QUANTITATIVE EQUIDISTRIBUTION AND THE LOCAL STATISTICS OF THE SPECTRUM OF A FLAT TORUS

By

ELON LINDENSTRAUSS,* AMIR MOHAMMADI† AND ZHIREN WANG‡

Dedicated with admiration and gratitude to Peter Sarnak on occasion of his 70th birthday

Abstract. We show that a pair correlation function for the spectrum of a flat 2-dimensional torus satisfying an explicit Diophantine condition agrees with those of a Poisson process with a polynomial error rate.

The proof is based on a quantitative equidistribution theorem and tools from geometry of numbers.

Contents

*E. L. acknowledges support by ERC 2020 grant HomDyn (grant no. 833423).

†A. M. acknowledges support by the NSF, grants DMS-1764246 and 2055122.

‡Z. W. acknowledges support by the NSF, grant DMS-1753042.

JOURNAL D'ANALYSE MATHÉMATIQUE, Vol. 151 (2023)
DOI 10.1007/s11854-023-0332-x

1 Introduction

Let $\Delta \subset \mathbb{R}^2$ be a lattice. The eigenvalues of the Laplacian of the corresponding flat torus $\mathsf{M} = \mathbb{R}^2/\Delta$ are the values of the quadratic form

$$B_{\mathsf{M}}(\mathsf{x}, \mathsf{y}) = 4\pi^2 \|\mathsf{x}v_1 + \mathsf{y}v_2\|^2 \tag{1.1}$$

at integer points, where $\{v_1, v_2\}$ is a basis for the dual lattice Δ^*.

Let

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \cdots$$

be the corresponding eigenvalues counted with multiplicity. By the Weyl's law we have

$$\#\{j : \lambda_j \leq T\} \sim \frac{\mathrm{vol}(\mathsf{M})}{4\pi} T.$$

The set of eigenvalues has a clear symmetry; let us write $j \sim k$ if $\lambda_j = B_{\mathsf{M}}(u)$ and $\lambda_k = B_{\mathsf{M}}(\pm u)$. Let $\alpha < \beta$, and define the **pair correlation function**

$$R_{\mathsf{M}}(\alpha, \beta, T) = \frac{\#\{(j, k) : j \nsim k, \lambda_j, \lambda_k \leq T,\ \alpha \leq \lambda_j - \lambda_k \leq \beta\}}{T}.$$

The following was proved by Eskin, Margulis, and Mozes [EMM05].

1.1 Theorem ([EMM05], Theorem 1.7)**.** *Let* M *be a two-dimensional flat torus, and let*

$$B_{\mathsf{M}}(\mathsf{x}, \mathsf{y}) = \mathsf{a}\mathsf{x}^2 + 2\mathsf{b}\mathsf{x}\mathsf{y} + \mathsf{c}\mathsf{y}^2$$

be the associated quadratic form giving the Laplacian spectrum of M*, normalized so that* $\mathsf{ac} - \mathsf{b}^2 = 1$*. Suppose there exist* $A \geq 1$ *such that for all* $(p_1, p_2, q) \in \mathbb{Z}^3$ *with* $q \geq 2$*, we have*

$$\left|\frac{\mathsf{b}}{\mathsf{a}} - \frac{p_1}{q}\right| + \left|\frac{\mathsf{c}}{\mathsf{a}} - \frac{p_2}{q}\right| > q^{-A}. \tag{1.2}$$

Then for any interval $[\alpha, \beta]$ *with* $0 \notin [\alpha, \beta]$*, we have*

$$\lim_{T\to\infty} R_{\mathsf{M}}(\alpha, \beta, T) = \pi^2(\beta - \alpha). \tag{1.3}$$

Prior to [EMM05], Sarnak [Sar97] showed that (1.3) holds on a set of full measure in the space of flat tori. The case of inhomogeneous forms, which correspond to eigenvalues of quasi-periodic eigenfunctions, was also studied by Marklof [Mar03, Mar02], and by Margulis and the second-named author [MM11]. More recently, Blomer, Bourgain, Radziwill, and Rudnick [BBRlR17] studied consecutive spacing for certain families of rectangular tori, i.e., $\mathsf{b} = 0$. We also refer to the work of Strömbergsson and Vishe [SV20] where an effective version of [Mar03] is obtained.

In this paper, we prove a polynomially effective version of Theorem 1.1, i.e., we provide a polynomial error term for $R_{\mathsf{M}}(\alpha, \beta, T)$.

1.2 Theorem. *Let* M *be a two-dimensional flat torus,*

$$B_{\mathsf{M}}(\mathsf{x}, \mathsf{y}) = \mathsf{a}\mathsf{x}^2 + 2\mathsf{b}\mathsf{x}\mathsf{y} + \mathsf{c}\mathsf{y}^2$$

the associated quadratic form giving the Laplacian spectrum of M *normalized so that* $\mathsf{ac} - \mathsf{b}^2 = 1$*, and let* $A \geq 10^3$*. Then there are absolute constants* δ_0 *and* N*, some* A' *depending on* A*, and* C *and* T_0 *depending on* A*,* a*,* b*, and* c*, and for every* $0 < \delta \leq \delta_0$*, a* $\kappa = \kappa(\delta, A)$ *so that the following holds.*

Let $T \geq T_0$*, assume that for all* $(p_1, p_2, q) \in \mathbb{Z}^3$ *with* $T^{\delta/A'} < q < T^{\delta}$ *we have*

$$\left|\frac{\mathsf{b}}{\mathsf{a}} - \frac{p_1}{q}\right| + \left|\frac{\mathsf{c}}{\mathsf{a}} - \frac{p_2}{q}\right| > q^{-A}. \tag{1.4}$$

Then if

$$|R_{\mathsf{M}}(\alpha, \beta, T) - \pi^2(\beta - \alpha)| > C(1 + |\alpha| + |\beta|)^N T^{-\kappa},$$

then there are two primitive vectors $u_1, u_2 \in \mathbb{Z}^2$ *so that*

$$\|u_1\|, \|u_2\| \leq T^{\delta/A} \quad \text{and} \quad |B_{\mathsf{M}}(u_1, u_2)| \leq T^{-1+\delta} \tag{1.5}$$

and moreover

$$R_{\mathsf{M}}(\alpha, \beta, T) - \pi^2(\beta - \alpha) = \frac{M_T(u_1, u_2)}{T} + O((1 + |\alpha| + |\beta|)^N T^{-\kappa})$$

with

$$M_T(u_1, u_2) = \#\Big\{(\ell_1, \ell_2) \in \frac{1}{2}\mathbb{Z}^2 : \ell_1 u_1 \pm \ell_2 u_2 \in \mathbb{Z}^2,\ B_{\mathsf{M}}(\ell_1 u_1 \pm \ell_2 u_2) \leq T,$$
$$4B_{\mathsf{M}}(u_1, u_2)\ell_1\ell_2 \in [\alpha, \beta]\Big\}.$$

Note that our result does not require the assumption that $0 \notin [\alpha, \beta]$ (a restriction that appears in the work of Eskin, Margulis and Mozes, and is needed in order for Theorem 1.1 to hold). The proof of Theorem 1.2 is effective, and for all of the above implicit constants, one can give explicit expressions if desired.

Remark. Let us now elaborate on the term $M_T(u_1, u_2)$ in the statement of Theorem 1.2: Let $u_1, u_2 \in \mathbb{Z}^2$ be two primitive vectors satisfying

$$0 < \|u_i\| \leq T^{\delta/A} \quad \text{and} \quad |B_{\mathsf{M}}(u_1, u_2)| \leq T^{-1+\delta}.$$

Then for all $(\ell_1, \ell_2) \in \frac{1}{2}\mathbb{Z}^2$, we have

$$B_{\mathsf{M}}(\ell_1 u_1 + \ell_2 u_2) - B_{\mathsf{M}}(\ell_1 u_1 - \ell_2 u_2) = 4B_{\mathsf{M}}(u_1, u_2)\ell_1\ell_2.$$

In particular, if $T^{-1-\delta} \le |B_{\mathsf{M}}(u_1, u_2)| \le T^{-1+\delta}$, then there would be $\gg T^{1-10\delta}$ pairs of integers ℓ_1, ℓ_2 of size $\ll T^{\frac{1}{2}(1-\delta)}$ (so that $B_{\mathsf{M}}(\ell_1 u_1 + \ell_2 u_2) \le T$), such that

$$\ell_1\ell_2 \in \Big[\frac{\alpha}{4B_{\mathsf{M}}(u_1, u_2)}, \frac{\beta}{4B_{\mathsf{M}}(u_1, u_2)}\Big]$$

as the last interval is of length $\gg T^{1-\delta}$. All such pairs contribute to $M_T(u_1, u_2)$, making $\frac{M_T(u_1,u_2)}{T} \gg T^{-10\delta}$, which is bigger than any fixed polynomial error term. Moreover, even if (1.4) holds, such pairs $u_1, u_2 \in \mathbb{Z}^2$ can definitely exist.

If (1.4) holds, up to changing the order such a pair u_1, u_2 is unique—see Lemma 2.5—hence there is no need for additional error terms. The subspaces of $\mathbb{R}^4$ spanned by pairs (u_1, u_2) as above are called exceptional. In Section 6 we introduce a Margulis function that accounts for all the contributions towards pairs R_{M} out of exceptional spaces and show that exceptional subspaces are the only source of large error terms.

We now state a corollary of Theorem 1.2. A rectangular torus has extra multiplicities in the spectrum built in, so to accommodate that we consider the modified pair correlation function

$$R'_{\mathsf{M}}(\alpha, \beta, T) = \frac{\#\{(j,k) : \lambda_j \ne \lambda_k < T,\ \alpha \le \lambda_j - \lambda_k \le \beta\}}{T}.$$

1.3 Corollary. *Let* M *be a two dimensional flat torus, and let*

$$B_{\mathsf{M}}(\mathsf{x}, \mathsf{y}) = \mathsf{a}\mathsf{x}^2 + 2\mathsf{b}\mathsf{x}\mathsf{y} + \mathsf{c}\mathsf{y}^2$$

be normalized so that $\mathsf{ac} - \mathsf{b}^2 = 1$.

(1) *Suppose there exist* $A \ge 1$ *and* $q > 0$ *such that for all* $(m, n, k) \in \mathbb{Z}^3 \setminus \{0\}$,

$$|\mathsf{a}m + \mathsf{b}n + \mathsf{c}k| > q\|(m, n, k)\|^{-A}. \tag{1.6}$$

Then

$$|R_{\mathsf{M}}(\alpha, \beta, T) - \pi^2(\beta - \alpha)| \le C(1 + |\alpha| + |\beta|)^N T^{-\kappa}.$$

(2) *Let* M *be a rectangular torus, i.e.,* $\mathsf{b} = 0$. *Assume there exist* $A \ge 1$ *and* $q > 0$ *such that for all* $(m, n) \in \mathbb{Z}^2 \setminus \{0\}$ *we have*

$$|\mathsf{a}^2 m + n| > q\|(m, n)\|^{-A}.$$

Then

$$|R'_{\mathsf{M}}(\alpha, \beta, T) - \pi^2(\beta - \alpha)| \le C(1 + |\alpha| + |\beta|)^N T^{-\kappa},$$

where N *is absolute,* κ *depends on* A, *and* C *depends on* a, b, c, A, *and* q.

Indeed, under (1.6), pairs u_1, u_2 of primitive integer vectors as in Theorem 1.2 do not exist, and if M is a rectangular torus the unique (up to order) pair of primitive vectors is given by $e_1 = (1, 0)$, $e_2 = (0, 1)$, for which the contribution of $M_T(e_1, e_2)$ can be accounted for by looking at $R'_{\mathsf{M}}(\alpha, \beta, T)$ instead of $R_{\mathsf{M}}(\alpha, \beta, T)$.

Note that in part (2), though the modified pair correlation function $R'_{\mathsf{M}}(\alpha, \beta, T)$ avoids counting zero values, the interval $[\alpha, \beta]$ is still allowed to contain 0. This is slightly stronger than assuming $0 \notin [\alpha, \beta]$, as Corollary 1.3.(2) in particular gives effective bounds on the number of extremely close eigenvalues.

The general strategy of the proof of Theorem 1.2 is similar to [EMM98] and [EMM05]. That is, we deduce the above theorems from an equidistribution theorem for certain unbounded functions in homogeneous spaces. Unlike [EMM98] and [EMM05], where the analysis takes place in the space of unimodular lattices in $\mathbb{R}^4$, the homogeneous space in question here is

$$X = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})/\Gamma'$$

where Γ' is a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$.

This reduction is carried out in §3. The lower bound estimate will be proved using the following effective equidistribution theorem that relies on [LMW22, Thm. 1.1]:

Let $G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$. For all $h \in \mathrm{SL}_2(\mathbb{R})$, we let $\Delta(h)$ denote the element $(h, h) \in G$, and let $H = \Delta(\mathrm{SL}_2(\mathbb{R}))$. For every $t \in \mathbb{R}$ and every $\theta \in [0, 2\pi]$, let

$$a_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \quad \text{and} \quad r_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

1.4 Theorem. *Assume Γ is an arithmetic lattice in G. For every $x_0 \in X = G/\Gamma$, and large enough R (depending explicitly on X and the injectivity radius at x_0), for any $e^t \geq R^D$, at least one of the following holds.*

(1) *For every $\varphi \in C_c^\infty(X)$ and 2π-periodic smooth function ξ on $\mathbb{R}$, we have*

$$\left| \int_0^{2\pi} \varphi(\Delta(a_t r_\theta)x_0)\xi(\theta)\,\mathrm{d}\theta - \int_0^{2\pi} \xi(\theta)\,\mathrm{d}\theta \int \varphi\,\mathrm{d}m_X \right| \leq \mathcal{S}(\varphi)\mathcal{S}(\xi)R^{-\kappa_0}$$

where we use $\mathcal{S}(\cdot)$ to denote an appropriate Sobolev norm on both X and $\mathbb{R}$, respectively.

(2) *There exists $x \in X$ such that Hx is periodic with $\mathrm{vol}(Hx) \leq R$, and*

$$d_X(x, x_0) \leq R^D t^D e^{-t}.$$

The constants D and κ_0 are positive and depend on X but not on x_0, and d_X is a fixed metric on X.

This is a variant of [LMW22, Thm. 1.1]. Indeed, instead of expanding an orbit segment of the unipotent flow $\Delta(u_s)$ where

$$u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix},$$

here we expand an orbit of the compact group $\{\Delta(r_\theta)\}$. The deduction of Theorem 1.4 from [LMW22, Thm. 1.1] is given in §5 using a fairly simple and standard argument.

To prove the upper bound estimate, in addition to Theorem 1.4, we also need to analyze Margulis functions à la [EMM98, EMM05]; our analysis simplifies substantially thanks to the simpler structure of the cusp in $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})/\Gamma'$ compared to that in $\mathrm{SL}_4(\mathbb{R})/\mathrm{SL}_4(\mathbb{Z})$. This is the content of §6. Indeed Proposition 6.1 reduces the analysis to special subspaces, see Definition 2.3, that are closely connected to the pairs of almost B_{M}-orthogonal vectors discussed above. We study these special subspaces using the elementary Lemma 2.2; in particular, using this lemma we establish Lemma 2.4, which shows that under (1.4) there are at most two special subspaces. Finally, Lemma 2.6 shows that even for special subspaces, only the range asserted in (1.5) can produce enough solutions to affect the error term.

Acknowledgments. We would like to thank Jens Marklof for helpful conversations. We also thank the anonymous referee for valuable comments.

This paper is dedicated to Peter Sarnak on the occasion of his 70th birthday. Peter's deep and remarkably broad work touches many areas in mathematics, and in particular this work is strongly connected to his work: whether directly [Sar97] or indirectly through the paper of Eskin, Margulis and Mozes [EMM05] where he is thanked for encouragement and many helpful conversation, as well as in other ways—some we are aware of, and some that we will surely find out. Peter has also been consistently and actively encouraging the research towards effective equidistribution results for unipotent flows (indeed, he has authored one of the first papers in this direction [Sar81] already in 1981!). The three of us have been fortunate to benefit greatly from his profound knowledge and exceptional generosity, and it is a pleasure to dedicate this paper to him with our sincere gratitude for all he has done to help us and many, many others.

2 Notation and preliminaries

In this paper

$$G = \left\{ \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix} : g_1, g_2 \in \mathrm{SL}_2(\mathbb{R}) \right\} \quad \text{and} \quad H = \left\{ \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} : g \in \mathrm{SL}_2(\mathbb{R}) \right\}.$$

Let $\mathfrak{g} = \mathrm{Lie}(G)$ and $\mathfrak{h} = \mathrm{Lie}(H)$.

We identify G with $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ and H with

$$\{(g, g) : g \in \mathrm{SL}_2(\mathbb{R})\} \subset \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}).$$

Indeed, to simplify the notation, we will often denote

$$g = \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix} \in G$$

by (g_1, g_2). Given $v = (\mathsf{x}_1, \mathsf{y}_1, \mathsf{x}_2, \mathsf{y}_2) \in \mathbb{R}^4$, we write $g.v = (g_1 v_1, g_2 v_2)$ where $v_i = (\mathsf{x}_i, \mathsf{y}_i) \in \mathbb{R}$ for $i = 1, 2$ (for purely typographical reasons, we prefer to work with row vectors even though representing these as column vectors would be more consistent).

For all $h \in \mathrm{SL}_2(\mathbb{R})$, we let $\Delta(h) = (h, h) \in H$. In particular, for every $t \in \mathbb{R}$ and every $\theta \in [0, 2\pi]$, $\Delta(a_t)$ and $\Delta(r_\theta)$ denote the images of

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

in H, respectively.

2.1 Quadratic forms. Let Q_0 denote the *determinant* form on $\mathbb{R}^4$:

$$Q_0(\mathsf{x}_1, \mathsf{y}_1, \mathsf{x}_2, \mathsf{y}_2) = \mathsf{x}_1 \mathsf{y}_2 - \mathsf{x}_2 \mathsf{y}_1.$$

Note that $H = G \cap \mathrm{SO}(Q_0)$.

Let $\Delta \subset \mathbb{R}^2$ be a lattice and let Δ^* be the dual lattice. We normalize Δ^* to have covolume $(2\pi)^{-2}$ and fix $g_{\mathsf{M}} \in \mathrm{SL}_2(\mathbb{R})$ so that

$$2\pi \Delta^* = g_{\mathsf{M}} \mathbb{Z}^2.$$

The eigenvalues of the Laplacian on $\mathbb{R}^2/\Delta$ are $\|v\|^2$ for $v \in 2\pi\Delta^*$. Therefore, given two eigenvalues $\lambda_i = \|v_i\|^2$, $i = 1, 2$, we have

$$\begin{aligned} \lambda_1 - \lambda_2 &= (\|v_1\|^2 - \|v_2\|^2) = (v_1 + v_2) \cdot (v_1 - v_2) \\ &= Q_0(v_1 + v_2, \omega(v_1 - v_2)) \end{aligned} \tag{2.1}$$

where $\omega = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

Recall that $G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}) \subset \mathrm{SL}_4(\mathbb{R})$. Define

$$\Lambda = \{(v_1 + v_2, \omega(v_1 - v_2)) : v_1, v_2 \in \mathbb{Z}^2\} \subset \mathbb{R}^4.$$

Then $\{(v_1 + v_2, \omega(v_1 - v_2)) : v_1, v_2 \in 2\pi\Delta^*\} = (g_{\mathsf{M}}, -\omega g_{\mathsf{M}}\omega)\Lambda$.

Let Γ' be the maximal subgroup of $\mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ which preserves Λ. More explicitly,

$$\Gamma' = \{(\gamma_1, \gamma_2) \in \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) : \gamma_1 \equiv \omega\gamma_2\omega \pmod 2\}.$$

Let $X = G/\Gamma'$.

Möbius transformations. In this section, we prove an elementary lemma concerning Möbius transformations. This lemma will be used to complete the proof of Lemma 2.5; it also will be used in the proof of Lemma 6.4.

Let $\mathcal{P}$ denote the set of primitive vectors in $\mathbb{Z}^2$. For every $t \geq 1$, let

$$\mathcal{P}(t) = \{v \in \mathcal{P} : \|v\| < e^t\}.$$

2.2 Lemma. *Let $A \geq 10^3$, $s > 0$ and $0 < \eta^A < e^{-s/100}$. Assume that for $i = 1, 2$ there are $v_i, v_i', v_i'' \in \mathcal{P}(s)$ satisfying*

$$1 \leq |Q_0(v, w)| \ll \eta^{-4}, \quad \text{for } v, w \in \{v_i, v_i', v_i''\}. \tag{2.2}$$

Also suppose there are $h \in \mathrm{PGL}_2(\mathbb{R})$ and $C > 0$ so that

$$hv_1 = \mu v_2 + w_{1,2}, \quad hv_1' = \mu' v_2' + w_{1,2}', \quad hv_1'' = \mu'' v_2'' + w_{1,2} \tag{2.3}$$

where $|\mu|, |\mu'|, |\mu''| \geq C^{-1}$ and $\|w\| \leq C\eta^A e^{-s}$ for $w \in \{w_{1,2}, w_{1,2}', w_{1,2}''\}$.

Then there exists $Q \in \mathrm{Mat}_2(\mathbb{Z})$ with $\|Q\| \leq \eta^{-100}$ and $\lambda \in \mathbb{R}$ so that

$$\|h - \lambda Q\| \leq C'\eta^{A-50},$$

where C' depends on C and polynomially on $\|h\|$.

Proof. Let us write $v_i = (\mathsf{x}_i, \mathsf{y}_i)$, $v_i' = (\mathsf{x}_i', \mathsf{y}_i')$, and $v_i'' = (\mathsf{x}_i'', \mathsf{y}_i'')$. The matrix

$$M_1 = \begin{pmatrix} \mathsf{y}_1 & -\mathsf{x}_1 \\ \mathsf{y}_1'\mathsf{z}_1 & -\mathsf{x}_1'\mathsf{z}_1 \end{pmatrix} \quad \text{for } \mathsf{z}_1 = \frac{\mathsf{x}_1''\mathsf{y}_1 - \mathsf{x}_1\mathsf{y}_1''}{\mathsf{x}_1''\mathsf{y}_1' - \mathsf{x}_1'\mathsf{y}_1''}$$

acting on $\mathbb{P}^1$ takes $(\mathsf{x}_1 : \mathsf{y}_1)$ to $(0 : 1)$, $(\mathsf{x}_1' : \mathsf{y}_1')$ to $(1 : 0)$ and $(\mathsf{x}_1'' : \mathsf{y}_1'')$ to $(1 : 1)$. The matrix

$$M_2 = \begin{pmatrix} -\mathsf{x}_2'\mathsf{z}_2 & \mathsf{x}_2 \\ -\mathsf{y}_2'\mathsf{z}_2 & \mathsf{y}_2 \end{pmatrix} \quad \text{for } \mathsf{z}_2 = \frac{\mathsf{x}_2''\mathsf{y}_2 - \mathsf{x}_2\mathsf{y}_2''}{\mathsf{x}_2''\mathsf{y}_2' - \mathsf{x}_2'\mathsf{y}_2''}$$

in turn takes $(0:1)$ to $(\mathsf{x}_2:\mathsf{y}_2)$, $(1:0)$ to $(\mathsf{x}_2':\mathsf{y}_2')$ and $(1:1)$ to $(\mathsf{x}_2'':\mathsf{y}_2'')$. By (2.2), we have that $r = |\det(M_1)\det(M_2)|^{-1}$ is a rational number of height $\ll \eta^{-20}$. Thus by (2.3)

$$h = \pm\sqrt{r} M_2 M_1 + O(\eta^{A-50}) \quad \text{or}$$
$$h = \pm\sqrt{r}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} M_2 M_1 + O(\eta^{A-50}).$$

Since the denominators of the entries of M_1 and M_2 are bounded by η^{-4}, and since all our implicit constants are allowed to depend on $\|h\|$, we may conclude the claim. □

We draw some corollaries of Lemma 2.2.

