DEVIATIONS OF ERGODIC SUMS FOR TORAL TRANSLATIONS
II. BOXES.

DMITRY DOLGOPYAT AND BASSAM FAYAD

Abstract. We study the Kronecker sequence $\{n\alpha\}_{n\leq N}$ on the torus $\mathbb{T}^d$ when α is uniformly distributed on $\mathbb{T}^d$. We show that the discrepancy of the number of visits of this sequence to a random box, normalized by $\ln^d N$, converges as $N \to \infty$ to a Cauchy distribution. The key ingredient of the proof is a Poisson limit theorem for the Cartan action on the space of $d+1$ dimensional lattices.

Contents

1. Introduction

1.1. **Equidistribution of Kronecker sequences on $\mathbb{T}^d$.** It is well known that the orbits of a non resonant translation on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ are uniformly distributed. A quantitative measure of uniform distribution is given by the discrepancy function: for a set $\mathcal{C} \subset \mathbb{T}^d$ let

$$D(\alpha, x, \mathcal{C}, N) = \sum_{n=0}^{N-1} \mathbb{1}_{\mathcal{C}}(x + k\alpha) - N\nu(\mathcal{C})$$

where $(\alpha, x) \in \mathbb{T}^d \times \mathbb{T}^d$, $\mathbb{1}_{\mathcal{C}}$ is the characteristic function of the set $\mathcal{C}$ and ν is the Haar measure on the torus. (We will sometimes write ν_d if we want to emphasize the dimension of the torus). Uniform distribution of the sequence $x + k\alpha$ on $\mathbb{T}^d$ is equivalent to the fact that, for regular sets $\mathcal{C}$, $D(\alpha, x, \mathcal{C}, N)/N \to 0$ as $N \to \infty$. A step further is the study of the rate of convergence to 0 of $D(\alpha, x, \mathcal{C}, N)/N$.

Already with $d = 1$, it is clear that if $\alpha \in \mathbb{T} - \mathbb{Q}$ is fixed, the discrepancy $D(\alpha, x, \mathcal{C}, N)$ displays an oscillatory behavior according to the position of N with respect to the denominators of the best rational approximations of α. A great deal of work in Diophantine approximation has been done on estimating the oscillations of the discrepancy function in relation with the arithmetic properties of $\alpha \in \mathbb{T}$, and more generally for $\alpha \in \mathbb{T}^d$. It is of common knowledge that in studying the discrepancies in dimension 1 the continued fraction algorithm provides crucial help, and that the absence of an analogue in higher dimensions makes the study of discrepancies much harder.

In particular, let

$$\overline{D}(\alpha, N) = \sup_{\mathcal{C} \in \mathbb{B}} D(\alpha, 0, \mathcal{C}, N)$$

where the supremum is taken over all sets $\mathcal{C}$ in some natural class of sets $\mathbb{B}$, for example balls or boxes. The case of (straight) boxes was extensively studied, and growth properties of the sequence $\overline{D}(\alpha, N)$ were obtained with a special emphasis on their relations with the Diophantine approximation properties of α. In particular, following earlier advances of [22, 12, 25, 18, 29] and others, [2] proves that for arbitrary positive increasing function $\phi(n)$

$$(1.1) \qquad \sum_n \frac{1}{\phi(n)} < \infty \iff \frac{\overline{D}(\alpha, N)}{(\ln N)^d \phi(\ln \ln N)} \text{ is bounded for almost every } \alpha \in \mathbb{T}^d.$$

In dimension $d = 1$, this result is the content of Khinchine theorems obtained in the early 1920's [18], and follows easily from well-known results from the metrical theory of continued fractions (see for example the introduction of [2]). The higher dimensional case is significantly more difficult and (1.1) was only obtained in the 1990s.

The bound in (1.1) focuses on the worst case scenario, that is on how bad can the discrepancy become along a subsequence of N, for a fixed α in a full measure set.

