\Gamma\backslash G)$, we have*

$$\left|\widetilde{f}_n\left(\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right)\right| \ll_m \frac{\|f\|_{\mathrm{C}_{\mathrm{b}}^m}}{n^m(c^2+d^2)^{\frac{m}{2}}}, \qquad \forall \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in G'. \tag{18}$$

Proof. The left invariant differential operator corresponding to $X \in \mathfrak{g}$ is given by $Xf(g) = \lim_{t\to 0}(f(g\exp(tX)) - f(g))/t$. In particular, since $\exp(tX_4) = (1_2,(t,0))$ and $\exp(tX_5) = (1_2,(0,t))$ (cf. (10)), we find that if we parametrize G by $(1_2,(x_1,x_2))\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$ then

$$X_4 = d\frac{\partial}{\partial x_1} - b\frac{\partial}{\partial x_2} \qquad \text{and} \qquad X_5 = -c\frac{\partial}{\partial x_1} + a\frac{\partial}{\partial x_2}.$$

Now

$$\widetilde{f}_n\left(\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right) = \int_{\mathbb{T}^2} f\left((1_2,(x_1,x_2))\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right) e(-nx_1)\,dx_2\,dx_1,$$

and hence by repeated integration by parts we have

$$(2\pi i n d)^m \cdot \widetilde{f}_n\left(\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right) = \int_{\mathbb{T}^2} [X_4^m f]\left((1_2,(x_1,x_2))\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right) e(-nx_1)\,dx_2\,dx_1,$$

and

$$(-2\pi i n c)^m \cdot \widetilde{f}_n\left(\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right) = \int_{\mathbb{T}^2} [X_5^m f]\left((1_2,(x_1,x_2))\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right) e(-nx_1)\,dx_2\,dx_1.$$

Hence

$$\max(|c|^m,|d|^m)\cdot\left|\widetilde{f}_n\left(\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right)\right| \leq (2\pi n)^{-m}\|f\|_{\mathrm{C}_{\mathrm{b}}^m},$$

and this implies that (18) holds. ☐

Using Lemma 4.2 we immediately also obtain bounds on *derivatives* of $\widetilde{f}_n$. To make this explicit, let us embed $\mathfrak{sl}(2,\mathbb{R})$ as a subalgebra of $\mathfrak{g}$ through $X \mapsto (X,\mathbf{0})$ (using our notation $\mathfrak{g} \cong \mathfrak{sl}(2,\mathbb{R})\oplus\mathbb{R}^2$). Then each $X \in \mathfrak{sl}(2,\mathbb{R})$, and more generally any element D in the universal enveloping algebra $\mathcal{U}(\mathfrak{sl}(2,\mathbb{R}))$, gives rise to a left invariant differential operator both on G' and on G.

Lemma 4.3. *For any $m \in \mathbb{Z}_{\geq 0}$, $n \in \mathbb{Z}^+$, any $D \in \mathcal{U}(\mathfrak{sl}(2,\mathbb{R}))$ of order $\leq k$, and any $f \in \mathrm{C}_{\mathrm{b}}^{m+k}(\Gamma\backslash G)$, we have*

$$\left|D\widetilde{f}_n\left(\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\right)\right| \ll_m \frac{\|Df\|_{\mathrm{C}_{\mathrm{b}}^m}}{n^m(c^2+d^2)^{\frac{m}{2}}}, \qquad \forall \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in G'. \tag{19}$$

Proof. Set $g := Df \in \mathrm{C}_{\mathrm{b}}^m(\Gamma\backslash G)$; then by differentiation under the integration sign in (14) we have $D\widetilde{f}_n = \widetilde{g}_n$. Hence the lemma follows from Lemma 4.2 applied to g. ☐

We will often consider the function $\widetilde{f}_n$ in Iwasawa coordinates, that is we write (by a slight abuse of notation)

$$\widetilde{f}_n(u,v,\theta) := \widetilde{f}_n\left(\begin{pmatrix} 1 & u \\ 0 & 1\end{pmatrix}\begin{pmatrix} \sqrt{v} & 0 \\ 0 & 1/\sqrt{v}\end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}\right), \tag{20}$$

for $u \in \mathbb{R}$, $v > 0$, $\theta \in \mathbb{R}/2\pi\mathbb{Z}$.

Lemma 4.4. *For any $m \in \mathbb{Z}_{\geq 0}$, $n \in \mathbb{Z}^+$, $k_1,k_2,k_3 \in \mathbb{Z}_{\geq 0}$ and $f \in \mathrm{C}_{\mathrm{b}}^{m+k}(\Gamma\backslash G)$, where $k = k_1+k_2+k_3$, we have*

$$\left|\Big(\frac{\partial}{\partial u}\Big)^{k_1}\Big(\frac{\partial}{\partial v}\Big)^{k_2}\Big(\frac{\partial}{\partial \theta}\Big)^{k_3}\widetilde{f}_n(u,v,\theta)\right| \ll_{m,k} \|f\|_{\mathrm{C}_{\mathrm{b}}^{m+k}} n^{-m} v^{\frac{m}{2}-k_1-k_2}. \tag{21}$$

Proof. Let $\mathbf{k}_\theta := \left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{R})$, and write X_1, X_2, X_3 for the $\mathfrak{sl}(2,\mathbb{R})$-elements $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, respectively. This is consistent with (10) and our fixed embedding of $\mathfrak{sl}(2,\mathbb{R})$ in $\mathfrak{g}$. Also let $\mathrm{Ad} : \mathrm{SL}(2,\mathbb{R}) \to \mathrm{Aut}(\mathfrak{sl}(2,\mathbb{R}))$ be the adjoint representation. Then we compute, in the parametrization (20),

$$\mathrm{Ad}(\mathbf{k}_{-\theta})X_1 = v\frac{\partial}{\partial u}; \qquad \mathrm{Ad}(\mathbf{k}_{-\theta})X_3 = 2v\frac{\partial}{\partial v}; \qquad X_2 - X_1 = \frac{\partial}{\partial\theta}.$$

But $\mathrm{Ad}(\mathbf{k}_{-\theta})X_1$ and $\mathrm{Ad}(\mathbf{k}_{-\theta})X_3$ belong to a fixed compact subset of $\mathfrak{sl}(2,\mathbb{R})$; in fact one checks by a quick computation that these elements always lie in $\{c_1X_1 + c_2X_2 + c_3X_3 \,:\, c_1, c_2, c_3 \in [-1,1]\}$. Hence we have, at every point $(u,v,\theta) \in \mathbb{R} \times \mathbb{R}_{>0} \times (\mathbb{R}/2\pi\mathbb{Z})$,

$$\left|\Big(v\frac{\partial}{\partial u}\Big)^{k_1}\Big(2v\frac{\partial}{\partial v}\Big)^{k_2}\Big(\frac{\partial}{\partial\theta}\Big)^{k_3}\widetilde{f}_n(u,v,\theta)\right| \leq \sum_{\mathrm{ord}(D)=k}\big|D\widetilde{f}_n(u,v,\theta)\big|,$$

where the sum is taken over all the 3^k monomials in X_1, X_2, X_3 of degree k. Now the desired bound follows immediately from Lemma 4.3 and the preceding discussion, if we also note that $c^2 + d^2 = v^{-1}$ holds whenever $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & u \\ 0 & 1 \end{smallmatrix}\right) a(v)\mathbf{k}_\theta$ (the matrix in (20)), and that $v^{k_2}\partial_v^{k_2}$ can be expressed as a linear combination of $(v\partial_v)^j$ for $j = 1, \ldots, k_2$. □

5. The leading term; horocycle equidistribution in $X' = \mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$

Our task is to study the integral

$$\int_{\mathbb{R}} f\Big(\Gamma\,(1_2,\boldsymbol{\xi})\,U^x a(y)\Big)\nu(x)\,dx = \int_{\mathbb{R}} f\left(\Gamma\,(1_2,\boldsymbol{\xi})\begin{pmatrix}\sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y}\end{pmatrix}\right)\nu(x)\,dx.$$

Decomposing f as in Lemma 4.1 we get

$$\begin{aligned}(22)\qquad &= \int_{\mathbb{R}} \widetilde{f}_0\left(\begin{pmatrix}\sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y}\end{pmatrix}\right)\nu(x)\,dx \\ &\quad + \sum_{n=1}^{\infty}\sum_{(c,d)\in\widehat{\mathbb{Z}}^2} e(n(d\xi_1 - c\xi_2))\int_{\mathbb{R}}\widetilde{f}_n\left(\begin{pmatrix} * & * \\ c & d\end{pmatrix}\begin{pmatrix}\sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y}\end{pmatrix}\right)\nu(x)\,dx.\end{aligned}$$

Recall that $\widetilde{f}_0$ is invariant under Γ'; hence the first integral in (22) is simply a weighted average along a closed horocycle in $X' = \Gamma'\backslash G' = \mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$, a case which has been thoroughly studied in the literature. One can prove, either through a careful study of the cohomological equation and invariant distributions for the horocycle flow, as in Flaminio and Forni, [11], or more directly from the representation theory of $\mathrm{SL}(2,\mathbb{R})$ as in Burger [2], that

(23)
$$\int_{\mathbb{R}} \widetilde{f}_0\left(\begin{pmatrix}\sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y}\end{pmatrix}\right)\nu(x)\,dx = \int_{X'}\widetilde{f}_0\,d\mu'\int_{\mathbb{R}}\nu\,dx + O\Big(\|\widetilde{f}_0\|_{\mathrm{C}_\mathrm{b}^4}\,\|\nu\|_{\mathrm{W}^{1,1}}\,y^{\frac12}\log^3(2+1/y)\Big).$$

(See [47] for how to extend [2] to the case of a non-cocompact but cofinite group such as $\mathrm{SL}(2,\mathbb{Z})$. In particular (23) follows easily from [47, Thm. 1, Rem. 3.4].) In (23), note that

$$(24)\qquad \int_{X'}\widetilde{f}_0\,d\mu' = \int_{\Gamma'\backslash G'}\int_{\mathbb{T}^2} f((1_2,\boldsymbol{\xi})M)\,d\boldsymbol{\xi}\,d\mu'(M) = \int_{\Gamma\backslash G} f\,d\mu.$$

Hence (23) accounts for the leading term in (12) in Theorem 3.1. We also note that the error term in (23) is subsumed by the error term in (12), since $\|\widetilde{f}_0\|_{\mathrm{C}_\mathrm{b}^4} \leq \|f\|_{\mathrm{C}_\mathrm{b}^4} \leq \|f\|_{\mathrm{C}_\mathrm{b}^8}$.

6. Initial discussion of the main error contribution

It now remains to treat the sum over $n \in \mathbb{Z}^+$ in (22).

The contribution from the terms with $c = 0$ can be bounded easily. Indeed, for each n there are two such terms, for which we can take $\left(\begin{smallmatrix} * & * \\ c & d \end{smallmatrix}\right)$ to be 1_2 and -1_2, respectively, and by Lemma 4.2 we have

$$\int_{\mathbb{R}} \widetilde{f}_n\left(\pm\begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y}\end{pmatrix}\right)\nu(x)\,dx \ll \|\nu\|_{\mathrm{L}^1}\|f\|_{\mathrm{C}_\mathrm{b}^2}\frac{y}{n^2}. \tag{25}$$

Adding this over all $n \in \mathbb{Z}^+$ we conclude that the contribution from all terms with $c = 0$ in (22) is $O(\|\nu\|_{\mathrm{L}^1}\|f\|_{\mathrm{C}_\mathrm{b}^2} y)$, which is clearly subsumed by the error term in (12).

Hence from now on we focus on the terms with $c \neq 0$. The following lemma expresses the integral appearing in the second line of (22) in the Iwasawa parametrization (cf. (20)). Note that in this notation, the fact that $\widetilde{f}_n$ is left Γ'_∞-invariant (cf. Lemma 4.1) means that $\widetilde{f}_n(u+1,v,\theta) \equiv \widetilde{f}_n(u,v,\theta)$.

Lemma 6.1. *For any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma'$ with $c > 0$, and any $n \in \mathbb{Z}^+$, $y > 0$, we have*

$$\begin{aligned}&\int_{\mathbb{R}} \widetilde{f}_n\left(\begin{pmatrix} a & b \\ c & d\end{pmatrix}\begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y}\end{pmatrix}\right)\nu(x)\,dx \\ &\qquad = \int_0^\pi \widetilde{f}_n\left(\frac{a}{c} - \frac{\sin 2\theta}{2c^2y}, \frac{\sin^2\theta}{c^2y}, \theta\right)\nu\left(-\frac{d}{c} + y\cot\theta\right)\frac{y\,d\theta}{\sin^2\theta}.\end{aligned} \tag{26}$$

Remark 6.1. In the case $c < 0$ one obtains exactly the same formula, except that $\int_0^\pi$ is replaced by $\int_{-\pi}^0$ in the right hand side of (26).

Proof. By a quick computation identifying matrix entries, we find that for any $x \in \mathbb{R}$, the unique $u \in \mathbb{R}$, $v > 0$, $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ satisfying

$$\begin{pmatrix} a & b \\ c & d\end{pmatrix}\begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y}\end{pmatrix} = \begin{pmatrix} 1 & u \\ 0 & 1\end{pmatrix}\begin{pmatrix} \sqrt{v} & 0 \\ 0 & 1/\sqrt{v}\end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix},$$

are given by

$$u = \frac{a}{c} - \frac{cx+d}{c((cx+d)^2+(cy)^2)}; \qquad v = \frac{y}{(cx+d)^2+(cy)^2}; \qquad \theta = \arg(cx+d+i(cy)).$$

(Thus $\cos\theta = \frac{cx+d}{\sqrt{(cx+d)^2+(cy)^2}}$ and $\sin\theta = \frac{cy}{\sqrt{(cx+d)^2+(cy)^2}}$.) In particular θ is a smooth and strictly decreasing function of $x \in \mathbb{R}$, with $\theta \to \pi$ as $x \to -\infty$ and $\theta \to 0$ as $x \to \infty$. We may thus take θ as a new variable of integration. Then

$$\cot\theta = \frac{cx+d}{cy}, \qquad \text{so that} \qquad x = -\frac{d}{c} + y\cot\theta,$$

and furthermore

$$u = \frac{a}{c} - \frac{(\sin\theta)(\cos\theta)}{c^2y} = \frac{a}{c} - \frac{\sin 2\theta}{2c^2y}, \qquad v = \frac{\sin^2\theta}{c^2y}.$$

Hence we obtain the stated identity. □

Note that the map $T \mapsto -T$ gives a bijection from $\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma' \,:\, c > 0\}$ onto $\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma' \,:\, c < 0\}$. Also note that for any matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma'$ with $c \neq 0$ we have $a \equiv d^* \bmod c$, where $d^* \in \mathbb{Z}$ denotes a multiplicative inverse of d modulo c. Hence, by Lemma 6.1 and Remark 6.1, the sum in the second line of (22), excluding all terms with $c = 0$, can be expressed as

(27)

$$\sum_{n=1}^\infty\sum_{c=1}^\infty\int_{-\pi}^\pi \sum_{\substack{d\in\mathbb{Z}\\(c,d)=1}} \nu\left(-\frac{d}{c} + y\cot\theta\right)\widetilde{f}_n\left(\frac{d^*}{c} - \frac{\sin 2\theta}{2c^2y}, \frac{\sin^2\theta}{c^2y}, \theta\right)e\big(n(\operatorname{sgn}\theta)(d\xi_1 - c\xi_2)\big)\frac{y\,d\theta}{\sin^2\theta}.$$

Of course, $\widetilde{f}_n(\frac{d^*}{c} - \frac{\sin 2\theta}{2c^2y}, \frac{\sin^2\theta}{c^2y}, \theta)$ is independent of the choice of d^* since $\widetilde{f}_n$ is periodic with period 1 in its first variable.

It is clear from the way in which we have obtained (27), and also easy to check directly, that if we try to bound (27) by simply inserting absolute values and using our bounds on $\widetilde{f}_n$ proved in Section 4 together with the fact that ν has compact support and bounded L^∞-norm, we obtain that (27) stays *bounded* as $y \to 0$ (for fixed $f, \nu, \boldsymbol{\xi}$). Hence to reach our goal of proving that (27) tends to zero as $y \to 0$, it suffices to establish *any* systematic cancellation in this expression.

Remark 6.2. Our approach, working with the sum in (27), has close similarities to the following method of proving equidistribution of pieces of closed horocycles in X'.

Let f be a function on X', which for simplicity we assume to be smooth and compactly supported, i.e. $f \in \mathrm{C}_c^\infty(X')$. Any such f can be expressed as

$$f(\Gamma' g) = \sum_{\gamma \in \Gamma'} \eta(\gamma g) \qquad (\forall g \in G'), \tag{28}$$

for some $\eta \in \mathrm{C}_c^\infty(G')$. We wish to study the weighted average of f along a closed horocycle in X', $\int_{\mathbb{R}} f(\Gamma' U^x a(y))\nu(x)\,dx$, in the limit $y \to 0$. To do so we use (28), and change order of summation and integration. The contribution from all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $c = 0$ is seen to vanish for y small, since η has compact support. The remaining terms are handled by expressing η in Iwasawa coordinates (cf. (20)), applying an analogue of Lemma 6.1, and then introducing $\widetilde{\eta}(u, v, \theta) := \sum_{n \in \mathbb{Z}} \eta(u + n, v, \theta)$, a function on $(\mathbb{R}/\mathbb{Z}) \times \mathbb{R}_{>0} \times (\mathbb{R}/2\pi\mathbb{Z})$:

$$\begin{aligned}
\int_{\mathbb{R}} f(\Gamma' U^x a(y))\nu(x)\,dx &= \sum_{c=1}^{\infty} \sum_{\substack{a,d \in \mathbb{Z} \\ ad \equiv 1 \bmod c}} \int_{-\pi}^{\pi} \nu\Big(-\frac{d}{c} + y\cot\theta\Big)\eta\left(\frac{a}{c} - \frac{\sin 2\theta}{2c^2y}, \frac{\sin^2\theta}{c^2y}, \theta\right) \frac{y\,d\theta}{\sin^2\theta} \\
&= \int_{-\pi}^{\pi} \sum_{c=1}^{\infty} \sum_{\substack{d \in \mathbb{Z} \\ (d,c)=1}} \nu\Big(-\frac{d}{c} + y\cot\theta\Big)\widetilde{\eta}\left(\frac{d^*}{c} - \frac{\sin 2\theta}{2c^2y}, \frac{\sin^2\theta}{c^2y}, \theta\right) \frac{y\,d\theta}{\sin^2\theta} \\
&\approx \int_{-\pi}^{\pi} \sum_{c=1}^{\infty} \varphi(c) \int_{\mathbb{R}} \nu(x)\,dx \int_{\mathbb{R}/\mathbb{Z}} \widetilde{\eta}\Big(u, \frac{\sin^2\theta}{c^2y}, \theta\Big)\,du\,\frac{y\,d\theta}{\sin^2\theta} \\
&\approx \int_{-\pi}^{\pi} \int_0^{\infty} \frac{6}{\pi^2} r \int_{\mathbb{R}/\mathbb{Z}} \widetilde{\eta}\Big(u, \frac{\sin^2\theta}{r^2y}, \theta\Big)\,du\,dr\,\frac{y\,d\theta}{\sin^2\theta} \int_{\mathbb{R}} \nu\,dx \\
&= \frac{3}{\pi^2} \int_{-\pi}^{\pi} \int_{\mathbb{R}} \int_0^{\infty} \eta(u, v, \theta)\,\frac{dv\,du\,d\theta}{v^2} \int_{\mathbb{R}} \nu\,dx = \int_{G'} \eta\,d\mu' \int_{\mathbb{R}} \nu\,dx = \int_{X'} f\,d\mu' \int_{\mathbb{R}} \nu\,dx.
\end{aligned}$$

Here the approximate equality between the second and third lines holds since, for any large c, as d varies through a not too small interval of integers, the multiplicative inverse d^* becomes approximately equidistributed $\mathbb{Z}/c\mathbb{Z}$. The next approximate equality holds since $\varphi(c)$ for large c behaves like $\frac{6}{\pi^2}c$ on average. The last equality in the above computation follows using (28) and standard unfolding. The errors in the approximations can be bounded using Lemma 7.1 below and [17, Thm. 330] (together with summation by parts); in this way one obtains, with some work, that the total difference between the horocycle average $\int_{\mathbb{R}} f(\Gamma' U^x a(y))\nu(x)\,dx$ and the volume average $\int_{X'} f\,d\mu' \int_{\mathbb{R}} \nu\,dx$ is bounded by $O_{f,\nu,\varepsilon}(y^{\frac14 - \varepsilon})$ as $y \to 0$. This falls short of the optimal error bound $y^{\frac12 - \varepsilon}$ which we pointed out in Section 5; but it is comparable with the "non-Diophantine" part of the error bound in Theorem 3.1.