Definition 2.3. Let $g = (g_1, g_2) \in G$. A two-dimensional $g\mathbb{Z}^4$-rational linear subspace $L \subset \mathbb{R}^4$ is called (ρ, A, t)-exceptional if there are $(v_1, 0), (0, v_2) \in \mathbb{Z}^4$ satisfying

$$\|g_1 v_1\|, \|g_2 v_2\| \le e^{\rho t} \quad \text{and} \quad |Q_0(g_1 v_1, g_2 v_2)| \le e^{-A\rho t} \tag{2.4}$$

so that $L \cap g\mathbb{Z}^4$ is spanned by $\{(g_1 v_1, 0), (0, g_2 v_2)\}$.

Given a (ρ, A, t)-special subspace L, we will refer to $\{(g_1 v_1, 0), (0, g_2 v_2)\}$ as a *spanning set* for L.

2.4 Lemma. *Let $A \ge 10^3$, and let $g = (g_1, g_2) \in G$. Let $\rho \le A/100$. Then for all t large enough, depending on $\|g\|$, at least one of the following holds:*

(1) *There are at most two different (ρ, A, t)-exceptional subspaces.*
(2) *There exists $Q \in \mathrm{Mat}_2(\mathbb{Z})$ whose entries are bounded by $e^{100\rho t}$ and $\lambda \in \mathbb{R}$ satisfying $\|g_2^{-1} g_1 - \lambda Q\| \le e^{-(A-100)\rho t}$.*

Proof. We begin by proving the first assertion in the lemma. Let $\eta = e^{-\rho t}$ and $s = \rho t$. Indeed assume there are three different (ρ, A, t)-special subspaces in $\mathbb{R}^4$, and let $v_i, v_i', v_i'' \in \mathcal{P}_s$, $i = 1, 2$, be the corresponding spanning vectors. Then

$$1 < |Q_0(v, w)| \ll e^{2\rho t}, \quad \text{for } v, w \in \{v_1, v_1', v_1''\}.$$

That is, $\{v_1, v_1', v_1''\}$ satisfies (2.2) with $\eta = e^{-\rho t}$ so long as t is large enough to account for the implied constant. Moreover, if we put $h = g_2^{-1} g_1$, then

$$h v_1 = \mu v_2 + w_{1,2}$$

where $\mu \in \mathbb{R}$ satisfies $|\mu| \gg 1$ and $\|w_{1,2}\| \ll e^{-A\rho t} = \eta^{(A-1)} e^{-s}$ (recall that the implicit constants in these inequalities are allowed to depend polynomially on $\|g_1\|$ and $\|g_2\|$). Similarly,

$$h v_1' = \mu' v_2' + w_{1,2}' \quad \text{and} \quad h v_1'' = \mu'' v_2'' + w_{1,2}''$$

where $\mu', \mu'' \in \mathbb{R}$ satisfy $|\mu'|, |\mu''| \gg 1$ and $\|w'_{1,2}\|, \|w''_{1,2}\| \ll e^{-A\rho t}$. Therefore, $\{v_2, v'_2, v''_2\}$ also satisfy (2.2). Moreover, $h = g_2^{-1} g_1$ satisfies (2.3) with $A - 1$, η, and s, in view of the above discussion. Hence, Lemma 2.2 implies that the assertion in part (2) of this lemma holds so long as t is large enough. □

Special subspaces and the spectrum of flat tori. Using the discussion in §2.1, we translate the conclusion of Lemma 2.4 to a similar statement about the quadratic form B_{M}.

2.5 Lemma. *Let $A \geq 10^4$, and let $\rho \leq A/100$. Recall that*

$$B_{\mathsf{M}}(\mathsf{x}, \mathsf{y}) = \mathsf{a}\mathsf{x}^2 + 2\mathsf{b}\mathsf{x}\mathsf{y} + \mathsf{c}\mathsf{y}^2$$

is renormalized so that $\mathsf{ac} - \mathsf{b}^2 = 1$. *Then for all $t \geq t_0$, depending on ρ, $|\mathsf{a}|$, $|\mathsf{b}|$, and $|\mathsf{c}|$, at least one of the following holds:*

(1) *There is a unique, up to change of order, pair of primitive vectors $u_1, u_2 \in \mathbb{Z}^2 \setminus \{0\}$ satisfying*

$$\|u_i\| \leq e^{\rho t} \quad \text{and} \quad |B_{\mathsf{M}}(u_1, u_2)| \leq e^{-A\rho t}.$$

(2) *There exists $Q \in \mathrm{Mat}_2(\mathbb{Z})$ whose entries are bounded by $e^{100\rho t}$ and $\lambda \in \mathbb{R}$ satisfying*

$$\left\| \begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix} - \lambda Q \right\| \leq e^{-(A-100)\rho t}.$$

In particular, if M is a rectangular torus, i.e., $\mathsf{b} = 0$, then t_0 may be chosen so that if part (2) is not satisfied, then (up to changing the order) $u_1 = (1, 0)$ and $u_2 = (0, 1)$.

Proof. Let t_1 be large enough so that Lemma 2.4 holds for all $t \geq t_1$. Since B_{M} is positive definite, there exists t'_0 so that if $t \geq t'_0$, then

$$|B_{\mathsf{M}}(u_1, u_2)| < e^{-A\rho t}$$

implies that $\{u_1, u_2\}$ is linearly independent.

Let $t_0 = \max(t_1, t'_0)$ and let $t \geq t_0$. Put

$$g = (g_1, 1) = \left(\begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix}, 1 \right).$$

Note that if part (2) in Lemma 2.4 holds, then part (2) in this lemma holds and the proof is complete. Thus let us assume that part (1) in Lemma 2.4 holds.

Let $u_i = (\mathsf{x}_i, \mathsf{y}_i) \in \mathbb{Z}^2 \setminus \{0\}$. Then

$$\begin{aligned} B_{\mathsf{M}}(u_1, u_2) &= (\mathsf{x}_1, \mathsf{y}_1) \begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix} \begin{pmatrix} \mathsf{x}_2 \\ \mathsf{y}_2 \end{pmatrix} \\ &= (\mathsf{a}\mathsf{x}_1 + \mathsf{b}\mathsf{y}_1)\mathsf{x}_2 + (\mathsf{b}\mathsf{x}_1 + \mathsf{c}\mathsf{y}_1)\mathsf{y}_2 \\ &= \left(\begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix} \begin{pmatrix} \mathsf{x}_1 \\ \mathsf{y}_1 \end{pmatrix} \right) \wedge \begin{pmatrix} -\mathsf{y}_2 \\ \mathsf{x}_2 \end{pmatrix} = Q_0\big(g_1(\mathsf{x}_1, \mathsf{y}_1), (-\mathsf{y}_2, \mathsf{x}_2)\big). \end{aligned}$$

Thus if u_1, u_2 satisfy part (1), then $(g_1(\mathsf{x}_1, \mathsf{y}_1), (0, 0))$ and $((0, 0), (-\mathsf{y}_2, \mathsf{x}_2))$ span a (ρ, A, t)-special subspace for $g\mathbb{Z}^4$.

By Lemma 2.4, there is at most two such subspaces. Moreover, since $B_{\mathsf{M}}(\ ,\ )$ is symmetric, we conclude that

$$Q_0(g_1(\mathsf{x}_2, \mathsf{y}_2), (-\mathsf{y}_1, \mathsf{x}_1)) = Q_0\big(g_1(\mathsf{x}_1, \mathsf{y}_1), (-\mathsf{y}_2, \mathsf{x}_2)\big).$$

This implies the two special subspaces are spanned by

$$\{(g_1(\mathsf{x}_1, \mathsf{y}_1), 0, 0), (0, 0, -\mathsf{y}_2, \mathsf{x}_2)\} \quad \text{or} \quad \{(g_1(\mathsf{x}_2, \mathsf{y}_2), 0, 0), (0, 0, -\mathsf{y}_1, \mathsf{x}_1)\}.$$

This shows part (1) in this lemma holds.

Assume now that $\mathsf{b} = 0$, and suppose part (2) does not hold. Let u_i be as in part (1). Then

$$|B_{\mathsf{M}}(u_1, u_2)| = |\mathsf{a}\mathsf{x}_1\mathsf{x}_2 + \mathsf{a}^{-1}\mathsf{y}_1\mathsf{y}_2| \le e^{-A\rho t}. \tag{2.5}$$

Unless $\mathsf{y}_1\mathsf{y}_2 = 0$, the above contradicts that part (2) does not hold. Therefore, we have $\mathsf{y}_1\mathsf{y}_2 = 0$. Assuming t is large enough so that the right side in (2.5) is $< |\mathsf{a}|$, we conclude $\mathsf{x}_1\mathsf{x}_2 = 0$ and the claim follows. □

The following lemma further investigates the contribution of special subspaces, or more precisely, vectors u_1, u_2 satisfying part (1) in Lemma 2.5. We note that condition (2.6) is (1.5) in Theorem 1.2.

2.6 Lemma. *Let $A \ge 10^3$ and $0 < \rho < 1/(100A)$. Let*

$$B_{\mathsf{M}}(\mathsf{x}, \mathsf{y}) = \mathsf{a}\mathsf{x}^2 + 2\mathsf{b}\mathsf{x}\mathsf{y} + \mathsf{c}\mathsf{y}^2$$

which is normalized so that $\mathsf{a}\mathsf{c} - \mathsf{b}^2 = 1$. The following holds for all large enough t, depending on ρ, $|\mathsf{a}|$, $|\mathsf{b}|$, and $|\mathsf{c}|$. Let $u_1, u_2 \in \mathbb{Z}^2 \setminus \{0\}$ satisfy

$$\|u_i\| \le e^{\rho t} \quad \textit{and} \quad |B_{\mathsf{M}}(u_1, u_2)| \le e^{-A\rho t}.$$

Assume further that

$$|B_{\mathsf{M}}(u_1, u_2)| > e^{(-2+2\rho)t}. \tag{2.6}$$

Let $C > 0$, then

$$\#\Big\{(\ell_1, \ell_2) \in \frac{1}{2}\mathbb{Z}^2 : |\ell_i| \le Ce^t 4B_{\mathsf{M}}(u_1, u_2)\ell_1\ell_2 \in [\alpha, \beta]\Big\} \ll \max(|\alpha|, |\beta|)e^{(2-\rho)t}$$

where the implied constant depends on C, a, b, and c.

Proof. Let (ℓ_1, ℓ_2) satisfy that $|\ell_i| \le Ce^t$ and

$$4B_{\mathsf{M}}(u_1, u_2)\ell_1\ell_2 \in [\alpha, \beta]. \tag{2.7}$$

Then the number of solutions with $\ell_1 = 0$ or $\ell_2 = 0$ is $\ll e^t$. Therefore, we assume $\ell_i \neq 0$ for $i = 1, 2$ for the rest of the argument.

Assume that

$$|B_{\mathsf{M}}(u_1, u_2)| > e^{(-2+2\rho)t}.$$

Then (2.7) implies that

$$0 < 4|\ell_1\ell_2| \le \max(|\alpha|, |\beta|)e^{(2-2\rho)t}. \tag{2.8}$$

The number of $(\ell_1, \ell_2) \in \mathbb{Z}^2$ with $0 < |\ell_1| \le Ce^t$ so that (2.8) holds is

$$\ll \max(|\alpha|, |\beta|)te^{(2-2\rho)t} \ll \max(|\alpha|, |\beta|)e^{(2-\rho)t}$$

as we claimed. □

3 Circular averages and values of quadratic forms

In this section, we state an equidistribution result for the action of $\mathrm{SO}(Q_0)$. Theorem 1.2 will be deduced from this equidistribution theorem in §4 using some preparatory lemmas which will be established in this section.

Let f_i be compactly supported bounded Borel functions on $\mathbb{R}^2$, and define f on $\mathbb{R}^4$ by $f(w_1, w_2) = f_1(w_1)f_2(w_2)$. For any $g' \in G$, let

$$\hat{f}(g'\Gamma') = \sum_{v \in g'\Lambda_{\mathrm{nz}}} f(v) \tag{3.1}$$

where

$$\begin{aligned}
\Lambda &= \{(v_1 + v_2, \omega(v_1 - v_2)) : v_1, v_2 \in \mathbb{Z}^2\} \subset \mathbb{R}^4,\\
\Lambda_{\mathrm{nz}} &= \{(w_1, w_2) \in \Lambda : w_1 \neq 0 \text{ and } w_2 \neq 0\},\\
\Gamma' &= \{(\gamma_1, \gamma_2) \in \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) : \gamma_1 \equiv \omega\gamma_2\omega \pmod 2\},
\end{aligned}$$

and $\omega = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Note that Γ' preserves Λ and Λ_{nz}.

Let $X = G/\Gamma'$, and let m_X denote the G-invariant probability measure on X.

3.1 Theorem. *For every $A \geq 10^4$ and $0 < \rho \leq 10^{-4}$, there exist $\hat{A}$ (depending on A) and δ_1, δ_2 (depending on ρ and A) with*

$$\rho/\hat{A} \leq \delta_1/A \leq \rho/100,$$

so that for all $g = (g_1, g_2) \in G$ and all large enough t, depending linearly on $\log(\|g_i\|)$, the following holds.

Assume that for every $Q \in \mathrm{Mat}_2(\mathbb{Z})$ with $e^{\rho t/\hat{A}} \leq \|Q\| \leq e^{\rho t}$ and all $\lambda \in \mathbb{R}$, we have

$$\|g_2^{-1}g_1 - \lambda Q\| > \|Q\|^{-A/1000}. \tag{3.2}$$

There exists some C' depending on A and polynomially on $\|g_i\|$ so that the following holds. For any 2π-periodic smooth function ξ on $\mathbb{R}$, if

$$\left| \int_0^{2\pi} \hat{f}(\Delta(a_t r_\theta) g \Gamma') \xi(\theta)\, \mathrm{d}\theta - \int_0^{2\pi} \xi \, \mathrm{d}\theta \int_X \hat{f} \, \mathrm{d}m_X \right| > C' \mathcal{S}(f) \mathcal{S}(\xi) e^{-\delta_2 t}$$

then there are at least one, and at most two, $(\delta_1/A, A, t)$-exceptional subspaces, say L and L' (for notational convenience, if there is only one exceptional subspace, set $L' = L$). Moreover

$$\int_0^{2\pi} \hat{f}(\Delta(a_t r_\theta) g \Gamma') \xi(\theta)\, \mathrm{d}\theta = \int_0^{2\pi} \xi \, \mathrm{d}\theta \int_X \hat{f} \, \mathrm{d}m_X + \mathcal{M} + O(\mathcal{S}(f)\mathcal{S}(\xi) e^{-\delta_2 t})$$

where

$$\mathcal{M} = \int_{\mathcal{C}} \hat{f}_{\mathrm{sp}}(\theta) \xi(\theta)\, \mathrm{d}\theta$$

with

$$\hat{f}_{\mathrm{sp}}(\theta) = \sum_{v \in g\Lambda_{\mathrm{nz}} \cap (L \cup L')} f(\Delta(a_t r_\theta) v),$$

$$\mathcal{C} = \{\theta \in [0, 2\pi] : \hat{f}_{\mathrm{sp}}(\theta) \geq e^{\delta_1 t}\}.$$

The proof of Theorem 3.1 will be completed in §7; it relies on results in §5 and §6. Notice that, even though the functions f_1, f_2 are bounded on $\mathbb{R}^2$, the resulting function $\hat{f}$ is unbounded on G/Γ'. It is well-recognized that such unboundedness can be overcome with the use of cusp functions and their contracting functions; see, e.g., [EM22]. The adaptation of this method to our setting where exceptional subspaces are present will be contained in §6.

The goal in the remaining parts of this section and §4 is to complete the proof of Theorem 1.2 using Theorem 3.1. We will also explicate the proof of Corollary 1.3 at the end of §4.

Before proceeding, however, let us record an a priori, i.e., without assuming (3.2), upper bound for $\int_0^{2\pi} \hat{f}(\Delta(a_t r_\theta) g \Gamma')\, \mathrm{d}\theta$.

3.2 Lemma. *For every* $0 < \eta < 1$*, there exists* $t_\eta \ll |\log\eta|$ *so that the following hold. Let* $g = (g_1, g_2) \in G$ *and* $R \geq 1$*; assume that* $\|g_i\| \leq R$*. Let* f_i *be the characteristic function of* $\{w \in \mathbb{R}^2 : \|w\| \leq R\}$*, and put* $f = f_1 f_2$*.*

(1) *For every* $t \geq t_\eta$ *we have*

$$\int_0^{2\pi} \hat{f}(\Delta(a_t r_\theta) g\Gamma')\,\mathrm{d}\theta \ll e^{\eta t}.$$

(2) *Let* $t \geq t_\eta$*. Let* $L \subset \mathbb{R}^4$ *be a two-dimensional subspace so that* $L \cap g\mathbb{Z}^4$ *is spanned by* $\{(g_1 v_1, 0), (0, g_2 v_2)\}$ *for* $(v_1, 0), (0, v_2) \in \mathbb{Z}^4 \setminus \{0\}$*. Then*

$$\int_{[0,2\pi]\setminus\mathcal{C}_L} \hat{f}_L(\theta)\,\mathrm{d}\theta \ll e^{(-1+\eta)t}$$

where $\hat{f}_L(\theta) = \sum_{v \in g\Lambda_{\mathrm{nz}} \cap L} f(\Delta(a_t r_\theta) v)$ *and*

$$\mathcal{C}_L = \{\theta \in [0, 2\pi] : \hat{f}_L(\theta) \geq e^{\eta t}\}.$$

The implied constants depend polynomially on R*.*

We postpone the proof of this lemma to the end of §6. Part (1) in this Lemma should be compared with [EMM98, Lemma 5.13]; indeed in loc. cit. the integral appearing as part (1) in Lemma 3.2 is bounded by $O(t)$ (vs. $e^{o(t)}$ that we give here) which is sharp. The above however suffices for our needs.

3.3 A linear algebra lemma. The goal in the remaining parts of this section is to relate the circular integrals as they appear in Theorem 3.1 to the counting problem in Theorem 1.2. This is the content of Lemma 3.4 which should be compared with [EMM98, Lemma 3.6] and [EM01, Lemma 3.4]. We will also establish a certain upper bound estimate in Lemma 3.9 which will be used in the proof of Theorem 1.2.

Let us begin by fixing some notation which will be used in Lemma 3.5 and Lemma 3.4. Let $\alpha < \beta$, $R \geq \max\{1, |\alpha|, |\beta|\}$, $R^{-1} \leq q \leq R$, and $0 < \varepsilon < R^{-4}$. Let $\varrho : \mathbb{R} \to [0, 1]$ be a smooth function supported on $[q - \varepsilon, q]$. Let f_1 be a smooth function on $\mathbb{R}^2$ satisfying

$$1_{[-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]}(\mathsf{x}) \cdot \varrho(\mathsf{y}) \leq f_1(\mathsf{x}, \mathsf{y}) \leq 1_{[\frac{-\varepsilon-\varepsilon^2}{2}, \frac{\varepsilon+\varepsilon^2}{2}]}(\mathsf{x}) \cdot \varrho(\mathsf{y}); \tag{3.3}$$

we chose ϱ and f_1 so that their partial derivatives are $\ll_R \varepsilon^{-10}$.

For an interval $I = [a, b]$ and $\delta > 0$, put

$$\begin{aligned} I_\delta &= [a - \delta, b + \delta] \supset I, \\ I_{-\delta} &= [a + \delta, b - \delta] \subset I. \end{aligned} \tag{3.4}$$

Given two intervals $I \subset [-R^2, R^2]$ and $I' \subset [0, R]$, let $f_{I,I'}$ be a smooth function with partial derivatives $\ll_R \varepsilon^{-10}$ satisfying

$$1_{I^{(1)}}(\mathsf{x}) \cdot 1_{I'^{(1)}}(|\mathsf{y}|) \leq f_{I,I'}(\mathsf{x}, \mathsf{y}) \leq 1_{I^{(2)}}(\mathsf{x}) \cdot 1_{I'^{(2)}}(|\mathsf{y}|), \tag{3.5}$$

where we write $I^{(k)} = I_{10kR^3\varepsilon}$ (in the formula above we used $k = 1, 2$, but later also large values of k will be used).

For any function h on $\mathbb{R}^2$, define

$$J_h(\mathsf{y}) = \int_{\mathbb{R}} h(\mathsf{x}, \mathsf{y})\, \mathrm{d}\mathsf{x}.$$

Note that if f_1 is as in (3.3), then

$$J_{f_1}(\mathsf{y}) = \varrho(\mathsf{y})(\varepsilon + O(\varepsilon^2)). \tag{3.6}$$

Let f_1 be as above (for this q and some ϱ) and let

$$f_2 = f_{I_0, I_1}$$

(for $I_0 = [-q^{-1}\beta, -q^{-1}\alpha]$ and some $I_1 \subset [0, R]$). Define f on $\mathbb{R}^4$ by

$$f(v_1, v_2) = f_1(v_1) f_2(v_2).$$

We will work with a slight variant of polar coordinates in $\mathbb{R}^2$: $0 \neq w \in \mathbb{R}^2$ is denoted by $(\theta_w, \|w\|)$ where $\theta_w \in [0, 2\pi]$ is so that $r_{\theta_w} w = (0, \|w\|)$.

3.4 Lemma. *Let the notation be as above. Let $t > \log(4R^3\varepsilon^{-2})$, and let ξ be a 2π-periodic non-negative smooth function. Let $v = (v_1, v_2) \in \mathbb{R}^4$ with $\|v_i\| \geq R^{-1}$. Then*

$$\begin{aligned} &qe^{2t} \int_0^{2\pi} f(\Delta(a_t r_\theta)v)\xi(\theta)\, \mathrm{d}\theta \\ &\leq \begin{cases} (1 + O(\varepsilon))J_{f_1}(e^{-t}\|v_1\|)\xi(\theta_1) + O(\mathrm{Lip}(f_1)\,\mathrm{Lip}(\xi)e^{-2t}), & \textit{if } (3.8) \textit{ holds}, \\ 0, & \textit{otherwise}, \end{cases} \end{aligned} \tag{3.7}$$

where

$$(-q^{-1}Q_0(v), e^{-t}\|v_2\|) \in I_0^{(3)} \times I_1^{(3)} \quad \textit{and} \quad \|v_1\| \leq 2Re^t. \tag{3.8}$$

If we moreover assume that $e^{-t}\|v_2\| \in I_1$ and $Q_0(v) \in [\alpha, \beta]$, then

$$\begin{aligned} &qe^{2t} \int_0^{2\pi} f(\Delta(a_t r_\theta)v)\xi(\theta)\, \mathrm{d}\theta \\ &\qquad = (1 + O(\varepsilon))J_{f_1}(e^{-t}\|v_1\|)\xi(\theta_1) f_2(-q^{-1}Q_0(v), e^{-t}\|v_2\|) \\ &\qquad\quad + O(\mathrm{Lip}(f_1)\,\mathrm{Lip}(\xi)e^{-2t}). \end{aligned} \tag{3.9}$$

The implied constants depend polynomially on R.

Analogous statements hold with the roles of v_1 and v_2 switched.

The proof is based on a direct computation which we will carry out in the next lemma.

3.5 Lemma. *Let the notation be as in Lemma 3.4. Let $t > \log(4R^3\varepsilon^{-2})$. If*

$$f(\Delta(a_t r_\theta)v) \neq 0$$

for some $v = (v_1, v_2) \in \mathbb{R}^4$ with $\|v_i\| \geq R^{-1}$ and some $\theta \in [0, 2\pi]$, then all of the following properties hold

(1) $q(1-2\varepsilon) \leq e^{-t}\|v_1\| \leq q(1+\varepsilon)$,
(2) $|\theta - \theta_{v_1}| \leq 2R\varepsilon e^{-2t}$,
(3) $e^{-t}\|v_2\| \in I_1^{(2)}$, *and*
(4) $-q^{-1}Q_0(v) \in I_0^{(3)}$.