The restriction on α is necessary, since given any $\epsilon_n \to 0$ it is easy to see that for $\alpha \in \mathbb{T}$ *sufficiently Liouville*, the discrepancy (relative to intervals) can be as bad as $N_n \epsilon_n$ along a suitable sequence N_n (large multiples of denominators of very good rational approximations). It is conjectured that for any α the discrepancy can be as bad as $(\ln N)^d$ but not much is known better than the general lower bound $(\ln N)^{d/2}$ that holds for every sequence on $\mathbb{T}^d$ ([27]). Here again, due to the use of continued fractions the latter conjecture can be easily verified in dimension 1 (cf. discussion in [2]).

In another direction, but still studying the discrepancy for a fixed α and along subsequences of N, [15] obtains a Central Limit Theorem in the one dimensional case of circle rotations. The results of [15] apply either for a set of α of zero measure (so called badly approximable numbers) and a set of times of large density, or for all α but for a small set of times (in both cases, the time sets depend on α).

By contrast, if one lets α and x be random then it is possible to obtain asymptotic distributions of the adequately normalized discrepancy for *all* N.

This is the approach adopted by Kesten in [16, 17] (see also [3]) where he studied the distribution of the discrepancies related to circular rotations as α and x are randomly distributed over the circle. He proved the following result.

Theorem [16, 17]. *Let $0 < a < b < 1$ and define*

$$D(\alpha, x, [a,b], N) = \sum_{n=0}^{N-1} \mathbb{1}_{[a,b]}(x + k\alpha) - N(b-a).$$

There is a number $\rho = \rho(b-a)$ such that if (α, x) is uniformly distributed on $\mathbb{T}^2$ then $\frac{D(\alpha,x,[a,b],N)}{\rho \ln N}$ converges to the standard Cauchy distribution, that is,

$$\nu_2 \left\{ (\alpha, x) : \frac{D(\alpha, x, [a,b], N)}{\rho \ln N} \leq z \right\} \to \mathfrak{C}(z)$$

where

$$\mathfrak{C}(z) = \frac{\tan^{-1} z}{\pi} + \frac{1}{2}. \tag{1.2}$$

Moreover $\rho(b-a) \equiv \rho_0$ is independent of $b-a$ if $b-a \notin \mathbb{Q}$ and it has non-trivial dependence on $b-a$ if $b-a \in \mathbb{Q}$.

Our goal is to extend this result to higher dimensions. As in the case of other results related to discrepancies of Kronecker sequences, the main difficulty comes from the absence of a continued fraction algorithm that was also the main tool in Kesten's proof.

Before we describe our approach, let us mention that there are two natural counterparts to intervals in higher dimension: balls and boxes. In [8] we considered the case where $\mathcal{C}$ is analytic and strictly convex and showed that $D(\alpha, x, \mathcal{C}, N)/N^{(d-1)/2d}$ has a limiting distribution (which however depends on $\mathcal{C}$).

Here we address the case where $\mathcal{C}$ is a box and show that $\frac{D(\alpha,x,\mathcal{C},N)}{(\ln N)^d}$ converges to a Cauchy distribution. To avoid the irregular behavior of the limiting distribution on the size of the considered box, as is the case in Kesten's result for example, we introduce an additional randomness to the parameters, by letting the lengths of the box's sides fluctuate. For a reason that will be explained in the sequel we also have to apply (arbitrarily small) random linear deformations on the boxes.

More precisely, for $u = (u_1, \dots, u_d)$ with $0 < u_i < 1/2$ for every i, we define a *box* on the d-torus by $C_u = [-u_1, u_1] \times \dots [-u_d, u_d]$. For any small $\eta > 0$ let

$$G_\eta = \{(a_{ij}) \in \mathrm{SL}_d(\mathbb{R}) : |a_{i,i} - 1| < \eta, \forall i \text{ and } |a_{i,j}| < \eta \ \forall j \neq i\}.$$

We denote by MC_u the image of C_u by a matrix $M \in G_\eta$. Next, each length u_i is assumed to be uniformly distributed in a segment $[v_i, w_i]$ where v_i, w_i are fixed such that $0 < v_i < w_i < 1/2$ for every i.