The main difference between the above computation and our proof of Theorem 3.1 is that we will establish *cancellation* in (27), caused by the oscillating factor $e(n(\operatorname{sgn}\theta)(d\xi_1 - c\xi_2))$. For ξ_1 nicely Diophantine, cancellation can be established already in the inner sum over d (cf. Section 8); however when ξ_1 is well-approximable by rational numbers we will collect certain main contributions from the inner sum and establish cancellation when these are added over c (cf. Section 9).

7. Cancellation in an exponential sum

The following lemma is a standard application of Weil's bound on Kloosterman sums.

Lemma 7.1. *Let $0 < \eta < 1$, $\alpha \in \mathbb{R}$, $c \in \mathbb{Z}^+$, let $g_1 \in \mathrm{C}^2(\mathbb{R})$ with compact support and $g_2 \in \mathrm{C}^2(\mathbb{R}/\mathbb{Z})$, and let N be an arbitrary subset of $\mathbb{Z}$. Then*

$$\sum_{\substack{d\in\mathbb{Z}\\(c,d)=1}} g_1\Big(\frac{d}{c}\Big)e(d\alpha)g_2\Big(\frac{d^*}{c}\Big)$$

$$(29)\quad = \sum_{k\in N}\Big(\int_{\mathbb{R}} g_1(x)e\big((c\alpha-k)x\big)\,dx\Big)\Big(\int_{\mathbb{R}/\mathbb{Z}} g_2\Big)\mu\Big(\frac{c}{(c,k)}\Big)\frac{\varphi(c)}{\varphi(c/(c,k))}$$
$$+O_\eta\Big(\|g_1\|_{\mathrm{W}^{1,1}}^{1-\eta}\|g_1\|_{\mathrm{W}^{2,1}}^{\eta}\Big)\Big\{\|g_2\|_{\mathrm{L}^1(\mathbb{R}/\mathbb{Z})}\sum_{k\in\mathbb{Z}\setminus N}\frac{(c,k)}{1+|k-c\alpha|^{1+\eta}}+\|g_2''\|_{\mathrm{L}^1(\mathbb{R}/\mathbb{Z})}\,\sigma(c)\sqrt{c}\Big\}.$$

Proof. Set

$$h(x) = \sum_{m\in\mathbb{Z}} g_1(x+m)e(c\alpha(x+m)).$$

Then $h \in \mathrm{C}^2(\mathbb{R})$ and $h(x+n) = h(x)$ for all $n \in \mathbb{Z}$; hence we may view h as a function in $\mathrm{C}^2(\mathbb{R}/\mathbb{Z})$. Let the Fourier expansions of $h(x)$ and $g_2(x)$ be

$$h(x) = \sum_{n\in\mathbb{Z}} a_n e(nx) \qquad \text{and} \qquad g_2(x) = \sum_{m\in\mathbb{Z}} b_m e(mx).$$

Here

$$a_n = \int_{\mathbb{R}/\mathbb{Z}} h(x)e(-nx)\,dx = \int_{\mathbb{R}} g_1(x)e((c\alpha-n)x)\,dx, \tag{30}$$

and thus, by integration by parts, $|a_n| \leq \|g_1^{(j)}\|_{\mathrm{L}^1}(2\pi|c\alpha-n|)^{-j}$ for $j = 0,1,2$. Hence, making use of the general inequality $\min(A, By) \leq A^{1-\eta}B^\eta y^\eta$ (true for all $A, B, y \geq 0$) with $A = \|g_1'\|_{\mathrm{L}^1}$, $B = \|g_1''\|_{\mathrm{L}^1}$ and $y = (2\pi|c\alpha-n|)^{-1}$, we conclude:

$$|a_n| \ll \frac{\|g_1\|_{\mathrm{W}^{1,1}}^{1-\eta}\|g_1\|_{\mathrm{W}^{2,1}}^{\eta}}{1+|c\alpha-n|^{1+\eta}}, \qquad \forall n \in \mathbb{Z}.$$

Similarly, using $b_m = \int_{\mathbb{R}/\mathbb{Z}} g_2(x)e(-mx)\,dx$, we have $|b_0| \leq \|g_2\|_{\mathrm{L}^1(\mathbb{R}/\mathbb{Z})}$ and $|b_m| \leq \|g_2''\|_{\mathrm{L}^1(\mathbb{R}/\mathbb{Z})}\,|m|^{-2}$ for $m \neq 0$. Now the sum in the left hand side of (29) can be expressed as

$$\sum_{d\in(\mathbb{Z}/c\mathbb{Z})^\times} h\Big(\frac{d}{c}\Big)g_2\Big(\frac{d^*}{c}\Big) = \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} a_n b_m S(n,m;c), \tag{31}$$

where we use standard notation for Kloosterman sums; $S(n,m;c) := \sum_{d\in(\mathbb{Z}/c\mathbb{Z})^\times} e(n\frac{d}{c}+m\frac{d^*}{c})$.

For $m = 0$, $S(n,m;c)$ is a Ramanujan sum;

$$S(n,0;c) = \sum_{d\in(\mathbb{Z}/c\mathbb{Z})^\times} e\Big(n\frac{d}{c}\Big) = \mu\Big(\frac{c}{(c,n)}\Big)\frac{\varphi(c)}{\varphi(c/(c,n))},$$

and in particular $|S(n,0;c)| \leq (c,n)$ (cf., e.g., [21, Sec. 3.2]). Hence the contribution from all terms with $m = 0$ in (31) is

$$= \sum_{n\in N} a_n b_0 S(n,0;c) + O\bigg(\|g_1\|_{\mathrm{W}^{1,1}}^{1-\eta}\|g_1\|_{\mathrm{W}^{2,1}}^{\eta}\|g_2\|_{\mathrm{L}^1}\sum_{n\in\mathbb{Z}\setminus N}\frac{(c,n)}{1+|c\alpha-n|^{1+\eta}}\bigg),$$

and the sum over $n \in N$ expands to give the first line in the right hand side of (29).

Next for $m \neq 0$ we use Weil's bound, $|S(n,m;c)| \leq \sigma(c)\gcd(n,m,c)^{1/2}\sqrt{c}$ (cf. [51], and [21, Ch. 11.7]), and $\gcd(n,m,c)^{1/2} \leq |m|^{1/2}$, to see that the contribution from all terms with $m \neq 0$ in (31) is

$$\ll \|g_1\|_{\mathrm{W}^{1,1}}^{1-\eta}\|g_1\|_{\mathrm{W}^{2,1}}^{\eta}\|g_2''\|_{\mathrm{L}^1}\sigma(c)\sqrt{c}\sum_{n\in\mathbb{Z}}\frac{1}{1+|c\alpha-n|^{1+\eta}}\sum_{m\in\mathbb{Z}\setminus\{0\}}|m|^{-3/2}$$
$$\ll_\eta \|g_1\|_{\mathrm{W}^{1,1}}^{1-\eta}\|g_1\|_{\mathrm{W}^{2,1}}^{\eta}\|g_2''\|_{\mathrm{L}^1}\sigma(c)\sqrt{c}.$$

This completes the proof of the lemma. □

8. Proof of a weaker version of Theorem 3.1

In this section we go through the first steps of the proof of Theorem 3.1; the outcome of this is a version of Theorem 3.1 which only involves the Diophantine properties of ξ_1 and not those of ξ_2; see Proposition 8.3 below. This result is strong enough to imply that the error term in Theorem 3.1 (as well as the left hand side in Theorem 1.2) decays like $y^{\min(\frac{1}{4},\frac{1}{K})-\varepsilon}$ in the case of ξ_1 irrational of Diophantine type K (see Remark 8.1); however for $\boldsymbol{\xi}$ with $\xi_1 \in \mathbb{Q}$ but $\xi_2 \notin \mathbb{Q}$, Proposition 8.3 does not imply any equidistribution whatsoever.

Recall that our task is to bound the sum in (27). We write $\omega = \operatorname{sgn}(\theta)$, where we always assume $\theta \neq 0$ so that $\omega \in \{-1,1\}$. Applying Lemma 7.1 and replacing k by ωk we get the following estimate valid for any $\theta \in (-\pi,\pi)\setminus\{0\}$ and $n,c \in \mathbb{Z}^+$:

$$\sum_{\substack{d\in\mathbb{Z}\\(c,d)=1}}\nu\Big(-\frac{d}{c}+y\cot\theta\Big)\,\widetilde{f}_n\Big(\frac{d^*}{c}-\frac{\sin 2\theta}{2c^2y},\frac{\sin^2\theta}{c^2y},\theta\Big)e\big(n\omega(d\xi_1-c\xi_2)\big)$$
$$=\sum_{k\in N}\frac{\mu(\frac{c}{(c,k)})\varphi(c)}{\varphi(\frac{c}{(c,k)})}e\big(-n\omega c\xi_2\big)\int_{\mathbb{R}}\nu(-x+y\cot\theta)e\big((cn\xi_1-k)\omega x\big)\,dx\int_{\mathbb{R}/\mathbb{Z}}\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\,du$$
$$+O_\eta\Big(\|\nu\|_{\mathrm{W}^{1,1}}^{1-\eta}\|\nu\|_{\mathrm{W}^{2,1}}^{\eta}\Big)\bigg\{\Big(\int_{\mathbb{R}/\mathbb{Z}}\Big|\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\Big|\,du\Big)\sum_{k\in\mathbb{Z}\setminus N}\frac{(c,k)}{1+|k-cn\xi_1|^{1+\eta}}$$
$$+\Big(\int_{\mathbb{R}/\mathbb{Z}}\Big|\frac{\partial^2}{\partial u^2}\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\Big|\,du\Big)\sigma(c)\sqrt{c}\bigg\}. \tag{32}$$

Here N is a subset of $\mathbb{Z}$ which we are free to choose (it may depend on n, c, θ). In the present section, we will in fact make the simple choice $N=\emptyset$! Thus the first row in the right hand side of (32) *vanishes.* In order to bound the remaining expressions, note that by Lemma 4.2, for any $m \in \mathbb{Z}_{\geq 0}$ we have

$$\int_{\mathbb{R}/\mathbb{Z}}\Big|\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\Big|\,du \ll_m \|f\|_{\mathrm{C}_{\mathrm{b}}^m}\Big(\frac{|\sin\theta|}{nc\sqrt{y}}\Big)^m. \tag{33}$$

Using this bound for both $m=0$ and a general $m \in \mathbb{Z}_{\geq 0}$ gives

$$\int_{\mathbb{R}/\mathbb{Z}}\Big|\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\Big|\,du \ll_m \|f\|_{\mathrm{C}_{\mathrm{b}}^m}\min\Big(1,\Big(\frac{|\sin\theta|}{nc\sqrt{y}}\Big)^m\Big). \tag{34}$$

Similarly, by Lemma 4.4, we have for any $\ell \in \mathbb{Z}_{\geq 4}$:

$$\int_{\mathbb{R}/\mathbb{Z}}\Big|\frac{\partial^2}{\partial u^2}\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\Big|\,du \ll_\ell \|f\|_{\mathrm{C}_{\mathrm{b}}^{\ell+2}}n^{-4}\min\Big(1,\Big(\frac{|\sin\theta|}{nc\sqrt{y}}\Big)^{\ell-4}\Big). \tag{35}$$

We also compute that, for $m \geq 2$ and any real $a>0$,

$$\int_{-\pi}^{\pi}\min\Big(1,\big(a^{-1}|\sin\theta|\big)^m\Big)\frac{d\theta}{\sin^2\theta} \asymp_m a^{-1}(1+a)^{1-m}. \tag{36}$$

Using the bounds (33)–(35) in (32) and then applying (36) with $a = nc\sqrt{y}$, *assuming from now on that* $m \geq 2$ *and* $\ell \geq 6$, we conclude that (27) is

$$\ll_{m,\ell,\eta} \|\nu\|_{\mathrm{W}^{1,1}}^{1-\eta} \|\nu\|_{\mathrm{W}^{2,1}}^{\eta} \Big\{ \|f\|_{\mathrm{C}_{\mathrm{b}}^m} \, y^{1-\frac{m}{2}} \sum_{n=1}^{\infty} n^{-m} \sum_{c=1}^{\infty} c^{-1}\big((n\sqrt{y})^{-1} + c\big)^{1-m} \sum_{k\in\mathbb{Z}} \frac{(c,k)}{1+|k-cn\xi_1|^{1+\eta}}$$

$$+\|f\|_{\mathrm{C}_{\mathrm{b}}^{\ell+2}} \, y^{3-\frac{\ell}{2}} \sum_{n=1}^{\infty} n^{-\ell} \sum_{c=1}^{\infty} \big((n\sqrt{y})^{-1} + c\big)^{5-\ell} \frac{\sigma(c)}{\sqrt{c}} \Big\}. \tag{37}$$

Lemma 8.1. *For any* $X > 0$ *and* $m \in \mathbb{Z}^+$ *we have*

$$\sum_{c=1}^{\infty} (X+c)^{-m} \frac{\sigma(c)}{\sqrt{c}} \ll_m \begin{cases} X^{\frac{1}{2}-m} \log(1+X) & \text{if } X \geq 1 \\ 1 & \text{if } X < 1. \end{cases}$$

Proof. This follows by using $\sum_{1\leq c\leq x} \sigma(c) \ll x\log(1+x)$, $\forall x \geq 1$ (cf., e.g., [21, (1.75)]), and integration by parts. □

It follows from Lemma 8.1 and a simple summation over n that the expression in the second line of (37) is $\ll \|f\|_{\mathrm{C}_{\mathrm{b}}^{\ell+2}} y^{\frac{1}{4}} \log(1+y^{-1})$.

When bounding the double sum over c and k appearing in the first line of (37), it is natural to introduce the following majorant function.

$$\mathfrak{M}_\alpha(X) := \sum_{\ell=1}^{\infty} \min\Big(\frac{1}{\ell^2}, \frac{1}{X\ell\langle \ell\alpha\rangle}\Big) \qquad (X > 0,\ \alpha \in \mathbb{R}). \tag{38}$$

Clearly this is a decreasing function of X for fixed α, but it is never very rapidly decreasing; in short we have

$$\mathfrak{M}_\alpha(X_2) \leq \mathfrak{M}_\alpha(X_1) \leq \min\Big(\frac{X_2}{X_1}\mathfrak{M}_\alpha(X_2), \zeta(2)\Big), \qquad \forall 0 < X_1 \leq X_2. \tag{39}$$

The proof is immediate, using (38).

Lemma 8.2. *Fix* $\eta > 0$ *and* $m \in \mathbb{Z}_{\geq 3}$*. Then for any* $\alpha \in \mathbb{R}$ *and* $X > 0$ *we have*

$$\sum_{c=1}^{\infty} c^{-1}(X+c)^{1-m} \sum_{k\in\mathbb{Z}} \frac{(c,k)}{1+|k-c\alpha|^{1+\eta}} \ll_{\eta,m} \begin{cases} X^{2-m}\,\mathfrak{M}_\alpha(X) & \text{if } X \geq 1 \\ 1 & \text{if } X < 1. \end{cases} \tag{40}$$

Proof. For given positive integers d and $c = \ell d$ ($\ell \in \mathbb{Z}^+$), we will bound the sum over k in (40) when further restricted by the condition $(c,k) = d$. Denote by k_0 the unique integer in the interval $c\alpha - \frac{1}{2}d < k_0 \leq c\alpha + \frac{1}{2}d$ which is divisible by d. Then $|k_0 - c\alpha|$ equals the distance from $c\alpha$ to the point set $d\mathbb{Z}$, viz. $|k_0 - c\alpha| = d\langle \frac{c}{d}\alpha\rangle$. Note that the set $\{k \in \mathbb{Z} : (c,k) = d\}$ is contained in $d\mathbb{Z} = k_0 + d\mathbb{Z}$. This gives

$$\sum_{\substack{k\in\mathbb{Z} \\ ((c,k)=d)}} \frac{(c,k)}{1+|k-c\alpha|^{1+\eta}} \leq \sum_{k\in k_0+d\mathbb{Z}} \frac{d}{1+|k-c\alpha|^{1+\eta}} \ll_\eta \frac{d}{1+(d\langle\frac{c}{d}\alpha\rangle)^{1+\eta}},$$

since the contribution from all terms with $k \neq k_0$ is $\ll d^{-\eta}\sum_{v=1}^{\infty} v^{-1-\eta} \ll_\eta d^{-\eta}$. Hence the left hand side of (40) is

$$\ll_\eta \sum_{\ell=1}^{\infty} \ell^{-1} \sum_{d=1}^{\infty} \frac{(X+\ell d)^{1-m}}{1+(d\langle\ell\alpha\rangle)^{1+\eta}} \asymp_{m,\eta} \sum_{\ell=1}^{\infty} \ell^{-1} \left\{ \begin{matrix} X^{2-m}\ell^{-1} & \text{if } 1 \leq X/\ell \leq \langle\ell\alpha\rangle^{-1} \\ X^{1-m}\langle\ell\alpha\rangle^{-1} & \text{if } \langle\ell\alpha\rangle^{-1} < X/\ell \\ \ell^{1-m} & \text{if } X/\ell < 1 \end{matrix} \right\},$$

where we used $m \geq 3$ in the last step. If $X < 1$ then the above sum is $\sum_{\ell=1}^{\infty} \ell^{-m} \ll 1$. On the other hand if $X \geq 1$ then we get

$$= X^{2-m} \sum_{1\leq\ell\leq X} \min\Big(\frac{1}{\ell^2}, \frac{1}{X\ell\langle\ell\alpha\rangle}\Big) + \sum_{\ell>X} \ell^{-m} \ll X^{2-m}\,\mathfrak{M}_\alpha(X).$$

□

In Lemma 8.2, of course the bound $X^{2-m}\mathfrak{M}_\alpha(X)$ is valid also when $X<1$, albeit wasteful. We now get in (37), *assuming from now on* $m\geq 3$,

$$y^{1-\frac{m}{2}}\sum_{n=1}^\infty n^{-m}\sum_{c=1}^\infty c^{-1}\big((n\sqrt{y})^{-1}+c\big)^{1-m}\sum_{k\in\mathbb{Z}}\frac{(c,k)}{1+|k-cn\xi_1|^{1+\eta}}$$
$$\ll_{m,\eta} y^{1-\frac{m}{2}}\sum_{n=1}^\infty n^{-m}(n\sqrt{y})^{m-2}\mathfrak{M}_{n\xi_1}\Big(\frac{1}{n\sqrt{y}}\Big)=\sum_{n=1}^\infty\sum_{\ell=1}^\infty\min\Big(\frac{1}{(n\ell)^2},\frac{\sqrt{y}}{n\ell\langle n\ell\xi_1\rangle}\Big)=\widetilde{\mathfrak{M}}_{\xi_1}(y^{-1/2}),$$

where

$$\widetilde{\mathfrak{M}}_{\xi_1}(X):=\sum_{k=1}^\infty\sigma(k)\min\Big(\frac{1}{k^2},\frac{1}{Xk\langle k\xi_1\rangle}\Big). \tag{41}$$

We have now proved:

Proposition 8.3. *Let $0<\eta<1$ be fixed. Then for any $f\in\mathrm{C}_\mathrm{b}^8(\Gamma\backslash G)$, any $\nu\in\mathrm{C}^2(\mathbb{R})$ with compact support, and any $\boldsymbol{\xi}\in\mathbb{R}^2$, $0<y<1$,*

$$\int_\mathbb{R} f\Big(\Gamma\,(1_2,\boldsymbol{\xi})\,U^xa(y)\Big)\,\nu(x)\,dx=\int_{\Gamma\backslash G}f\,d\mu\int_\mathbb{R}\nu\,dx$$
$$+O_\eta\Big(\|\nu\|_{\mathrm{W}^{1,1}}^{1-\eta}\|\nu\|_{\mathrm{W}^{2,1}}^{\eta}\Big)\Big\{\|f\|_{\mathrm{C}_\mathrm{b}^8}\,y^{\frac14}\log(1+y^{-1})+\|f\|_{\mathrm{C}_\mathrm{b}^3}\widetilde{\mathfrak{M}}_{\xi_1}(y^{-1/2})\Big\}. \tag{42}$$

Remark 8.1. Note that for fixed ξ_1, $\lim_{y\to0}\widetilde{\mathfrak{M}}_{\xi_1}(y^{-1/2})=0$ if and only if $\xi_1\notin\mathbb{Q}$. Hence Proposition 8.3 gives an effective version of Theorem 1.1 in the special case of ξ_1 irrational.