Proof. The definitions of f_1 and f_2 imply that

$$\text{if} \quad \|v_i\| > (R + 20R^3\varepsilon)e^t, \quad \text{then} \quad f(\Delta(a_t r_\theta)v) = 0$$

and there is nothing to prove. We thus assume that $\|v_i\| \leq (R + 20R^3\varepsilon)e^t$ for the rest of the argument.

For convenience, we will write $\theta_1 = \theta_{v_1}$. Since $\theta \in [0, 2\pi]$ satisfies

$$a_t r_\theta v_1 \in \left[\frac{-\varepsilon - \varepsilon^2}{2}, \frac{\varepsilon + \varepsilon^2}{2}\right] \times [q - \varepsilon, q]$$

only if

$$|\theta - \theta_1| \leq \frac{3}{2}\varepsilon e^{-t}\|v_1\|^{-1} \leq 2R\varepsilon e^{-2t}, \tag{3.10}$$

we see that when

$$q(1-2\varepsilon) \leq e^{-t}\|v_1\| \leq q(1+\varepsilon) \tag{3.11}$$

fails, $f(\Delta(a_t r_\theta)v) = 0$.

Thus, assume that (3.10) and (3.11) hold for the rest of the argument, which is to say the conditions (1) and (2) in the lemma are satisfied if $f(\Delta(a_t r_\theta)v) \neq 0$. We now show (3) and (4) must also hold.

Let us write

$$r_{\theta_1} v_2 = (\overline{\mathsf{x}}_2, \overline{\mathsf{y}}_2).$$

Recall that $\|v_i\| \leq (R + 20R^3\varepsilon)e^t$ and that θ is in the range (3.10), and write

$$r_\theta v_1 = (\mathsf{x}'_1, \mathsf{y}'_1) \quad \text{and} \quad r_\theta v_2 = (\mathsf{x}'_2, \mathsf{y}'_2).$$

Then $|\mathsf{x}_1'| \leq 4R\varepsilon e^{-t}$, $|\mathsf{y}_1' - \|v_1\|| \leq 4R\varepsilon e^{-t}$,

$$|\mathsf{x}_2' - \overline{\mathsf{x}}_2|, |\mathsf{y}_2' - \overline{\mathsf{y}}_2| \leq 3R\varepsilon e^{-t}\|v_1\|^{-1}\|v_2\| \leq 4R^3\varepsilon e^{-t}; \tag{3.12}$$

in the last inequality we used $\varepsilon < R^{-4}$, $\|v_2\| \leq (R + 20R^3\varepsilon)e^t$, and (3.11).

Thus, we conclude that

$$a_t r_\theta v_2 = (e^t\mathsf{x}_2', e^{-t}\mathsf{y}_2') = (e^t\overline{\mathsf{x}}_2 + \mathsf{x}_{2,\theta}, e^{-t}\overline{\mathsf{y}}_2 + \mathsf{y}_{2,\theta})$$

where $|\mathsf{x}_{2,\theta}| \leq 4R^3\varepsilon$ and $|\mathsf{y}_{2,\theta}| \leq 4R^3\varepsilon e^{-2t}$.

In view of the definition of f_2, we conclude that $f_2(a_t r_\theta v_2) = 0$, unless

$$e^t\mathsf{x}_2' \in (I_0^{(1)})_{20R^3\varepsilon} \quad \text{and} \quad e^{-t}\mathsf{y}_2' \in (I_1^{(1)})20R^3\varepsilon + \varepsilon$$

These and the bound on $\mathsf{x}_{2,\theta}$ imply that

$$e^t\overline{\mathsf{x}}_2 \in (I_0^{(1)})_{24R^3\varepsilon} \tag{3.13}$$

and hence using the upper bound on $|\overline{\mathsf{x}}_2|$ implied by (3.13), we get

$$||\overline{\mathsf{y}}_2| - \|v_2\|| \leq \frac{R^4e^{-2t}}{\|v_2\|}. \tag{3.14}$$

Since $e^{-t}\mathsf{y}_2' \in (I_1^{(1)})_{20R^3\varepsilon+\varepsilon}$ and $|\mathsf{y}_{2,\theta}| \leq 4R^3\varepsilon e^{-2t}$, we conclude from (3.14) that if $f_2(a_t r_\theta v_2) \neq 0$, then

$$e^{-t}\|v_2\| \in (I_1^{(1)})_{21R^3\varepsilon}$$

which establishes (3) in the lemma.

Finally, combining (3.13) and (3.11), we conclude that

$$q^{-1}\|v_1\|\overline{\mathsf{x}}_2 \in (I_0^{(1)})_{30R^3\varepsilon}.$$

Since $\Delta(r_\theta) \in \mathrm{SO}(Q_0)$ for all θ and $\Delta(r_{\theta_1})v = (0, \|v_1\|, \overline{\mathsf{x}}_2, \overline{\mathsf{y}}_2)$, we get

$$-q^{-1}Q_0(v) = -q^{-1}Q_0(\Delta(r_{\theta_1}v)) = q^{-1}\|v_1\|\overline{\mathsf{x}}_2 \in I_0^{(2)},$$

as it was claimed in (4). □

We now turn to the proof of Lemma 3.4.

Proof of Lemma 3.4. For convenience we write $\theta_1 = \theta_{v_1}$. By Lemma 3.5 if $f(\Delta(a_t r_\theta)v) \neq 0$, then all the following hold true:

$$q(1 - 2\varepsilon) \leq e^{-t}\|v_1\| \leq q(1 + \varepsilon), \tag{3.15a}$$

$$|\theta - \theta_1| \leq 2R\varepsilon e^{-2t}, \tag{3.15b}$$

$$e^{-t}\|v_2\| \in I_1^{(3)}, \tag{3.15c}$$

$$-q^{-1}Q_0(v) \in I_0^{(3)}. \tag{3.15d}$$

We begin with the following computation which will be used in the proof of both (3.7) and (3.9):

$$\int_0^{2\pi} f_1(a_t r_\theta v_1)\,\mathrm{d}\theta = \int_0^{2\pi} f_1(-e^t\|v_1\|\sin\theta, e^{-t}\|v_1\|\cos\theta)\,\mathrm{d}\theta.$$

Performing the change of variable $\mathsf{z} = -e^t\|v_1\|\sin\theta$, the above integral equals

$$\begin{aligned} &\frac{e^{-t}}{\|v_1\|}\int_{-\infty}^{\infty} f_1\Big(\mathsf{z}, e^{-t}\|v_1\|\sqrt{1-(e^{-t}\mathsf{z}/\|v_1\|)^2}\Big)\frac{1}{\sqrt{1-(e^{-t}\mathsf{z}/\|v_1\|)^2}}\,\mathrm{d}\mathsf{z} \\ &\quad= \frac{e^{-t}}{\|v_1\|}\int_{-\infty}^{\infty} f_1(\mathsf{z}, e^{-t}\|v_1\|)\,\mathrm{d}\mathsf{z} + O(R^2\operatorname{Lip}(f_1)e^{-4t}) \\ &\quad= q^{-1}(1+O(\varepsilon))e^{-2t}J_{f_1}(e^{-t}\|v_1\|) + O(R^2\operatorname{Lip}(f_1)e^{-4t}) \end{aligned} \tag{3.16}$$

where in the last equality we used (3.15a) and (3.6).

Let us now begin the proof of (3.7). We can restrict the integration in (3.7) to θ satisfying (3.15b). In this range

$$|\xi(\theta) - \xi(\theta_1)| \le 2R\varepsilon e^{-2t}\operatorname{Lip}(\xi). \tag{3.17}$$

Since $0 \le f_1, f_2 \le 1$ and ξ is non-negative, we have

$$\int_0^{2\pi} f(\Delta(a_t r_\theta)v)\xi(\theta)\,\mathrm{d}\theta \le \int_0^{2\pi} f_1(a_t r_\theta)\xi(\theta)\,\mathrm{d}\theta. \tag{3.18}$$

Moreover, in view of (3.17), we have

$$f_1(a_t r_\theta)\xi(\theta) = f_1(a_t r_\theta v_1)\xi(\theta_1) + O(R^2\operatorname{Lip}(\xi)\varepsilon e^{-2t}).$$

This, (3.18), and the fact that the range of integration is (3.15b) implies

$$e^{2t}\int_0^{2\pi} f(\Delta(a_t r_\theta)v)\xi(\theta)\,\mathrm{d}\theta \le \xi(\theta_1)e^{2t}\int_0^{2\pi} f_1(a_t r_\theta v_1) + O(R^2\operatorname{Lip}(\xi)\varepsilon e^{-2t}).$$

This and (3.16) imply that

$$\begin{aligned} &e^{2t}\int_0^{2\pi} f(\Delta(a_t r_\theta)v)\xi(\theta)\,\mathrm{d}\theta \\ &\qquad\le q^{-1}(1+O(\varepsilon))J_{f_1}(e^{-t}\|v_1\|)\xi(\theta_1) + O(R^2\operatorname{Lip}(f_1)\operatorname{Lip}(\xi)\varepsilon e^{-2t}). \end{aligned} \tag{3.19}$$

Thus (3.7) follows from (3.19) in view of (3.15c) and (3.15d).

Note that the claim regarding $\mathcal{E}$ follows as well, indeed if either (3.15a), (3.15c) or (3.15d) fails, both the left and right side of (3.7) equal zero.

The proof of (3.9) is similar. Indeed one argues as in the proof of Lemma 3.5 to show that if $e^{-t}\|v_2\| \in I_1$ and $Q_0(v) \in [\alpha, \beta]$, then for all θ in the range (3.15b), one has

$$f_2(a_t r_\theta v_2) = 1.$$

One then repeats the above argument and obtains (3.9). □

3.6 A smooth cell decomposition. Let

$$\Omega = \{(w_1 + w_2, \omega(w_1 - w_2)) : \|w_k\| \le 1\},$$
$$\mathsf{D} = \{(v_1, v_2) : \|v_k\| \le 1\}.$$

As before, write $v = (v_1, v_2) \in \mathbb{R}^4$ where $v_k \in \mathbb{R}^2$. Let $\pi_1(v) = (v_1, 0)$ and $\pi_2(v) = (0, v_2)$; abusing the notation, we also consider $\pi_k(\Omega) \subset \mathbb{R}^2$.

Write $\Omega \setminus \mathsf{D} = \Omega_1 \cup \Omega_2$ where

$$\Omega_1 := \{(v_1, v_2) \in \Omega : \|v_1\| > 1\} \quad \text{and}$$
$$\Omega_2 := \{(v_1, v_2) \in \Omega : \|v_1\| \le 1, \|v_2\| > 1\}.$$

A direct computation shows that $(v_1, v_2) \in \Omega$ if and only if

$$\|v_2\|^2 \le 4 - \|v_1\|^2 - 2|Q_0(v_1, v_2)|.$$

It follows that for every $v_1 \in \pi_1(\Omega_1)$, we have

$$\{\|\lambda v_1\| : (v_1, \lambda v_1) \in \Omega_1\} = [0, \sqrt{4 - \|v_1\|^2}], \tag{3.20}$$

and for $v_2 \in \pi_2(\Omega_2)$, we have

$$\{\|\lambda v_2\| : (\lambda v_2, v_2) \in \Omega_2\} = [0, \min(1, \sqrt{4 - \|v_2\|^2})].$$

Fix some $R \ge 10^3$ and let $0 < \varepsilon < R^{-20}$. Let $E \in \mathbb{N}$ be so that

$$\frac{1}{E} \le 100R^{10}\varepsilon \le \frac{1}{E-1},$$

and put

$$I_i = \Big[\frac{i-1}{E}, \frac{i}{E}\Big] \quad \text{for all } 1 \le i \le E.$$

Fix two families of smooth functions $\{\xi_i^-\}$ and $\{\xi_i^+\}$ with C^1 norm $\ll \varepsilon^{-10}$ satisfying the following:

(ξ-1) For all i, $0 \le \xi_i^- \le \xi_i^+ \le 1$,

$$\xi_i^+ \equiv 1 \text{ on } 2\pi I_i, \qquad \operatorname{supp}(\xi_i^+) \subset 2\pi(I_i)_{\varepsilon^2},$$
$$\xi_i^- \equiv 1 \text{ on } 2\pi(I_i)_{-4\varepsilon^2}, \quad \operatorname{supp}(\xi_i^-) \subset 2\pi(I_i)_{-2\varepsilon^2}$$

(here we use the notation (3.4)). We extend $\xi_i^\pm$ to 2π-periodic functions on $\mathbb{R}$.

Similarly, let $E' \in \mathbb{N}$ be so that $\frac{1}{E'} \le 100R^9\varepsilon \le \frac{1}{E'-1}$, and let

$$I'_j = \Big[\frac{j-1}{E'}, \frac{j}{E'}\Big] \quad \text{for all } 1 \le j \le E'.$$

Fix two families of functions $\{\varrho_j^+\}$ and $\{\varrho_j^-\}$ with C^1 norm $\ll \varepsilon^{-10}$ so that

(ϱ-1) For all i, $0 \le \varrho_j^- \le \varrho_j^+ \le 1$,

$$\begin{aligned} &\varrho_j^+ \equiv 1 \text{ on } RI_j, && \operatorname{supp}(\varrho_j^+) \subset R(I_i)_{\varepsilon^2}, \\ &\varrho_j^- \equiv 1 \text{ on } R(I_j)_{-4\varepsilon^2}, && \operatorname{supp}(\varrho_j^-) \subset R(I_i)_{-2\varepsilon^2}. \end{aligned}$$

Extend $\varrho_j^\pm$ to $\mathbb{R}$ by defining them to equal 0 outside their supports.

Define

$$\varphi_{i,j}^+(\theta, r) = \xi_i^+(\theta)\varrho_j^+(r) \quad \text{and} \quad \varphi_{i,j}^-(\theta, r) = \xi_i^-(\theta)\varrho_j^-(r).$$

We will consider $\varphi_{i,j}^\pm$ as functions on $\mathbb{R}^2$ using our slightly non-standard polar coordinate system where any $0 \ne w \in \mathbb{R}^2$ corresponds to $(\theta_w, \|w\|)$ if $r_{\theta_w} w = (0, \|w\|)$. Let

$$\begin{aligned} \mathcal{I}_1^+ &= \{(i,j) : \operatorname{supp}(\varphi_{i,j}^+) \cap \pi_1(\boldsymbol{\Omega}_1) \ne \emptyset\}, \\ \mathcal{I}_1^- &= \{(i,j) : \operatorname{supp}(\varphi_{i,j}^-) \subset \pi_1(\boldsymbol{\Omega}_1)\}. \end{aligned} \tag{3.21}$$

We define $\mathcal{I}_2^\pm$ similarly with $\boldsymbol{\Omega}_2$ and π_2 in lieu of $\boldsymbol{\Omega}_1$ and π_1. Note that for $k = 1, 2$ and $\sigma = \pm$

$$\left| \operatorname{area}(\pi_k(\boldsymbol{\Omega}_k)) - \sum_{(i,j)\in\mathcal{I}_k^\sigma} \int \varphi_{i,j}^\sigma \right| \ll \varepsilon.$$

We will work with $k = 1$ for the remainder of this section; similar analysis applies to $k = 2$ with the role of v_1 and v_2 switched. For all $(i, j) \in \mathcal{I}_1^+$, let

$$\boldsymbol{\Omega}_{i,j}^+ = \{(v_1, v_2 + w) : (v_1, v_2) \in \boldsymbol{\Omega}_1, \varphi_{i,j}^+(v_1) = 1, \|w\| \le 3R\varepsilon\}.$$

We will also define $\boldsymbol{\Omega}_{i,j} \subset \boldsymbol{\Omega}_{i,j}^+$ as follows. In view of (3.20), we will call the pair (i, j) **typical** if

$$\inf\{\sqrt{4 - \|v_1\|^2} : v_1 \in \operatorname{supp}(\varphi_{i,j}^+) \cap \pi_1(\boldsymbol{\Omega}_1)\} \ge \sqrt{\varepsilon}.$$

Let $\mathring{\mathcal{I}}_1^-$ denote the set of $(i, j) \in \mathcal{I}_1^-$ where (i, j) is typical and for every

$$(v_1, \lambda v_1) \in \boldsymbol{\Omega}_1 \cap (\operatorname{supp}(\varphi_{i,j}^-) \times \mathbb{R}^2) \quad \text{with } \|\lambda v_1\| \in ([0, \sqrt{4 - \|v_1\|^2}])_{-20R\varepsilon}$$

we have $(v_1, \lambda v_1 + w) \in \boldsymbol{\Omega}_1$ for all $w \in \mathbb{R}^2$ with $\|w\| \le 10R\varepsilon$.

For any $(i, j) \in \mathring{\mathcal{I}}_1^-$, set

$$\begin{aligned} \boldsymbol{\Omega}_{i,j} := \{(v_1, v_2 + w) : (v_1, v_2) &\in \boldsymbol{\Omega}_1 \cap (\operatorname{supp}(\varphi_{i,j}^-) \times \mathbb{R}^2), \\ &w \in \mathbb{R}^2, \|w\| \le \varepsilon\} \cap \boldsymbol{\Omega}_1. \end{aligned} \tag{3.22}$$

Since $\operatorname{supp}(\varphi_{i,j}^-) \subset \{w : \varphi_{i,j}^+(w) = 1\}$, we have $\boldsymbol{\Omega}_{i,j} \subset \boldsymbol{\Omega}_{i,j}^+$. Moreover, since $\{\operatorname{supp}(\varphi_{i,j}^-)\}$ is a disjoint collection, $\{\boldsymbol{\Omega}_{i,j}\}$ is a disjoint collection.

In view of (ξ-1), (ϱ-1), and the above definitions,

$$1_{\Omega_1} \leq \sum_{\mathcal{I}_1^+} 1_{\Omega_{i,j}^+} \leq 4 \cdot 1_{\{(v_1,v_2):\|v_k\|\leq 3\}}, \tag{3.23a}$$

$$\sum_{\mathring{\mathcal{I}}_1^-} 1_{\Omega_{i,j}} \leq 1_{\Omega_1}. \tag{3.23b}$$

The intervals $I_{i,j}^+$ and $I_{i,j}^-$. In our application of Lemma 3.4, $\xi_i^\pm$ will play the role of ξ; we will also work with $f = f_1 f_2$ where f_1 is defined using $\varrho_j^\pm$ above and f_2 is defined using $I_0 = [-q^{-1}\beta, -q^{-1}\alpha]$ (for some $R^{-1} \leq q \leq R$) and intervals $I_{i,j}^\pm$ which we now define. Put

$$\begin{aligned} I'_{i,j,+} &= [0, b_{i,j}^+], \quad b_{i,j}^+ = \sup\{\sqrt{4-\|v_1\|^2} : v_1 \in \operatorname{supp}(\varphi_{i,j}^+) \cap \pi_1(\Omega_1)\}, \\ I'_{i,j,-} &= [0, b_{i,j}^-], \quad b_{i,j}^- = \inf\{\sqrt{4-\|v_1\|^2} : v_1 \in \operatorname{supp}(\varphi_{i,j}^+) \cap \pi_1(\Omega_1)\}. \end{aligned} \tag{3.24}$$

If (i,j) is typical, i.e., if $b_{i,j}^- \geq \sqrt{\varepsilon}$, put

$$I_{i,j}^+ = (I'_{i,j,+})_{10\varepsilon} \quad \text{and} \quad I_{i,j}^- = (I'_{i,j,-})_{-200R^{10}\varepsilon}. \tag{3.25}$$

Since $\operatorname{supp}(\varphi_{i,j}^\pm)$ has diameter $\leq 200R^{10}\varepsilon$ and $\varepsilon < R^{-20}$, if (i,j) is not typical, then $b_{i,j}^+ \leq 2\sqrt{\varepsilon}$. In this case, put $I_{i,j}^\pm = [0, 3\sqrt{\varepsilon}]$.

We have the following lemma.

3.7 Lemma. *Assume $R \geq \max\{10^3, |\alpha|, |\beta|\}$ and let $R^{-1} \leq q \leq R$. Let $t \geq \log(R^2\varepsilon^{-1})$, where as before $0 < \varepsilon < R^{-20}$.*

(1) *Let $I_0 = [-q^{-1}\beta, -q^{-1}\alpha]$. Let $(i,j) \in \mathring{\mathcal{I}}_1^-$ and let f_1 satisfy (3.3) with ϱ_j^- (and with $\varepsilon' = 200R^{10}\varepsilon$ instead of ε). If*

$$J_{f_1}(e^{-t}\|v_1\|)\xi_i^-(\theta_{v_1})1_{I_0^{(3)}}(-q^{-1}Q_0(v))1_{(I_{i,j}^-)^{(3)}}(e^{-t}\|v_2\|) \neq 0$$

for some $v = (v_1, v_2) \in \mathbb{R}^4$, then all the following hold

(a) *$Q_0(v) \in ([\alpha,\beta])_{30R^4\varepsilon}$, and*

(b) *$e^{-t}v_1 \in \operatorname{supp}(\varphi_{i,j}^-)$, and*

(c) *$e^{-t}v \in \Omega_{i,j}$.*

(2) *Let $(i,j) \in \mathcal{I}_1^+$. If $v = (v_1, v_2) \in e^t\Omega_{i,j}^+$ satisfies $Q_0(v) \in [\alpha,\beta]$, then*

$$e^{-t}\|v_2\| \in I_{i,j}^+.$$

Proof. We first prove part (1). If $Q_0(v) \notin ([\alpha,\beta])_{30R^4\varepsilon}$, then

$$-q^{-1}Q_0(v) \notin (I_0)_{30R^3\varepsilon} = I_0^{(3)},$$

hence

$$1_{I_0^{(3)}}(-q^{-1}Q_0(v)) = 0.$$

Moreover, if we put $\overline{v_1} := e^{-t}v_1$, then $\theta_{\overline{v_1}} = \theta_{v_1}$, and $\overline{v_1} \notin \mathrm{supp}(\varphi^-_{i,j})$ would imply that $\varrho^-_j(e^{-t}\|v_1\|)\xi^-_i(\theta_{v_1}) = 0$. This in turn yields

$$0 \le f_1(\mathsf{x}, e^{-t}\|v_1\|)\xi^-_i(\theta_{v_1}) \le \varrho^-_j(e^{-t}\|v_1\|)\xi^-_i(\theta_{v_1}) = 0,$$

see (3.3); thus, $J_{f_1}(e^{-t}\|v_1\|)\xi^-_i(\theta_{v_1}) = 0$. In conclusion, we may assume that

$$J_{f_1}(e^{-t}\|v_1\|)\xi^-_i(\theta_{v_1})1_{I_0^{(3)}}(-q^{-1}Q_0(v))1_{(I^-_{i,j})^{(3)}}(e^{-t}\|v_2\|) \ne 0, \tag{3.26}$$

and that

$$Q_0(v) \in ([\alpha, \beta])_{30R^4\varepsilon} \quad \text{and} \quad \overline{v_1} \in \mathrm{supp}(\varphi^-_{i,j}).$$

We need to show that (c) is also satisfied.

Since $Q_0(v) \in ([\alpha, \beta])_{30R^4\varepsilon}$, where $R \ge \max\{10^3, |\alpha|, |\beta|\}$ and $\varepsilon < R^{-20}$, and $\|v_1\| \ge e^t$, there is $\lambda \in \mathbb{R}$ so that

$$v_2 = \lambda v_1 + w, \quad \text{where } w \perp v_1 \text{ and } \|w\| \le 2R\|v_1\|^{-1} \le 2Re^{-t}. \tag{3.27}$$

Thus $e^{-t}v_2 = \lambda e^{-t}v_1 + e^{-t}w = \lambda\overline{v_1} + e^{-t}w$.