Let

$$X = \Big\{(\alpha, x, u, (a_{ij})) \in \mathbb{T}^d \times \mathbb{T}^d \times ([v_1, w_1] \times \dots [v_d, w_d]) \times G_\eta\Big\}$$

and denote by λ the normalized Lebesgue measure on X. For $\vartheta = (\alpha, x, u, M) \in X$, define the following discrepancy function

$$D(\vartheta, N) = \#\{1 \leq m \leq N : (x + m\alpha) \bmod 1 \in MC_u\} - 2^d \left(\Pi_i u_i\right) N.$$

Theorem 1.1. *Let* $\boldsymbol{\rho} = \frac{1}{d!}\left(\frac{2}{\pi}\right)^{2d+2}$. *For any* $z \in \mathbb{R}$ *we have*

$$\lim_{N\to\infty} \lambda\{\vartheta \in X \ / \ \frac{D(\vartheta, N)}{(\ln N)^d} \leq z\} = \mathfrak{C}(\boldsymbol{\rho} z) \tag{1.3}$$

where $\mathfrak{C}$ *is defined by* (1.2)

As will be clear from the proof, the same statement holds if λ is replaced by any probability measure on X with smooth density. Actually, we could replace the two perturbations of the box, the fluctuation of the sides' lengths and the application of an $\mathrm{SL}_d(\mathbb{R})$ matrix, by a single random linear perturbation, or by rM with r smoothly distributed in a neighborhood of 1 and $M \in G_\eta$. We prefer to keep the perturbations split because their roles in the proof are quite different.

As it is alluded in the title, the discrepancy is a special case of ergodic sums $\sum_{n=0}^{N-1} A(x + n\alpha)$. We refer the reader to a recent survey [9] for more results and open questions on this subject.

Our proof of Theorem 1.1 shows that for typical α, a *quenched* limit (that is, with fixed α, and x uniformly distributed on $\mathbb{T}^d$) of $D(\alpha, x, \mathcal{C}, N)$ does not exist even if we would allow the normalizing sequence to depend on α. The reason is that the main contribution to the discrepancy comes from a small set of so called *small denominators* and, at different scales, different small denominators become important. Also, the number of the small denominators of a given size fluctuates. Therefore there is a sequence of times when the discrepancy is dominated by a single small denominator, so, after a proper normalization we get limiting distribution of compact support. On the other hand, we can consider a sequence of times when there are many small denominators of approximately equal strength, in which case the limiting distribution will be Gaussian. Since we can obtain different limit distributions along different sequences, no limit exists as $N \to \infty$. We note that the absence of quenched limits is often observed in zero entropy systems [4, 8, 24].

1.2. **Plan of the paper.** We now give a description of the paper's content and of the main ingredients in the proofs.

Section 2 contains preliminaries and reminders. In Section 2.1 we recall the representation of the Cauchy distribution in terms of a Poisson process. In section 2.2 we present Rogers formulas that allow to compute the average and higher moments for the number of points of a random lattice in a given domain.

In Section 3, harmonic analysis of the discrepancy's Fourier series allows to bound the frequencies that have essential contributions to the discrepancy and show that they must be resonant with α. After eliminating a small measure set of vectors α, for which the resonances are too strong we obtain that the good normalization for the discrepancy is $(\ln N)^d$. The main outcome of Section 3 is to reduce the proof of Theorem 1.1 to that of Theorem 4.1 establishing a Poisson limit theorem for the distribution of the small denominators that appear in the (resonant) Fourier terms that contribute to the discrepancy.