In order to compare Proposition 8.3 and Theorem 1.2 (or Theorem 3.1) in the y-aspect, we point out that

$$\widetilde{\mathfrak{M}}_{\xi_1}(y^{-1/2})\ll_\varepsilon\big(b_{\xi_1}^1(y)+y^{\frac14}\big)^{1-\varepsilon},\qquad\text{where }\ b_{\xi_1}^1(y):=\max_{q\in\mathbb{Z}^+}\min\Big(\frac{1}{q^2},\frac{\sqrt{y}}{q\langle q\xi_1\rangle}\Big). \tag{43}$$

This can be proved by following the same argument as we will use later below (74), and again below (79). Note that $\widetilde{\mathfrak{M}}_{\xi_1}(y^{-1/2})\geq b_{\xi_1}^1(y)$ holds trivially from the definition (41), and so the bound (43) is essentially sharp whenever $b_{\xi_1}^1(y)\gg y^{1/4}$.

On the other hand for ξ_1 Diophantine of type $K\geq2$, $\widetilde{\mathfrak{M}}_{\xi_1}(y^{-1/2})\ll_{\varepsilon,\xi_1}y^{\frac1K-\varepsilon}$ (cf. Lemma 8.4); in particular if $K<4$ then $\widetilde{\mathfrak{M}}_{\xi_1}(y^{-1/2})$ decays more rapidly than $y^{\frac14}$ as $y\to0$. We expect that for $\boldsymbol{\xi}$ satisfying an appropriate Diophantine condition, the error bound in (42) can be improved to $O_{f,\nu,\boldsymbol{\xi},\varepsilon}(y^{\frac12-\varepsilon})$. This is the rate which one obtains when f is a lift of a function on X' (cf. [11], [47]); furthermore the exponent $\frac12$ corresponds to the exponential rate of mixing for the flow $\Phi^\mathbb{R}$ on X, when acting on sufficiently smooth vectors in $\mathrm{L}^2(X)$ (cf. [19, Thm. 3.3.10] as well as [4]). In our approach we are stuck at the exponent $\frac14$ since in (27) we bound the absolute value of the sum over d individually for each c using the Weil bound; cf. Lemma 7.1.

Lemma 8.4. *Let $\varepsilon>0$, $X\geq1$ and $\xi\in\mathbb{R}$, and assume $\langle n\xi\rangle\geq cn^{1-K}$ for all $n\in\mathbb{Z}^+$ and some fixed $c>0$ and $K\geq2$. Then*

$$\widetilde{\mathfrak{M}}_\xi(X)\ll_\varepsilon c^{-\frac2K}X^{\varepsilon-\frac2K}. \tag{44}$$

(This bound is essentially optimal. Indeed, if $\langle n\xi\rangle\leq cn^{1-K}$ holds for some n then already the single term $\sigma(n)\min(\frac{1}{n^2},\frac{1}{Xn\langle n\xi\rangle})$ equals $\sigma(n)(cX)^{-\frac2K}$ when $X=n^K/c$.)

Proof. We assume $cX>1$ since otherwise the stated bound is trivial. Let $\frac{p_j}{q_j}$ for $j\in\mathbb{Z}_{\geq0}$ be the jth convergent of the (simple) continued fraction expansion of ξ (cf., e.g., [17, Ch. X]).

Thus $1 = q_0 \leq q_1 < q_2 < \cdots$. Now for any $\ell \geq 1$ we have

$$\sum_{1\leq n\leq q_\ell/2} \frac{\sigma(n)}{n\langle n\xi\rangle} \ll_\varepsilon q_\ell^\varepsilon \sum_{j=1}^{\ell} \sum_{q_{j-1}/2<n\leq q_j/2} \frac{1}{n\langle n\xi\rangle} \ll q_\ell^\varepsilon \sum_{j=1}^{\ell} q_{j-1}^{-1} \sum_{1\leq n\leq q_j/2} \frac{1}{\langle n\xi\rangle} \ll q_\ell^\varepsilon \sum_{j=1}^{\ell} \frac{q_j \log q_j}{q_{j-1}},$$

where the last bound follows from [35, Lemma 4.8], since $|\xi - \frac{p_j}{q_j}| < \frac{1}{q_j q_{j+1}}$ ([17, Thm. 171]). But for every $j \geq 1$ we have $c q_{j-1}^{1-K} \leq \langle q_{j-1}\xi\rangle < q_j^{-1}$, i.e. $q_j < c^{-1} q_{j-1}^{K-1}$. Hence

(45)

$$\sum_{1\leq n\leq q_\ell/2} \frac{\sigma(n)}{Xn\langle n\xi\rangle} \ll_\varepsilon (cX)^{-1} q_\ell^\varepsilon \sum_{j=0}^{\ell-1} q_j^{K-2} \log q_{j+1} \ll_\varepsilon (cX)^{-1} q_\ell^{2\varepsilon} \ell q_{\ell-1}^{K-2} \ll_\varepsilon (cX)^{-1} q_\ell^{3\varepsilon} q_{\ell-1}^{K-2},$$

where we used the fact that q_ℓ is bounded below by the ℓth Fibonacci number.

Next note that for any $\ell \geq 1$ and $h \geq 1$, by [35, Lemma 4.9],

$$\sum_{hq_\ell+1\leq n\leq (h+1)q_\ell} \sigma(n) \min\Big(\frac{1}{n^2}, \frac{1}{Xn\langle n\xi\rangle}\Big) \ll_\varepsilon X^{-1}(hq_\ell)^{\varepsilon-1} \sum_{r=1}^{q_\ell} \min\Big(\frac{X}{hq_\ell}, \frac{1}{\langle (hq_\ell + r)\xi\rangle}\Big) \ll (hq_\ell)^\varepsilon \Big(\frac{1}{(hq_\ell)^2} + \frac{\log q_\ell}{Xh}\Big). \tag{46}$$

Similarly also

$$\sum_{q_\ell/2<n\leq q_\ell} \sigma(n) \min\Big(\frac{1}{n^2}, \frac{1}{Xn\langle n\xi\rangle}\Big) \ll_\varepsilon X^{-1} q_\ell^{\varepsilon-1} \sum_{r=1}^{q_\ell} \min\Big(\frac{X}{q_\ell}, \frac{1}{\langle r\xi\rangle}\Big) \ll q_\ell^\varepsilon \Big(\frac{1}{q_\ell^2} + \frac{\log q_\ell}{X}\Big). \tag{47}$$

Adding (47) and (46) for all $h \leq X/q_\ell$ we obtain

$$\sum_{q_\ell/2<n\leq X} \sigma(n) \min\Big(\frac{1}{n^2}, \frac{1}{Xn\langle n\xi\rangle}\Big) \ll_\varepsilon X^\varepsilon \Big(\frac{1}{q_\ell^2} + \frac{\log q_\ell}{X}\Big). \tag{48}$$

(This is valid, trivially, also if $X \leq q_\ell/2$.) Now choose $\ell \geq 1$ so that $q_{\ell-1} \leq (cX)^{\frac{1}{K}} < q_\ell$. Then $q_\ell < c^{-1} q_{\ell-1}^{K-1} \leq (cX)^{-\frac{1}{K}} X < X$. Now (44) follows by adding (45), (48) and the bound $\sum_{n>X} \sigma(n) n^{-2} \ll_\varepsilon X^{\varepsilon-1}$, replacing ε by $\frac{1}{3}\varepsilon$, and using $(cX)^{-\frac{2}{K}} > X^{-\frac{2}{K}} \geq X^{-1}$. $\square$

9. Proof of Theorem 3.1

We will now make a choice of the set N in (32) which will allow us to reach a reasonable bound also when ξ_1 is rational or well-approximable by rational numbers, provided that ξ_2 has good Diophantine properties. Given any irrational number α, let $\frac{p_j}{q_j}$ ($j \in \mathbb{Z}_{\geq 0}$) be the jth convergent of the (simple) continued fraction expansion of α (cf., e.g., [17, Ch. X]; thus $1 = q_0 \leq q_1 < q_2 < \cdots$), and set, for each $c \in \mathbb{Z}^+$,

$$N_c^{(\alpha)} := \Big\{k \in \mathbb{Z} \,:\, \frac{k}{c} \in \Big\{\frac{p_0}{q_0}, \frac{p_1}{q_1}, \ldots\Big\}\Big\}. \tag{49}$$

We will choose $N = N_c^{(n\xi_1)}$ in (32). In order for this to make sense we have to assume that ξ_1 *is irrational.* This assumption is made merely for notational convenience, to ensure that the continued fraction expansion of $n\xi_1$ is not finite. Note that the assumption can be made without loss of generality: if (12) holds whenever ξ_1 is irrational then it must also hold when ξ_1 is rational, because all expressions involved depend continuously on ξ_1. (There is some flexibility in the possible choices of the set N in (32) which make the proof work; cf. Remark 9.2 below; however the choice made here is notationally convenient.)

We will use the following lemma to bound the contribution from the sum over $k \in \mathbb{Z} \setminus N$ in (32) to the expression in (27).

Lemma 9.1. *Fix $\eta>0$ and $m\in\mathbb{Z}_{\geq 3}$. Then for any irrational $\alpha\in\mathbb{R}$, and any $X>0$,*

$$\sum_{c=1}^{\infty}\sum_{k\in\mathbb{Z}\setminus N_c^{(\alpha)}} c^{-1}(X+c)^{1-m}\frac{(c,k)}{1+|k-c\alpha|^{1+\eta}}\ll_{m,\eta}\begin{cases}X^{\frac32-m} & \text{if } X\geq 1\\ 1 & \text{if } X<1.\end{cases}\tag{50}$$

Proof. Introduce d, ℓ, k_0 as in the proof of Lemma 8.2. Note that if $|k_0-c\alpha|<\frac{d}{2\ell}$ then $|\frac{k_0/d}{\ell}-\alpha|<\frac{1}{2\ell^2}$, which implies that $\frac{k_0}{c}=\frac{k_0/d}{\ell}$ is a convergent of the continued fraction expansion of α [17, Thm. 184], viz. $k_0\in N_c^{(\alpha)}$. Hence

$$\sum_{\substack{k\in\mathbb{Z}\setminus N_c^{(\alpha)}\\ ((c,k)=d)}}\frac{(c,k)}{1+|k-c\alpha|^{1+\eta}}\leq\sum_{\substack{v\in\mathbb{Z}\\ v\neq 0\text{ or }|k_0-c\alpha|\geq d/2\ell}}\frac{d}{1+|k_0+dv-c\alpha|^{1+\eta}}$$
$$\ll\left\{\begin{matrix}\frac{d}{1+|k_0-c\alpha|^{1+\eta}} & \text{if } |k_0-c\alpha|\geq d/2\ell\\ 0 & \text{otherwise}\end{matrix}\right\}+d^{-\eta}\sum_{v=1}^{\infty}v^{-1-\eta}\ll_\eta\frac{d}{1+(d/\ell)^{1+\eta}}\ll\frac{d\ell}{d+\ell}.$$

Hence the left hand side of (50) is (using $m\geq 3$):

$$\ll_\eta\sum_{d=1}^{\infty}\sum_{\ell=1}^{\infty}\frac{(X+\ell d)^{1-m}}{d+\ell}\ll\sum_{d=1}^{\infty}\sum_{\ell=d}^{\infty}\frac{(X+\ell d)^{1-m}}{\ell}\ll\begin{cases}X^{\frac32-m} & \text{if } X\geq 1\\ 1 & \text{if } X<1.\end{cases}$$

□

By following the steps leading to (37), with $m=3$ and $\ell=6$, and also using Lemma 9.1 and Lemma 8.1, it follows that the contribution from the last two lines of (32) to the expression in (27) is:

$$\ll\|\nu\|_{\mathrm{W}^{1,1}}^{1-\eta}\|\nu\|_{\mathrm{W}^{2,1}}^{\eta}\|f\|_{\mathrm{C}_{\mathrm{b}}^8}\,y^{\frac14}\log(1+y^{-1}).$$

Remark 9.1. The simple bound in Lemma 9.1 is wasteful in the X-aspect for any α of Diophantine type $K<4$; cf. Lemma 8.2 and Lemma 8.4; however this does not matter for us, since the end result is anyway subsumed by the $y^{\frac14}\log(1+y^{-1})$ bound coming from the last line of (32). The fact that, in this paper, we are not aiming to get below the exponent $\frac14$, will also be convenient at certain steps later in our discussion; cf. pp. 27–28.

It remains to bound the contribution from the first line in the right hand side of (32) to the expression in (27). This contribution equals:

$$=\sum_{n=1}^{\infty}\sum_{c=1}^{\infty}\int_{-\pi}^{\pi}\sum_{k\in N_c^{(n\xi_1)}}\Big(\int_{\mathbb{R}}\nu(-x+y\cot\theta)e\big((cn\xi_1-k)\omega x\big)\,dx\int_{\mathbb{R}/\mathbb{Z}}\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\,du\Big)$$
$$\times\mu\Big(\frac{c}{(c,k)}\Big)\frac{\varphi(c)}{\varphi(c/(c,k))}e\big(-n\omega c\xi_2\big)\frac{y\,d\theta}{\sin^2\theta}.\tag{51}$$

In order to bound this, we first fix $n\in\mathbb{Z}^+$ and $\theta\in(-\pi,\pi)$ (assuming $\theta\neq 0$), and write $a=y\cot\theta$, $\alpha_1:=n\xi_1$ and $\alpha_2:=-n\omega\xi_2$ ($\omega=\operatorname{sgn}(\theta)$ as before). Now

$$\sum_{c=1}^{\infty}\sum_{k\in N_c^{(\alpha_1)}}\Big(\int_{\mathbb{R}}\nu(a-x)e\big((c\alpha_1-k)\omega x\big)\,dx\int_{\mathbb{R}/\mathbb{Z}}\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{c^2y},\theta\Big)\,du\Big)\mu\Big(\frac{c}{(c,k)}\Big)\frac{\varphi(c)}{\varphi(c/(c,k))}e(c\alpha_2)$$
$$=\sum_{j=0}^{\infty}\frac{\mu(q_j)}{\varphi(q_j)}\sum_{k=1}^{\infty}\varphi(kq_j)e(kq_j\alpha_2)\Big(\int_{\mathbb{R}}\nu(a-x)e\big(k(q_j\alpha_1-p_j)\omega x\big)\,dx\int_{\mathbb{R}/\mathbb{Z}}\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{(kq_j)^2y},\theta\Big)\,du\Big),\tag{52}$$

where $\frac{p_j}{q_j}$ are the convergents of α_1. We will treat this double sum using integration by parts (cf. (64) below), and the key task then is to bound the following sum:

$$B_{n,\theta}(X) := \sum_{j=0}^{\infty} \frac{\mu(q_j)}{\varphi(q_j)} \sum_{1 \le k \le X/q_j} \varphi(kq_j) e(kq_j\alpha_2) \int_{\mathbb{R}} \nu\big(a-x\big) e\big(k(q_j\alpha_1 - p_j)\omega x\big)\, dx. \tag{53}$$

Lemma 9.2. *For any $q \in \mathbb{Z}^+$, $\alpha \in \mathbb{R}$ and $X \ge 1$,*

$$\sum_{1 \le k \le X} \varphi(kq) e(k\alpha) \ll \sigma_1(q) X^2 \sum_{1 \le j \le X} \min\Big(\frac{1}{j^2}, \frac{1}{Xj\langle j\alpha\rangle}\Big). \tag{54}$$

Proof. For any $k \in \mathbb{Z}^+$ we have

$$\varphi(kq) = \sum_{d \mid kq} \mu\Big(\frac{kq}{d}\Big) d = \sum_{d_1 \mid k} \sum_{\substack{d_2 \mid q \\ (d_2, k/d_1)=1}} \mu\Big(\frac{kq}{d_1 d_2}\Big) d_1 d_2 = \sum_{\substack{d_1 \mid k \\ (q,k/d_1)=1}} \sum_{d_2 \mid q} \mu\Big(\frac{kq}{d_1 d_2}\Big) d_1 d_2,$$

where the third equality holds since the conditions $d_1 \mid k$, $d_2 \mid q$, $(d_2, k/d_1) = 1$ and $\mu(\frac{kq}{d_1 d_2}) \neq 0$ together imply $(q, k/d_1) = 1$. Using this formula and then substituting $k = jd_1$, we get

$$\sum_{1 \le k \le X} \varphi(kq) e(k\alpha) = \sum_{d_2 \mid q} d_2 \sum_{\substack{1 \le j \le X \\ (q,j)=1}} \mu\Big(\frac{jq}{d_2}\Big) \sum_{1 \le d_1 \le X/j} d_1 e(d_1 j\alpha). \tag{55}$$

But for any $j, n \in \mathbb{Z}^+$ with $j\alpha \notin \mathbb{Z}$ we have

$$\sum_{d_1=1}^{n} d_1 e(d_1 j\alpha) = \frac{ne((n+2)j\alpha) - (n+1)e((n+1)j\alpha) + e(j\alpha)}{(e(j\alpha)-1)^2} \ll \min\Big(n^2, \frac{n}{\langle j\alpha\rangle} + \frac{1}{\langle j\alpha\rangle^2}\Big)$$
$$\ll \min\Big(n^2, \frac{n}{\langle j\alpha\rangle}\Big),$$

and the last bound is valid also when $j\alpha \in \mathbb{Z}$. Using this bound in (55) we obtain (54). $\square$

Lemma 9.3. *Let $q \in \mathbb{Z}^+$, $\alpha, \beta \in \mathbb{R}$, $Y \ge 1$ and $g \in \mathrm{C}_c(\mathbb{R})$. Let L and L' be positive real numbers such that* $\mathrm{supp}(g) \subset [A, A+L] \subset [-L', L']$, *for some $A \in \mathbb{R}$. Then*

$$\sum_{1 \le k \le Y} \varphi(kq) e(k\alpha) \int_{\mathbb{R}} g(x) e(k\beta x)\, dx$$
$$\ll \|g\|_{\mathrm{L}^\infty} \sigma_1(q) \begin{cases} Y \min\big(LY, |\beta|^{-1}\big)\big(1 + \log^+\big(LY|\beta|\big)\big)^2 \,\mathfrak{M}_\alpha\big(\frac{\min(LY,|\beta|^{-1})}{L'}\big) & \text{if } L|\beta| \le 10; \\ LY(1+\log Y)^2 & \text{if } L|\beta| \ge \frac{1}{10}. \end{cases} \tag{56}$$

(Thus both bounds are valid when $\frac{1}{10} \le L|\beta| \le 10$.)