Moreover, by (3.26), we have $e^{-t}\|v_2\| \in (I^-_{i,j})^{(3)} = (I^-_{i,j})_{30R^3\varepsilon}$, where

$$I^-_{i,j} = (I'_{i,j,-})_{-200R^{10}\varepsilon} \quad \text{and} \quad I'_{i,j,-} \subset [0, \sqrt{4 - \|\overline{v_1}\|^2}],$$

see (3.24) and (3.25). Since $\|e^{-t}w\| \le 2Re^{-2t}$, we conclude that

$$\|\lambda\overline{v_1}\| \in ([0, \sqrt{4 - \|\overline{v_1}\|^2}])_{-20R\varepsilon}.$$

In particular,

$$(\overline{v_1}, \lambda\overline{v_1}) \in \Omega_1 \cap (\mathrm{supp}(\varphi^-_{i,j}) \times \mathbb{R}^2),$$

and $v = e^t(\overline{v_1}, \lambda\overline{v_1} + e^{-t}w)$ where $\|e^{-t}w\| \le 2Re^{-2t}$. By the definition of $\mathring{\mathfrak{I}}^-_1$ and $\Omega_{i,j}$, we conclude that $e^{-t}v \in \Omega_{i,j}$. Thus, (c) also holds.

The proof of (2) is similar to the proof of (c), see in particular (3.27). $\square$

3.8 Upper bound estimates. Before starting the proof of Theorem 1.2, we record a weaker (but more explicit) version of [EMM98, Thm. 2.3], which will be used in the sequel—see also the very recent work of Kelmer, Kontorovich, and Lutsko [KKL23].

For every $R > 0$, let

$$\mathsf{D}(R) = \{(v_1, v_2) : \|v_k\| \le R\}.$$

Then $\mathsf{D}(R) \setminus \mathsf{D}(e^{-1}R) = \mathsf{D}(R)_1 \cup \mathsf{D}(R)_2$, where

$$\mathsf{D}_1(R) = \{(v_1, v_2) \in \mathsf{D}(R) : e^{-1}R < \|v_1\| \le R\} \quad \text{and}$$
$$\mathsf{D}_2(R) = \{(v_1, v_2) \in \mathsf{D}(R) : \|v_1\| \le e^{-1}R, e^{-1}R < \|v_2\| \le R\}.$$

We constructed smooth cell decomposition for $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ in §3.6; in the following lemma we will use a similar construction (without repeating this construction) for $\mathsf{D}_1(R)$ and $\mathsf{D}_2(R)$.

3.9 Lemma. *Let $g = (g_1, g_2) \in G$ and put $\Lambda' = g\Lambda$. Let*

$$R \ge \max\{10^3, |\alpha|, |\beta|, \|g_1\|^{\pm 1}, \|g_2\|^{\pm 1}\},$$

and let $0 < \eta < 1$. There exists $t_0 \ll |\log \eta|$ so that if $t \ge t_0$, then

$$\#\{v = (v_1, v_2) \in \Lambda' : \max(\|v_1\|, \|v_2\|) \le Re^t, \alpha \le Q_0(v) \le \beta\} \ll e^{(2+\eta)t}$$

where the implied constant depends polynomially on R.

Proof. The following basic lattice point estimate will be used:

$$\#\{v \in \Lambda' \cap e^{t/2}\mathsf{D}(R)\} \ll e^{2t} \tag{3.28}$$

where the implied constant depends polynomially on R.

Since R is fixed, we will denote $\mathsf{D}_k(R)$ by D_k ($k = 1, 2$) for the rest of the proof. Let $\varepsilon = 10^{-6}R^{-20}$. Apply the construction in §3.6 for $\pi_1(\mathsf{D}_1)$ with this R and ε. In particular, the functions ξ_i^+ are defined as in (ξ-1) with

$$I_i = \left[\frac{i-1}{E}, \frac{i}{E}\right] \quad \text{for all } 1 \le i \le E \text{ where } \frac{1}{E} \le 100R^{10}\varepsilon \le \frac{1}{E-1},$$

and ϱ_j^+ are defined as in (ϱ-1) with

$$I'_j = \left[\frac{j-1}{E'}, \frac{j}{E'}\right] \quad \text{for all } 1 \le j \le E' \text{ where } \frac{1}{E'} \le 100R^{9}\varepsilon \le \frac{1}{E'-1}.$$

For all i, j as above, let $\xi_i = \xi_i^+$, $\varrho_j = \varrho_j^+$, and let $\varphi_{i,j} = \xi_i \varrho_j$. Put

$$\mathcal{I}_1^+ = \{(i, j) : \operatorname{supp}(\varphi_{i,j}) \cap \pi_1(\mathsf{D}_1) \ne \emptyset\};$$

for all $(i, j) \in \mathcal{I}_1^+$, we have $\operatorname{supp}(\varrho_j) \subset [e^{-2}R, R] \subset [R^{-1}, R]$.

For all $(i, j) \in \mathcal{I}_1^+$, put

$$\hat{\mathsf{D}}_{i,j} = \{(v_1, v_2) \in \mathbb{R}^4 : \varphi_{i,j}(v_1) = 1, \|v_2\| \le R\}.$$

Then $1_{\mathsf{D}_1} \le \sum_{\mathcal{I}_1^+} 1_{\hat{\mathsf{D}}_{i,j}} \le 4_{\mathsf{D}(2R)_1}$.

Define f_1 as in (3.3) for q and ϱ_j, and with $200R^{10}\varepsilon$ instead of ε. Let

$$f_2 = f_{[-q^{-1}\beta, -q^{-1}\alpha],[0,R]}, \tag{3.29}$$

see (3.5). Put $f_{i,j} = f_1 f_2$. By the choice of R, we have $\sum f_{i,j} \le 4_{\mathsf{D}(2R)}$.

By Lemma 3.4, for any $v = (v_1, v_2) \in e^t\hat{\mathsf{D}}_{i,j}$ with $Q_0(v) \in [\alpha, \beta]$, we have

$$\begin{aligned} &e^{2t}\int_0^{2\pi} f_{i,j}(\Delta(a_t r_\theta)v)\xi_i(\theta)\,\mathrm{d}\theta \\ &\qquad = q^{-1}(1+O(\varepsilon))J_{f_1}(e^{-t}\|v_1\|)\xi_i(\theta_{v_1})f_2(-q^{-1}Q_0(v), e^{-t}\|v_2\|) \\ &\qquad\quad + O(\mathrm{Lip}(f_1)\,\mathrm{Lip}(\xi_i)e^{-2t}) \end{aligned} \tag{3.30}$$

where the implied constant depends on R.

First note that, if t is large enough compared to R, we have

$$O(\mathrm{Lip}(f_1)\,\mathrm{Lip}(\xi_i)e^{-2t}) \ll \varepsilon^{-20}e^{-2t} \le \varepsilon^2. \tag{3.31}$$

Furthermore, for any $v = (v_1, v_2) \in e^t\hat{\mathsf{D}}_{i,j}$ so that $Q_0(v) \in [\alpha, \beta]$, we have $f_2(q^{-1}Q_0(v), e^{-t}\|v_2\|) = 1$. Thus, using (3.6), we have

$$J_{f_1}(e^{-t}\|v_1\|)\xi_i(\theta_{v_1})f_2(q^{-1}Q_0(v), e^{-t}\|v_2\|) = \varepsilon + O(\varepsilon^2). \tag{3.32}$$

Put $x = g\Gamma'$. Summing (3.30) over all $v \in \Lambda' \cap e^t\hat{\mathsf{D}}_{i,j}$ so that $Q_0(v) \in [\alpha, \beta]$ and using (3.30) and (3.32), we conclude that

$$\varepsilon(\#\{v \in \Lambda' \cap e^t\hat{\mathsf{D}}_{i,j} : \alpha \le Q_0(v) \le \beta\}) \ll q e^{2t}\int_0^{2\pi} \hat{f}_{i,j}(\Delta(a_t r_\theta)x)\,\mathrm{d}\theta, \tag{3.33}$$

where we used $0 \le \xi_i \le 1$ and replaced $\varepsilon^2 + \varepsilon + O(\varepsilon^2)$ obtained from adding (3.31) and (3.32) by $O(\varepsilon)$.

Summing (3.33) over all $(i, j) \in \mathcal{I}_1^+$ and using $\sum_{i,j} f_{i,j} \le 4_{\mathsf{D}(2R)}$, we get

$$\#\{v \in \Lambda' \cap e^t\mathsf{D}_1 : \alpha \le Q_0(v) \le \beta\} \ll \varepsilon^{-1}q e^{2t}\int_0^{2\pi} \hat{1}_{\mathsf{D}(2R)}(\Delta(a_t r_\theta)x)\,\mathrm{d}\theta.$$

One obtains a similar bound for the number $v \in \Lambda' \cap e^t\mathsf{D}_2$ with $Q_0(v) \in [\alpha, \beta]$. Since $\mathsf{D} \setminus e^{-1}\mathsf{D} = \mathsf{D}_1 \cup \mathsf{D}_2$ and $\varepsilon = 10^{-6}R^{-20}$, we conclude that

$$\begin{aligned} &\#\{v \in \Lambda' \cap e^t(\mathsf{D} \setminus e^{-1}\mathsf{D}) : \alpha \le Q_0(v) \le \beta\} \\ &\qquad\qquad \ll \varepsilon^{-1}q e^{2t}\int_0^{2\pi} \hat{1}_{\mathsf{D}(2R)}(\Delta(a_t r_\theta)x)\,\mathrm{d}\theta. \end{aligned}$$

Let t_η be as in Lemma 3.2 applied with η and $2R$, and let $t > 10t_\eta$. Then by Lemma 3.2,

$$\#\{v \in \Lambda' \cap e^t(\mathsf{D} \setminus e^{-1}\mathsf{D}) : \alpha \le Q_0(v) \le \beta\} \ll e^{(2+\eta)t}.$$

We may repeat the above with $t-\ell$ for all $0 \le \ell \le t/2$, and obtain

$$\#\{v \in \Lambda' \cap e^{t-\ell}(\mathsf{D} \setminus e^{-1}\mathsf{D}) : \alpha \le Q_0(v) \le \beta\} \ll e^{(2+\eta)(t-\ell)}; \tag{3.34}$$

we also used $t-\ell \ge t/2 \ge t_\eta$ when applying Lemma 3.2 with $t-\ell$.

Since $e^t(e^{-\ell}\mathsf{D}) = e^{t-\ell}\mathsf{D}$, summing (3.34) over $0 \le \ell \le t/2$, we conclude that

$$\#\{v \in \Lambda' \cap e^{t}(\mathsf{D} \setminus e^{-t/2}\mathsf{D}) : \alpha \le Q_0(v) \le \beta\} \ll e^{(2+\eta)t}. \tag{3.35}$$

The lemma follows from (3.35) and (3.28). □

4 Proof of Theorem 1.2

The proof relies on Theorem 3.1 and will be completed in some steps. Recall that $\mathsf{M} = \mathbb{R}^2/\Delta$ and that Δ^* denotes the dual lattice. In view of our normalization, $2\pi\Delta^* = g_{\mathsf{M}}\mathbb{Z}^2$ where $g_{\mathsf{M}} \in \mathrm{SL}_2(\mathbb{R})$. Let

$$g = (g_{\mathsf{M}}, -\omega g_{\mathsf{M}}\omega) = (g_1, g_2) \in G \quad \text{where } \omega = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{4.1}$$

4.1 Passage to Q_0. As it was observed in (2.1), if $\lambda_i = \|v_i\|^2$, where for $i = 1, 2$, $v_i \in 2\pi\Delta^*$ is an eigenvalue of the Laplacian of M, then

$$\lambda_1 - \lambda_2 = Q_0(v_1 + v_2, \omega(v_1 - v_2)). \tag{4.2}$$

Define $\Omega = \{(v_1 + v_2, \omega(v_1 - v_2)) : \|v_i\| \le 1\}$; and let

$$\Lambda' = \{(v_1 + v_2, \omega(v_1 - v_2)) : v_1, v_2 \in 2\pi\Delta^*\} = g\Lambda$$

where $\Lambda = \{(v_1 + v_2, \omega(v_1 - v_2)) : v_1, v_2 \in \mathbb{Z}^2\}$.

Let T be a (large) parameter, and put $t = \frac{1}{2}\log T$. In view of (4.2),

$$R_{\mathsf{M}}(\alpha, \beta, T) = \#\{v \in \Lambda'_{\mathrm{nz}} \cap e^t\Omega : \alpha \le Q_0(v) \le \beta\}; \tag{4.3}$$

recall that $\Lambda'_{\mathrm{nz}} = \{(w_1, w_2) \in \Lambda' : w_i \ne 0\}$.

Let A and δ be as in Theorem 1.2. Without loss of generality, we assume $A \ge 10^5$ and $0 < \delta < 10^{-5}$. Let $\hat{A}$ be given by Theorem 3.1 applied with $10^3 A$. We will show the claim in Theorem 1.2 holds with $A' = 10\hat{A}$. To simplify the notation, write $\bar{A} = 10^3 A$ for the rest of the proof.

Thus let us assume (1.4) holds for A': for $T \ge T_0$ (T_0 is a yet to be determined large constant) and all $(p_1, p_2, q) \in \mathbb{Z}^3$ with $T^{\delta/A'} < q < T^\delta$,

$$\left|\frac{\mathsf{b}}{\mathsf{a}} - \frac{p_1}{q}\right| + \left|\frac{\mathsf{c}}{\mathsf{a}} - \frac{p_2}{q}\right| > q^{-A}. \tag{4.4}$$

This implies that so long as $t = \frac{1}{2}\log T$ is large enough (depending on a, b, and c), we have

$$g_2^{-1}g_1 = -\omega g_{\mathsf{M}}^{-1}\omega g_{\mathsf{M}} = \begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix} \tag{4.5}$$

satisfies (3.2) with t, $\rho = \delta/10$, $\hat{A}$. That is: for every $Q \in \mathrm{Mat}_2(\mathbb{Z})$ with $e^{\rho t/\hat{A}} \leq \|Q\| \leq e^{\rho t}$ and all $\lambda \in \mathbb{R}$, we have

$$\|g_2^{-1}g_1 - \lambda Q\| > \|Q\|^{-A} = \|Q\|^{-\bar{A}/1000}. \tag{4.6}$$

4.2 Lemma. *There are at most two $g\mathbb{Z}^4$-rational two-dimensional subspaces L, L' so that if for some $2t/5 \leq s \leq t$, L_s is a $(\delta_1/\bar{A}, \delta_1, s)$-exceptional subspace, then $L_s = L$ or L'.*

Proof. Let $2t/5 \leq s \leq t$. Recall that a $(\delta_1/\bar{A}, \delta_1, s)$-exceptional subspace is spanned by two vectors $(g_1 w_1, 0), (0, g_2 w_2) \in g\mathbb{Z}^4$ satisfying

$$\begin{aligned} &0 < \|g_i w_i\| \leq e^{\delta_1 s/\bar{A}}, \quad \text{and} \\ &|Q_0(g_1 w_1, g_2 w_2)| \leq e^{-\delta_1 s}. \end{aligned} \tag{4.7}$$

We also note that

$$e^{\delta_1 s/\bar{A}} \leq e^{\delta_1 t/\bar{A}} \quad \text{and} \quad e^{-\delta_1 s} \leq e^{-2\delta_1 t/5}$$

for any $2t/5 \leq s \leq t$.

Assume now that there are three pairs (possibly corresponding to different values of $2t/5 \leq s \leq t$) so that (4.7) is satisfied. Then Lemma 2.4, applied with $\delta_1/\bar{A}$ and $2\bar{A}/5$, implies that there is $Q \in \mathrm{Mat}_2(\mathbb{Z})$ with $\|Q\| \leq e^{100\delta_1 t/\bar{A}}$ so that

$$\begin{aligned} \|g_2^{-1}g_1 - \lambda Q\| = \left\| \begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix} - \lambda Q \right\| &\leq e^{-(\frac{2\bar{A}}{5} - 100)(\delta_1/\bar{A})} \\ &\leq \max\{\|Q\|^{-\bar{A}/1000}, 100e^{-\rho\bar{A}t/(1000\hat{A})}\}. \end{aligned}$$

Since $\rho/\hat{A} \leq \delta_1/\bar{A} \leq \rho/100$, this contradicts the fact that $g_2^{-1}g_1$ satisfies (4.6) with t, ρ, $\hat{A}$—note that if $\|Q\| \leq e^{\rho t/\hat{A}}$, we may replace Q by an integral multiple nQ with $e^{\rho/\hat{A}} \leq \|nQ\| \leq 2e^{\rho t/\hat{A}}$. The proof is complete. $\square$

Let L and L' be as in Lemma 4.2. For a set $\mathsf{E} \subset \mathbb{R}^4$ and $s > 0$ we let

$$\begin{aligned} N_s(\mathsf{E}) &:= \#\{v \in \Lambda'_{\mathrm{nz}} \cap e^s\mathsf{E} : \alpha \leq Q_0(v) \leq \beta\}, \\ N'_s(\mathsf{E}) &:= \#\{v \in (\Lambda'_{\mathrm{nz}} \setminus (L_s \cup L'_s)) \cap e^s\mathsf{E} : \alpha \leq Q_0(v) \leq \beta\}. \end{aligned}$$

4.3 Counting and circular averages. For the rest of the proof, we fix $\varepsilon = e^{-\eta' t}$ for some $0 < \eta' < 1/100$ which is small and will be optimized later. We will also assume $\beta - \alpha \geq \varepsilon$ otherwise Theorem 1.2 holds trivially.

Recall that

$$\Omega = \{(w_1 + w_2, \omega(w_1 - w_2)) : \|w_i\| \leq 1\},$$

and that $\Omega \setminus \mathsf{D} = \Omega_1 \cup \Omega_2$ where $\mathsf{D} = \{(v_1, v_2) \in \mathbb{R}^4 : \|v_k\| \leq 1\}$, and

$$\Omega_1 = \{(v_1, v_2) \in \Omega : \|v_1\| > 1\} \quad \text{and}$$
$$\Omega_2 = \{(v_1, v_2) \in \Omega : \|v_1\| \leq 1, \|v_2\| > 1\}.$$

Let R be a large constant (we will always assume $R < \varepsilon^{-1/20}$, hence, R is much smaller that e^t), satisfying

$$R \geq \max\{10^3, |\alpha|, |\beta|, |\mathsf{a}|, |\mathsf{b}|, |\mathsf{c}|\};$$

note that $\pi_k(\Omega) \subset B(0, R)$.

Apply the construction in §3.6 for $\pi_k(\Omega_k)$ with ε and R here. The analysis for $k = 1$ and 2 are similar, thus, let $k = 1$ until further notice. Let

$$\varphi^{\pm}_{i,j} = \xi^{\pm}_i \varrho^{\pm}_j \quad \text{for } (i, j) \in \mathfrak{I}^{\pm}_1.$$

Note that $\operatorname{supp}(\varrho^{\pm}_j) \subset [q - 200R^{10}\varepsilon, q] \subset [R^{-1}, R]$ for some $R^{-1} \leq q \leq R$, see (ϱ-1)—indeed in the case at hand, we have $1 \leq q \leq 2$.

For $\sigma = \pm$, define f^{σ}_1 as in (3.3) for q and ϱ^{σ}_j. Let

$$f^{\sigma}_2 = f_{I^{\sigma}_0, I^{\sigma}_{i,j}}, \tag{4.8}$$

where $I^+_0 = [-q^{-1}\beta, -q^{-1}\alpha]$ and $I^-_0 = (I^+_0)_{-100R^5\varepsilon}$, see (3.5) and (3.4). Put

$$f^{\sigma}_{i,j} = f^{\sigma}_1 f^{\sigma}_2.$$

4.4 Lemma. *Let the notation be as above, and let L and L' denote $(\delta_1/\overline{A}, \delta_1, t)$-exceptional subspaces if they exist.*

If $(i, j) \in \mathring{\mathfrak{I}}^-_1$, then

$$\begin{aligned} qe^{2t} \sum_{v \in \Lambda'_{\mathrm{nz}} \setminus (L \cup L')} \int_0^{2\pi} f^-_{i,j}(\Delta(a_t r_\theta)v)\xi^-_i(\theta)\, \mathrm{d}\theta \\ \leq (\varepsilon + O(\varepsilon^2)) \cdot N'_t(\Omega_{i,j}) + O(\varepsilon^{-21}). \end{aligned} \tag{4.9}$$

Moreover for every $(i, j) \in \mathfrak{I}^+_1$, we have

$$(\varepsilon + O(\varepsilon^2)) \cdot N'_t(\Omega^+_{i,j}) \leq qe^{2t} \sum_{v \in \Lambda'_{\mathrm{nz}} \setminus (L \cup L')} \int_0^{2\pi} f^+_{i,j}(\Delta(a_t r_\theta)v)\xi^+_i(\theta)\, \mathrm{d}\theta. \tag{4.10}$$

The implied constants depend polynomially on R.

The proof is similar to the proof of Lemma 3.9. More precisely, we will use (3.7) for $f^-_{i,j}$ and (3.9) for $f^+_{i,j}$; let us now turn to the details.

Proof. When there is no confusion we drop i, j from the notation and denote $f^{\pm}_{i,j}$ by $f^{\pm}$, $\xi^{\pm}_i$ by $\xi^{\pm}$, etc. Also, we will put $I_0 = I^-_0$ and $I_1 = I^-_{i,j}$, but will keep the more cumbersome notation for I^+_0 and $I^+_{i,j}$.

By (3.7) in Lemma 3.4 applied with $f^- = f^-_{i,j}$, for any $v \in \mathbb{R}^4$, we have

$$\begin{aligned} qe^{2t} &\int_0^{2\pi} f^-(\Delta(a_t r_\theta)v)\xi^-(\theta)\,\mathrm{d}\theta \\ &\le (1+O(\varepsilon))J_{f^-_1}(e^{-t}\|v_1\|)\xi^-(\theta_{v_1})1_{I_0^{(3)}}(-q^{-1}Q_0(v))1_{I_1^{(3)}}(e^{-t}\|v_2\|)+\mathcal{E}, \end{aligned} \tag{4.11}$$

where $I^{(k)} = I_{10kR^3\varepsilon}$ and

$$\mathcal{E} = O(\mathrm{Lip}(f^-_1)\,\mathrm{Lip}(\xi^-)e^{-2t}); \tag{4.12}$$

furthermore, $\mathcal{E} = 0$ if

$$(-q^{-1}Q_0(v), e^{-t\|v_2\|}) \notin I_0^{(3)} \times I_1^{(3)} \quad \text{or} \quad \|v_1\| > 2Re^t.$$

By (3.9) in Lemma 3.4 applied with $f^+ = f^+_{i,j}$, for any $v \in \mathbb{R}^4$ with $e^{-t}\|v_2\| \in I^+_{i,j}$ and $Q_0(v) \in [\alpha, \beta]$, we have

$$\begin{aligned} qe^{2t} &\int_0^{2\pi} f^+(\Delta(a_t r_\theta)v)\xi^+(\theta)\,\mathrm{d}\theta \\ &= (1+O(\varepsilon))J_{f^+_1}(e^{-t}\|v_1\|)\xi^+(\theta_{v_1})f^+_2(-q^{-1}Q_0(v), e^{-t}\|v_2\|) \\ &\quad + O(\mathrm{Lip}(f^+_1)\,\mathrm{Lip}(\xi^+)e^{-2t}). \end{aligned} \tag{4.13}$$

In particular, (4.13) holds for all $v \in e^t\mathbf{\Omega}^+_{i,j}$ with $Q_0(v) \in [\alpha, \beta]$ thanks to part (2) in Lemma 3.7.

Before analysing (4.11) and (4.13) further, we record the following:

$$O(\mathrm{Lip}(f^{\pm}_1)\,\mathrm{Lip}(\xi^{\pm})e^{-2t}) = O(\varepsilon^{-20}e^{-2t}) \ll \varepsilon^3, \tag{4.14}$$

so long as t is large enough (recall that the implied constants depend polynomially on R).

Let us now begin with (4.13). In view of (3.6), for any $v = (v_1, v_2) \in e^t\mathbf{\Omega}_1$ so that $\alpha \le Q_0(v) \le \beta$, we have

$$\begin{aligned} &J_{f^+_1}(e^{-t}\|v_1\|)\xi^+(\theta_{v_1})f^+_2(-q^{-1}Q_0(v), e^{-t}\|v_2\|) \\ &\quad = (\varepsilon + O(\varepsilon^2))\varrho^+(e^{-t}\|v_1\|)\xi^+(\theta_{v_1})f^+_2(-q^{-1}Q_0(v), e^{-t}\|v_2\|). \end{aligned} \tag{4.15}$$

Moreover, for every $v \in e^t\mathbf{\Omega}^+_{i,j}$, satisfying $\alpha \le Q_0(v) \le \beta$,

$$f^+_2(-q^{-1}Q_0(v), e^{-t}\|v_2\|) = 1, \quad \xi^+(\theta_{v_1}) = 1, \quad \text{and} \quad \varrho^+(e^{-t}\|v_1\|) = 1;$$

from this and (4.15), we conclude that

$$J_{f_1^+}(e^{-t}\|v_1\|)\xi^+(\theta_{v_1})f_2^+(-q^{-1}Q_0(v), e^{-t}\|v_2\|) = (\varepsilon + O(\varepsilon^2)).$$

Together with (4.13) and (4.14), this implies that

$$qe^{2t}\int_0^{2\pi} f^+(\Delta(a_t r_\theta)v)\xi^+(\theta)\,\mathrm{d}\theta = \varepsilon + O(\varepsilon^2) \tag{4.16}$$

for every $v \in e^t\Omega^+_{i,j}$ with $\alpha \le Q_0(v) \le \beta$.

Summing (4.16) over all such $v \in \Lambda'_{\mathrm{nz}} \setminus (L \cup L')$, we obtain

$$(\varepsilon + O(\varepsilon^2)) \cdot N'_t(\Omega^+_{i,j}) \le qe^{2t} \sum_{v\in\Lambda'_{\mathrm{nz}}\setminus(L\cup L')} \int_0^{2\pi} f^+(\Delta(a_t r_\theta)v)\xi^+(\theta)\,\mathrm{d}\theta. \tag{4.17}$$

This establishes (4.10).

Let us now assume $(i,j) \in \mathring{\mathfrak{I}}_1^-$ and obtain a lower bound for $N_t(\Omega_{i,j})$. For this, we investigate the term appearing in the second line of (4.11).

We first claim that

$$J_{f_1^-}(e^{-t}\|v_1\|)\xi^-(\theta_{v_1})1_{I_0^{(3)}}(-q^{-1}Q_0(v))1_{I_1^{(3)}}(e^{-t}\|v_2\|) \neq 0,$$

then $Q_0(v) \in [\alpha, \beta]$ and $v \in e^t\Omega_{i,j}$.

To see the claim, recall that by part (1) in Lemma 3.7, for any $v \in \mathbb{R}^4$,

$$J_{f_1^-}(e^{-t}\|v_1\|)\xi^-(\theta_{v_1})1_{I_0^{(3)}}(-q^{-1}Q_0(v))1_{I_1^{(3)}}(e^{-t}\|v_2\|) = 0$$

unless all the following are satisfied

$$Q_0(v) \in [\alpha + 50R^5\varepsilon, \beta - 50R^5\varepsilon], \tag{4.18a}$$

$$v_1 \in e^t\,\mathrm{supp}(\varphi^-_{i,j}), \tag{4.18b}$$

and

$$v \in e^t\Omega_{i,j}. \tag{4.18c}$$

In deducing (4.18a) from Lemma 3.7, we used the definitions

$$I_0^{(3)} = (I_0^-)_{30R^3\varepsilon} \quad\text{and}\quad I_0^- = ([-q^{-1}\beta, -q^{-1}\alpha])_{-100R^5\varepsilon}.$$

We conclude from (4.18a) that $Q_0(v) \in [\alpha, \beta]$. Using the definition of $\Omega_{i,j}$ in (3.22) and since $2R^3e^{-2t} < \varepsilon$, (4.18c) implies that $v \in e^t\Omega_{i,j}$, and completes the proof of the claim.

We now return to the proof of the lemma. Recall that

$$\begin{aligned}J_{f_1^-}(e^{-t}\|v_1\|)\xi^-(\theta_{v_1})1_{I_0^{(3)}}(-q^{-1}Q_0(v))1_{I_1^{(3)}}(e^{-t}\|v_2\|)\\ \le J_{f_1^-}(e^{-t}\|v_1\|) = (\varepsilon + O(\varepsilon^2))\varrho^-(e^{-t}\|v_1\|) \le \varepsilon + O(\varepsilon^2).\end{aligned}$$

This and the above claim imply that

$$\begin{aligned}(4.19)\quad \sum_{v\in\Lambda'_{\rm nz}\setminus(L\cup L')} & J_{f_1^-}(e^{-t}\|v_1\|)\xi^-(\theta_{v_1})1_{I_0^{(3)}}(-q^{-1}Q_0(v))1_{I_1^{(3)}}(e^{-t}\|v_2\|)\\ &\le (\varepsilon+O(\varepsilon^2))\cdot N'_t(\Omega_{i,j}).\end{aligned}$$

Moreover, since $(-q^{-1}Q_0(v), e^{-t\|v_2\|}) \notin I_0^{(3)}\times I_1^{(3)}$ or $\|v_1\| > 2Re^t$ imply $\mathcal{E}=0$, we conclude from Lemma 3.9 applied with $\eta=\eta'/10$ that

$$\sum_{v\in\Lambda'}\mathcal{E} \ll \varepsilon^{-20}e^{-2t}e^{(2+\eta)t} \ll \varepsilon^{-21};$$

we used $\mathrm{Lip}(f_1^-)\,\mathrm{Lip}(\xi^-)e^{-2t} \ll \varepsilon^{-20}e^{-2t}$, see (4.14), and $\varepsilon = e^{-\eta' t}$. This, (4.19) and (4.11) imply that

$$qe^{2t}\sum_{v\in\Lambda'_{\rm nz}\setminus L\cup L'}\int_0^{2\pi} f^-(\Delta(a_tr_\theta)v)\xi(\theta)\,\mathrm{d}\theta + O(\varepsilon^{-21}) \le (\varepsilon+O(\varepsilon^2))\cdot N'_t(\Omega_{i,j}),$$

as we claimed in (4.9). □

We will use Theorem 3.1 to reduce both (4.9) and (4.10) to the study of $\int_X \hat f^{\pm}_{i,j}\,\mathrm{d}m_X$, see (3.1). Let us begin with computing this integral.

4.5 Lemma. *For $\sigma=\pm$ let $f^\sigma_{i,j} = f_1^\sigma f_2^\sigma$, where for $k=1,2$, f_k^σ is as in §4.3. There is an absolute constant c_Λ so that*

$$q\int_X \hat f^\sigma_{i,j}\,\mathrm{d}m_X = c_\Lambda\varepsilon(\beta-\alpha)|I^\sigma_{i,j}|\int\varrho^\sigma_j + O(\varepsilon^2)(\beta-\alpha)|I^\sigma_{i,j}|\int\varrho^\sigma_j. \tag{4.20}$$

Proof. We have

$$\begin{aligned}\int_X \hat f^\sigma_{i,j}\,\mathrm{d}m_X &= c_\Lambda\int_{\mathbb{R}^2} f_1^\sigma\int_{\mathbb{R}^2} f_2^\sigma\\ &= c_\Lambda\varepsilon\int_{\mathbb{R}}\varrho^\sigma_j\int_{\mathbb{R}^2} f_2^\sigma + O(\varepsilon^2)\int_{\mathbb{R}}\varrho^\sigma_j\int_{\mathbb{R}^2} f_2^\sigma\end{aligned}$$

where c_Λ is absolute and the implied constants depend only on R.

Since f_2 is defined as in (4.8), we conclude that

$$\int f_2^\sigma = q^{-1}(\beta-\alpha)|I^\sigma_{i,j}| + O(q^{-1}\varepsilon(\beta-\alpha)|I^\sigma_{i,j}|);$$

again the implied constant depends only on R. The lemma follows. □

4.6 Lemma. *Let the notation be as in Lemma 4.5. In particular,*

$$f^{\pm}_{i,j} = f^{\pm}_1 f^{\pm}_2,$$

where $f^{\pm}_k$ are as in §4.3. Also put

$$\Upsilon^{\pm}_{i,j} = c_\Lambda(\beta - \alpha)|I^{\pm}_{i,j}| \int \xi^{\pm}_i \int \varrho^{\pm}_j.$$

If $(i,j) \in \mathring{\mathcal{I}}^{-}_1$, then

$$e^{2t}(\Upsilon^{-}_{i,j} + O(\mathcal{S}(f^{-}_{i,j})\mathcal{S}(\xi^{-}_i)e^{-\delta_2 t})) \le (1 + O(\varepsilon)) \cdot N'_t(\Omega_{i,j}). \tag{4.21}$$

Moreover, for every $(i,j) \in \mathcal{I}^{+}_1$, we have

$$(1 + O(\varepsilon)) \cdot N'_t(\Omega^{+}_{i,j}) \le e^{2t}(\Upsilon^{+}_{i,j} + O(\mathcal{S}(f^{+}_{i,j})\mathcal{S}(\xi^{+}_i)e^{-\delta_2 t})), \tag{4.22}$$

where the implied constants depends polynomially on R.

Proof. We will prove the lemma using Lemma 4.4 and Theorem 3.1. Let us begin with restating the main conclusion of Theorem 3.1 in the form which will be used here. When there is no confusion, we drop i, j from the notation and denote $f^{\pm}_{i,j}$ by $f^{\pm}$, $\xi^{\pm}_i$ by $\xi^{\pm}$, etc.

Recall that $\Lambda' = g\Lambda$ where $g = (g_1, g_2)$ is as in (4.1). Let L and L' be as in Lemma 4.2 if they exist. For $\sigma = \pm$, put

$$\hat{f}^{\sigma}_{\mathrm{sp}}(\theta) = \sum_{v \in \Lambda' \cap (L \cup L')} f^{\sigma}_{\mathrm{sp}}(\Delta(a_t r_\theta)v),$$

$$\mathcal{C}_\sigma = \{\theta \in [0, 2\pi] : \hat{f}^{\sigma}_{\mathrm{sp}}(\theta) \ge e^{\delta_1 t}\},$$

and define

$$\hat{f}^{\sigma}_{\mathrm{mod}}(\theta) = \begin{cases} \hat{f}^{\sigma}(\theta) - \hat{f}^{\sigma}_{\mathrm{sp}}(\theta), & \theta \in \mathcal{C}_\sigma, \\ \hat{f}^{\sigma}(\theta), & \text{otherwise}, \end{cases}$$

where we write $\hat{f}^{\sigma}(\theta) = \hat{f}^{\sigma}(\Delta(a_t r_\theta)g\Gamma')$.

Since g satisfies (4.6), Theorem 3.1 and the definition of $\hat{f}^{\sigma}_{\mathrm{mod}}(\theta)$ imply

$$\int_0^{2\pi} \hat{f}^{\sigma}_{\mathrm{mod}}(\theta)\xi^{\sigma}(\theta)\,\mathrm{d}\theta = \int \xi^{\sigma}\mathrm{d}\theta \int_X \hat{f}^{\sigma}\mathrm{d}m_X + O(\mathcal{S}(f^{\sigma})\mathcal{S}(\xi^{\sigma})e^{-\delta_2 t}). \tag{4.23}$$

With this established, we first show (4.21). Let $\sigma = -$. Assuming η' in the definition of $\varepsilon = e^{-\eta' t}$ is small enough, we have

$$O(\mathcal{S}(f^{-})\mathcal{S}(\xi^{-})e^{-\delta_2 t}) < \varepsilon^4(\beta - \alpha).$$

Recall from §3.6 that $\int \varrho^{-}_j \ge \varepsilon$ and that $|I^{-}_{i,j}| \ge \sqrt{\varepsilon}$. Thus (4.23), together with the above and Lemma 4.5, implies that

$$\int_0^{2\pi} \hat{f}^{-}_{\mathrm{mod}}(\theta)\xi^{-}(\theta)\,\mathrm{d}\theta \gg \varepsilon^3(\beta - \alpha). \tag{4.24}$$

Moreover, by part (2) in Lemma 3.2 applied with δ_1, L, and L', we have

$$\int_{[0,2\pi]\setminus\mathcal{C}_-} \hat{f}^-_{\mathrm{sp}}(\theta)\,\mathrm{d}\theta \ll e^{(-1+\delta_1)t}. \tag{4.25}$$

Recall that $\delta_1 < 1/100$, hence, if $\eta' < 1/100$, then $e^{(-1+\delta_1)t} < \varepsilon^4(\beta-\alpha)$. Thus, we get from (4.24) and (4.25)

$$\begin{aligned}
&\sum_{v\in\Lambda'_{\mathrm{nz}}\setminus(L\cup L')} \int_0^{2\pi} f^-(\Delta(a_t r_\theta)v)\xi^-(\theta)\,\mathrm{d}\theta \\
&\qquad = \int_0^{2\pi} \hat{f}^-_{\mathrm{mod}}(\theta)\xi^-(\theta)\,\mathrm{d}\theta - \int_{[0,2\pi]\setminus\mathcal{C}_-} \hat{f}^-_{\mathrm{sp}}(\theta)\xi_i^-(\theta)\,\mathrm{d}\theta \\
&\qquad = (1+O(\varepsilon)) \int_0^{2\pi} \hat{f}^-_{\mathrm{mod}}(\theta)\xi^-(\theta)\,\mathrm{d}\theta.
\end{aligned} \tag{4.26}$$

In view of (4.9) in Lemma 4.4,

$$\begin{aligned}
qe^{2t} \sum_{v\in\Lambda'_{\mathrm{nz}}\setminus(L\cup L')} \int_0^{2\pi} f^-(\Delta(a_t r_\theta)v)\xi^-(\theta)\,\mathrm{d}\theta + O(\varepsilon^{-21})& \\
\le (\varepsilon + O(\varepsilon^2))\cdot N'_t(\Omega_{i,j})&.
\end{aligned}$$

Using this and (4.26) (multiplied by qe^{2t}), we conclude that

$$qe^{2t}(1+O(\varepsilon)) \int_0^{2\pi} \hat{f}^-_{\mathrm{mod}}(\theta)\xi^-(\theta)\,\mathrm{d}\theta + O(\varepsilon^{-21}) \le (\varepsilon+O(\varepsilon^2))\cdot N'_t(\Omega_{i,j}).$$

This, (4.23) and (4.20) yield

$$e^{2t}(\Upsilon^-_{i,j} + O(\mathcal{S}(f^-)\mathcal{S}(\xi^-)e^{-\delta_2 t}) + O(\varepsilon^{-21}) \le (1+O(\varepsilon))\cdot N'_t(\Omega_{i,j}). \tag{4.27}$$

Assuming η' is small enough and t large, we have

$$\varepsilon^{-23} < e^{2t}\cdot\left(c_\Lambda(\beta-\alpha)|I_{i,j,-}| \int_{\mathbb{R}} \varrho_j^- \int_{\mathbb{R}} \xi_i^-\right).$$

Hence, (4.21) follows from (4.27).

We now show (4.22); the argument is similar and simpler. By (4.10),

$$\begin{aligned}
(\varepsilon+O(\varepsilon^2))\cdot N'_t(\Omega^+_{i,j}) &\le qe^{2t} \sum_{v\in\Lambda'_{\mathrm{nz}}\setminus(L\cup L')} \int_0^{2\pi} f^+(\Delta(a_t r_\theta)v)\xi^+(\theta)\,\mathrm{d}\theta \\
&\le qe^{2t} \int_0^{2\pi} \hat{f}^+_{\mathrm{mod}}(\theta)\xi^+(\theta)\,\mathrm{d}\theta.
\end{aligned} \tag{4.28}$$

Thus, (4.22) follows from (4.28), (4.23) and (4.20), applied with $\sigma = +$. □

4.7 Lemma. *There exists η depending on η' and some $\overline{C}_1$ so that*

$$N_t(\Omega \setminus \mathsf{D}) = \overline{C}_1(\beta - \alpha)e^{2t} + \mathcal{M}_0 + O((1 + |\alpha| + |\beta|)^N e^{(2-2\eta)t}) \tag{4.29}$$

where N is absolute, the implied constants depend on R and

$$\mathcal{M}_0 = \#\{v \in \Lambda'_{\mathrm{nz}} \cap (L \cup L') \cap e^t(\Omega \setminus \mathsf{D}) : \alpha \le Q_0(v) \le \beta\}.$$

A similar assertion holds with $\Omega \setminus \mathsf{D}$ replaced by $\mathsf{D} \setminus e^{-1}\mathsf{D}$.

Proof. We will prove the assertion for $\Omega \setminus \mathsf{D}$, the proof for $\mathsf{D} \setminus e^{-1}\mathsf{D}$ is similar. Recall that $\Omega \setminus \mathsf{D} = \Omega_1 \cup \Omega_2$ where

$$\Omega_1 = \{(v_1, v_2) \in \Omega : \|v_1\| > 1\}$$

and

$$\Omega_2 = \{(v_1, v_2) \in \Omega : \|v_1\| \le 1, \|v_2\| > 1\}.$$

Fix $k = 1$ or 2. By (4.21), for all $(i, j) \in \mathring{\mathcal{I}}_k^-$,

$$\begin{aligned} &e^{2t}(\Upsilon_{i,j}^- + O(\mathcal{S}(f_{i,j}^\sigma)\mathcal{S}(\xi_i^+)e^{-\delta_2 t})) \\ &\qquad \le (1 + O(\varepsilon)) \cdot N'_t(\Omega_{i,j}) \le (1 + O(\varepsilon)) \cdot N_t(\Omega_{i,j}^+), \end{aligned} \tag{4.30}$$

where we used $\Omega_{i,j} \subset \Omega_{i,j}^+$ in the second inequality, (3.22).

Also by (4.22), for all $\varphi_{i,j}^+ \in \mathcal{I}_k^+$, we have

$$(1 + O(\varepsilon)) \cdot N'_t(\Omega_{i,j}^+) \le e^{2t}(\Upsilon_{i,j}^+ + O(\mathcal{S}(f_{i,j}^+)\mathcal{S}(\xi_i^+)e^{-\delta_2 t})). \tag{4.31}$$

Thus summing (4.30) over all $(i, j) \in \mathring{\mathcal{I}}_k^-$,

$$\begin{aligned} &e^{2t}\sum_{\mathring{\mathcal{I}}_k^-}(\Upsilon_{i,j}^- + O(\mathcal{S}(f_{i,j}^+)\mathcal{S}(\xi_i^+)e^{-\delta_2 t})) \\ &\qquad \le (1 + O(\varepsilon))\sum_{\mathring{\mathcal{I}}_k^-} N'_t(\Omega_{i,j}) \le (1 + O(\varepsilon))\sum_{\mathring{\mathcal{I}}_k^-} N'_t(\Omega_{i,j}^+). \end{aligned} \tag{4.32}$$

Moreover, summing (4.31) over all $(i, j) \in \mathcal{I}_k^+$, we get the following:

$$\begin{aligned} (1 + O(\varepsilon))\sum_{\mathring{\mathcal{I}}_k^-} N'_t(\Omega_{i,j}^+) &\le (1 + O(\varepsilon))\sum_{\mathcal{I}_k^+} N'_t(\Omega_{i,j}^+) \\ &\le e^{2t}\sum_{\mathcal{I}_k^+}(\Upsilon_{i,j}^+ + O(\mathcal{S}(f_{i,j}^+)\mathcal{S}(\xi_i^+)e^{-\delta_2 t}). \end{aligned} \tag{4.33}$$

By (3.23a) and (3.23b), $\Omega_{i,j} \subset \Omega_k$ are disjoint and $\Omega_k \subset \bigcup_{\mathcal{I}_k^+} \Omega_{i,j}^+$. Hence, (4.32) implies that

$$(I) \le (1 + O(\varepsilon))N'_t(\Omega_k) \le (II), \tag{4.34}$$

where (I) is the first line in (4.32) and (II) is the last line in (4.33).

Recall from Lemma 4.6 that

$$\Upsilon^{\pm}_{i,j} = c_\Lambda(\beta-\alpha)|I_{i,j,\pm}|\int_{\mathbb{R}}\xi_i^{\pm}\int_{\mathbb{R}}\varrho_j^{\pm};$$

in view of (ξ-1), (ϱ-1), and (3.25), the above implies that

$$\sum_{\mathcal{I}_k^+}\Upsilon^+_{i,j} = (1+O(\varepsilon))\sum_{\mathring{\mathcal{I}}_k^-}\Upsilon^-_{i,j} = (1+O(\varepsilon))(\beta-\alpha)\bar{C}_{k,1}$$

where $\bar{C}_{k,1}$ is absolute and the implied constants depend on R.

Furthermore, using $\varepsilon = e^{-\eta'}$, we conclude that

$$\begin{aligned}\sum_{i,j}\mathcal{S}(f^{\pm}_{i,j,})\mathcal{S}(\xi_i^{\pm})e^{-\delta_2 t} &\ll (1+|\alpha|+|\beta|)^N\varepsilon^{-N}e^{-\delta_2 t}\\ &\ll (1+|\alpha|+|\beta|)^N e^{-\delta_2 t/2},\end{aligned}$$

where the implied constant depends on R and we assume η' is small enough so that $\delta_2 - N\eta' > \delta_2/2$.

Altogether, there is some $\eta > 0$ so that for $k = 1, 2$, we have

$$N'_t(\Omega_k) = \bar{C}_{k,1}(\beta-\alpha)e^{2t} + (1+|\alpha|+|\beta|)^N e^{(2-2\eta)t}.$$

Since $\Omega\setminus\mathsf{D} = \Omega_1\cup\Omega_2$ is a disjoint union, we conclude that

$$N'_t(\Omega\setminus\mathsf{D}) = \bar{C}_1(\beta-\alpha)e^{2t} + (1+|\alpha|+|\beta|)^N e^{(2-2\eta)t} \tag{4.35}$$

where $\bar{C}_1 = \bar{C}_{1,1} + \bar{C}_{2,1}$.

The lemma follows from (4.35) and the definition of $\mathcal{M}_0$. □

Proof of Theorem 1.2. We will again use the following:

$$\#\{v\in\Lambda'\cap e^{\frac{2t}{5}}\mathsf{D}\} \le C_1' e^{\frac{8t}{5}} \tag{4.36}$$

where C_1' depends on R, see (3.28).

First apply Lemma 4.7, with t and $\Omega\setminus\mathsf{D}$. Then

$$\begin{aligned}&N'_t(\Omega\setminus\mathsf{D})\\ &\quad= \bar{C}_1(\beta-\alpha)e^{2(t-\ell)} + \mathcal{M}' + O((1+|\alpha|+|\beta|)^N e^{(2-2\eta)(t-\ell)})\end{aligned} \tag{4.37}$$

where

$$\mathcal{M}' = \#\{v\in\Lambda'_{\mathrm{nz}}\cap(L\cup L')\cap e^t(\Omega\setminus\mathsf{D}) : \alpha\le Q_0(v)\le\beta\}.$$

We now control the contribution of $\Lambda'\cap e^t\mathsf{D}$ to the count. Recall our notation $\mathsf{D}(e^{-\ell}) = e^{-\ell}\mathsf{D}$. Then $e^t\mathsf{D}(e^{-\ell}) = e^{t-\ell}\mathsf{D}$, and

$$e^{t-\ell}(\mathsf{D}\setminus e^{-1}\mathsf{D}) = e^t(\mathsf{D}(e^{-\ell})\setminus(e^{-1}\mathsf{D}(e^{-\ell}))).$$

Applying Lemma 4.7 with $t - \ell$ (instead of t) for $\ell \le 3t/5$ and $\mathsf{D} \setminus e^{-1}\mathsf{D}$,

$$N'_t(\mathsf{D}(e^{-\ell}) \setminus e^{-1}\mathsf{D}(e^{-\ell})) = \bar{\bar{C}}_1(\beta - \alpha)e^{2(t-\ell)} + \mathcal{M}_\ell + O((1 + |\alpha| + |\beta|)^N e^{(2-2\eta)(t-\ell)}) \tag{4.38}$$

where

$$\mathcal{M}_\ell = \#\{v \in \Lambda'_{\mathrm{nz}} \cap (L \cup L') \cap e^t(\mathsf{D}(e^{-\ell}) \setminus e^{-1}\mathsf{D}(e^{-\ell})) : \alpha \le Q_0(v) \le \beta\}$$

and L, L' are as in Lemma 4.2.