Proof. Changing order of summation and integration and applying Lemma 9.2, we have

$$\sum_{1 \le k \le Y} \varphi(kq) e(k\alpha) \int_{\mathbb{R}} g(x) e(k\beta x)\, dx \ll \sigma_1(q) Y^2 \sum_{1 \le j \le Y} \int_{\mathbb{R}} \big|g(x)\big| \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx. \tag{57}$$

Here

$$\int_{\mathbb{R}} \big|g(x)\big| \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx \le \|g\|_{\mathrm{L}^\infty} \int_A^{A+L} \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx$$
$$\le \|g\|_{\mathrm{L}^\infty} \int_{-L}^{L} \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j\beta x\rangle}\Big)\, dx, \tag{58}$$

where the last inequality is holds since the function $x \mapsto \min(\frac{1}{j^2}, \frac{1}{Yj\langle j\beta x\rangle})$ is even, periodic with period $\frac{1}{j|\beta|}$, and decreasing in $[0, \frac{1}{2j|\beta|}]$ (if $\beta \neq 0$). If $Lj|\beta| \geq \frac{1}{10}$ then (58) is (using $1 \leq j \leq Y$):

$$\ll \|g\|_{\mathrm{L}^\infty} L \int_{\mathbb{R}/\mathbb{Z}} \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle y\rangle}\Big)\, dy \ll \frac{\|g\|_{\mathrm{L}^\infty} L}{jY}\Big(1+\log(Y/j)\Big). \tag{59}$$

On the other hand if $Lj|\beta| \leq 1$ then (58) is

$$\ll \|g\|_{\mathrm{L}^\infty} \frac{1}{j|\beta|} \int_0^{Lj|\beta|} \min\Big(\frac{1}{j^2}, \frac{1}{Yjy}\Big)\, dy \asymp \frac{\|g\|_{\mathrm{L}^\infty}}{j^2} \min\Big(L, \frac{1}{Y|\beta|}\Big)\Big(1+\log^+\big(LY|\beta|\big)\Big). \tag{60}$$

We also note an alternative bound in a special case: If $\langle j\alpha\rangle \geq 2jL'|\beta|$ then for all x in the support of g we have $|j\beta x| \leq \frac{1}{2}\langle j\alpha\rangle$ and thus $\langle j(\alpha+\beta x)\rangle \geq \frac{1}{2}\langle j\alpha\rangle$; therefore

$$\int_{\mathbb{R}} |g(x)| \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx \ll \|g\|_{\mathrm{L}^\infty} L \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j\alpha\rangle}\Big). \tag{61}$$

If $L|\beta| \geq \frac{1}{10}$ then (57) and (59) immediately imply that the second bound in (56) holds. Hence from now on we assume $L|\beta| \leq 10$, and our task is to prove the first bound in (56).

Case I: $LY|\beta| \geq 1$**.** In this case we split our sum over $1 \leq j \leq Y$ into two parts corresponding to $j < (L|\beta|)^{-1}$ and $j \geq (L|\beta|)^{-1}$, respectively. For $1 \leq j < (L|\beta|)^{-1}$ we have, by (60) and (61):

$$\begin{aligned}\int_{\mathbb{R}} |g(x)| \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx &\ll \|g\|_{\mathrm{L}^\infty} \left\{\begin{matrix} \frac{1}{Y|\beta|j^2}(1+\log(LY|\beta|)) & \text{(in general)} \\ \frac{L}{Yj\langle j\alpha\rangle} & \text{(if } \langle j\alpha\rangle \geq 2jL'|\beta|) \end{matrix}\right\} \\ &\ll \|g\|_{\mathrm{L}^\infty} \frac{1}{Y|\beta|} \min\Big(\frac{1}{j^2}, \frac{L'|\beta|}{j\langle j\alpha\rangle}\Big)\Big(1+\log(LY|\beta|)\Big),\end{aligned}$$

where we used $L \leq 2L'$. Hence

$$\sum_{1\leq j<(L|\beta|)^{-1}} \int_{\mathbb{R}} |g(x)| \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx \ll \|g\|_{\mathrm{L}^\infty} \frac{1+\log(LY|\beta|)}{Y|\beta|} \sum_{j=1}^{\infty} \min\Big(\frac{1}{j^2}, \frac{L'|\beta|}{j\langle j\alpha\rangle}\Big).$$

For the remaining sum we have, by (59),

$$\begin{aligned}\sum_{(L|\beta|)^{-1}\leq j\leq Y} \int_{\mathbb{R}} |g(x)| \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx &\ll \frac{\|g\|_{\mathrm{L}^\infty} L}{Y} \sum_{(L|\beta|)^{-1}\leq j\leq Y} \frac{1}{j}\Big(1+\log(Y/j)\Big) \\ &\ll \|g\|_{\mathrm{L}^\infty} \frac{L}{Y}\Big(1+\log(LY|\beta|)\Big)^2.\end{aligned}$$

Now (56) follows from (57) and the last two bounds, since we are assuming $Y^{-1} \leq L|\beta| \leq 10$.

Case II: $LY|\beta| < 1$**.** Then for all $1 \leq j \leq Y$, by (60) and (61),

$$\begin{aligned}\int_{\mathbb{R}} |g(x)| \min\Big(\frac{1}{j^2}, \frac{1}{Yj\langle j(\alpha+\beta x)\rangle}\Big)\, dx &\ll \|g\|_{\mathrm{L}^\infty} \left\{\begin{matrix} \frac{L}{j^2} & \text{(in general)} \\ \frac{L}{Yj\langle j\alpha\rangle} & \text{(if } \langle j\alpha\rangle \geq 2jL'|\beta|) \end{matrix}\right\} \\ &\ll \|g\|_{\mathrm{L}^\infty} L \min\Big(\frac{1}{j^2}, \frac{L'/L}{Yj\langle j\alpha\rangle}\Big).\end{aligned}$$

Hence again (56) follows from (57). □

Recall that we wish to bound $B_{n,\theta}(X)$ for $X \geq 1$ (cf. (53)). For each $j \geq 0$ such that $q_j \leq X$, we apply Lemma 9.3 with $g(x) = \nu(a-x)$, $\beta = (q_j\alpha_1 - p_j)\omega$, $\alpha = q_j\alpha_2$, $q = q_j$ and $Y = X/q_j$. Note that $\mathrm{supp}(g) \subset [a-L, a+L]$, since $\mathrm{supp}(\nu) \subset [-L, L]$ by assumption. Also

$(2q_{j+1})^{-1} < |\beta| < q_{j+1}^{-1}$ (cf., e.g., [17, Ch. X]). Hence Lemma 9.3 implies

$$\sum_{1\le k\le X/q_j} \varphi(kq_j)e(kq_j\alpha_2)\Big(\int_{\mathbb{R}} \nu\big(a-x\big)e\big(k(q_j\alpha_1-p_j)\omega x\big)\,dx\Big)$$
$$\ll \|\nu\|_{\mathrm{L}^\infty}\sigma_1(q_j)\begin{cases} \frac{X}{q_j}\min(\frac{LX}{q_j},q_{j+1})\big(1+\log^+(\frac{LX}{q_jq_{j+1}})\big)^2\mathfrak{M}_{q_j\alpha_2}\big(\frac{\min(LX/q_j\,,\,q_{j+1})}{L+|a|}\big) & \text{if } q_{j+1}\ge L\\ \frac{LX}{q_j}(1+\log(\frac{X}{q_j}))^2 & \text{if } q_{j+1}<L.\end{cases}$$

In order to get a bound on $B_{n,\theta}(X)$, we multiply the last bound with $|\mu(q_j)|\varphi(q_j)^{-1}$, and then add over all $j\ge 0$ for which $q_j\le X$. We split the set of these j into three disjoint parts, according to the following conditions:

$$\begin{cases}\text{(i) } q_{j+1}<L;\\ \text{(ii) } q_{j+1}\ge L \text{ and } q_jq_{j+1}\le LX;\\ \text{(iii) } q_{j+1}\ge L \text{ and } q_jq_{j+1}> LX.\end{cases} \tag{62}$$

We thus obtain

$$\begin{aligned} B_{n,\theta}(X) \ll \|\nu\|_{\mathrm{L}^\infty}\Big\{ &\sum_j^{(i)} \frac{|\mu(q_j)|\sigma_1(q_j)}{\varphi(q_j)}\frac{LX}{q_j}\Big(1+\log\Big(\frac{X}{q_j}\Big)\Big)^2\\ &+\sum_j^{(ii)} \frac{|\mu(q_j)|\sigma_1(q_j)}{\varphi(q_j)}\frac{Xq_{j+1}}{q_j}\Big(1+\log\Big(\frac{LX}{q_jq_{j+1}}\Big)\Big)^2\mathfrak{M}_{q_j\alpha_2}\Big(\frac{q_{j+1}}{L+|a|}\Big)\\ &+\sum_j^{(iii)} \frac{|\mu(q_j)|\sigma_1(q_j)}{\varphi(q_j)}\frac{LX^2}{q_j^2}\mathfrak{M}_{q_j\alpha_2}\Big(\frac{LX}{q_j(L+|a|)}\Big)\Big\}. \end{aligned} \tag{63}$$

Let us first record a trivial bound on $B_{n,\theta}(X)$.

Lemma 9.4. *For any positive integer q, $\frac{|\mu(q)|\sigma_1(q)}{\varphi(q)} \ll (\log\log(q+2))^4$.*

Proof. The bound is trivial unless $\mu(q)\ne 0$, i.e. q is squarefree. For squarefree q,

$$\frac{|\mu(q)|\sigma_1(q)}{\varphi(q)} = \prod_{p|q}\frac{p+1}{p-1} \le \prod_{p|q}(1+4p^{-1}) \le \prod_{p|q}(1+p^{-1})^4 \ll (\log\log(q+2))^4,$$

e.g. by [20, Theorems 4 and 7]. □

Lemma 9.5. *For all $X\ge 1$, $B_{n,\theta}(X) \ll \|\nu\|_{\mathrm{L}^\infty} LX^2$.*

Proof. Using Lemma 9.4 and (for (ii)) the fact that $u(1+\log(LX/u))^2 \ll LX$ for all $u\in[1,LX]$, we obtain (for any fixed $\varepsilon>0$)

$$B_{n,\theta}(X) \ll_\varepsilon \|\nu\|_{\mathrm{L}^\infty}\Big\{LX\big(1+\log X\big)^2\sum_j^{(i)} q_j^{-1+\varepsilon} + LX^2\sum_j^{(ii)} q_j^{\varepsilon-2} + LX^2\sum_j^{(iii)} q_j^{\varepsilon-2}\Big\}.$$

Here the first sum is bounded using the fact that the sequence $\{q_j\}$ grows geometrically in the precise sense that $q_{j+2}\ge q_{j+1}+q_j\ge 2q_j$ for all j (cf. [17, (10.2.2)]); the remaining two sums are bounded trivially. This gives the stated bound. □

Using (52), (53) and (33) with $m=3$, together with the fact that $B_{n,\theta}(X) \ll_{\nu,L} X^2$ as $X\to\infty$, by Lemma 9.5, we see that the expression in (51) can be rewritten as:

$$-\sum_{n=1}^\infty \int_{-\pi}^{\pi}\int_1^\infty\Big(\frac{\partial}{\partial X}\int_{\mathbb{R}/\mathbb{Z}}\widetilde{f}_n\Big(u,\frac{\sin^2\theta}{X^2y},\theta\Big)\,du\Big)\,B_{n,\theta}(X)\,dX\,\frac{y\,d\theta}{\sin^2\theta}. \tag{64}$$

By Lemma 4.4 we have, for any fixed $m \in \mathbb{Z}_{\geq 0}$,

$$\frac{\partial}{\partial X} \int_{\mathbb{R}/\mathbb{Z}} \widetilde{f}_n\Big(u, \frac{\sin^2\theta}{X^2 y}, \theta\Big)\, du \ll_m \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}} X^{-1} \min\Big(1, \Big(\frac{|\sin\theta|}{nX\sqrt{y}}\Big)^m\Big).$$

Hence the expression in (51) is

$$\ll_m \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}}\, y \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} \int_1^{\infty} \big|B_{n,\theta}(X)\big|\, \min\Big(1, \Big(\frac{|\sin\theta|}{nX\sqrt{y}}\Big)^m\Big) \frac{dX\, d\theta}{X \sin^2\theta}. \tag{65}$$

The following three lemmas will allow us to further simplify the bound.

Lemma 9.6. *For any $m \geq 2$, $0 < \sigma < 1$ and $0 < y < 1$ we have*

$$\sum_{n=1}^{\infty} \int_{-\pi}^{\pi} \int_1^{\infty} X^{1+\sigma} \min\Big(1, \Big(\frac{|\sin\theta|}{nX\sqrt{y}}\Big)^m\Big) \frac{dX\, d\theta}{X \sin^2\theta} \ll_{m,\sigma} y^{-\frac{1}{2}(1+\sigma)}.$$

Proof. Using (36) we see that the given expression is

$$\ll_m \sum_{n=1}^{\infty} \int_1^{\infty} (nX\sqrt{y})^{-1}(1 + nX\sqrt{y})^{1-m}\, X^{\sigma}\, dX \ll_m \sum_{n=1}^{\infty} \begin{cases} (n\sqrt{y})^{-m} & \text{if } n\sqrt{y} \geq 1 \\ (n\sqrt{y})^{-1-\sigma} & \text{if } n\sqrt{y} \leq 1, \end{cases}$$

and this gives the stated bound. □

The next lemma generalizes (36).

Lemma 9.7. *For $m \geq 2$, $a > 0$, and $0 \leq \delta \leq 1$, we have*

$$\int_{\substack{\theta \in (-\pi,\pi) \\ |\sin\theta| < \delta}} \min\Big(1, \big(a^{-1}|\sin\theta|\big)^m\Big) \frac{d\theta}{\sin^2\theta} \asymp_m \left\{\begin{matrix} a^{-m}\delta^{m-1} & \textit{if } \delta \leq a \\ a^{-1} & \textit{if } \delta \geq a \end{matrix}\right\} = a^{-1} \min(1, (\delta/a)^{m-1}).$$

Proof. This is seen by a direct computation. □

Lemma 9.8. *Let $\sigma \in [0, \frac{1}{2}]$, $m \geq 3$ and $0 < y < 1$. If the integral over θ in* (65) *is restricted by the condition $|\sin\theta| \leq y^{\sigma}$, the resulting expression is $\ll_m \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}} \|\nu\|_{\mathrm{L}^\infty}\, L y^{\sigma}$.*

Proof. Using Lemma 9.7 and Lemma 9.5 we see that the expression in question is

$$\ll \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}} \|\nu\|_{\mathrm{L}^\infty}\, Ly \sum_{n=1}^{\infty} \frac{1}{n\sqrt{y}} \int_1^{\infty} \min\Big(1, \Big(\frac{y^{\sigma}}{nX\sqrt{y}}\Big)^{m-1}\Big)\, dX$$
$$\ll \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}} \|\nu\|_{\mathrm{L}^\infty}\, Ly \sum_{n=1}^{\infty} \frac{1}{n\sqrt{y}} \left\{\begin{matrix} y^{\sigma-\frac{1}{2}}/n & \text{if } n \leq y^{\sigma-\frac{1}{2}} \\ (y^{\sigma-\frac{1}{2}}/n)^{m-1} & \text{if } n \geq y^{\sigma-\frac{1}{2}} \end{matrix}\right\} \ll \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}} \|\nu\|_{\mathrm{L}^\infty}\, Ly^{\sigma}.$$

□

We saw in the proof of Lemma 9.5 that the $\sum^{(i)}$-sum in (63) is $\ll_\varepsilon LX^{1+\varepsilon}$; hence by Lemma 9.6 the total contribution from the $\sum^{(i)}$-sum in (63) to the expression in (65) is $\ll_{m,\varepsilon} \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}} \|\nu\|_{\mathrm{L}^\infty} L y^{\frac{1}{2}-\varepsilon}$. We also use Lemma 9.8 with $\sigma = \frac{1}{2}$, and note that for $|\sin\theta| > y^{\frac{1}{2}}$ we have $|a| = y|\cot\theta| < y^{\frac{1}{2}} < 1$ and $L + |a| \leq 2L$ in (63); hence we conclude that the whole expression in (65) is

$$\ll_{m,\varepsilon} \|f\|_{\mathrm{C}_\mathrm{b}^{m+1}} \|\nu\|_{\mathrm{L}^\infty} \Big\{ Ly^{\frac{1}{2}-\varepsilon} + y \sum_{n=1}^{\infty} \int_1^{\infty} B_n(X) \int_{-\pi}^{\pi} \min\Big(1, \Big(\frac{|\sin\theta|}{nX\sqrt{y}}\Big)^m\Big) \frac{d\theta}{\sin^2\theta} \frac{dX}{X} \Big\}, \tag{66}$$

where (keeping from now on $\alpha_2 := n\xi_2$, and using the fact that $\mathfrak{M}_\alpha(X) \equiv \mathfrak{M}_{-\alpha}(X)$)

$$B_n(X) = \sum_j^{(ii)} \frac{|\mu(q_j)|\sigma_1(q_j)}{\varphi(q_j)} \frac{Xq_{j+1}}{q_j}\Big(1+\log\Big(\frac{LX}{q_jq_{j+1}}\Big)\Big)^2 \mathfrak{M}_{q_j\alpha_2}\Big(\frac{q_{j+1}}{2L}\Big) + \sum_j^{(iii)} \frac{|\mu(q_j)|\sigma_1(q_j)}{\varphi(q_j)} \frac{LX^2}{q_j^2}\mathfrak{M}_{q_j\alpha_2}\Big(\frac{X}{2q_j}\Big). \tag{67}$$

Using (36) to bound the integral over θ we conclude that the expression in (66) is

$$\ll_{m,\varepsilon} \|f\|_{\mathrm{C}_{\mathrm{b}}^{m+1}}\|\nu\|_{\mathrm{L}^\infty}\Big\{Ly^{\frac12-\varepsilon} + y^{\frac12}\sum_{n=1}^\infty \frac1n \int_1^\infty \frac{B_n(X)}{(1+n\sqrt{y}X)^{m-1}X^2}\,dX\Big\}. \tag{68}$$

Now note that

$$\int_1^\infty \frac{B_n(X)}{(1+n\sqrt{y}X)^{m-1}X^2}\,dX = \sum_{\substack{j\geq 0\\(q_{j+1}\geq L)}} \frac{|\mu(q_j)|\sigma_1(q_j)}{\varphi(q_j)}\Big\{\frac{L}{q_j^2}\int_1^{q_jq_{j+1}/L}\frac{\mathfrak{M}_{q_j\alpha_2}(\frac{X}{2q_j})}{(1+n\sqrt{y}X)^{m-1}}\,dX + \frac{q_{j+1}}{q_j}\mathfrak{M}_{q_j\alpha_2}\Big(\frac{q_{j+1}}{2L}\Big)\int_{q_jq_{j+1}/L}^\infty \frac{(1+\log(\frac{LX}{q_jq_{j+1}}))^2}{(1+n\sqrt{y}X)^{m-1}X}\,dX\Big\}. \tag{69}$$

Let us write $\mathfrak{M}^1_\alpha(X)$ for the integral of $\mathfrak{M}_\alpha(X)$:

$$\mathfrak{M}^1_\alpha(X) := \int_0^X \mathfrak{M}_\alpha(Y)\,dY = \sum_{n=1}^\infty \min\Big(\frac{X}{n^2},\frac{1}{n\langle n\alpha\rangle}\Big)\Big(1+\log^+\Big(\frac{X\langle n\alpha\rangle}{n}\Big)\Big). \tag{70}$$

We have the bound

$$\int_1^{q_jq_{j+1}/L}\frac{\mathfrak{M}_{q_j\alpha_2}(\frac{X}{2q_j})}{(1+n\sqrt{y}X)^{m-1}}\,dX \ll \int_0^{q_jq_{j+1}/L}\mathfrak{M}_{q_j\alpha_2}\Big(\frac{X}{2q_j}\Big)\,dX = 2q_j\mathfrak{M}^1_{q_j\alpha_2}\Big(\frac{q_{j+1}}{2L}\Big).$$

If $q_jq_{j+1}/L > (n\sqrt{y})^{-1}$ then the same integral can also be bounded more sharply as (assuming $m\geq 3$, and using the fact that $\mathfrak{M}_\alpha(X) \leq X^{-1}\mathfrak{M}^1_\alpha(X)$ for all X)

$$\ll \int_0^{(n\sqrt{y})^{-1}}\mathfrak{M}_{q_j\alpha_2}\Big(\frac{X}{2q_j}\Big)\,dX + \int_{(n\sqrt{y})^{-1}}^\infty \frac{\mathfrak{M}_{q_j\alpha_2}(\frac{1}{2q_jn\sqrt{y}})}{(n\sqrt{y}X)^{m-1}}\,dX \ll 2q_j\mathfrak{M}^1_{q_j\alpha_2}\Big(\frac{1}{2q_jn\sqrt{y}}\Big).$$