Summing (4.38) over $0 \le \ell \le 3t/5$, we get

$$N_t(\mathsf{D} \setminus e^{-3t/5}\mathsf{D}) = \bar{\bar{C}}_1(\beta - \alpha)e^{2t} + \mathcal{M}'' + O((1 + |\alpha| + |\beta|)^N e^{(2-\eta)t})$$

where $\mathcal{M}'' = \sum \mathcal{M}_\ell$. This, (4.37) and (4.36) thus imply

$$N_t(\Omega) = C_1(\beta - \alpha)e^{2t} + \mathcal{M} + O((1 + |\alpha| + |\beta|)^N e^{(2-\eta)t}) \tag{4.39}$$

where $\mathcal{M} = \#\{v \in \Lambda'_{\mathrm{nz}} \cap (L \cup L') \cap e^t\Omega : \alpha \le Q_0(v) \le \beta\}$.

To conclude the proof, we rewrite (4.39) in the notation of Theorem 1.2 and further analyze $\mathcal{M}$. Recall that $t = \frac{1}{2}\log T$, hence, by (4.3) and (4.39),

$$R_{\mathsf{M}}(\alpha, \beta, T) = C_1(\beta - \alpha)T + \mathcal{M} + O((1 + |\alpha| + |\beta|)^N T^{1-\frac{\eta}{2}}), \tag{4.40}$$

we now turn to the term $\mathcal{M}$. Since

$$Q_0(g_1w_1, g_2w_2) = Q_0(g_2^{-1}g_2w_1, w_2) \quad \text{and} \quad g_2^{-1}g_2 = \begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix}.$$

We conclude, as in the proof of Lemma 2.5, that if we put $w_1 = (\mathsf{x}_1, \mathsf{y}_1)$ and $w_2 = (-\mathsf{y}_2, \mathsf{x}_2)$, then $u_i = (\mathsf{x}_i, \mathsf{y}_i)$ satisfy

$$\begin{aligned} &\|u_i\| \le \max\{\|g_1^{\pm 1}\|, \|g_2^{\pm 1}\|\}e^{\delta_1 t/\bar{A}} \le e^{2\delta_1 t/\bar{A}} \quad \text{and} \\ &|B_{\mathsf{M}}(u_1, u_2)| = |Q_0(g_2^{-1}g_1w_1, w_2)| \le e^{-2\delta_1 t/5}, \end{aligned} \tag{4.41}$$

where we assumed t is large in the second inequality of the first line. Thus by Lemma 2.5, the pair (w'_1, w'_2) is obtained from (u_2, u_1) using the above relation, that is, $w'_1 = (\mathsf{x}_2, \mathsf{y}_2)$ and $w'_2 = (-\mathsf{y}_1, \mathsf{x}_1)$.

Let $v \in \Lambda' \cap (L \cup L') \cap e^t\Omega$ satisfy that $\alpha \le Q_0(v) \le \beta$. For simplicity, let us assume that $v \in L$ and write $v = \ell_1(g_1w_1, 0) + \ell_2(0, g_2w_2)$. Then,

$$v = (v_1 + v_2, \omega(v_1 - v_2)) = (\ell_1 g_1 w_1, \ell_2 g_2 w_2)$$

where $v_i \in 2\pi\Delta^*$ and $\|v_i\| \le e^t$. Recall also that $(g_1, g_2) = (g_{\mathsf{M}}, -\omega g_{\mathsf{M}}\omega)$ and $g_{\mathsf{M}}\mathbb{Z}^2 = 2\pi\Delta^*$, hence,

$$\begin{aligned} v_1 &= g_{\mathsf{M}}\frac{\ell_1 w_1 - \ell_2\omega w_2}{2} = g_{\mathsf{M}}\frac{\ell_1 u_1 + \ell_2 u_2}{2},\\ v_2 &= g_{\mathsf{M}}\frac{\ell_1 w_1 + \ell_2\omega w_2}{2} = g_{\mathsf{M}}\frac{\ell_1 u_1 - \ell_2 u_2}{2}; \end{aligned}$$

changing L to L' yields $v_1 = g_{\mathsf{M}}\frac{\ell_1 u_1 + \ell_2 u_2}{2}$ and $v_2 = g_{\mathsf{M}}\frac{-\ell_1 u_1 + \ell_2 u_2}{2}$.

Altogether, (4.2) implies that

$$\begin{aligned} \mathcal{M} = \#\Big\{(\ell_1, \ell_2) : g_{\mathsf{M}}\frac{\ell_1 u_1 + \ell_2 u_2}{2} = v_1,\, g_{\mathsf{M}}\frac{\ell_1 u_1 - \ell_2 u_2}{2} = v_2\\ v_i \in 2\pi\Delta^*, \|v_i\| \le e^t, \alpha \le \|v_1\|^2 - \|v_2\|^2 \le \beta\Big\}. \end{aligned}$$

By Lemma 2.6, applied with $2\delta_1/\overline{A}$ and $\overline{A}/5$, we conclude that

$$\mathcal{M} \ll \max(|\alpha|, |\beta|)e^{(2-\frac{2\delta_1}{A})t} = \max(|\alpha|, |\beta|)T^{1-\frac{\delta_1}{A}},$$

where the implied constant depends on a, b, and c unless

$$|B_{\mathsf{M}}(u_1, u_2)| \le e^{(-2+\frac{2\delta_1}{A})t} \le T^{-1+\delta}.$$

Let $\kappa = \min\{\eta/2, \delta_1/\overline{A}\}$. Altogether, we conclude that

$$R_{\mathsf{M}}(\alpha, \beta, T) = C_1(\beta - \alpha)T + O((1 + |\alpha| + |\beta|)^N T^{1-\kappa})$$

unless $\{u_1, u_2\}$ satisfy (1.5), in which case we have

$$R_{\mathsf{M}}(\alpha, \beta, T) = C_1(\beta - \alpha)T + \mathcal{M} + O((1 + |\alpha| + |\beta|)^N T^{1-\kappa}). \tag{4.42}$$

We now show that $\mathcal{M} = M_T(u_1, u_2)$. Let (ℓ_1, ℓ_2) be as in the definition of $\mathcal{M}$, then

$$B_{\mathsf{M}}\Big(\frac{\ell_1 u_1 + \ell_2 u_2}{2}\Big) = \Big\|g_{\mathsf{M}}\frac{\ell_1 u_1 + \ell_2 u_2}{2}\Big\|^2 = \|v_1\|^2 \le e^{2t} = T.$$

Similarly for $v_2 = \frac{\ell_1 u_1 - \ell_2 u_2}{2}$. Moreover, we have

$$\begin{aligned} \|v_1\|^2 - \|v_2\|^2 &= B_{\mathsf{M}}\Big(\frac{\ell_1 u_1 + \ell_2 u_2}{2}\Big) - B_{\mathsf{M}}\Big(\frac{\ell_1 u_1 - \ell_2 u_2}{2}\Big)\\ &= B_{\mathsf{M}}(u_1, u_2)\ell_1\ell_2 \in [\alpha, \beta]. \end{aligned}$$

Thus $(\ell_1/2, \ell_2/2)$ satisfies the conditions in the definition $M_T(u_1, u_2)$. Similarly if (ℓ'_1, ℓ'_2) satisfies the conditions in the definition $M_T(u_1, u_2)$, then $(2\ell'_1, 2\ell'_2)$ satisfy the conditions in the definition of $\mathcal{M}$.

The proof is complete. $\square$

Proof of Corollary 1.3. We first prove part (1). Recall our assumption that there exist $A, q > 0$ so that for all $(m, n, k) \in \mathbb{Z}^3$ we have

$$|\mathsf{a}m + \mathsf{b}n + \mathsf{c}k| > q\|(m, n, k)\|^{-A}. \tag{4.43}$$

This implies that (1.4) holds for some A', depending on A, and all $T \geq T_0(A, q)$. Furthermore, in view of (4.43), for $u_i = (\mathsf{x}_i, \mathsf{y}_i) \in \mathbb{Z}^2$, we have

$$\begin{aligned}|B_{\mathsf{M}}(u_1, u_2)| &= |\mathsf{a}\mathsf{x}_1\mathsf{x}_2 + \mathsf{b}(\mathsf{y}_1\mathsf{x}_2 + \mathsf{x}_1\mathsf{y}_2) + \mathsf{c}\mathsf{y}_1\mathsf{y}_2| \\ &> q\|(\mathsf{x}_1\mathsf{x}_2, \mathsf{y}_1\mathsf{x}_2 + \mathsf{x}_1\mathsf{y}_2, \mathsf{y}_1\mathsf{y}_2)\|^{-A},\end{aligned}$$

which implies (1.5) does not hold so long as δ is small enough. In view of Theorem 1.2, this finishes the proof of part (1).

The proof of part (2) is similar. Recall that $\mathsf{b} = 0$ and $\mathsf{ac} = 1$. By our assumption there exist $A, q > 0$ so that for all $(m, n) \in \mathbb{Z}^2$, we have

$$|\mathsf{a}^2 m + n| > q\|(m, n)\|^{-A}. \tag{4.44}$$

As in the previous case, we conclude that (1.4) holds for some A', depending on A, and all $T \geq T_0(A, q)$. Hence, by Theorem 1.2, either

$$|R_{\mathsf{M}}(\alpha, \beta, T) - \pi^2(\beta - \alpha)| \leq C(1 + |\alpha| + |\beta|)^N T^{-\kappa},$$

which implies the claim in this part, or there are $u_1, u_2 \in \mathbb{Z}^2 \setminus \{0\}$ so that

$$\|u_1\|, \|u_2\| \leq T^{\delta/A} \quad \text{and} \quad |B_{\mathsf{M}}(u_1, u_2)| \leq T^{-1+\delta} \tag{4.45}$$

and moreover

$$R_{\mathsf{M}}(\alpha, \beta, T) - \pi^2(\beta - \alpha) = \frac{M_T(u_1, u_2)}{T} + O(C(1 + |\alpha| + |\beta|)^N T^{-\kappa}) \tag{4.46}$$

where

$$\begin{aligned}M_T(u_1, u_2) = \#\Big\{&(\ell_1, \ell_2) \in \tfrac{1}{2}\mathbb{Z}^2 : \ell_1 u_1 \pm \ell_2 u_2 \in \mathbb{Z}^2, \\ &B_{\mathsf{M}}(\ell_1 u_1 \pm \ell_2 u_2) \leq T, 4B_{\mathsf{M}}(u_1, u_2)\ell_1\ell_2 \in [\alpha, \beta]\Big\}.\end{aligned}$$

By Lemma 2.5, if T_0 is large enough, then $B_{\mathsf{M}}(u_1, u_2) = 0$. Hence $M_T(u_1, u_2)$ does not contribute to $R'_{\mathsf{M}}(\alpha, \beta)$. This and (4.46) finish the proof of this case and of the corollary. $\square$

5 Equidistribution of expanding circles

In this section we prove an effective equidistribution result for circular averages; the proof is based on [LMW22].

Let $G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ and let $\Gamma \subset G$ be a lattice; put $X = G/\Gamma$. Let m_X denote the G-invariant probability measure on X.

We fix a right invariant metric on G using the Killing form and the maximal compact subgroup $\mathrm{SO}(2) \times \mathrm{SO}(2)$, and let d_X denote the induced metric on X. There exists D' so that for all $\tau \geq 2$ and all $\theta \in \mathbb{R}$,

$$d_X(x, x') \leq e^{D'\tau} d_X(\Delta(a_\tau r_\theta)x, \Delta(a_\tau r_\theta)x'). \tag{5.1}$$

For the convenience of the reader, we give again the statement of Theorem 1.4:

1.4 Theorem. *Assume Γ is arithmetic. For every $x_0 \in X$, and large enough R (depending explicitly on X and the injectivity radius at x_0), for any $e^t \geq R^D$, at least one of the following holds.*

(1) *For every $\phi \in C_c^\infty(X)$ and 2π-periodic smooth function ξ on $\mathbb{R}$, we have*

$$\left| \int_0^{2\pi} \phi(\Delta(a_t r_\theta)x_0)\xi(\theta)\,\mathrm{d}\theta - \int_0^{2\pi} \xi(\theta)\,\mathrm{d}\theta \int \phi\,\mathrm{d}m_X \right| \leq \mathcal{S}(\phi)\mathcal{S}(\xi)R^{-\kappa_0}$$

where we use $\mathcal{S}(\cdot)$ to denote an appropriate Sobolev norm on both X and $\mathbb{R}$ respectively.

(2) *There exists $x \in X$ such that Hx is periodic with $\mathrm{vol}(Hx) \leq R$, and*

$$d_X(x, x_0) \leq R^D t^D e^{-t}.$$

The constants D and κ_0 are positive and depend on X but not on x_0.

Proof. Fix $0 < \zeta_0 < 1/10$ such that the U^-AU decomposition is an analytic diffeomorphism on the identity neighborhood of radius $2\zeta_0$ in $\mathrm{SL}_2(\mathbb{R})$, where U^- is the subgroup of lower triangular unipotent matrices, U is the subgroup of upper triangular unipotent matrices, and A is the subgroup of diagonal matrices. In particular, there are analytic diffeomorphisms s^-, τ, s from $(-\zeta_0, \zeta_0)$ to neighborhoods of 0 in $(-1, 1)$, such that $r_\zeta = u^-_{s^-(\zeta)} a_{\tau(\zeta)} u_{s(\zeta)}$. Note that

$$\tau(\zeta) = O(\zeta^2),\ s(\zeta) = \zeta + O(\zeta^2),\ s^-(\zeta) = -\zeta + O(\zeta^2), \tag{5.2}$$

and $\frac{\mathrm{d}}{\mathrm{d}\zeta}s = 1 + O(\zeta)$.

Using this we approximate the circular average (on small intervals) with unipotent average. First note that

$$\begin{aligned}\Delta(a_t r_{\hat\zeta+\zeta})x_0 &= \Delta(a_t u^-_{s^-(\zeta)} a_{\tau(\zeta)} u_{s(\zeta)} r_{\hat\zeta})x_0 \\ &= \Delta(a_t u^-_{s^-(\zeta)} a_{-t} a_{\tau(\zeta)})\Delta(a_t u_{s(\zeta)} r_{\hat\zeta})x_0\end{aligned}$$

is within distance $O(e^{-2t}s^{-}(\zeta)+\tau(\zeta)) = O(e^{-2t}\zeta+\zeta^2)$ from $\Delta(a_t u_{s(\zeta)} r_{\hat\zeta})x_0$. Therefore for all $0 \le \zeta \le \zeta_0$ we have

$$\begin{aligned}
&\frac{1}{\zeta}\int_0^{\zeta} \phi(\Delta(a_t r_{\hat\zeta+\theta})x_0)\,\mathrm{d}\theta \\
&\qquad = \frac{1}{\zeta}\int_0^{\zeta} \phi(\Delta(a_t u_{s(\theta)} r_{\hat\zeta})x_0)\,\mathrm{d}\theta + O(\mathcal{S}(\phi)(e^{-2t}\zeta+\zeta^2)) \\
&\qquad = \frac{1}{\zeta}\int_0^{s(\zeta)} \phi(\Delta(a_t u_{\theta} r_{\hat\zeta}(x_0)(s^{-1}(\theta))'\,\mathrm{d}\theta + O(\mathcal{S}(\phi)(e^{-2t}\zeta+\zeta^2))
\end{aligned}$$

where we used the above estimate in the first equality and a change of variable in the second equality.

Since $s(\zeta)-\zeta = O(\zeta^2)$, see (5.2), we conclude that

$$\frac{1}{\zeta}\int_0^{\zeta} \phi(\Delta(a_t r_{\hat\zeta+\theta})x_0)\,\mathrm{d}\theta = \frac{1}{\zeta}\int_0^{\zeta} \phi(\Delta(a_t u_{\theta} r_{\hat\zeta})x_0)(s^{-1}(\theta))'\,\mathrm{d}\,\theta + O(\mathcal{S}(\phi)\zeta)$$

where we used $e^{-2t}\zeta+\zeta^2 \le 2\zeta$.

Similarly, using $\sup_{\theta\in(0,\zeta)} |(s^{-1}(\theta))'-1| \ll \zeta$ and a change of variable,

$$\begin{aligned}
\frac{1}{\zeta}\int_0^{\zeta} \phi(\Delta(a_t r_{\hat\zeta+\theta})x_0)\,\mathrm{d}\theta &= \frac{1}{\zeta}\int_0^{\zeta} \phi(\Delta(a_t u_{\theta} r_{\hat\zeta})x_0)\,\mathrm{d}\theta + O(\mathcal{S}(\phi)\zeta) \\
&= \int_0^{1} \phi(\Delta(a_t u_{\zeta s} r_{\hat\zeta})x_0)\,\mathrm{d}s + O(\mathcal{S}(\phi)\zeta).
\end{aligned} \tag{5.3}$$

Let $\tau = -(\log\zeta)/2$. Then

$$\begin{aligned}
\int_0^{1} \phi(\Delta(a_t u_{\zeta s} r_{\hat\zeta})x_0)\,\mathrm{d}s &= \int_0^{1} \phi(\Delta(a_{t-\tau} a_\tau u_{\zeta s} a_{-\tau} a_\tau r_{\hat\zeta})x_0)\,\mathrm{d}s \\
&= \int_0^{1} \phi(\Delta(a_{t-\tau} u_s a_\tau r_{\hat\zeta})x_0)\,\mathrm{d}s.
\end{aligned} \tag{5.4}$$

Let D_1 and κ_1 be the constants given by [LMW22, Thm. 1.1] applied with X (D_1 denotes A in [LMW22, Thm. 1.1]). We will show the proposition holds with

$$D = D_1 + D' + 1$$

where D' is as in (5.1).

Let $T = e^{t-\tau}$ and $R = e^{D''\tau}$ for some $D'' \ge 1$ which will be explicated momentarily. Assume $e^t \ge R^D$, then

$$T = e^{t-\tau} = e^t R^{-1/D''} \ge R^{D-1} \ge R^{D_1}. \tag{5.5}$$

Apply [LMW22, Thm. 1.1], with $x_{\hat\zeta} := \Delta(a_\tau r_{\hat\zeta})x_0$, $T \ge R^{D_1}$, see (5.5), then so long as D'' is large enough, at least one of the following holds:

Case 1: For every $\hat{\zeta} \in [0, 2\pi]$ and all $\phi \in C_c^\infty(X)$,

$$\left| \int_0^1 \phi(\Delta(a_{\log T} u_s) x_{\hat{\zeta}})\,\mathrm{d}s - \int \phi\,\mathrm{d}m_X \right| \le \mathcal{S}(\phi) R^{-\kappa_1}. \tag{5.6}$$

Case 2: For some $\hat{\zeta} \in [0, 2\pi]$, there exists $x \in X$ such that Hx is periodic with $\mathrm{vol}(Hx) \le R$ and

$$d_X(x, x_{\hat{\zeta}}) \le R^{D_1} (\log T)^{D_1} T^{-1}. \tag{5.7}$$

We will show that part 1 in the proposition holds if case 1 holds and part 2 in the proposition holds if case 2 holds.

Let us first assume that case 1 holds. We begin with the following computation.

$$\begin{aligned}\int_0^{2\pi} \phi(\Delta(a_t r_\theta) x_0) \xi(\theta)\,\mathrm{d}\theta &= \frac{1}{\zeta} \int_{\hat{\zeta}=0}^{2\pi} \int_0^{\zeta} \phi(\Delta(a_t r_{\hat{\zeta}+\theta}) x_0) \xi(\hat{\zeta}+\theta)\,\mathrm{d}\theta\,\mathrm{d}\hat{\zeta} \\ &= \frac{1}{\zeta} \int_{\hat{\zeta}=0}^{2\pi} \left(\int_0^{\zeta} \phi(\Delta(a_t r_{\hat{\zeta}+\theta}) x_0)\,\mathrm{d}\theta \right) \xi(\hat{\zeta})\,\mathrm{d}\hat{\zeta} \\ &\quad + O\Big(\sup|\phi| \cdot \sup_{\hat{\zeta}\in[0,2\pi],\theta\in[0,\zeta]} |\xi(\hat{\zeta}+\theta) - \xi(\hat{\zeta})| \Big).\end{aligned}$$

Thus, we conclude

$$\begin{aligned}&\int_0^{2\pi} \phi(\Delta(a_t r_\theta) x_0) \xi(\theta)\,\mathrm{d}\theta \\ &\qquad = \frac{1}{\zeta} \int_{\hat{\zeta}=0}^{2\pi} \left(\int_0^{\zeta} \phi(\Delta(a_t r_{\hat{\zeta}+\theta}) x_0)\,\mathrm{d}\theta \right) \xi(\hat{\zeta})\,\mathrm{d}\hat{\zeta} + O(\mathcal{S}(\phi)\mathcal{S}(\xi)\zeta).\end{aligned} \tag{5.8}$$

Furthermore, by (5.3) and (5.4), we have

$$\frac{1}{\zeta} \int_0^{\zeta} \phi(\Delta(a_t r_{\hat{\zeta}+\theta}) x_0)\,\mathrm{d}\theta = \int_0^1 \phi(\Delta(a_{\log T} u_s) x_{\hat{\zeta}})\,\mathrm{d}s + O(\mathcal{S}(\phi)\zeta). \tag{5.9}$$

Altogether, using (5.6), (5.8), and (5.9), we conclude that

$$\begin{aligned}&\left| \int_0^{2\pi} \phi(\Delta(a_t\, r_\theta) x_0) \xi(\theta)\,\mathrm{d}\theta - \int_0^{2\pi} \xi(\theta)\,\mathrm{d}\theta \int \phi\,\mathrm{d}m_X \right| \\ &\qquad\qquad \le \mathcal{S}(\phi)\mathcal{S}(\xi) R^{-\kappa_0},\end{aligned} \tag{5.10}$$

where $\kappa_0 = \min\{\kappa_1, 2/D''\}$—we used $\zeta^{-1} = e^{2\tau} = R^{2/D''}$. Thus, part 1 in the proposition holds if case 1 holds.

Let us now assume that case 2 holds and let $x_{\hat{\zeta}} = \Delta(a_\tau r_{\hat{\zeta}}) x_0$ be as in (5.7). Then by (5.1), we have

$$\begin{aligned}d_X(\Delta(a_\tau r_{\hat{\zeta}})^{-1} x, x_0) &\le e^{D'\tau} R^{D_1} (\log T)^{D_1} T^{-1} \\ &\le e^{(1+D')\tau} R^{D_1} t^{D_1} e^{-t} \le R^D t^D e^{-t}.\end{aligned}$$

Furthermore, $\Delta(a_\tau r_{\hat{\zeta}})^{-1}x$ has a periodic H-orbit of volume $\leq R$. Thus part 2 in the proposition holds in this case. The proof is complete. □

6 Cusp functions of Margulis and the upper bound

In this section, we put

$$\Gamma = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \subset G.$$

Recall the following definition.