Also by an easy computation,

$$\int_{q_jq_{j+1}/L}^\infty \frac{(1+\log(\frac{LX}{q_jq_{j+1}}))^2}{(1+n\sqrt{y}X)^{m-1}X}\,dX \ll \begin{cases} (1+\log(\frac{L}{q_jq_{j+1}n\sqrt{y}}))^3 & \text{if } q_jq_{j+1}/L \leq (n\sqrt{y})^{-1} \\ (\frac{L}{q_jq_{j+1}n\sqrt{y}})^{m-1} & \text{if } (n\sqrt{y})^{-1} \leq q_jq_{j+1}/L. \end{cases} \tag{71}$$

Adding up the bounds (using the fact that if $(n\sqrt{y})^{-1} \leq q_jq_{j+1}/L$ then $\mathfrak{M}_{q_j\alpha_2}(\frac{q_{j+1}}{2L}) \leq 2q_jn\sqrt{y}\,\mathfrak{M}^1_{q_j\alpha_2}(\frac{1}{2q_jn\sqrt{y}})$), we conclude

$$\int_1^\infty \frac{B_n(X)}{(1+n\sqrt{y}X)^{m-1}X^2}\,dX \ll L\sum_{\substack{j\geq 0\\(q_{j+1}\geq L)}} \frac{|\mu(q_j)|\sigma_1(q_j)}{q_j\varphi(q_j)}\Big(1+\log^+\Big(\frac{L}{q_jq_{j+1}n\sqrt{y}}\Big)\Big)^3\mathfrak{M}^1_{q_j\alpha_2}\Big(\min\Big(\frac{q_{j+1}}{L},\frac{1}{q_jn\sqrt{y}}\Big)\Big). \tag{72}$$

Using also $q_{j+1} < \langle q_j\alpha_1\rangle^{-1} < 2q_{j+1}$ we get

$$\ll L\sum_{q=1}^\infty \frac{|\mu(q)|\sigma_1(q)}{q\varphi(q)}\Big(1+\log^+\Big(\frac{L\langle q\alpha_1\rangle}{qn\sqrt{y}}\Big)\Big)^3\mathfrak{M}^1_{q\alpha_2}\Big(\min\Big(\frac{1}{L\langle q\alpha_1\rangle},\frac{1}{qn\sqrt{y}}\Big)\Big). \tag{73}$$

Adding now over n (recalling $\alpha_1 = n\xi_1$, $\alpha_2 = n\xi_2$) we get, after substituting $k = qn$ and using $\sum_{q|k} \frac{|\mu(q)|\sigma_1(q)}{\varphi(q)} \ll_\varepsilon k^\varepsilon$,

$$y^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{1}{n} \int_1^\infty \frac{B_n(X)}{(1+n\sqrt{y}X)^{m-1}X^2}\, dX$$

$$\ll_\varepsilon Ly^{\frac{1}{2}} \sum_{k=1}^{\infty} k^{\varepsilon-1} \Big(1 + \log^+\Big(\frac{L\langle k\xi_1\rangle}{k\sqrt{y}}\Big)\Big)^3 \mathfrak{M}^1_{k\xi_2}\Big(\min\Big(\frac{1}{L\langle k\xi_1\rangle}, \frac{1}{k\sqrt{y}}\Big)\Big). \tag{74}$$

Remark 9.2. In (73) we overestimate a sum over $\{q_j\}$ by a sum over *all* $q \in \mathbb{Z}^+$; the (simple) reason why this is not very wasteful will be seen in the next paragraph. It is clear from this that there is some flexibility in choosing the set N in (32) so as to make our proof work.

As we will see, the vast majority of the terms in the sum in (74) can be bounded trivially. First of all, using $\mathfrak{M}^1_{k\xi_2}(X) \ll X$ and the fact that for any $k, y > 0$ the function $f(\delta) = (1 + \log^+(\frac{\delta}{k\sqrt{y}}))^3 \min(\delta^{-1}, \frac{1}{k\sqrt{y}})$ satisfies $f(\delta_2) \ll f(\delta_1)$ for all $0 < \delta_1 \le \delta_2$, it follows that the contribution from all k with $\langle k\xi_1\rangle \ge (2k)^{-1}$ in the right hand side of (74) is

$$\ll Ly^{\frac{1}{2}} \sum_{k=1}^{\infty} k^{\varepsilon-1}\Big(1 + \log^+\Big(\frac{L}{k^2\sqrt{y}}\Big)\Big)^3 \min\Big(\frac{k}{L}, \frac{1}{k\sqrt{y}}\Big) \ll_\varepsilon L^{\frac{1}{2}+\varepsilon} y^{\frac{1}{4}-\varepsilon}.$$

It remains to consider the contribution from all k with $\langle k\xi_1\rangle < (2k)^{-1}$. From now on, let $\frac{p_j}{q_j}$ for $j \in \mathbb{Z}_{\ge 0}$ be the jth convergent of the continued fraction expansion of ξ_1. Then each $k \ge 1$ satisfying $\langle k\xi_1\rangle < (2k)^{-1}$ is known to be of the form $k = \ell q_j$ for some $j \in \mathbb{Z}_{\ge 0}$ and some $\ell \in \mathbb{Z}^+$ which is so small that $\langle k\xi_1\rangle = \ell\langle q_j\xi_1\rangle$ and thus $\frac{\ell}{2q_{j+1}} < \langle k\xi_1\rangle < \frac{\ell}{q_{j+1}}$ (cf., e.g., [17, Thm. 184]). Hence the total contribution from all such k to the right hand side of (74) is

$$\ll Ly^{\frac{1}{2}} \sum_{j=0}^{\infty}\sum_{\ell=1}^{\infty} (\ell q_j)^{\varepsilon-1}\Big(1 + \log^+\Big(\frac{L}{q_jq_{j+1}\sqrt{y}}\Big)\Big)^3 \mathfrak{M}^1_{\ell q_j\xi_2}\Big(\ell^{-1}\min\Big(\frac{q_{j+1}}{L}, \frac{1}{q_j\sqrt{y}}\Big)\Big). \tag{75}$$

Recall the summation formula for $\mathfrak{M}^1_\alpha(X)$; cf. (70). Let us write $\mathfrak{M}^{1,\varepsilon}_\alpha(X)$ for the analogous sum with an extra factor n^ε in each term:

$$\mathfrak{M}^{1,\varepsilon}_\alpha(X) := \sum_{n=1}^{\infty} n^\varepsilon \min\Big(\frac{X}{n^2}, \frac{1}{n\langle n\alpha\rangle}\Big)\Big(1 + \log^+\Big(\frac{X\langle n\alpha\rangle}{n}\Big)\Big). \tag{76}$$

Lemma 9.9. *For any $X > 0$, $\alpha \in \mathbb{R}$, $\varepsilon > 0$,*

$$\sum_{\ell=1}^{\infty} \ell^{\varepsilon-1}\mathfrak{M}^1_{\ell\alpha}(\ell^{-1}X) \ll_\varepsilon \mathfrak{M}^{1,2\varepsilon}_\alpha(X), \tag{77}$$

Proof. Using (70) and substituting $k = \ell n$ we obtain

$$\sum_{\ell=1}^{\infty} \ell^{\varepsilon-1}\mathfrak{M}^1_{\ell\alpha}(\ell^{-1}X) = \sum_{k=1}^{\infty}\Big(\sum_{\ell|k}\ell^\varepsilon\Big)\min\Big(\frac{X}{k^2}, \frac{1}{k\langle k\alpha\rangle}\Big)\Big(1 + \log^+\Big(\frac{X\langle k\alpha\rangle}{k}\Big)\Big).$$

Now the desired bound follows using $\sum_{\ell|k}\ell^\varepsilon \ll_\varepsilon k^{2\varepsilon}$. ☐

Lemma 9.10. *For any $X > 0$, $\alpha \in \mathbb{R}$, $0 < \varepsilon \le \frac{1}{2}$ we have $\mathfrak{M}^{1,\varepsilon}_\alpha(X) \ll X$.*

Proof. Simply note $\mathfrak{M}^{1,\varepsilon}_\alpha(X) \le (\sum_{n=1}^\infty n^{\varepsilon-2})X$. ☐

Using Lemma 9.9 we see that (75) is

$$\ll_\varepsilon Ly^{\frac{1}{2}} \sum_{j=0}^{\infty} q_j^{\varepsilon-1}\Big(1 + \log^+\Big(\frac{L}{q_jq_{j+1}\sqrt{y}}\Big)\Big)^3 \mathfrak{M}^{1,2\varepsilon}_{q_j\xi_2}\Big(\min\Big(\frac{q_{j+1}}{L}, \frac{1}{q_j\sqrt{y}}\Big)\Big). \tag{78}$$

Recall that $0 < y < 1$. Now let $j_0 \geq 0$ be the unique index satisfying

$$q_{j_0}^4 < y^{-\frac{1}{2}} \leq q_{j_0+1}^4.$$

Then the contribution from all $j < j_0$ in (78) is (using Lemma 9.10, keeping $\varepsilon \leq \frac{1}{4}$)

$$\ll Ly^{\frac{1}{2}} \sum_{j<j_0} (1+\log(Ly^{-1}))^3 \frac{q_{j+1}}{L} \ll y^{\frac{1}{2}} q_{j_0} (1+\log(Ly^{-1}))^3 < y^{\frac{3}{8}} (1+\log(Ly^{-1}))^3.$$

Also if $\frac{q_{j_0+1}}{Lq_{j_0}} \leq y^{-\frac{1}{4}}$ then the contribution from $j = j_0$ in (78) is

$$\ll Ly^{\frac{1}{2}} q_{j_0}^{\varepsilon} \Big(1 + \log(y^{-1}) + \log^+\Big(\frac{L}{q_{j_0+1}}\Big)\Big)^3 \frac{q_{j_0+1}}{Lq_{j_0}} \ll_\varepsilon Ly^{\frac{1}{4}-\varepsilon}.$$

Finally the contribution from all $j > j_0$ to (78) is

$$\ll Ly^{\frac{1}{2}} (1+\log(y^{-1}))^3 \sum_{j=j_0+1}^{\infty} q_j^{\varepsilon-1} \left\{ \begin{matrix} \frac{q_{j+1}}{L}(1+\log(\frac{L}{q_{j+1}}))^3 & \text{if } q_{j+1} < L \\ \frac{1}{q_j\sqrt{y}} & \text{if } q_{j+1} \geq L \end{matrix} \right\}$$

$$\ll Ly^{\frac{1}{2}} (1+\log(y^{-1}))^3 \Big(1 + y^{-\frac{1}{2}} \sum_{j=j_0+1}^{\infty} q_j^{\varepsilon-2}\Big) \ll_\varepsilon Ly^{\frac{1}{4}-\varepsilon}.$$

Now there remains *at most one* j to consider in the bound in (78): namely that $j \geq 0$, if any, which satisfies $q_j^4 < y^{-\frac{1}{2}} < (\frac{q_{j+1}}{Lq_j})^2$. In the case when such a j exists, let us write $q := q_j$ and $q' := q_{j+1}$. Collecting our bounds and recalling the definition of $\mathfrak{M}^{1,2\varepsilon}_{q\xi_2}(X)$, we have now proved that the right hand side of (74) is $\ll_\varepsilon Ly^{\frac{1}{4}-\varepsilon}$ if the special denominator q does not exist, and otherwise it is

(79)

$$\ll_\varepsilon Ly^{\frac{1}{4}-\varepsilon} + Ly^{\frac{1}{2}} q^{\varepsilon-1}\Big(1+\log^+\Big(\frac{L}{q'q\sqrt{y}}\Big)\Big)^3 \sum_{n=1}^{\infty} n^{2\varepsilon-1}\Big(1+\log^+\Big(\frac{X\langle nq\xi_2\rangle}{n}\Big)\Big) \min\Big(\frac{X}{n}, \frac{1}{\langle nq\xi_2\rangle}\Big)$$

where

$$X := \min\Big(\frac{1}{q\sqrt{y}}, \frac{q'}{L}\Big) > 1.$$

For the rest of this discussion we will assume that the special denominator q exists. In close analogy to what we have shown for (74), we will see that the vast majority of the terms in the sum in (79) can be bounded trivially. First, note that the total contribution to (79) from all $n \in \mathbb{Z}^+$ with $\langle nq\xi_2\rangle \geq (2n)^{-1}$ is

$$\ll_\varepsilon Ly^{\frac{1}{2}} q^{\varepsilon-1}\Big(1+\log^+\Big(\frac{L}{q'q\sqrt{y}}\Big)\Big)^3 (1+\log X) X^{\frac{1}{2}+\varepsilon} \ll_\varepsilon Ly^{\frac{1}{4}-\varepsilon}, \tag{80}$$

where the last relation holds since $\log^+(\frac{L}{q'q\sqrt{y}}) \leq \log(y^{-1})$ and $X \leq y^{-\frac{1}{2}}$. From now on we assume, without loss of generality, that ξ_2 *is irrational* (just as we did for ξ_1 on p. 19). Let $\frac{r_j}{s_j}$ for $j \in \mathbb{Z}_{\geq 0}$ be the jth convergent of the continued fraction expansion of $q\xi_2$. Then each $n \in \mathbb{Z}^+$ satisfying $\langle nq\xi_2\rangle < (2n)^{-1}$ is of the form $n = \ell s_j$ for some $j \in \mathbb{Z}_{\geq 0}$ and some $\ell \in \mathbb{Z}^+$ which is so small that $\langle nq\xi_2\rangle = \ell\langle s_j q\xi_2\rangle$ and thus $\frac{\ell}{2s_{j+1}} < \langle nq\xi_2\rangle < \frac{\ell}{s_{j+1}}$. Hence the total contribution from all these n to the sum in (79) is

$$\ll \sum_{j=0}^{\infty}\sum_{\ell=1}^{\infty} (\ell s_j)^{2\varepsilon-1}\Big(1+\log^+\Big(\frac{X}{s_j s_{j+1}}\Big)\Big)\ell^{-1}\min\Big(\frac{X}{s_j}, s_{j+1}\Big)$$

$$\ll \sum_{j=0}^{\infty} s_j^{2\varepsilon-1}\Big(1+\log^+\Big(\frac{X}{s_j s_{j+1}}\Big)\Big)\min\Big(\frac{X}{s_j}, s_{j+1}\Big). \tag{81}$$

Next let $j_0 \geq 0$ be the unique index satisfying $s_{j_0}^4 < X \leq s_{j_0+1}^4$. Then the contribution from all $j < j_0$ in (81) is $\leq (1+\log X)\sum_{j<j_0} s_{j+1} \ll_\varepsilon X^{\frac14+\varepsilon}$. Also if $\frac{s_{j_0+1}}{s_{j_0}} \leq \sqrt{X}$ then the contribution from $j = j_0$ in (81) is $\leq (1+\log X)s_{j_0}^{2\varepsilon}\frac{s_{j_0+1}}{s_{j_0}} \ll_\varepsilon X^{\frac12+\varepsilon}$. Finally the contribution from all $j > j_0$ in (81) is

$$\leq X(1+\log X)\sum_{j=j_0+1}^{\infty} s_j^{2\varepsilon-2} \ll X(1+\log X)s_{j_0+1}^{2\varepsilon-2} \ll_\varepsilon X^{\frac12+\varepsilon}.$$

Now there remains *at most one* j to consider in (81): namely that $j \geq 0$, if any, which satisfies $s_j^4 < X < (\frac{s_{j+1}}{s_j})^2$. In the case when such a j exists, let us write $s := s_j$ and $s' := s_{j+1}$. Collecting our bounds (recalling also the last relation in (80)), we have now proved that (79) is $\ll_\varepsilon Ly^{\frac14-\varepsilon}$ if the special denominator s does not exist, and otherwise it is

$$\ll_\varepsilon Ly^{\frac14-\varepsilon} + Ly^{\frac12}q^{\varepsilon-1}\Big(1+\log^+\Big(\frac{L}{q'q\sqrt{y}}\Big)\Big)^3 s^{2\varepsilon-1}\Big(1+\log^+\Big(\frac{X}{ss'}\Big)\Big)\min\Big(\frac{X}{s},s'\Big).$$

Recalling the definition of X, and writing $U = \frac{L}{q'q\sqrt{y}}$, $V = \frac{1}{qss'\sqrt{y}}$, the above is seen to be

$$\leq Ly^{\frac14-\varepsilon} + L(sq)^{2\varepsilon-2}\frac{(1+\log(\max(1,U,V)))^4}{\max(1,U,V)} \ll_\varepsilon Ly^{\frac14-\varepsilon} + L\big((sq)^{-2}\min(1,U^{-1},V^{-1})\big)^{1-\varepsilon}$$
$$\leq L\big(y^{\frac14-\varepsilon} + b_{\boldsymbol{\xi},L}(y)^{1-\varepsilon}\big), \tag{82}$$

Here the last relation follows from (3), since $q' < \langle q\xi_1\rangle^{-1} \leq s\langle sq\xi_1\rangle^{-1}$ and $s' < \langle sq\xi_2\rangle^{-1}$ and therefore, writing $q_0 := qs$, we have $(qs)^{-2}U^{-1} < \frac{\sqrt{y}}{Lq_0\langle q_0\xi_1\rangle}$ and $(qs)^{-2}V^{-1} < \frac{\sqrt{y}}{q_0\langle q_0\xi_2\rangle}$. Taking $m = 3$ it follows that (68) is $\ll L\|\nu\|_{\mathrm{L}^\infty}\|f\|_{\mathrm{C}_{\mathrm{b}}^4}\big(y^{\frac14} + b_{\boldsymbol{\xi},L}(y)\big)^{1-4\varepsilon}$. Hence, replacing ε by $\frac14\varepsilon$, we have now completed the proof of Theorem 3.1. □ □ □

Remark 9.3. Let us note that the last step in (82) is essentially sharp. Indeed, if $b_{\boldsymbol{\xi},L}(y) > 2y^{\frac14}$ (and $y < 1$, $L \geq 1$) then the "special denominators" q, s introduced above do in fact exist, and

$$b_{\boldsymbol{\xi},L}(y) = \min\Big(\frac{1}{(sq)^2}, \frac{\sqrt{y}}{Ls^2q\langle q\xi_1\rangle}, \frac{\sqrt{y}}{sq\langle sq\xi_2\rangle}\Big) \tag{83}$$

(and here $\frac12 < q'\langle q\xi_1\rangle < 1$ and $\frac12 < s'\langle sq\xi_2\rangle < 1$, so that the expression in (83) is comparable with $(sq)^{-2}\min(1,U^{-1},V^{-1})$ in the notation used in (82)).

To prove this claim, assume $b_{\boldsymbol{\xi},L}(y) > 2y^{\frac14}$ and let q_0 be a positive integer for which the maximum in (3) is attained. Then $\langle q_0\xi_1\rangle < \frac{y^{1/4}}{2Lq_0} < \frac{1}{2q_0}$; therefore there exist $q, s \in \mathbb{Z}^+$ such that $q_0 = sq$, q is a denominator of a convergent of the continuous fraction expansion of ξ_1 and $\langle q_0\xi_1\rangle = s\langle q\xi_1\rangle$. It follows that (83) holds for these q, s. Next note that also $\langle sq\xi_2\rangle < \frac{y^{1/4}}{2sq} < \frac{1}{2s}$; therefore there exist $\widetilde{s}, k \in \mathbb{Z}^+$ such that $s = k\widetilde{s}$, $\widetilde{s}$ is a denominator of a convergent of $q\xi_2$, and $\langle sq\xi_2\rangle = k\langle\widetilde{s}q\xi_2\rangle$; and now the assumption that the maximum in (3) is attained at q_0 forces $k = 1$, i.e. s itself is a denominator of a convergent of $q\xi_2$. Finally note that since (83) is larger than $2y^{\frac14}$, we have $(sq)^4 < \frac14 y^{-\frac12}$, $(\frac{1}{Lq\langle q\xi_1\rangle})^2 > 4\frac{s^4}{\sqrt{y}}$ and $(\frac{1}{s\langle sq\xi_2\rangle})^2 > 4\frac{q^2}{\sqrt{y}}$, and these inequalities imply that q, s are in fact the "special denominators" introduced above.