Definition 2.3. Let $g = (g_1, g_2) \in G$. A two-dimensional $g\mathbb{Z}^4$-rational linear subspace $L \subset \mathbb{R}^4$ is called (ρ, A, t)-exceptional if there are $(v_1, 0), (0, v_2) \in \mathbb{Z}^4$ satisfying

$$\|g_1 v_1\|, \|g_2 v_2\| \leq e^{\rho t} \quad \text{and} \quad |Q_0(g_1 v_1, g_2 v_2)| \leq e^{-A\rho t} \tag{6.1}$$

so that $L \cap g\mathbb{Z}^4$ is spanned by $\{(g_1 v_1, 0), (0, g_2 v_2)\}$.

Given a (ρ, A, t)-special subspace L, we will refer to $\{(g_1 v_1, 0), (0, g_2 v_2)\}$ as a **spanning set** for L.

Let $f_i \in C_c(\mathbb{R}^2)$, and define f on $\mathbb{R}^4$ by $f(w_1, w_2) = f_1(w_1)f_2(w_2)$. For every $h \in \mathrm{SL}_2(\mathbb{R})$, let

$$\tilde{f}_{\rho,A,t}(h; g\Gamma) = \sum_{v \in \mathcal{N}_t(g\mathbb{Z}^4)} f(\Delta(h)v), \tag{6.2}$$

where $\mathcal{N}_t(g\mathbb{Z}^4)$ denotes the set of vectors in $g\mathbb{Z}^4$ not contained in any (ρ, A, t)-special subspace L and also not contained in $\mathbb{R}^2 \times \{0\} \cup \{0\} \cup \mathbb{R}^2$. In the sequel, we will often drop the dependence on A, ρ, and t from the notation and denote $\tilde{f}_{\rho,A,t}(h; g\Gamma)$ by $\tilde{f}(h; g\Gamma)$.

The following is one of the main results of this section.

6.1 Proposition. *For all $A_1 \geq 10^3$ we have the following: Let $(g_1, g_2) \in G$. Then for all small enough ρ and all large enough t at least one of the following holds:*

(1) *Let $\mathcal{C}_t = \{\theta \in [0, 2\pi] : \tilde{f}(a_t r_\theta; g\Gamma) \geq e^{A_1 \rho t}\}$. Then*

$$\int_{\mathcal{C}_t} \tilde{f}(a_t r_\theta; g\Gamma)\, \mathrm{d}\theta \ll e^{-\rho^3 t/A_1},$$

where $\tilde{f}(h; g\Gamma) = \tilde{f}_{\rho,A_1,t}(h; g\Gamma)$, see (6.2).

(2) *There exists $Q \in \mathrm{Mat}_2(\mathbb{Z})$ whose entries are bounded by $e^{100\rho t}$ and $\lambda \in \mathbb{R}$ satisfying $\|g_2^{-1} g_1 - \lambda Q\| \ll e^{-(A_1 - 100)\rho t}$.*

The implied constants depend polynomially on $\|g_1\|$ and $\|g_2\|$.

The proof of this proposition occupies most of this section.

The cusp functions. Let $\mathcal{P}$ denote the set of primitive vectors in $\mathbb{Z}^2$. For any $h \in \mathrm{SL}_2(\mathbb{R})$, define

$$\omega(h\mathrm{SL}_2(\mathbb{Z})) = \sup\{1/\|hv\| : v \in \mathcal{P}\}. \tag{6.3}$$

We begin with the following lemma.

6.2 Lemma (cf. Lemma 7.4 [EM01]). *For every $0 < p < 2$, there exists t_p and b_p so that the following holds. For every $x \in \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ and all $t \geq t_p$, we have*

$$\int_0^{2\pi} \omega(a_t r_\theta x)^p \, \mathrm{d}\theta \leq 2^{-t/t_p}\omega(x)^p + b_p.$$

Proof. This is well known by now, see, e.g., [EM22]. □

The sets $\Theta_t(\delta)$ and $\Theta_t'(\delta)$. To put an emphasis on the product structure of G and X, we will often write $X = G_1/\Gamma_1 \times G_2/\Gamma_2$ where $G_i = \mathrm{SL}_2(\mathbb{R})$ and $\Gamma_i = \mathrm{SL}_2(\mathbb{Z})$. Moreover, given $g = (g_1, g_2) \in G$, we write

$$\omega_i(g_i\Gamma_i) := \omega(g_i\mathrm{SL}_2(\mathbb{Z})). \tag{6.4}$$

For $i = 1, 2$, let $x_i \in G_i/\Gamma_i$. For all $t \geq 0$ and every $0 < \delta \leq 1/10$, let

$$\Theta_t(\delta) = \{\theta \in [0, 2\pi] : \omega_2(a_t r_\theta x_2)^{1-2\delta} \leq \omega_1(a_t r_\theta x_1) \leq \omega_2(a_t r_\theta x_2)^{1+2\delta}\} \tag{6.5}$$

and let $\Theta_t'(\delta) = [0, 2\pi] \setminus \Theta_t(\delta)$.

We have the following:

6.3 Lemma. *Let $0 < \delta < 1/10$, and put*

$$p_1 = (2 - 2\delta)(1 + \frac{1}{2}\delta) \quad \textit{and} \quad p_2 = \frac{(2 + 2\delta)(1 + \frac{1}{2}\delta)}{1 + 2\delta};$$

note that $p_1, p_2 < 2$. Let $t(\delta) = \max(t_{p_1}, t_{p_2})$ and $b(\delta) = \max(b_{p_1}, b_{p_2})$ where the notation is as in Lemma 6.2. Then for all $(x_1, x_2) \in X$ and all $t \geq t(\delta)$

$$\int_{\Theta_t'(\delta)} (\omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2))^{1+\frac{1}{2}\delta} \, \mathrm{d}\theta \leq 2^{-t/t(\delta)}(\omega_1(x_1) + \omega_2(x_2)) + 2b(\delta).$$

Proof. Let us write $\Theta_t'(\delta) = \Theta_{t,1}'(\delta) \cup \Theta_{t,2}'(\delta)$, where

$$\Theta_{t,1}'(\delta) = \{\theta \in [0, 2\pi] : \omega_2(a_t r_\theta x_1) < \omega_1(a_t r_\theta x_2)^{1-2\delta}\}$$
$$\Theta_{t,2}'(\delta) = \{\theta \in [0, 2\pi] : \omega_2(a_t r_\theta x_1) > \omega_1(a_t r_\theta x_2)^{1+2\delta}\}.$$

Using Lemma 6.2, for every $t > t_{p_1}$ we have

$$\begin{aligned}\int_{\Theta'_{t,1}(\delta)} (\omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2))^{1+\frac{1}{2}\delta}\,\mathrm{d}\theta &\le \int_0^{2\pi} \omega_1(a_t r_\theta x_1)^{p_1}\,\mathrm{d}\theta \\ &\le 2^{-t/t_{p_1}}\omega_2(x_2) + b_{p_1}.\end{aligned}$$

Similarly, for every $t > t_{p_2}$, we have

$$\begin{aligned}\int_{\Theta'_{t,2}(\delta)} (\omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2))^{1+\frac{1}{2}\delta}\,\mathrm{d}\theta &\le \int_0^{2\pi} \omega_2(a_t r_\theta x_2)^{p_2}\,\mathrm{d}\theta \\ &\le 2^{-t/t_{p_2}}\omega_1(x_1) + b_{p_2}.\end{aligned}$$

The claim follows from these two estimates. □

A Diophantine condition. The following lemma is a crucial input in the proof of Proposition 6.1.

For every $t \ge 1$, let

$$\begin{aligned}\mathcal{P}_t &= \{v \in \mathcal{P} : e^{t-1} \le \|v\| < e^t\},\\ \mathcal{P}(t) &= \{v \in \mathcal{P} : \|v\| < e^t\}.\end{aligned}$$

6.4 Lemma. *The following holds for all $A \ge 10^3$ and all $\rho \le 1/(100A)$. Let $(g_1, g_2) \in G$. There exist $t_1 \ge 1$, depending on ρ and polynomially on $\|g_i\|$, so that if $t \ge t_1$, then at least one of the following holds:*

(1) *We have*

$$\#\{v_1 \in \mathcal{P}_t : \exists v_2 \in \mathcal{P}(t), |Q_0(g_1 v_1, g_2 v_2)| \le e^{-A\rho t}\} \ll e^{(2-\rho)t}$$

where the implied constant depends polynomially on $\|g_i\|$.

(2) *There exist $Q \in \mathrm{Mat}_2(\mathbb{Z})$ whose entries are bounded by $e^{100\rho t}$ and $\lambda \in \mathbb{R}$ satisfying $\|g_2^{-1}g_1 - \lambda Q\| \le e^{-(A-100)\rho t}$.*

Proof. For simplicity in the notation, let us write $\eta = e^{-\rho t}$. Let $A \ge 10^3$, and assume that

$$\begin{aligned}\#\{v_1 \in \mathcal{P}_t : \exists v_2 \in \mathcal{P}(t), |Q_0(g_1 v_1, g_2 v_2)| \le \eta^A\}&\\ > E(\|g_1\|\|g_2\|)^E \eta e^{2t}.&\end{aligned} \tag{6.6}$$

We will show that if E is large enough, then part (2) holds.

Let us write

$$h := g_2^{-1}g_1 = \begin{pmatrix} a & b \\ c & d\end{pmatrix}.$$

Then (6.6) and the fact that for any $q \in \mathrm{SL}(2,\mathbb{R})$, $\Delta(q) \in \mathrm{SO}(Q_0)$ implies that if t is large enough, depending on $\|h\|$, for $\gg \eta e^{2t}$ many $v_1 = (\mathsf{x}_1, \mathsf{y}_1) \in \mathcal{P}_t$ both of the following hold:

- We have $|c\mathsf{x}_1 + d\mathsf{y}_1| \geq \eta^2 e^t$.
- There exists at least one $(\mathsf{x}_2, \mathsf{y}_2) \in \mathcal{P}(t)$ so that

$$|Q_0(h(\mathsf{x}_1, \mathsf{y}_1), (\mathsf{x}_2, \mathsf{y}_2))| \leq \eta^A. \tag{6.7}$$

Moreover, the fact that there are $\gg \eta e^{2t}$ vectors satisfying these two conditions implies that there are $v_1, v_1', v_1'' \in \mathcal{P}_t$ satisfying the above two conditions so that

$$1 \leq |Q_0(v, w)| \ll \eta^{-4}, \quad \text{for } v, w \in \{v_1, v_1', v_1''\}. \tag{6.8}$$

Let us fix three vectors v_1, v_1', v_1'' satisfying (6.8), and let v_2, v_2', v_2'' be the corresponding vectors in $\mathcal{P}(t)$ satisfying (6.7), respectively. Then

$$hv_1 = \mu v_2 + w_{1,2} \tag{6.9}$$

where $\mu \in \mathbb{R}$ satisfies $|\mu| \asymp 1$ and $\|w_{1,2}\| \ll \eta^A e^{-t}$ (recall that the implicit constants in these inequalities are allowed to depend polynomially on $\|h\|$). Similarly,

$$hv_1' = \mu' v_2' + w_{1,2}' \quad \text{and} \quad hv_1'' = \mu'' v_2'' + w_{1,2}''$$

where $\mu', \mu'' \in \mathbb{R}$ satisfy $|\mu'|, |\mu''| \asymp 1$ and $\|w_{1,2}'\|, \|w_{1,2}''\| \ll \eta^A e^{-t}$.

With this notation we have

$$h(v_1\ v_1') = (v_2\ v_2') \begin{pmatrix} \mu & 0 \\ 0 & \mu' \end{pmatrix} + O(\eta^A e^{-t}) \tag{6.10}$$

and similarly for v_1, v_1'' and v_1', v_1''. Thus by (6.8)

$$1 \leq |Q_0(v_2, v_2')|, |Q_0(v_2, v_2'')|, |Q_0(v_2', v_2'')| \ll \eta^{-4}. \tag{6.11}$$

In view of (6.8), (6.9), (6.10) and (6.11) the conditions in Lemma 2.2 hold. The claim thus follows from Lemma 2.2 so long as t is large enough to account for the constant C in that lemma. □

Proof of Proposition 6.1. Recall that $g = (g_1, g_2)$. Put

$$x_i = g_i \mathrm{SL}_2(\mathbb{Z}), \quad \text{for } i = 1, 2.$$

Let $A_1 \geq 10^4$, $0 < \rho < 10^{-4}$ (small), and $t \geq 1$ (large) be so that Lemma 6.4 holds for these choices. Put $\delta = 2\rho^2/A_1$, and define $\Theta_t(\delta)$ and $\Theta_t'(\delta)$ as in (6.5) with t and δ and x_i. That is,

$$\Theta_t(\delta) = \{\theta \in [0, 2\pi] : \omega_2(a_t r_\theta x_2)^{1-2\delta} \leq \omega_1(a_t r_\theta x_1) \leq \omega_2(a_t r_\theta x_2)^{1+2\delta}\},$$

and $\Theta_t'(\delta) = [0, 2\pi] \setminus \Theta_t(\delta)$.

Apply Lemma 6.4 with $A = A_1$ and ρ. If part (2) in that lemma holds, then part (2) in Proposition 6.1 holds and the proof is complete. Thus, assume for the rest of the argument that part (1) in Lemma 6.4 holds. We will show that part (1) in the Proposition 6.1 holds.

Motivated by the definition of $\tilde{f}$ and Lemma 2.4, define

$$\tilde{\omega}(a_t r_\theta; g\Gamma) = \sup\{(\|a_t r_\theta g_1 v_1\| \|a_t r_\theta g_2 v_2\|)^{-1} : (v_1, v_2) \in \mathcal{P}^2(g)\} \tag{6.12}$$

where $\mathcal{P}$ is the set of primitive vectors in $\mathbb{Z}^2$ and $\mathcal{P}^2(g)$ denotes the set of $(v_1, v_2) \in \mathcal{P}^2$ so that $\{(g_1 v_1, 0), (0, g_2 v_2)\}$ is not a spanning set for any (ρ, A_1, t)-special subspace of $g\mathbb{Z}^4$, see Definition 2.3.

It follows from the definition that

$$\tilde{\omega}(a_t r_\theta; g\Gamma) \le \omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2). \tag{6.13}$$

Put $\mathcal{B}_t = \{\theta \in [0, 2\pi] : \tilde{\omega}(a_t r_\theta; g\Gamma) < \omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2)\}$.

By a variant of Schmidt's Lemma, see also [EMM98, Lemma 3.1], and the definition of $\tilde{f}$, we have

$$\tilde{f}(a_t r_\theta; g\Gamma) \ll \tilde{\omega}(a_t r_\theta; g\Gamma). \tag{6.14}$$

Put $\tilde{\mathcal{C}}_t = \{\theta \in [0, 2\pi] : \tilde{\omega}(a_t r_\theta; g\Gamma) \ge e^{A_1 \rho t}\}$. In view of (6.14) and with this notation, it suffices to show that

$$\int_{\tilde{\mathcal{C}}_t} \tilde{\omega}(a_t r_\theta; g\Gamma)\, \mathrm{d}\theta \ll e^{-\rho^2 t/A_1}. \tag{6.15}$$

Contribution of $\mathcal{B}_t$. Recall that if $\omega(h\mathrm{SL}_2(\mathbb{Z})) \ge 2$ for some $h \in \mathrm{SL}_2(\mathbb{R})$, then there is some $v_h \in \mathcal{P}$ so that

$$\|hv_h\|^{-1} = \omega(h\mathrm{SL}_2(\mathbb{Z})) \quad \text{and} \quad \|hv\| > 1/2 \text{ for all } v_h \ne v \in \mathcal{P}. \tag{6.16}$$

Let $\theta \in \mathcal{B}_t$. By the definition of $\tilde{\omega}$, there exist $v_1, v_2 \in \mathcal{P}$ so that

$$\tilde{\omega}(a_t r_\theta; g\Gamma) = \|a_t r_\theta g_1 v_1\|^{-1} \|a_t r_\theta g_2 v_2\|^{-1}.$$

Since $\tilde{\omega}(a_t r_\theta; g\Gamma) < \omega_1(a_t r_\theta g_1 \Gamma_1)\omega_2(a_t r_\theta g_2 \Gamma_2)$, we conclude that

$$\min\{\|a_t r_\theta g_1 v_1\|^{-1}, \|a_t r_\theta g_2 v_2\|^{-1}\} \le 2.$$

Therefore, for all such θ, we have

$$\tilde{\omega}(a_t r_\theta; g\Gamma) \le 2\max\{\omega_1(a_t r_\theta g_1 \Gamma_1), \omega_2(a_t r_\theta g_2 \Gamma_2)\}.$$

Thus using Lemma 6.2, we have

$$\begin{aligned}\int_{\mathcal{B}_t\cap\tilde{\mathcal{C}}_t} \tilde{\omega}(a_t r_\theta; g\Gamma)\,\mathrm{d}\theta &\le e^{-\frac{A_1\rho t}{2}} \int_{\mathcal{B}_t} \tilde{\omega}(a_t r_\theta; g\Gamma)^{3/2}\,\mathrm{d}\theta \\ &\le 2e^{-\frac{A_1\rho t}{2}} \int_0^{2\pi} \omega_1(a_t r_\theta x_1)^{\frac{3}{2}} + \omega_1(a_t r_\theta x_2)^{\frac{3}{2}}\,\mathrm{d}\theta \\ &\ll e^{-\frac{A_1\rho t}{2}}.\end{aligned} \tag{6.17}$$

Let $\Theta_t(\theta)$ and $\Theta'_t(\delta)$ be as above, and put

$$\tilde{\mathcal{C}}_t(\delta) := \tilde{\mathcal{C}}_t \cap \mathcal{B}_t^{\complement} \cap \Theta_t(\delta) \quad \text{and} \quad \tilde{\mathcal{C}}'_t(\delta) := \tilde{\mathcal{C}}_t \cap \mathcal{B}_t^{\complement} \cap \Theta'_t(\delta).$$

We consider the contribution of these two sets to $\int \tilde{\omega}$ separately—indeed, controling the contribution of $\tilde{\mathcal{C}}_t(\delta)$ occupies the bulk of the proof.

Contribution of $\tilde{\mathcal{C}}'_t(\delta)$. By Lemma 6.3, for all t large enough, we have

$$\int_{\Theta'_t(\delta)} (\omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2))^{1+\frac{1}{2}\delta}\,\mathrm{d}\theta \ll 1.$$

From this and (6.13), we conclude that

$$\begin{aligned}\int_{\tilde{\mathcal{C}}'_t(\delta)} \tilde{\omega}(a_t r_\theta; g\Gamma)\,\mathrm{d}\theta &\le \int_{\tilde{\mathcal{C}}'_t(\delta)} \omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2)\,\mathrm{d}\theta \\ &\le e^{-\delta\rho A_1 t/2} \int_{\Theta'_t(\delta)} (\omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2))^{1+\frac{1}{2}\delta}\,\mathrm{d}\theta \\ &\ll e^{-\rho^3 t}.\end{aligned} \tag{6.18}$$

Contribution of $\tilde{\mathcal{C}}_t(\delta)$. Recall that

$$\Theta_t(\delta) = \{\theta \in [0, 2\pi] : \omega_2(a_t r_\theta x_2)^{1-2\delta} \le \omega_1(a_t r_\theta x_1) \le \omega_2(a_t r_\theta x_2)^{1+2\delta}\},$$

and $\tilde{\mathcal{C}}_t(\delta) = \tilde{\mathcal{C}}_t \cap \mathcal{B}_t^{\complement} \cap \Theta_t(\delta)$. Note that the vectors which contribute to

$$\int_{\tilde{\mathcal{C}}_t(\delta)} \tilde{\omega}(a_t r_\theta; g\Gamma)\,\mathrm{d}\theta \tag{6.19}$$

satisfy $\{(g_1 v_1, g_2 v_2) : \|g_1 v_1\|, \|g_2 v_2\| \le e^t\}$. It is more convenient to consider the cases $\|g_1 v_1\| \ge \|g_2 v_2\|$ and $\|g_1 v_1\| \le \|g_2 v_2\|$ separately. As the arguments are similar in both cases, we assume $\|g_1 v_1\| \ge \|g_2 v_2\|$ for the rest of the proof.

Recall our notation: for $t \ge 1$

$$\mathcal{P}_t = \{v \in \mathcal{P} : e^{t-1} \le \|v\| < e^t\},$$

and $\mathcal{P}(t) = \{v \in \mathcal{P} : \|v\| \le e^t\}$.

For every $n \in \mathbb{N}$ with $n \le t + \log \|g_1\| + 1 =: t_1$, we investigate the contribution of $\mathcal{P}_n$ to (6.19). For any $v_1 \in \mathcal{P}_n$, let

$$I_{v_1} = \{\theta \in [0, 2\pi] : \|a_t r_\theta g_1 v_1\| \le 1/10\}.$$

Then the intervals I_{v_1} are disjoint. Let $\tilde{\mathcal{P}}_n = \{v_1 \in \mathcal{P}_n : I_{v_1} \cap \tilde{\mathcal{C}}_t(\delta) \ne \emptyset\}$.

Fix some $n \in \mathbb{N}$, $n \le t_1$. Let $v_1 \in \tilde{\mathcal{P}}_n$, and let $\theta \in I_{v_1} \cap \tilde{\mathcal{C}}_t(\delta)$. Then there exists $v_2 \in \mathcal{P}$ so that

$$\tilde{\omega}(a_t r_\theta; g\Gamma) = \frac{1}{\|a_t r_\theta g_1 v_1\| \, \|a_t r_\theta g_2 v_2\|}.$$

Since $\theta \in \mathcal{B}_t$, we have $\tilde{\omega}(a_t r_\theta; g\Gamma) = \omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2)$. Thus

$$\omega_i(a_t r_\theta x_i) = \|a_t r_\theta g_i v_i\|^{-1} \quad \text{for } i = 1, 2. \tag{6.20}$$

In view of (6.20), and the definitions of $\mathcal{B}_t$ and $\Theta_t(\theta)$, thus

$$\int_{\tilde{\mathcal{C}}_t(\delta)} \tilde{\omega}(a_t r_\theta; g\Gamma)\, \mathrm{d}\theta \le \sum_n \sum_{\tilde{\mathcal{P}}_n} \int_{I_{v_1}} \|a_t r_\theta g_1 v_1\|^{-2-2\delta}. \tag{6.21}$$

We also make some observations. Fix some $n \in \mathbb{N}$, $n \le t_1$. Let $v_1 \in \tilde{\mathcal{P}}_n$ and $\theta \in I_{v_1} \cap \tilde{\mathcal{C}}_t(\delta)$, and let $v_2 \in \mathcal{P}$ be so that (6.20) holds. That is,

$$\omega_i(a_t r_\theta x_i) = \|a_t r_\theta g_i v_i\|^{-1},$$

for $i = 1, 2$, and

$$\tilde{\omega}(a_t r_\theta; g\Gamma) = (\|a_t r_\theta g_1 v_1\| \, \|a_t r_\theta g_2 v_2\|)^{-1}.$$

Since $\theta \in \tilde{\mathcal{C}}_t$, we have $\tilde{\omega}(a_t r_\theta; g\Gamma) \ge e^{A_1 \rho t}$. This gives

$$\|a_t r_\theta g_1 v_1\| \, \|a_t r_\theta g_2 v_2\| \le e^{-A_1 \rho t},$$

which implies that

$$|Q_0(\Delta(a_t r_\theta)(g_1 v_1, g_2 v_2))| = |Q_0(a_t r_\theta g_1 v_1, a_t r_\theta g_2 v_2)| \le e^{-A_1 \rho t}.$$

Since $\Delta(a_t r_\theta) \in \mathrm{SO}(Q_0)$, we conclude from the above that

$$Q_0(g_1 v_1, g_2 v_2) \le e^{-A_1 \rho t}. \tag{6.22}$$

We claim:

$$\|g_1 v_1\| \ge e^{\rho t}. \tag{6.23}$$

Indeed if $\|g_1 v_1\| < e^{\rho t}$, then since $\|g_2 v_2\| \le \|g_1 v_1\|$, it follows from (6.22) that $\{(g_1 v_1, 0), (0, g_2 v_2)\}$ spans a (ρ, A_1, t)-special subspace. This contradicts the definition of $\tilde{\omega}$ and establishes (6.23).

Let us now return to estimating (6.21); we will estimate the sum on the right side of (6.21) using the following elementary fact.

Sublemma. *Let $t > 0$, and let $w \in \mathbb{R}^2$ be a non-zero vector. Then*

$$\int_0^{2\pi} \|a_t r_\theta w\|^{-2-2\delta}\, \mathrm{d}\theta \le \hat{C} e^{4\delta t} \|w\|^{-2-2\delta}$$

where $\hat{C}$ is absolute.

First note that (6.22), and the fact that part 1 in Lemma 6.4 holds, imply that there exist t_0 and C so that for all $t_0 \le n \le t_1$, we have

$$\#\tilde{\mathcal{P}}_n \le C e^{(2-\rho)n}. \tag{6.24}$$

Also recall from (6.23) that $\|g_1 v_1\| \ge e^{\rho t}$, which in particular implies that $\|v_1\| \gg e^{\rho t}$. Since $v_1 \in \mathcal{P}_n$, we conclude that $n \ge \rho t + O(1)$. Thus (6.24) and the Sublemma imply that

$$\begin{aligned} \sum_{v_1 \in \tilde{\mathcal{P}}_n} \int_{I_{v_1}} \|a_t r_\theta g_1 v_1\|^{-2-2\delta}\, \mathrm{d}\theta &\ll e^{(2-\rho)n} e^{4\delta t} e^{(-2-2\delta)n} \\ &\ll e^{-\rho^2 t} e^{4\delta t} \le e^{-2\delta t}; \end{aligned} \tag{6.25}$$

in the last inequality, we used $\rho^2 = A_1\delta/2 \ge 100\delta$ and assumed t is large.

We now sum over all $n \le t_1$ and get that

$$\sum_n \sum_{\tilde{\mathcal{P}}_n} \int_{I_{v_1}} \|a_t r_\theta g_1 v_1\|^{-2-2\delta} \ll t e^{-2\delta t} \ll e^{-\delta t}.$$

This and (6.21) complete the proof in this case.

In combination with (6.18) and (6.17), the proof is complete. □

Proof of the Sublemma. Without loss of generality, we may assume $w = (0, 1)$. Put

$$I = [e^{(-2+2\delta)t}, 2\pi - e^{(-2+2\delta)t}] \quad \text{and} \quad I' = [0, 2\pi] \setminus I.$$

Then

$$\begin{aligned} \int_0^{2\pi} \frac{\mathrm{d}\theta}{\|a_t r_\theta w\|^{2+2\delta}} &\ll \int_{I'} \frac{\mathrm{d}\theta}{\|a_t r_\theta w\|^{2+2\delta}} + \int_I \frac{\mathrm{d}\theta}{\|a_t r_\theta w\|^{2+2\delta}} \\ &\ll e^{(-2+2\delta)t} e^{(2+2\delta)t} + \int_I \frac{\mathrm{d}\theta}{\|a_t r_\theta w\|^{2+2\delta}} \\ &\le e^{4\delta t} + \int_I \frac{\mathrm{d}\theta}{\|a_t r_\theta w\|^{2+2\delta}}. \end{aligned}$$

We now compute the integral over I. Note that $\|a_t r_\theta w\|^{2+2\delta} \gg e^{(2+2\delta)t}\theta^{2+2\theta}$. Therefore,

$$\int_I \frac{\mathrm{d}\theta}{\|a_t r_\theta w\|^{2+2\delta}} \ll e^{(-2-2\delta)t}\int_I \theta^{-2-2\delta}\,\mathrm{d}\theta$$
$$\ll e^{(-2-2\delta)t}e^{(1+2\delta)(2-2\delta)t} \ll e^{-4\delta^2 t}.$$

The proof is complete. □

We end this section with the proof of Lemma 3.2.

Proof of Lemma 3.2. We begin with part (1). Recall that f_i is the characteristic function of $\{w \in \mathbb{R}^2 : \|w\| \le R\}$, and let $f = f_1 f_2$. Again by a variant of Schmidt's Lemma, we have

$$\hat{f}(\Delta(a_t r_\theta)g\Gamma') \le \omega_1(g_1\mathrm{SL}_2(\mathbb{Z}))\omega_2(g_2\mathrm{SL}_2(\mathbb{Z}))$$

Let $\delta = \eta/10$. As it was done in (6.5), define

$$\Theta_t(\delta) = \{\theta \in [0, 2\pi] : \omega_2(a_t r_\theta x_2)^{1-2\delta} \le \omega_1(a_t r_\theta x_1) \le \omega_2(a_t r_\theta x_2)^{1+2\delta}\}$$

and let $\Theta'_t(\delta) = [0, 2\pi] \setminus \Theta_t(\delta)$ where $x_i = g_i\mathrm{SL}_2(\mathbb{Z})$. Then by Lemma 6.3, we have for all $t \ge t(\delta)$

$$\int_{\Theta'_t(\delta)} \hat{f}(\Delta(a_t r_\theta)g\Gamma')\,\mathrm{d}\theta \le \int_{\Theta'_t(\delta)} (\omega_1(a_t r_\theta x_1)\omega_2(a_t r_\theta x_2))\,\mathrm{d}\theta \ll 1; \tag{6.26}$$

the implied constant depends polynomially on the injectivity radius of $g\Gamma'$.

We now find an upper bound for the integral over $\Theta_t(\delta)$:

$$\int_{\Theta_t(\delta)} \hat{f}(\Delta(a_t r_\theta)g\Gamma')\,\mathrm{d}\theta \le \int \omega_1(a_t r_\theta x_1)^{2+2\delta}\,\mathrm{d}\theta.$$

This, the sublemma, and standard arguments (which simplify significantly thanks to (6.16)), see, e.g., [EM22], imply that

$$\int_{\Theta_t(\delta)} \hat{f}(\Delta(a_t r_\theta)g\Gamma')\,\mathrm{d}\theta \ll e^{4\delta t}.$$

The claim in part (1) of the lemma follows.

We now turn to the proof of part (2). Let $(v_1, 0)$ and $(0, v_2)$ be as in the statement. For $i = 1, 2$ let $w_i = g_i v_i$. By a variant of Schmidt's Lemma,

$$\hat{f}(\theta) \le \|a_t r_\theta w_1\|^{-1}\|a_t r_\theta w_2\|^{-1}. \tag{6.27}$$

For $i = 1, 2$, set

$$I_i = \{\theta : R^{-1}e^{-\eta t}/10 \le \|a_t r_\theta w_i\|\}.$$

If $\theta \notin I_1 \cap I_2$, then $\hat{f}(\theta) > e^{\eta t}$. This, (6.27), and the definition of $\mathcal{C}_L$ imply

$$\int_{\mathcal{C}_L} \hat{f}(\theta) \leq \int_{I_1 \cap I_2} \frac{1}{\|a_t r_\theta w_1\| \|a_t r_\theta w_2\|}.$$

Thus, using Cauchy-Schwarz inequality, we need to find an upper bound for

$$\left(\int_{I_1} \frac{\mathrm{d}\theta}{\|a_t r_\theta w_1\|^2} \right)^{1/2} \left(\int_{I_2} \frac{\mathrm{d}\theta}{\|a_t r_\theta w_2\|^2} \right)^{1/2}.$$

The computation is similar to the one in the proof of the sublemma. Indeed, we may assume $w_i = (0, 1)$; then there is $R^{-1} \ll c < 1$ so that

$$I_i \subset [ce^{-(1+\eta)t}, 2\pi - ce^{-(1+\eta)t}].$$

From this, we conclude that

$$\int_{I_i} \frac{\mathrm{d}\theta}{\|a_t r_\theta w_i\|^2} \ll e^{(-1+\eta)t},$$

as claimed. □

7 Proof of Theorem 3.1

In this section, we will prove Theorem 3.1. The proof combines a lower bound estimate, which will be proved using Theorem 1.4, with an upper bound estimate, which follows from Proposition 6.1, as we now explicate.

Proof of Theorem 3.1. Recall that $f_i \in C_c^\infty(\mathbb{R}^2)$, and f is defined on $\mathbb{R}^4$ by $f(w_1, w_2) = f_1(w_1) f_2(w_2)$. We put

$$\hat{f}(g'\Gamma') = \sum_{v \in g'\Lambda_{\mathrm{nz}}} f(v) \tag{7.1}$$

where $\Lambda = \{(v_1 + v_2, \omega(v_1 - v_2)) : v_1, v_2 \in \mathbb{Z}^2\} \subset \mathbb{R}^4$,

$$\Gamma' = \{(\gamma_1, \gamma_2) \in \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) : \gamma_1 \equiv \omega\gamma_2\omega \pmod 2\}$$

stabilizes Λ, and $g' = (g_1', g_2') \in G$. We also put $X = G/\Gamma'$.

Let A and ρ be as in the statement, and let $t > 0$ be a parameter which is assumed to be large. Let $\hat{A}$ be a constant which will be explicated later, and let $g = (g_1, g_2) \in G$ satisfy the following: for every $Q \in \mathrm{Mat}_2(\mathbb{Z})$ with $e^{\rho t/\hat{A}} \leq \|Q\| \leq e^{\rho t}$ and all $\lambda \in \mathbb{R}$ we have

$$\|g_2^{-1} g_1 - \lambda Q\| > \|Q\|^{-A/1000}. \tag{7.2}$$

We claim that (7.2) implies the following:

Sublemma. *Let $g = (g_1, g_2)$ satisfy (7.2). There exists $A_1 \geq \max(4D, A)$, where D is as in Theorem 1.4 so that the following holds. For all t so that $t > 4D \log t$ and for every $x \in X$ with $\mathrm{vol}(Hx) \leq e^{\rho t/A_1}$, we have*

$$d(g\Gamma', x) > e^{-t/2}.$$

We first assume the sublemma and complete the proof of the theorem. In view of the sublemma, part (1) in Theorem 1.4 holds with $R = e^{\rho t/A_1}$ and t. Indeed, $D\rho/A_1 \leq 1/4$ and $t^D \leq e^{t/4}$, which imply

$$R^D t^D e^{-t} = e^{D\rho t/A_1} t^D e^{-t} \leq e^{-t/2};$$

hence, part (2) in Theorem 1.4 cannot hold.

For every S, let $1_{X_S} \leq \varphi_S \leq 1_{X_{S+1}}$ be a smooth function with $\mathcal{S}(\varphi_S) \ll S^\star$, where

$$X_\bullet = \{x = (x_1, x_2) \in X : \max(\omega_1(x_1), \omega_2(x_2)) \leq \bullet\},$$

see (6.4)—since Γ' is a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ this is well-defined. Put $\hat{f}_S = \varphi_S \hat{f}$; we let N be so that $\mathcal{S}(\hat{f}_S) \ll S^N \mathcal{S}(f)$.

Put $\eta = \kappa_0\rho/(2NA_1)$, where κ_0 is as in Theorem 1.4. We will show the claim in the theorem holds with

$$\hat{A} = 3NAA_1/\kappa_0, \quad \delta_1 = \eta, \quad \text{and} \quad \delta_2 = \eta^3/A^3.$$

First note that

$$\rho/\hat{A} = \kappa_0\rho/(3NAA_1') \leq \eta/A = \delta_1/A \leq \rho/100. \tag{7.3}$$

We now turn to the rest of the argument. Apply Lemma 2.4 with (g_1, g_2) and the triple $(\eta/A, A, t)$. In view of (7.3) and (7.2), Lemma 2.4 implies that there are at most two $(\eta/A, A, t)$-special subspaces.

Denote these subspaces by L and L' if they exist. For every $\theta \in [0, 2\pi]$, we write

$$\hat{f}(\Delta(a_t r_\theta)g\Gamma') = \hat{f}_S(\Delta(a_t r_\theta)g\Gamma') + \hat{f}_{\mathrm{cusp}}(\Delta(a_t r_\theta)g\Gamma') + \hat{f}_{\mathrm{sp}}(\Delta(a_t r_\theta)g\Gamma')$$

where $\hat{f}_S = \varphi_S\hat{f}$, $\hat{f}_{\mathrm{cusp}}$ is the contribution of $g\Lambda_{\mathrm{nz}} \setminus (L \cup L')$ to $\hat{f} - \hat{f}_S$, and $\hat{f}_{\mathrm{sp}}$ is the contribution of $g\Lambda_{\mathrm{nz}} \cap (L \cup L')$ to $\hat{f} - \hat{f}_S$.

By Theorem 1.4, applied with $R = e^{\rho t/A'}$, for any smooth function ξ on $[0, 2\pi]$ we have

$$\left| \int_0^{2\pi} \hat{f}_S(\Delta(a_t r_\theta)g\Gamma')\xi(\theta)\,\mathrm{d}\theta - \int_0^{2\pi} \xi\,\mathrm{d}\theta \int_X \hat{f}_S\,\mathrm{d}m_X \right| \\ \ll \mathcal{S}(\hat{f}_S)\mathcal{S}(\xi)e^{-\kappa_0\rho t/A'} \ll S^N\mathcal{S}(f)\mathcal{S}(\xi)e^{-\kappa_0\rho t/A'}. \tag{7.4}$$

If we choose $S = e^{\eta t} = e^{\kappa_0\rho t/(2NA')}$, the above is $\ll \mathcal{S}(f)\mathcal{S}(\xi)e^{-\eta t/2}$.

Moreover, by Lemma 6.2 applied with $p = 3/2$ and Chebyshev's inequality, we have

$$\int_{\{\theta:\Delta(a_t r_\theta)g\Gamma'\notin X_S\}} S\,\mathrm{d}\theta \ll S^{-3/2}S = S^{-1/2}. \tag{7.5}$$

This and (7.4) reduce the problem to investigating the integral of $\hat{f} - \hat{f}_S = \hat{f}_{\mathrm{cusp}} + \hat{f}_{\mathrm{sp}}$ over $\hat{\mathcal{C}} := \{\theta \in [0, 2\pi] : \hat{f} - \hat{f}_S \geq S\}$.

Let $\tilde{f}$ be as in (6.2) with η/A, A, and t. That is:

$$\tilde{f}(h; g\Gamma) = \sum_{v\in\mathcal{N}_t(g\mathbb{Z}^4)} f(\Delta(h)v)$$

where $\mathcal{N}_t(g\mathbb{Z}^4)$ denotes the set of vectors in $g\mathbb{Z}^4$ not contained in any $(\eta/A, A, t)$-special subspaces and also not contained in $\mathbb{R}^2 \times \{0\} \cup \{0\} \cup \mathbb{R}^2$.

Let $\tilde{\mathcal{C}}_t = \{\theta \in [0, 2\pi] : \tilde{f}(a_t r_\theta; g\Gamma) \geq e^{\eta t} = S\}$. By the definitions,

$$\int_{\hat{\mathcal{C}}} \hat{f}_{\mathrm{cusp}}(\Delta(a_t r_\theta)g\Gamma)\xi(\theta)\,\mathrm{d}\theta \leq \|\xi\|_\infty \int_{\tilde{\mathcal{C}}_t} \tilde{f}(a_t r_\theta; g\Gamma')\,\mathrm{d}\theta.$$

In view of (7.3), $e^{100\eta t/A}$ is in the range where (7.2) holds, thus Proposition 6.1, applied with η/A and A, implies

$$\int_{\tilde{\mathcal{C}}_t} \tilde{f}(\Delta(a_t r_\theta)g\Gamma')\,\mathrm{d}\theta \ll e^{-\eta^3 t/A^3}.$$

From these two, we conclude that

$$\int_{\hat{\mathcal{C}}} \hat{f}_{\mathrm{cusp}}(\Delta(a_t r_\theta)g\Gamma)\,\mathrm{d}\theta \ll \|\xi\|_\infty e^{-\eta^3 t/A^3}. \tag{7.6}$$

In view of (7.4), (7.5) and (7.6), we have

$$\left|\int_0^{2\pi} \hat{f}(\Delta(a_t r_\theta)g\Gamma)\xi(\theta)\,\mathrm{d}\theta - \int_0^{2\pi} \xi\,\mathrm{d}\theta \int_X \hat{f}_R\,\mathrm{d}m_X\right|$$
$$= \int_{\mathcal{C}} \hat{f}_{\mathrm{sp}}(\Delta(a_t r_\theta)g\Gamma)\xi(\theta)\,\mathrm{d}\theta + O(\mathcal{S}(f)\mathcal{S}(\xi)e^{-\eta^2 t/A^3})$$

where $\mathcal{C} = \{\theta : \hat{f}_{\mathrm{sp}}(\Delta(a_t r_\theta) > e^{\eta t}\}$.

This completes the proof if we let $\delta_1 = \eta$ and $\delta_2 = \eta^3/A^3$. □

Proof of the Sublemma. Let $x = (h_1, h_2)\Gamma'$ be so that Hx is periodic. In view of (the by now standard) non-divergence results, we may assume $\|h_i\| \ll 1$ where the implied constant is absolute, see, e.g., [LM21, §3].

Since Γ' is a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$, we conclude that

$$\{(h, h) : h \in \mathrm{SL}_2(\mathbb{R})\} \cap (h_1\mathrm{SL}_2(\mathbb{Z})h_1^{-1}) \times (h_2\mathrm{SL}_2(\mathbb{Z})h_2^{-1})$$

is a lattice in $\{(h,h) : h \in \mathrm{SL}_2(\mathbb{R})\}$. This implies that $h_1\mathrm{SL}_2(\mathbb{Z})h_1^{-1}$ and $h_2\mathrm{SL}_2(\mathbb{Z})h_2^{-1}$ are commensurable. Hence, $h_2^{-1}h_1$ belongs to the image of $\mathrm{GL}_2^+(\mathbb{Q})$ in $\mathrm{SL}_2(\mathbb{R})$, i.e., the commensurator of $\mathrm{SL}_2(\mathbb{Z})$ in $\mathrm{SL}_2(\mathbb{R})$.

Let $Q' \in \mathrm{Mat}_2(\mathbb{Z})$ be so that $h_2^{-1}h_1 = \lambda Q'$, where $\lambda = (\det Q')^{1/2}$. Since $\|h_i\| \ll 1$, we have

$$\|Q'\|^{A_2} \ll \mathrm{vol}(Hx) \ll \|Q'\|^{A_3}, \tag{7.7}$$

where $A_2 \le 1 \le A_3$ and the implied constants are absolute, see, e.g., [LMW22, Lemma 16.2].

We will show the sublemma holds with $A_1 = 4DA/A_2$. Assume now, contrary to our claim in the sublemma, that $\mathrm{vol}(Hx) \le e^{\rho t/A_1}$, for some A_1 which will be determined later, and that $d_X(g\Gamma', x) \le e^{-t/2}$.

Thus $g_1 = \epsilon_1 h_1\gamma_1$ and $g_2 = \epsilon_2 h_2 \gamma_2$ where $\|\epsilon_i\| \ll e^{-t/2}$ and $(\gamma_1, \gamma_2) \in \Gamma'$. Since $\|h_i\| \ll 1$, we conclude that $\|\gamma_i\| \ll \|g_i\|$. Moreover, we have

$$g_2^{-1}g_1 = \epsilon\gamma_2^{-1}h_2^{-1}h_1\gamma_1 \tag{7.8}$$

where $\|\epsilon\| \ll e^{-t/2}$ and the implied constants depend on $\|g_i\|$. Put $Q = \gamma_2^{-1}Q'\gamma_1$. Then

$$\|Q\| \ll \|Q'\| \ll e^{\rho t/A_1A_2} \le e^{\rho t/A}$$

where we used (7.7), $\mathrm{vol}(Hx) \ll e^{\rho t/A_1}$ and assumed t is large. Moreover, using (7.8) and (7.7), we conclude that

$$\|g_2^{-1}g_1 - \lambda Q\| \ll e^{-t/2}\|Q'\| \ll e^{-t/2} \cdot e^{\rho t/(A_1A_2)} \tag{7.9}$$

where the implied constants depend on $\|g_i\|$.

Assuming t is large enough to account for the implied constant and using $A_1 = 4DA/A_2$, the left side of (7.9) is $< e^{-\rho t}$. Thus (7.9) contradicts (7.2) and finishes the proof of the theorem. □

REFERENCES

[BBRlR17] V. Blomer, J. Bourgain, M. Radziwiłł and Z. Rudnick, *Small gaps in the spectrum of the rectangular billiard*, Ann. Sci. Éc. Norm. Supér. (4) **50** (2017), 1283–1300.

[EM01] A. Eskin and H. Masur, *Asymptotic formulas on flat surfaces*, Ergodic Theory Dynam. Systems **21** (2001), 443–478.

[EM22] A. Eskin and S. Mozes, *Margulis functions and their applications*, in *Dynamics, Geometry, Number Theory: The Impact of Margulis on Modern Mathematics*, The University of Chicago Press, Chicago, IL, 2022, pp. 342–361.

[EMM98] A. Eskin, G. Margulis and S. Mozes, *Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture*, Ann. of Math. (2) **147** (1998), 93–141.

[EMM05] A. Eskin, G. Margulis and S. Mozes, *Quadratic forms of signature* $(2, 2)$ *and eigenvalue spacings on rectangular 2-tori*, Ann. of Math. (2) **161** (2005), 679–725.

[KKL23] D. Kelmer, A. Kontorovich and C. Lutsko, *Mean square of Eisenstein series*, arXiv:2305.15162 [math.NT]

[LM21] E. Lindenstrauss and A. Mohammadi, *Polynomial effective density in quotients of* $\mathbb{H}^3$ *and* $\mathbb{H}^2 \times \mathbb{H}^2$, Invent. Math. **231** (2023), 1141–1237.

[LMW22] E. Lindenstrauss, A. Mohammadi and Z. Wang, *Effective equidistribution for some one parameter unipotent flows*, arXiv:2211.11099 [math.NT]

[Mar02] J. Marklof, *Pair correlation densities of inhomogeneous quadratic forms. II*, Duke Math. J. **115** (2002), 409–434.

[Mar03] J. Marklof, *Pair correlation densities of inhomogeneous quadratic forms*, Ann. of Math. (2) **158** (2003), 419–471.

[MM11] G. Margulis and A. Mohammadi, *Quantitative version of the Oppenheim conjecture for inhomogeneous quadratic forms*, Duke Math. J. **158** (2011), 121–160.

[Sar81] P. Sarnak, *Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series*, Comm. Pure Appl. Math. **34** (1981), 719–739.

[Sar97] P. Sarnak, *Values at integers of binary quadratic forms*, in *Harmonic Analysis and Number Theory (Montreal, PQ, 1996)*, American Mathematical Society, Providence, RI, 1997, pp. 181–203.

[SV20] A. Strömbergsson and P. Vishe, *An effective equidistribution result for* $\mathrm{SL}(2, \mathbb{R}) \ltimes (\mathbb{R}^2)^{\oplus k}$ *and application to inhomogeneous quadratic forms*, J. Lond. Math. Soc. (2) **102** (2020), 143–204.

Elon Lindenstrauss
The Einstein Institute of Mathematics
Edmond J. Safra Campus, Givat Ram
The Hebrew University of Jerusalem
Jerusalem 91904, Israel
email: elon.bl@mail.huji.ac.il

Amir Mohammadi
Department of Mathematics
University of California at San Diego
La Jolla, CA 92093, USA
email: ammohammadi@ucsd.edu

Zhiren Wang
Department of Mathematics
Pennsylvania State University
University Park, PA 16802, USA
email: zhirenw@psu.edu

(Received May 27, 2023 and in revised form October 9, 2023)