10. Basic properties of the majorant $b_{\boldsymbol{\xi},L}(y)$

In this section we will note some basic properties of the majorant $b_{\boldsymbol{\xi},L}(y)$ appearing in the bound in our Theorems 1.2 and 3.1. This is helpful for clarifying the content of those theorems in certain parameter regimes; we will also make use of the facts proved here in our treatment of general $U^{\mathbb{R}}$-orbits in the next section.

Lemma 10.1. *For any $\boldsymbol{\xi} \in \mathbb{R}^2$ and any $L_1, L_2, y_1, y_2 > 0$ we have*

$$\min\left(\tfrac{L_2}{L_1}, 1\right) \min\left(\tfrac{y_1}{y_2}, 1\right)^{\frac{1}{2}} b_{\boldsymbol{\xi},L_2}(y_2) \leq b_{\boldsymbol{\xi},L_1}(y_1) \leq \max\left(\tfrac{L_2}{L_1}, 1\right) \max\left(\tfrac{y_1}{y_2}, 1\right)^{\frac{1}{2}} b_{\boldsymbol{\xi},L_2}(y_2).$$

Proof. This is immediate from the definition, (3). □

In particular replacing L and/or y by numbers of the same order of magnitude does not change the order of magnitude of $b_{\boldsymbol{\xi},L}(y)$; we will use this fact several times in Section 11 without explicit mention.

Lemma 10.2. *For any $0 < y < 1$, $\boldsymbol{\xi} \in \mathbb{R}^2$, $L \geq 1$, and any integer $n \in [-L, L]$, we have $b_{\boldsymbol{\xi}U^n,L}(y) \asymp b_{\boldsymbol{\xi},L}(y)$.*

Proof. For every $q \in \mathbb{Z}^+$, $\langle q(n\xi_1 + \xi_2)\rangle \leq L\langle q\xi_1\rangle + \langle q\xi_2\rangle$; thus $\min(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\langle q(n\xi_1+\xi_2)\rangle}) \geq \frac{1}{2}\min(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\langle q\xi_2\rangle})$. Therefore $b_{\boldsymbol{\xi}U^n,L}(y) \geq \frac{1}{2} b_{\boldsymbol{\xi},L}(y)$. Similarly $b_{\boldsymbol{\xi},L}(y) \geq \frac{1}{2} b_{\boldsymbol{\xi}U^n,L}(y)$. □

The following lemma is in principle contained in the discussion on the last pages of Section 9. For clarity, we write out the short proof here.

Lemma 10.3. *Let $0 < y < 1$, $\boldsymbol{\xi} \in \mathbb{R}^2$ and $L \geq 1$, and assume that $b_{\boldsymbol{\xi},L}(y) > 2y^{\frac{1}{4}}$. Let q_0 be a positive integer where the maximum in (3) is attained. Then every positive integer q such that $\min(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\langle q\xi_2\rangle}) > 2y^{\frac{1}{4}}$ must be of the form $q = mq_0$ for some $m \in \mathbb{Z}^+$ which is so small that $\langle q\xi_1\rangle = m\langle q_0\xi_1\rangle$ and $\langle q\xi_2\rangle = m\langle q_0\xi_2\rangle$ (in particular q_0 is uniquely determined).*

Proof. We assume $\xi_1, \xi_2 \notin \mathbb{Q}$; the cases when $\xi_1 \in \mathbb{Q}$ or $\xi_2 \in \mathbb{Q}$ can then be treated by a limit argument. Let q be a positive integer satisfying $\min(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\langle q\xi_2\rangle}) > 2y^{\frac{1}{4}}$. Then $\langle q\xi_1\rangle < \frac{y^{1/4}}{2Lq} < \frac{1}{2q}$ and $\langle q\xi_2\rangle < \frac{y^{1/4}}{2q} < \frac{1}{2q}$. Therefore for both $j = 1, 2$, we have $q = m_j q_j$ for some $m_j, q_j \in \mathbb{Z}^+$ such that q_j is a denominator of a convergent of the continuous fraction expansion of ξ_j, and $\langle q\xi_j\rangle = m_j\langle q_j\xi_j\rangle$. Note that $q_j \leq q < y^{-\frac{1}{8}}$ and also, if we denote by q_j' the denominator of the "next" convergent of ξ_j, then $\frac{1}{2q_j'} < \langle q_j\xi_j\rangle < \frac{y^{1/4}}{2}$ and thus $q_j' > y^{-\frac{1}{4}} > y^{-\frac{1}{8}}$. Hence q_1 and q_2 are uniquely determined for our given $\boldsymbol{\xi}, L, y$; namely, q_j equals the largest number $< y^{-\frac{1}{8}}$ among all the denominators of the convergents of ξ_j. Let q_0 be the least common multiple of these two numbers q_1, q_2. It then follows that any number q as above must be of the form $q = mq_0$ for some $m \in \mathbb{Z}^+$ so small that $\langle q\xi_1\rangle = m\langle q_0\xi_1\rangle$ and $\langle q\xi_2\rangle = m\langle q_0\xi_2\rangle$; thus also q_0 is the unique number at which the maximum in (3) is attained. □

Let us define

$$\widetilde{b}_{\boldsymbol{\xi},L}(y) = b_{\boldsymbol{\xi},L}(y) + y^{\frac{1}{4}}. \tag{84}$$

Note that the obvious analogues of Lemmata 10.1 and 10.2 also hold for $\widetilde{b}$.

Lemma 10.4. *Let $0 < \eta < 1$. For any $0 < y < 1$, $\boldsymbol{\xi} \in \mathbb{R}^2$, $L \geq 1$ we have*

$$\sum_{|n| \leq L} \widetilde{b}_{\boldsymbol{\xi}U^n,1}(y)^\eta \asymp_\eta L\widetilde{b}_{\boldsymbol{\xi},L}(y)^\eta. \tag{85}$$

Proof. **Case 1.** Assume $b_{\boldsymbol{\xi},L}(y) \geq 8y^{\frac{1}{4}}$. Let q be the unique positive integer at which the maximum in (3) is attained (cf. Lemma 10.3). In particular then $q^2 \leq \frac{1}{8y^{1/4}}$, $\langle q\xi_1\rangle \leq \frac{y^{1/4}}{8Lq}$ and $\langle q\xi_2\rangle \leq \frac{y^{1/4}}{8q}$. We claim that, with the same q, for every integer n with $|n| \leq L$,

$$b_{\boldsymbol{\xi}U^n,1}(y) = \min\Big(\frac{1}{q^2}, \frac{\sqrt{y}}{q\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\langle q(n\xi_1 + \xi_2)\rangle}\Big). \tag{86}$$

To prove this, note that

$$\langle q(n\xi_1+\xi_2)\rangle \leq |n|\langle q\xi_1\rangle + \langle q\xi_2\rangle \leq L\langle q\xi_1\rangle + \langle q\xi_2\rangle \leq \frac{y^{1/4}}{4q}.$$

Hence the right hand side of (86) is $\geq 4y^{\frac14}$, and by Lemma 10.3, we have $q = mq_0$ for some $m \in \mathbb{Z}^+$, where q_0 is the positive integer for which the maximum for $b_{\boldsymbol{\xi} U^n,1}(y)$ is attained, and $\langle q\xi_1\rangle = m\langle q_0\xi_1\rangle$ and $\langle q(n\xi_1+\xi_2)\rangle = m\langle q_0(n\xi_1+\xi_2)\rangle$. Now

$$\langle q_0\xi_2\rangle = \langle q_0(n\xi_1+\xi_2-n\xi_1)\rangle \leq \frac{1}{m}\langle q(n\xi_1+\xi_2)\rangle + \frac{|n|}{m}\langle q\xi_1\rangle < \frac{1}{4m} + \frac{L}{8Lm} < \frac{1}{2m},$$

so that $\langle q\xi_2\rangle = m\langle q_0\xi_2\rangle$. Therefore $\min(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\langle q\xi_2\rangle}) = m^{-2}\min(\frac{1}{q_0^2}, \frac{\sqrt{y}}{Lq_0\langle q_0\xi_1\rangle}, \frac{\sqrt{y}}{q_0\langle q_0\xi_2\rangle})$, so that by our choice of q, $m = 1$ must hold. Now (86) is proved.

Case 1a: $\langle q\xi_2\rangle \geq 2L\langle q\xi_1\rangle$. Then $b_{\boldsymbol{\xi},L}(y) = \min(\frac{1}{q^2}, \frac{\sqrt{y}}{q\langle q\xi_2\rangle})$, and also for each $|n| \leq L$ we have $\frac12\langle q\xi_2\rangle \leq \langle q(n\xi_1+\xi_2)\rangle \leq \frac32\langle q\xi_2\rangle$ and $\frac23 b_{\boldsymbol{\xi},L}(y) \leq b_{\boldsymbol{\xi} U^n,1}(y) \leq 2b_{\boldsymbol{\xi},L}(y)$. Hence (85) holds.

Case 1b: $\langle q\xi_2\rangle < 2L\langle q\xi_1\rangle$. Then $\frac12\min(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}) \leq b_{\boldsymbol{\xi},L}(y) \leq \min(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle})$. Note also that there is $\omega \in \{\pm1\}$ such that $\langle q(n\xi_1+\xi_2)\rangle = \big|\omega n\langle q\xi_1\rangle + \langle q\xi_2\rangle\big|$ for every integer n with $|n| \leq L$; hence our task is to prove:

$$\sum_{|n|\leq L} \min\Big(\frac{1}{q^2}, \frac{\sqrt{y}}{q\langle q\xi_1\rangle}, \frac{\sqrt{y}}{q\big|n\langle q\xi_1\rangle + \langle q\xi_2\rangle\big|}\Big)^\eta \asymp_\eta L\min\Big(\frac{1}{q^2}, \frac{\sqrt{y}}{Lq\langle q\xi_1\rangle}\Big)^\eta. \tag{87}$$

Set $A = \max(0, \frac{\langle q\xi_2\rangle}{\langle q\xi_1\rangle} - L)$. Then $0 \leq A < L$, and (87) is

$$\asymp \sum_{n=1}^{L} \min\Big(\frac{1}{q^2}, \frac{\sqrt{y}}{(A+n)q\langle q\xi_1\rangle}\Big)^\eta = q^{-2\eta}\sum_{n=1}^{L}\min\Big(1, \frac{B}{A+n}\Big)^\eta,$$

with $B = \frac{\sqrt{y}q}{\langle q\xi_1\rangle}$. However the last expression is $\asymp_\eta q^{-2\eta}L\min\big(1,(B/L)^\eta\big)$, uniformly over all $A \in [0, L]$. Hence (87) holds.

Case 2: Assume $b_{\boldsymbol{\xi},L}(y) < 8y^{\frac14}$. Then the right hand side of (85) is $\asymp_\eta Ly^{\eta/4}$, and it now suffices to prove that $\sum_{|n|\leq L} b_{\boldsymbol{\xi} U^n,1}(y)^\eta \ll_\eta Ly^{\eta/4}$. For any integer n satisfying $|n| \leq L$ and $b_{\boldsymbol{\xi} U^n,1}(y) > 16y^{\frac14}$, we can do the following: Let $L' \geq 1$ be the largest number for which $b_{\boldsymbol{\xi} U^n,L'}(y) \geq 16y^{\frac14}$. Then in fact $L' < L$ and $b_{\boldsymbol{\xi} U^n,L'}(y) = 16y^{\frac14}$, since $b_{\boldsymbol{\xi} U^n,L}(y) \leq 2b_{\boldsymbol{\xi},L}(y) < 16y^{\frac14}$, by the proof of Lemma 10.2. Furthermore, by what we proved in Case 1,

$$\sum_{n-L'\leq m\leq n+L'} b_{\boldsymbol{\xi} U^m,1}(y)^\eta \ll_\eta L' b_{\boldsymbol{\xi} U^n,L'}(y)^\eta \asymp L' y^{\eta/4}.$$

It follows that for *every* integer n with $|n| \leq L$, there exist integers $a(n) \leq n \leq b(n)$ satisfying $b(n) - n = n - a(n) < L$, such that $\sum_{m=a(n)}^{b(n)} b_{\boldsymbol{\xi} U^m,1}(y)^\eta \ll_\eta (b(n)-a(n)+1)y^{\eta/4}$. (Indeed, if $b_{\boldsymbol{\xi} U^n,1}(y) \leq 16y^{\frac14}$ then take $a(n) = b(n) = n$.)

Now fix F to be any subset of $\mathbb{Z}\cap[-L,L]$ which is minimal with the property that $\cup_{n\in F}[a(n),b(n)]$ contains all $\mathbb{Z}\cap[-L,L]$. Let us write $F = \{n_1, n_2, \ldots, n_r\}$ where $n_1 < n_2 < \ldots < n_r$. Then $a(n_1) < a(n_2) < \ldots < a(n_r)$, for otherwise, if $a(n_j) \geq a(n_{j+1})$ for some $1 \leq j < r$, then $[a(n_j), b(n_j)] \subset [a(n_{j+1}), b(n_{j+1})]$ and so $\mathbb{Z}\cap[-L,L] \subset \cup_{n\in F\setminus\{n_j\}}[a(n),b(n)]$, contradicting the minimality of F. Similarly $b(n_1) < b(n_2) < \ldots < b(n_r)$. Next note that if $1 \leq j \leq r-2$ and $b(n_j) \geq a(n_{j+2})$, then $[a(n_{j+1}), b(n_{j+1})] \subset [a(n_j), b(n_j)] \cup [a(n_{j+2}), b(n_{j+2})]$, so that n_{j+1} could be removed from F, a contradiction. Hence, for every $j \in \{1, \ldots, r-2\}$ we have $b(n_j) < a(n_{j+2})$. It follows that $\sum_{\substack{1\leq j\leq r\\ j \text{ odd}}}(b(n_j)-a(n_j)+1) \leq b(n_r) - a(n_1) + 1 \ll L$, and the same bound holds for the sum over all even j. Hence

$$\sum_{|n|\le L} b_{\boldsymbol{\xi}U^n,1}(y)^\eta \le \sum_{j=1}^{r}\sum_{m=a(n_j)}^{b(n_j)} b_{\boldsymbol{\xi}U^m,1}(y)^\eta \ll_\eta \sum_{j=1}^{r}(b(n_j)-a(n_j)+1)y^{\eta/4} \ll Ly^{\eta/4}.$$

□

Remark 10.1. For any integer n we have

$$\int_\alpha^\beta f\big(\Gamma(1_2,\boldsymbol{\xi})U^x a(y)\big)\,dx = \int_{\alpha-n}^{\beta-n} f\big(\Gamma(1_2,\boldsymbol{\xi}U^n)U^x a(y)\big)\,dx,$$

since $U^{-n}\in\Gamma$. Hence in Theorem 1.2, (4) holds more generally with the right hand side replaced by

$$C\|f\|_{\mathrm{C}_\mathrm{b}^8}\frac{L_n}{\beta-\alpha}\,\widetilde{b}_{\boldsymbol{\xi}U^n,L_n}(y)^{1-\varepsilon}, \qquad \text{with } L_n := \max(1,|\alpha-n|,|\beta-n|),$$

where n is an arbitrary integer. It follows from Lemmata 10.1 and 10.2 that among the choices of n, the best bound (to within an absolute constant) is obtained for any n such that the point 0 lies within distance $\ll 1+|\beta-\alpha|$ from the interval $[\alpha-n,\beta-n]$.

Remark 10.2. Assume now that $|\beta-\alpha|$ is large, and that $[\alpha,\beta]$ has distance $\ll|\beta-\alpha|$ from 0, in line with Remark 10.1. One may then consider partitioning $[\alpha,\beta]$ into subintervals, applying Theorem 1.2 to each of these individually, and then adding the results. It follows from Lemma 10.4 (and Lemma 10.1) that the resulting error bound is never better (to within an absolute constant) than the original one, (4); and if each subinterval has length $\gg 1$ then the resulting error bound is in fact *equally good* as (4).

11. General orbits

We will now prove the effective equidistribution result for arbitrary $U^{\mathbb{R}}$-orbits in X, Theorem 1.6. The proof uses the technique of approximating nonclosed horocycles in X' by pieces of closed horocycles, in the precise form which was worked out in Sarnak and Ubis [41, Sec. 2]. We fix a left invariant Riemannian metric d_G on G. Recall that $G'=\mathrm{SL}(2,\mathbb{R})\subset G$ and $\Gamma'=\mathrm{SL}(2,\mathbb{Z})$.

Proposition 11.1. *[Sarnak and Ubis* [41]*.] There is an absolute constant $C_1>0$ such that the following holds. For every $M\in G'$ and $T\ge 2$ there is some $\gamma=\gamma_{M,T}\in\Gamma'$ and numbers $\alpha=\alpha_{M,T}\in\mathbb{R}$, $y=y_{M,T}>0$, $W=W_{M,T}\in\mathbb{R}$ and $\omega=\omega_{M,T}\in\{1,-1\}$ such that*

$$\frac{1}{C_1 y}\le T\le C_1|W| \tag{88}$$

and such that, writing $\ell(t)\equiv t$ if $\omega=1$ and $\ell(t)\equiv T-t$ if $\omega=-1$:

$$d_G\Big(\gamma^{-1}MU^{\ell(t)}\,,\,U^{\alpha+\frac{yW}{1-\omega t/W}}a\big(\tfrac{y}{(1-\omega t/W)^2}\big)\Big)\ll\frac{|W|^{-1}}{|1-\omega t/W|}, \qquad \forall t\in[0,T], \tag{89}$$

and

$$-\tfrac12<\Re\big(\gamma^{-1}MU^{\ell(0)}(i)\big)\le\tfrac12. \tag{90}$$

Proof. This is proved in [41, Sec. 2]. (Note that the restriction of d_G to G' is a left invariant Riemannian metric on G'. Note also that once (88) and (89) hold, we can make also (90) hold by replacing γ by γU^n and α by $\alpha-n$, for an appropriate $n\in\mathbb{Z}$.) □

Now for any $M\in G'$ and $T\ge 2$ we have defined both $y_{M,T}$ (in Proposition 11.1) and $y_M(T)$ (in (6)). These are in fact of the same order of magnitude:

Lemma 11.2. *$y_{M,T}\asymp y_M(T)$, uniformly over all $M\in G'$ and $T\ge 2$.*

Proof. Let $\gamma=\gamma_{M,T}$, $\alpha=\alpha_{M,T}$, $y=y_{M,T}$ and $W=W_{M,T}$ be as in Proposition 11.1. By (89), $d_G(\gamma^{-1}MU^{\ell(0)},U^{\alpha+yW}a(y))\ll|W|^{-1}$, and therefore $d_G(\gamma^{-1}MU^{\ell(0)}a(T),U^{\alpha+yW}a(y)a(T))\ll|W|^{-1}T\leq C_1$. Note here that $U^{\ell(0)}a(T)$ equals either $a(T)$ or $U^Ta(T)=a(T)U^1$. It follows that, if we set $g=\gamma^{-1}Ma(T)$ and consider the standard action of G' on the Poincaré upper half plane model of the hyperbolic plane, then $\Im g(i)\asymp\Im U^{\alpha+yW}a(y)a(T)(i)=yT\geq C_1^{-1}$. Hence the invariant height function used in [47] satisfies $\mathcal{Y}_{\Gamma'}(g)\ll\Im(g(i))\leq\mathcal{Y}_{\Gamma'}(g)$ and therefore $\mathcal{Y}_{\Gamma'}(g)\asymp yT$. The lemma follows from this, since $y_M(T)=T^{-1}\boldsymbol{\ell}(g)^{-2}=T^{-1}\mathcal{Y}_{\Gamma'}(g)$. □

Using Theorem 1.2 and Proposition 11.1 we will now prove:

Theorem 11.3. *Let $\varepsilon>0$ be fixed. For any $\boldsymbol{\xi}\in\mathbb{R}^2$, $M\in G'$, $T\geq2$, $f\in\mathrm{C}_{\mathrm{b}}^8(\Gamma\backslash G)$, and for any $y=y_{M,T}$ and $\gamma=\gamma_{M,T}$ as in Proposition 11.1, we have*

$$T^{-1}\int_0^T f(\Gamma(1_2,\boldsymbol{\xi})MU^t)\,dt=\int_{\Gamma\backslash G}f\,d\mu+O_\varepsilon\Big(\|f\|_{\mathrm{C}_{\mathrm{b}}^8}\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac12-\varepsilon}\Big).\tag{91}$$

Here $\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y):=b_{\boldsymbol{\xi}\gamma,yT}(y)+y^{\frac14}$ as in (84).

We will see below that Theorem 11.3 implies Theorem 1.6.

Proof of Theorem 11.3. Let $\boldsymbol{\xi},M,T,y,\gamma,f$ be as in the statement of the theorem; also fix corresponding numbers $\alpha=\alpha_{M,T}$, $W=W_{M,T}$, $\omega=\omega_{M,T}$ as in Proposition 11.1, and set $\ell(t)\equiv t$ if $\omega=1$, $\ell(t)\equiv T-t$ if $\omega=-1$. Note that (91) is trivial when $y\geq1$ (since then $\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)>1$); hence from now on we will assume $y<1$. We will partition the interval $[0,T]$ into smaller intervals $I_0,I_1,\ldots,I_m$, in a way which we make precise below. Using $\gamma^{-1}(1_2,\boldsymbol{\xi})M=(1_2,\boldsymbol{\xi}\gamma)\gamma^{-1}M$ we have

$$\int_0^T f\big(\Gamma(1_2,\boldsymbol{\xi})MU^t\big)\,dt=\sum_{j=0}^m\int_{I_j}f\big(\Gamma(1_2,\boldsymbol{\xi}\gamma)\gamma^{-1}MU^{\ell(t)}\big)\,dt.\tag{92}$$

For each j we set $\rho_j^{\max}=\sup_{t\in I_j}|1-\omega t/W|$, $\rho_j^{\min}=\inf_{t\in I_j}|1-\omega t/W|$. We also set $\tau_j=|I_j|$, the length of the interval I_j. Our partition will be such that I_0 contains those t for which $1-\omega t/W$ are closest to 0; in particular we will have $\rho_j^{\min}>0$ for all $j\geq1$. Using (89) together with $|f(\Gamma g_1)-f(\Gamma g_2)|\ll\|f\|_{\mathrm{C}_{\mathrm{b}}^1}d_G(g_1,g_2)$ $(\forall g_1,g_2\in G)$ and the fact that d_G is left invariant, we have, for each $j\geq1$,

$$\begin{aligned}\int_{I_j}f\big(\Gamma(1_2,\boldsymbol{\xi}\gamma)\gamma^{-1}MU^{\ell(t)}\big)\,dt=\int_{I_j}&f\Big(\Gamma(1_2,\boldsymbol{\xi}\gamma)U^{\alpha+\frac{yW}{1-\omega t/W}}a\big(\tfrac{y}{(1-\omega t/W)^2}\big)\Big)\,dt\\&+O\bigg(\|f\|_{\mathrm{C}_{\mathrm{b}}^1}\frac{\tau_j}{\rho_j^{\min}|W|}\bigg).\end{aligned}\tag{93}$$

We set $y_j^*=y/(\rho_j^{\min})^2$. Note that $a(\frac{y}{(1-\omega t/W)^2})=a(y_j^*)a\big(\frac{(1-\omega t/W)^2}{(\rho_j^{\min})^2}\big)^{-1}$ and $1\leq\frac{(1-\omega t/W)^2}{(\rho_j^{\min})^2}\leq1+\frac{2\rho_j^{\max}\tau_j}{(\rho_j^{\min})^2|W|}$ for all $t\in I_j$; therefore $d_G\big(a(\frac{y}{(1-\omega t/W)^2}),a(y_j^*)\big)\ll\frac{\rho_j^{\max}\tau_j}{(\rho_j^{\min})^2|W|}$ for all $t\in I_j$. We will choose the intervals $I_0,\ldots,I_m$ so that $\rho_j^{\max}\leq2\rho_j^{\min}$ for each $j\geq1$. Hence we may replace $a(\frac{y}{(1-\omega t/W)^2})$ by $a(y_j^*)$ in the integral in (93), without changing the error term. Next we take $s=\alpha+\frac{yW}{1-\omega t/W}$ as a new variable of integration. Let $S_j\subset\mathbb{R}$ be the s-interval which corresponds to I_j. Note that $|S_j|=\frac{y\tau_j}{\rho_j^{\min}\rho_j^{\max}}$, and $\frac{dt}{ds}=\omega y^{-1}(1-\omega t/W)^2=\omega\frac{\rho_j^{\min}\rho_j^{\max}}{y}+O(\frac{\rho_j^{\max}\tau_j}{y|W|})$ for $t\in I_j$. Hence (93) equals

$$\frac{\rho_j^{\min}\rho_j^{\max}}{y}\int_{S_j}f\big(\Gamma(1_2,\boldsymbol{\xi}\gamma)U^s\,a(y_j^*)\big)\,ds+O\bigg(\|f\|_{\mathrm{C}_{\mathrm{b}}^1}\frac{1+\tau_j^2}{\rho_j^{\min}|W|}\bigg).\tag{94}$$

We will choose the intervals $I_0, \ldots, I_m$ so that $y_j^* < 1$ for each $j \geq 1$. Take $n_j \in \mathbb{Z}$ so that $S_j - n_j$ intersects the interval $[0,1)$, and set $\gamma_j := \gamma U^{n_j}$. Applying Theorem 1.2 together with Remark 10.1 (with $n = n_j$), we conclude that

(95)

$$\int_{I_j} f\big(\Gamma(1_2, \boldsymbol{\xi}\gamma)\gamma^{-1} M U^{\ell(t)}\big)\, dt = \tau_j \int_{\Gamma\backslash G} f\, d\mu + O_\varepsilon\bigg(\|f\|_{\mathrm{C}_\mathrm{b}^8} \frac{L_j}{y_j^*} \widetilde{b}_{\boldsymbol{\xi}\gamma_j, L_j}(y_j^*)^{1-\varepsilon} + \|f\|_{\mathrm{C}_\mathrm{b}^1} \frac{1+\tau_j^2}{\rho_j^{\min}|W|}\bigg),$$

where $L_j = 1 + |S_j|$. We have $|S_j| = \frac{y\tau_j}{\rho_j^{\min}\rho_j^{\max}} \asymp y_j^*\tau_j$, and we will choose $I_0, \ldots, I_m$ in such a way that $y_j^*\tau_j \ll 1$ for all $j \geq 1$; hence (95) holds with L_j replaced by 1. We now wish to choose $I_0, \ldots, I_m$ in such a way that for each $j \geq 1$, τ_j takes a value which essentially minimizes τ_j^{-1} times the error term in (95), but subject to $y_j^*\tau_j \ll 1$.

The precise choice of $I_0, \ldots, I_m$ is made according to the following algorithm. Let the absolute constant $C_1 > 0$ be as in Proposition 11.1, and set $C_2 = \frac{1}{2(1+C_1)^3}$.

1. Set $j = 1$ and $T_1 = 0$.
2. If $1 - \omega T_j/W > 2y^{\frac14}$ then set $\rho_j = |1 - \omega T_j/W|$, $n_j = \lfloor \alpha + \frac{yW}{1-\omega T_j/W} \rfloor \in \mathbb{Z}$ and $\gamma_j = \gamma U^{n_j}$, and go to Step 3; otherwise change the value of T_j to $T_j = T$ and go to Step 4.
3. Set $\tau_j = \min\Big(\rho_j^{\frac32} y^{-\frac12} |W|^{\frac12} \widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y/\rho_j^2)^{\frac12}, C_2\rho_j^2 y^{-1}, T - T_j \Big)$, $T_{j+1} = T_j + \tau_j$ and $I_j = [T_j, T_{j+1}]$. If $T_{j+1} = T$, set $m = j$ and $I_0 = \emptyset$, and we are *done*; otherwise replace j by $j+1$ and go back to Step 2.
4. If $1 - \omega T_j/W < -2y^{\frac14}$ then set $\rho_j = |1 - \omega T_j/W|$, $n_j = \lfloor \alpha + \frac{yW}{1-\omega T_j/W} \rfloor \in \mathbb{Z}$ and $\gamma_j = \gamma U^{n_j}$, and go to Step 5; otherwise set $m = j - 1$ and $I_0 = [0,T] \setminus \cup_{i=1}^m I_i$ (this is an interval), and we are *done.*
5. Set $\tau_j = \min\Big(\rho_j^{\frac32} y^{-\frac12} |W|^{\frac12} \widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y/\rho_j^2)^{\frac12}, C_2\rho_j^2 y^{-1} \Big)$, $T_{j+1} = T_j - \tau_j$ and $I_j = [T_{j+1}, T_j]$. Then replace j by $j+1$ and go back to Step 4.

Note that $|1 - \omega t/W| \leq 1 + T/|W| \leq 1 + C_1$ for all $t \in [0,T]$; hence we always get $2y^{\frac14} < \rho_j \leq 1 + C_1$ in Steps 2 and 4. Using this and (88) we see that each time we set τ_j and I_j in Steps 3 and 5, we get $\tau_j \leq C_2\rho_j^2 y^{-1} \leq C_2 C_1^2 (1+C_1)\rho_j|W| < \frac12 \rho_j |W|$, and therefore $|1 - \omega t/W - \rho_j| < \frac12\rho_j$ for all $t \in I_j$. Hence for each such interval I_j we have $\rho_j^{\min} > 0$ and $\rho_j^{\max} < 2\rho_j^{\min}$, and also $y_j^* = y/(\rho_j^{\min})^2 < 4y/\rho_j^2 < y^{\frac12} < 1$. It also follows that for any interval I_j obtained in Step 3 (resp. Step 5) we have $1 - \omega t/W > y^{\frac14}$ (resp. $1 - \omega t/W < -y^{\frac14}$) for all $t \in I_j$; therefore the intervals constructed in Steps 2–3 do not overlap with those constructed in Steps 4–5. Hence the resulting $I_0, I_1, \ldots, I_m$ indeed form a partition of $[0,T]$ (after possibly removing one or both endpoints from some of the I_j's), satisfying all the conditions specified earlier.

For each $j \geq 1$ we have, because of the choice of τ_j in Steps 3 and 5,

$$\frac{1+\tau_j^2}{\rho_j^{\min}|W|} \ll (\rho_j^{\min})^{-1} + (\rho_j^{\min})^2 y^{-1} \widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y_j^*) \ll y_j^{*-\frac12} + y_j^{*-1}\widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y_j^*) \ll y_j^{*-1}\widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y_j^*).$$

Hence by (92) and (95) (with $L_j = 1$), we have (possibly with $m = 0$):

$$\int_0^T f\big(\Gamma(1_2, \boldsymbol{\xi}) M U^t\big)\, dt = \int_{I_0} f(\cdots)\, dt + \Big(\sum_{j=1}^m \tau_j\Big) \int_{\Gamma\backslash G} f\, d\mu + O_\varepsilon\bigg(\|f\|_{\mathrm{C}_\mathrm{b}^8} \sum_{j=1}^m y_j^{*-1} \widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y_j^*)^{1-\varepsilon} \bigg).$$

Next we note that $\int_{I_0} f\, dt = \tau_0 \int_{\Gamma\backslash G} f\, d\mu + O(\|f\|_{\mathrm{C}_\mathrm{b}^0}\tau_0)$ and $\tau_0 \ll y^{\frac14} T$. (Indeed, $\tau_0 \leq T$; also if $y^{\frac14} < \frac14$, say, and $I_0 \neq \emptyset$, then it follows from our construction that $|1 - \omega t/W| \leq 2y^{\frac14} < \frac12$

for all $t \in I_0$; hence $|W| < 2T$ and $\tau_0 = |I_0| \leq 4y^{\frac{1}{4}}|W| < 8y^{\frac{1}{4}}T$.) Therefore,

$$\int_0^T f\big(\Gamma(1_2,\boldsymbol{\xi})MU^t\big)\,dt = T\int_{\Gamma\backslash G} f\,d\mu + O_\varepsilon\Big(\|f\|_{\mathrm{C}_{\mathrm{b}}^8}\Big)\Big\{Ty^{\frac{1}{4}} + \sum_{j=1}^m y_j^{*-1}\,\widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y_j^*)^{1-\varepsilon}\Big\}. \tag{96}$$

Now for each $n \in \mathbb{Z}$, let J_n be the set of those $j \in \{1,\ldots,m\}$ for which $n_j = n$. By our choice of n_j, for each $j \in J_n$ there is some $t \in I_j$ such that $\frac{yW}{1-\omega t/W} \in [n-\alpha, n-\alpha+1)$. If $|n-\alpha| \geq 2$ then this forces $|1-\omega t/W| \asymp \frac{y|W|}{|n-\alpha|}$; on the other hand if $|n-\alpha| < 2$ then both $|1-\omega t/W| \asymp 1$ and $y|W| \asymp 1$, since $y|W| \geq C_1^{-2}$ and $|1-\omega t/W| \leq 1 + C_1$ by Prop. 11.1. It follows that $\rho_j^{\min} \asymp \rho_j^{\max} \asymp \rho(n) := \frac{y|W|}{1+|n-\alpha|}$ for each $j \in J_n$, and thus also $y_j^* \asymp y/\rho(n)^2$ and $|S_j| \asymp y\tau_j/\rho(n)^2$. But $\sum_{j\in J_n}|S_j| = |\cup_{j\in J_n} S_j| \ll 1$, since for each $j \in J_n$ we have $|S_j| \ll 1$ (by our choice of τ_j) and $S_j \cap [n, n+1) \neq \emptyset$. Hence $\sum_{j\in J_n}\tau_j \ll \rho(n)^2y^{-1}$. However for all except at most one $j \in J_n$ (the possible exception being $j = m$) we have $\tau_j \gg \min(\rho(n)^{\frac{3}{2}}y^{-\frac{1}{2}}|W|^{\frac{1}{2}}\widetilde{b}_{\boldsymbol{\xi}\gamma U^n,1}(y/\rho(n)^2)^{\frac{1}{2}}, \rho(n)^2/y)$. Hence

$$\#J_n \ll 1 + \rho(n)^{\frac{1}{2}}y^{-\frac{1}{2}}|W|^{-\frac{1}{2}}\,\widetilde{b}_{\boldsymbol{\xi}\gamma U^n,1}\big(y/\rho(n)^2\big)^{-\frac{1}{2}} \ll \widetilde{b}_{\boldsymbol{\xi}\gamma U^n,1}\big(y/\rho(n)^2\big)^{-\frac{1}{2}}, \tag{97}$$

and thus in (96) we have

$$\begin{aligned}\sum_{j=1}^m y_j^{*-1}\,\widetilde{b}_{\boldsymbol{\xi}\gamma_j,1}(y_j^*)^{1-\varepsilon} &\ll y^{-1}\sum_{\substack{n\in\mathbb{Z}\\(J_n\neq\emptyset)}}\rho(n)^2\,\widetilde{b}_{\boldsymbol{\xi}\gamma U^n,1}\big(y/\rho(n)^2\big)^{\frac{1}{2}-\varepsilon}\\ &\ll y^{-1}\sum_{\substack{n\in\mathbb{Z}\\(J_n\neq\emptyset)}}\rho(n)^{\frac{3}{2}}\,\widetilde{b}_{\boldsymbol{\xi}\gamma U^n,1}(y)^{\frac{1}{2}-\varepsilon},\end{aligned} \tag{98}$$

where we used Lemma 10.1 and the fact that $\rho(n) \ll 1$ for all n.

Let us first assume $|W| \geq 2T$. Then for every n with $J_n \neq \emptyset$ we have $\rho(n) \asymp 1$, and there is some $t \in [0,T]$ such that $n = \alpha + \frac{yW}{1-\omega t/W} + O(1)$; thus $n = \alpha + yW + O(yT) = O(yT)$, since $|\alpha + yW| \ll 1$ by (90) and (89). Hence by Lemma 10.4, (98) is $\ll T\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac{1}{2}-\varepsilon}$. Next assume instead $|W| < 2T$. Then $|W| \asymp T$, by (88). Given $\widetilde{\rho} \in (0,1]$, note that for every n with $\rho(n) = \frac{y|W|}{1+|n-\alpha|} \geq \widetilde{\rho}$ we have $|n-\alpha| < \widetilde{\rho}^{-1}y|W| \asymp \widetilde{\rho}^{-1}yT$, and since $|\alpha + yW| \ll 1$ this implies $|n| \ll \widetilde{\rho}^{-1}yT$. By Lemma 10.4, the sum of $\widetilde{b}_{\boldsymbol{\xi}\gamma U^n,1}(y)^{\frac{1}{2}-\varepsilon}$ over all these n is $\ll \widetilde{\rho}^{-1}yT\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac{1}{2}-\varepsilon}$. Hence the contribution from all n with $\rho(n) \geq \frac{1}{2}$ in (98) is $\ll T\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac{1}{2}-\varepsilon}$, and for each $k \in \mathbb{Z}^+$, the contribution from all n with $\rho(n) \in [2^{-k-1}, 2^{-k})$ in (98) is $\ll 2^{-\frac{1}{2}k}T\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac{1}{2}-\varepsilon}$. Adding over k we again conclude that (98) is $\ll T\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac{1}{2}-\varepsilon}$. In view of (96), this completes the proof of Theorem 11.3. □

We remark that the last step in (97) is in general wasteful, but leads to a simple result. Working instead with the first bound in (97) one obtains a variant of Theorem 11.3 with a more complicated but generally better error term:

Theorem 11.3′. *Let $\varepsilon > 0$ be fixed. For any $\boldsymbol{\xi} \in \mathbb{R}^2$, $M \in G'$, $T \geq 2$, $f \in \mathrm{C}_{\mathrm{b}}^8(\Gamma\backslash G)$, and for any $y = y_{M,T}$ and $\gamma = \gamma_{M,T}$ as in Proposition 11.1, we have*

$$T^{-1}\int_0^T f(\Gamma(1_2,\boldsymbol{\xi})MU^t)\,dt = \int_{\Gamma\backslash G} f\,d\mu + O_\varepsilon\Big(\|f\|_{\mathrm{C}_{\mathrm{b}}^8}\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac{1}{2}-\varepsilon}\Big(\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y)^{\frac{1}{2}} + \big(y|W|\big)^{-\frac{1}{2}}\Big)\Big).$$

It is worth noticing that in the special case $M = U^{\alpha_0}a(y_0)$ with $-\frac{1}{2} < \alpha_0 \leq \frac{1}{2}$, y_0 small and $T > C_1^{-1}y_0^{-1}$, we may take $\gamma_{M,T} = 1_2$ in Proposition 11.1, as well as $\alpha_{M,T} = \alpha_0 - y_0W_{M,T}$ and $W_{M,T}$ "very large" (that is, let $W_{M,T} \to \infty$ for our fixed M, T, so that (89) turns into an equality between two points in G). In this case, one may expect from the method of proof that Theorem 11.3 should recover the statement of Theorem 1.2, with $y = y_0$, $\alpha = \alpha_0$, $\beta = \beta_0 + y_0T$. This is indeed seen to be the case when we use the more precise error term of Theorem 11.3′. In this vein recall also Remarks 10.1, 10.2. In the case of $\beta - \alpha$ becoming *small* as $y \to 0$ in Theorem 1.2, we expect that Theorem 11.3′ should typically result in a *better* error term than that of Theorem 1.2.

Next we will reinterprete the error term in Theorem 11.3 and thereby deduce Theorem 1.6. Recall $\mathfrak{R}_L = [-L^{-1}, L^{-1}] \times [-1,1] \subset \mathbb{R}^2$. Let us first note that, for any $\boldsymbol{\xi} \in \mathbb{R}^2$, $L > 0$, $y > 0$,

$$b_{\boldsymbol{\xi},L}(y) = \inf\Big\{\delta > 0 \,:\, \Big[\forall q \in \mathbb{Z}^+_{\leq \delta^{-1/2}} :\, (q^{-1}\mathbb{Z}^2 + \boldsymbol{\xi}) \cap \frac{\sqrt{y}}{\delta q^2}\mathfrak{R}_L = \emptyset\Big]\Big\}. \tag{99}$$

Indeed, from the definition (3) we see that, given any $\delta > 0$ we have $b_{\boldsymbol{\xi},L}(y) \geq \delta$ if and only if there is some $q \in \mathbb{Z}^+$ such that $q \leq \delta^{-1/2}$, $\langle q\xi_1 \rangle \leq \frac{\sqrt{y}}{\delta L q}$ and $\langle q\xi_2 \rangle \leq \frac{\sqrt{y}}{\delta q}$; and the last two conditions hold if and only if $(\mathbb{Z}^2 + q\boldsymbol{\xi}) \cap \frac{\sqrt{y}}{\delta q}\mathfrak{R}_L \neq \emptyset$.

Using $\frac{\sqrt{y}}{\delta q^2}\mathfrak{R}_L = (\frac{1}{\delta q^2}\mathfrak{R}_{L/y})a(y)^{-1}$, the formula (99) may also be expressed as

$$b_{\boldsymbol{\xi},L}(y) = b_{(1_2,\boldsymbol{\xi})a(y)}(L/y), \tag{100}$$

where in the right hand side we use the notation introduced in (8).

Proof of Theorem 1.6. Let g, T, f be as in the statement of Theorem 1.6. Write $g = (1_2, \boldsymbol{\xi})M$; fix corresponding numbers $y = y_{M,T}$, $\alpha = \alpha_{M,T}$, $W = W_{M,T}$, $\omega = \omega_{M,T}$ and $\gamma = \gamma_{M,T} \in \Gamma'$ as in Proposition 11.1, and set $\ell(t) \equiv t$ if $\omega = 1$, $\ell(t) \equiv T - t$ if $\omega = -1$. By (89) we have $\gamma^{-1}MU^{\ell(0)} = U^{\alpha+yW}a(y)\eta$ for some $\eta \in G'$ in a $O(|W|^{-1})$-neighbourhood of 1_2. Hence for any $q \in \mathbb{Z}^+$,

$$(q^{-1}\mathbb{Z}^2)g = (q^{-1}\mathbb{Z}^2 + \boldsymbol{\xi})\gamma\gamma^{-1}M = (q^{-1}\mathbb{Z}^2 + \boldsymbol{\xi}\gamma)U^{\alpha+yW}a(y)\eta U^{-\ell(0)}.$$

Now assume that, for some $q \in \mathbb{Z}^+$ and $\delta > 0$, the lattice translate $q^{-1}\mathbb{Z}^2 + \boldsymbol{\xi}\gamma$ contains a point $(x_1, x_2) \in \frac{\sqrt{y}}{\delta q^2}\mathfrak{R}_{yT}$. Then $(q^{-1}\mathbb{Z}^2)g$ contains the point

$$(x_1, x_2)U^{\alpha+yW}a(y)\eta U^{-\ell(0)} = \Big(y^{1/2}x_1, y^{-1/2}((\alpha + yW)x_1 + x_2)\Big)\eta U^{-\ell(0)}.$$

But here $|\alpha + yW| \ll 1$ by (89), (90), and $yT \gg 1$ by (88); hence $|x_1| \leq \frac{1}{\delta q^2\sqrt{yT}} \ll \frac{\sqrt{y}}{\delta q^2}$, and the above point is

$$= \Big(O\Big(\frac{1}{\delta q^2 T}\Big), O\Big(\frac{1}{\delta q^2}\Big)\Big)\begin{pmatrix} 1 + O(|W|^{-1}) & O(|W|^{-1}) \\ O(|W|^{-1}) & 1 + O(|W|^{-1}) \end{pmatrix}\begin{pmatrix} 1 & O(T) \\ 0 & 1 \end{pmatrix} = \Big(O\Big(\frac{1}{\delta q^2 T}\Big), O\Big(\frac{1}{\delta q^2}\Big)\Big),$$

where we also used the fact that $|W| \gg T$. We have thus proved that there is an absolute constant $C_2 > 1$ such that, for any $q \in \mathbb{Z}^+$ and $\delta > 0$ for which $(q^{-1}\mathbb{Z}^2 + \boldsymbol{\xi}\gamma) \cap \frac{\sqrt{y}}{\delta q^2}\mathfrak{R}_{yT} \neq \emptyset$, we have $(q^{-1}\mathbb{Z}^2)g \cap \frac{C_2}{\delta q^2}\mathfrak{R}_T \neq \emptyset$. Hence by (99),

$$\begin{aligned} b_{\boldsymbol{\xi}\gamma,yT}(y) &= \inf\Big\{\delta > 0 \,:\, \Big[\forall q \in \mathbb{Z}^+_{\leq \delta^{-1/2}} :\, (q^{-1}\mathbb{Z}^2 + \boldsymbol{\xi}\gamma) \cap \frac{\sqrt{y}}{\delta q^2}\mathfrak{R}_{yT} = \emptyset\Big]\Big\} \\ &\leq \inf\Big\{\delta > 0 \,:\, \Big[\forall q \in \mathbb{Z}^+_{\leq \delta^{-1/2}} :\, (q^{-1}\mathbb{Z}^2)g \cap \frac{C_2}{\delta q^2}\mathfrak{R}_T = \emptyset\Big]\Big\} \\ &\leq C_2 \inf\Big\{\delta > 0 \,:\, \Big[\forall q \in \mathbb{Z}^+_{\leq \delta^{-1/2}} :\, (q^{-1}\mathbb{Z}^2)g \cap \frac{1}{\delta q^2}\mathfrak{R}_T = \emptyset\Big]\Big\} = C_2 b_g(T). \end{aligned}$$

Using this bound together with $\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y) = y^{\frac{1}{4}} + b_{\boldsymbol{\xi}\gamma,yT}(y)$ and $y = y_{M,T} \ll y_M(T) = y_g(T)$ (cf. Lemma 11.2), we see that $\widetilde{b}_{\boldsymbol{\xi}\gamma,yT}(y) \ll y_g(T)^{\frac{1}{4}} + b_g(T)$, so that Theorem 1.6 follows from Theorem 11.3. □

Finally let us prove that, generically, the error term in Theorem 1.6 decays like $T^{-\frac{1}{8}+\varepsilon}$ as $T \to \infty$.

Proposition 11.4. *Let $0 < \alpha < \frac{1}{2}$ and $M \in G'$ be given. Then for Lebesgue almost all $\boldsymbol{\xi} \in \mathbb{R}^2$, there is some $C > 0$ such that $b_{(M,\boldsymbol{\xi})}(T) < CT^{-\alpha}$ for all $T \geq 1$.*

Proof. It follows from the definition, (8), that for given $(M,\boldsymbol{\xi})\in G$ and $C\geq 1$, the inequality $b_{(M,\boldsymbol{\xi})}(T)<CT^{-\alpha}$ holds for all $T\geq 1$ if and only if, for every $q\in\mathbb{Z}^+$, the lattice translate $\boldsymbol{\xi}+q^{-1}\mathbb{Z}^2M$ is disjoint from the set

$$B_{Cq^2}=\bigcup_{T\geq(Cq^2)^{1/\alpha}}\frac{T^\alpha}{Cq^2}\mathcal{R}_T=\Big\{(x_1,x_2)\in\mathbb{R}^2\,:\,|x_1|\leq(Cq^2)^{-\frac{1}{\alpha}}\min\big(1,|x_2|^{1-\frac{1}{\alpha}}\big)\Big\}.$$

For given $M\in G'$ we write $L=\mathbb{Z}^2M$, so that the lattice translate in question is $\boldsymbol{\xi}+q^{-1}L$. Note that this point set only depends on the congruence class of $\boldsymbol{\xi}\bmod q^{-1}L$. Now

$$\int_{\mathbb{R}^2/L}I\big((\boldsymbol{\xi}+q^{-1}L)\cap B_{Cq^2}\neq\emptyset\big)\,d\boldsymbol{\xi}=q^{-2}\int_{\mathbb{R}^2/qL}I\big((\boldsymbol{\eta}+L)\cap qB_{Cq^2}\neq\emptyset\big)\,d\boldsymbol{\eta}$$
$$=\int_{\mathbb{R}^2/L}I\big((\boldsymbol{\eta}+L)\cap qB_{Cq^2}\neq\emptyset\big)\,d\boldsymbol{\eta}\leq\int_{\mathbb{R}^2/L}\sum_{\boldsymbol{m}\in L}I\big(\boldsymbol{\eta}+\boldsymbol{m}\in qB_{Cq^2}\big)\,d\boldsymbol{\eta}=\big|qB_{Cq^2}\big|=q^2\big|B_{Cq^2}\big|,$$

where we substituted $\boldsymbol{\xi}=q^{-1}\boldsymbol{\eta}$, and where $|\cdot|$ denotes Lebesgue measure on $\mathbb{R}^2$. Next note that, since $1-\frac{1}{\alpha}<-1$, we have $\big|B_{Cq^2}\big|=K(Cq^2)^{-\frac{1}{\alpha}}$ where $K>0$ is a constant which only depends on α. It follows that

$$\int_{\mathbb{R}^2/L}I\Big(\exists T\geq 1\,:\,b_{(M,\boldsymbol{\xi})}(T)\geq CT^{-\alpha}\Big)\,d\boldsymbol{\xi}\leq\sum_{q=1}^{\infty}q^2\big|B_{Cq^2}\big|=K\Big(\sum_{q=1}^{\infty}q^{2(1-\frac{1}{\alpha})}\Big)C^{-\frac{1}{\alpha}}.$$

The sum converges for our α, and the proposition follows since the last expression tends to zero as $C\to\infty$. $\square$

Remark 11.1. As we noted in the introduction, Proposition 11.4 implies that for μ-almost all $g\in G$, the right hand side in (9) in Theorem 1.6 decays more rapidly than $T^{\varepsilon-\frac{1}{8}}$ as $T\to\infty$ ($\forall\varepsilon>0$). On the other hand, using the fact that the flow $\{U^t\}$ is mixing on smooth vectors in $L^2(X)$ with a rate $t^{\varepsilon-1}$ as $t\to\infty$ (as follows from [4] combined with an argument as in [37, Lemma 2.3][2]), one can prove that for sufficiently nice test functions f on $\Gamma\backslash G$, and for μ-almost all $\Gamma g\in X$, the deviation of the ergodic average in the left hand side of (9) decays like $T^{\varepsilon-\frac{1}{2}}$ as $T\to\infty$; cf. [12]. In this last statement the μ-null set of exceptional points Γg is *non-explicit* and depends on f; furthermore the implied constant in the bound depends on both f and Γg in a non-explicit way; the strength of Theorem 1.6 lies of course in the fact that it gives a bound where all these dependencies are explicit. Nevertheless, the discussion suggests that it might be possible to improve Theorem 1.6 so as to yield a rate of decay $T^{\varepsilon-\frac{1}{2}}$ for any $\Gamma g\in X$ satisfying an appropriate Diophantine condition.

In this vein, we note that there are *two* steps in our proof of Theorem 1.6 which are clearly non-optimal, each of which causes a *halving* of the expected optimal exponent. The first is when we bound the d-sums in (27) individually for each c using the Weil bound, and the second is in (93), where we replace the integral over the given orbit with an integral over a nearby orbit which is a lift of a piece of a closed horocycle. We discussed the first of these in Remark 8.1. Regarding the second step, we note that a possible approach for an improved treatment might be to *rework* the proof of Theorem 1.2 for the case of an arbitrary U^t-orbit, choosing coordinates in a similar way as in the proofs of [47, Propositions 5.1 and 5.3].

References

[1] T. Browning and I. Vinogradov, Effective Ratner Theorem for $\mathrm{SL}(2,\mathbb{R})\ltimes\mathbb{R}^2$ and gaps in $\sqrt{n}$ modulo 1, arXiv:1311.6387

[2] M. Burger, Horocycle flow on geometrically finite surfaces, Duke Math. J. **61** (1990), 779–803.

[3] S. G. Dani, On uniformly distributed orbits of certain horocycle flows, Ergodic Theory Dynam. Systems **2** (1982), 139–158.

[2]Note that [37, p. 282, line -8] should be corrected to "$u(t)=r(\theta_2)a(\operatorname{arsinh}(t/2))r(\theta_1)$". Here Ratner's "$a(t)$" equals Φ^{2t} in our notation.

[4] S. Edwards, The Rate of Mixing for Diagonal Flows on Spaces of Affine Lattices, Master Thesis, Uppsala University, 2013. [Available at: http://www.math.uu.se/Research/Publications/Student+theses/]

[5] M. Einsiedler, G. Margulis, A. Venkatesh, Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces, Invent. Math. **177** (2009), 137–212.

[6] D. El-Baz, J. Marklof, I. Vinogradov, The distribution of directions in an affine lattice: two-point correlations and mixed moments, to appear in International Mathematics Research Notices.

[7] D. El-Baz, J. Marklof, I. Vinogradov, The two-point correlation function of the fractional parts of $\sqrt{n}$ is Poisson, to appear in Proceedings of the American Mathematical Society.

[8] N. D. Elkies and C. T. McMullen, Gaps in $\sqrt{n}$ mod 1 and ergodic theory. Duke Math. J. **123** (2004), 95–139, and a correction in Duke Math J. **129** (2005), 405–406.

[9] A. Eskin, Unipotent flows and applications, in *Homogeneous flows, moduli spaces and arithmetic*, Clay Math. Proc. **10**, pp. 71–129, Amer. Math. Soc., Providence, RI, 2010.

[10] A. Eskin and H. Oh, Ergodic theoretic proof of equidistribution of Hecke points, Ergodic Theory Dynam. Systems **26** (2006), 163–167.

[11] L. Flaminio and G. Forni, Invariant distributions and time averages for horocycle flows, Duke Math. J. **119** (2003), 465–526.

[12] V. F. Gaposhkin, The dependence of the rate of convergence in the strong law of large numbers for stationary processes on the rate of diminution of the correlation function, Teor. Veroyatnost. i Primenen. **26** (1981), 720–733; English transl. in *Theory Probab. Appl.* **26** (1981).

[13] A. Gorodnik and H. Oh, Orbits of discrete subgroups on a symmetric space and the Furstenberg boundary, Duke Math. J. **139** (2007), 483–525.

[14] B. Green and T. Tao, Linear equations in primes, Ann. of Math. **171** (2010), 1753–1850.

[15] B. Green and T. Tao, The quantitative behaviour of polynomial orbits on nilmanifolds, Ann. of Math. **175** (2012), 465–540.

[16] B. Green and T. Tao, The Möbius function is strongly orthogonal to nilsequences, Ann. of Math. **175** (2012), 541–566.

[17] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Clarendon Press, Oxford, 1938.

[18] D. A. Hejhal, On the uniform equidistribution of long closed horocycles, Asian J. Math. **4** (2000), 839–853.

[19] R. Howe and E. Tan, *Nonabelian harmonic analysis*, Springer-Verlag, New York, 1992.

[20] A. E. Ingham, *The Distribution of Prime Numbers*, Cambridge University Press, 1932.

[21] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, American Mathematical Society, 2004.

[22] V. Jarnik, Über die simultanen diophantischen Approximationen, Math. Z. **33** (1931), 505–543.

[23] A. W. Knapp, *Lie groups beyond an introduction*, Progress in Mathematics **140**, Birkhäuser Boston Inc., 2002.

[24] D. Kleinbock, N. Shah and A. Starkov, Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory. In: *Handbook of dynamical systems, Vol. 1A,* pp. 813–930. North-Holland, Amsterdam, 2002.

[25] E. Lindenstrauss and G. Margulis, Effective estimates on indefinite ternary forms, preprint 2013.

[26] G. Margulis, Problems and conjectures in rigidity theory, in *Mathematics: frontiers and perspectives*, AMS, Providence, RI, 2000, pp. 161–174

[27] G. Margulis, *On Some Aspects of the Theory of Anosov Systems*, Springer Monographs in Mathematics. Springer, Berlin, 2004. (A translation of Phd Thesis, Moscow State University, 1970.)

[28] G. Margulis and A. Mohammadi, Quantitative version of the Oppenheim conjecture for inhomogeneous quadratic forms, Duke Math. J. **158** (2011), 121–160.

[29] J. Marklof, Pair correlation densities of inhomogeneous quadratic forms. II. Duke Math. J. **115** (2002), 409–434.

[30] J. Marklof, Pair correlation densities of inhomogeneous quadratic forms, Annals of Math. **158** (2003), 419-471.

[31] J. Marklof, Mean square value of exponential sums related to the representation of integers as sums of squares, Acta Arith. **117** (2005), 353–370.

[32] J. Marklof and A. Strömbergsson, The distribution of free path lengths in the periodic Lorentz gas and related lattice point problems, Annals of Math. **172** (2010), 1949–2033.

[33] J. Marklof and A. Strömbergsson, Free Path Lengths in Quasicrystals, to appear in Communications in Mathematical Physics.

[34] A. Mohammadi, A special case of effective equidistribution with explicit constants, Ergodic Theory Dynam. Systems **32** (2012), 237–247.

[35] M. B. Nathanson, *Additive number theory*, Graduate Texts in Mathematics **164**, Springer-Verlag, 1996.

[36] M. S. Raghunathan, *Discrete subgroups of Lie groups*, Springer-Verlag, New York, 1972.

[37] M. Ratner, The rate of mixing for geodesic and horocycle flow, Ergod. Th. & Dynam. Sys. **7** (1987), 267–288.

[38] M. Ratner, On Raghunathan's measure conjecture, *Ann. of Math.* **134** (1991) 545-607.

[39] M. Ratner, Raghunathan's topological conjecture and distributions of unipotent flows, Duke Math. J. **63** (1991), 235–280.

[40] P. Sarnak, Asymptotic Behavior of Periodic Orbits of the Horocycle Flow and Eisenstein Series, Comm. Pure Appl. Math **34** (1981), 719–739.

[41] P. Sarnak and A. Ubis, The horocycle flow at prime times, arXiv:1110.0777

[42] N. Shah, Limit distributions of expanding translates of certain orbits on homogeneous spaces, Proc. Indian Acad. Sci. (Math. Sci.) **106** (1996), 105–125.

[43] N. Shah, Asymptotic evolution of smooth curves under geodesic flow on hyperbolic manifolds, Duke Math. J. **148** (2009), 281–304.

[44] N. Shah, Expanding translates of curves and Dirichlet-Minkowski theorem on linear forms, J. Amer. Math. Soc. **23** (2010), 563–589.

[45] Ya. G. Sinai, Statistics of gaps in the sequence $\{\sqrt{n}\}$, in *Dynamical systems and group actions*, Contemp. Math. **567**, pp. 185–189, Amer. Math. Soc. 2012.

[46] A. Strömbergsson, On the uniform equidistribution of long closed horocycles, Duke Math. J. **123** (2004), 507–547.

[47] A. Strömbergsson, On the deviation of ergodic averages for horocycle flows, Journal of Modern Dynamics, **7** (2013), 291–328.

[48] A. Strömbergsson and A. Venkatesh, Small solutions to linear congruences and Hecke equidistribution, Acta Arith., **118** (2005), 41–78.

[49] A. Strömbergsson and P. Vishe, Effective Ratner equidistribution for $\mathrm{SL}(2,\mathbb{R}) \ltimes (\mathbb{R}^2)^{\oplus k}$ and applications to quadratic forms, in preparation.

[50] A. Venkatesh, Sparse equidistribution problems, period bounds and subconvexity, Ann. of Math. **172** (2010), 989–1094.

[51] A. Weil, On some exponential sums, Proc. Nat. Acad. Sci. U.S.A. **34** (1948), 204–207.

[52] D. Witte Morris, *Ratner's Theorems on Unipotent Flows,* Chicago Lectures in Mathematics, University of Chicago Press, Chicago (2005). xii+203 pp.

Department of Mathematics, Box 480, Uppsala University, SE-75106 Uppsala, Sweden
astrombe@math.uu.se