Apart from Section 3, all our proofs are identical in any dimension and in the 2-dimensional case. We therefore present the proof of the Poisson limit theorem in dimension 2 in order to improve the readability of the paper. Thus, throughout Sections 5–8 we assume that $d = 2$ and use the notations $(\alpha, \beta), (x, y), (u, v)$,instead of $(\alpha_1, \alpha_2), (x_1, x_2), (u_1, u_2)$, and let

$$M_{a_1 b_1 a_2 b_2} \in G_\eta \ = \{ \ |a_1 - 1|, |b_2 - 1|, |a_2|, |b_1| < \eta; a_1 b_2 - b_1 a_2 = 1\}.$$

In dimension 2, we need to prove the Poisson limit theorem for the sequence

$$(\star) \qquad \big\{ \big(\ln^2 N (a_1 k + b_1 l)(a_2 k + b_2 l) \big\| k\alpha + l\beta \big\|, N(k\alpha + l\beta) \bmod (2), \\ \{(a_1 k + b_1 l)u\}, \{(a_2 k + b_2 l)v\}, \{kx + ly\}\big)\big\}$$

where (k, l) range over the resonant frequencies that contribute to the discrepancy $D(\vartheta, N)$, namely (k, l) such that

$$|a_1 k + b_1 l| > 1, \quad |a_2 k + b_2 l| > 1, \quad |(a_1 k + b_1 l)(a_2 k + b_2 l)| < N,$$
$$(a_1 k + b_1 l)(a_2 k + b_2 l) \big\| k\alpha + l\beta \big\| < \frac{1}{\bar{\epsilon} \ln^2 N}.$$

In section 5, we reduce the Poisson limit of the first two coordinates of $(\star)$ to a Poisson limit theorem (Theorem 5.3) for the number of visits to a cusp by orbits of the Cartan action on the space of three dimensional unimodular lattices $\mathcal{M} = SL_3(\mathbb{R})/SL_3(\mathbb{Z})$. To prove the Poisson limit for all components of $(\star)$, we need to show that the remaining components are asymptotically independent of the first two. This requires an extra work that is done in Proposition 5.2, where the argument is similar to the original analysis of Kesten.

The proof of Theorem 5.3 occupies Sections 6, 7, 8. In Section 6, using martingale methods, we establish an abstract Poisson limit theorem that is well adapted to variables coming from dynamical systems. Establishing Poisson limit theorems for dynamical systems is a subject with rich literature (see [1, 5, 6, 7, 11, 13, 14, 26] and the references therein). The most relevant work for our purposes is the paper [7] where a Poisson Limit Theorem is proven for partially hyperbolic systems assuming that the images of local unstable manifolds become equidistributed at sufficiently fast rate.

In the present setting there are two new difficulties. First, the geometry of the cusp is quite complicated (especially for large d), in the sense that we do not know what is the order of k and l that contribute to the resonances in ($\star$). However Rogers identities provide sufficiently strong control to handle this issue. Secondly, we need to consider the action of the full diagonal subgroup of $SL_3(\mathbb{R})$ because, for a typical resonance, $a_1k + b_2l$ and $a_2k + b_2l$ have very different sizes. For such higher rank actions there is no notion of "unstable manifold" because there is no notions of "future" and "past" and going to infinity in different Weyl chambers gives different expanding and contracting directions. In the present setting, we are able to prove a Poisson limit theorem using the fact that the long leaves of the Lyapunov foliations become uniformly distributed at a polynomial rate, except, possibly, for a small measure set.

The relevant equidistribution results for unipotent subgroups of $SL_3(\mathbb{R})$ acting on $\mathcal{M}$ are presented in Section 7.

Unfortunately, the possible existence of small exceptional sets, requires us to introduce additional parameters in the form of small affine deformations of the box. This allows to deal with a Poisson Limit Theorem for lattices having a smooth density on $\mathcal{M}$ whereas if we work with the straight boxes we would have to establish a Poisson Limit Theorem for lattices having a smooth distribution on a positive codimension submanifold of $\mathcal{M}$.

In Section 8 the conditions of the abstract theorem of Section 6 are verified for the Cartan action on $\mathcal{M}$ using the equidistribution results of Section 7.

In section 9 we discuss the discrepancy for the number of visits to boxes of small size $N^{-\gamma}$, $\gamma < 1/d$, and we obtain a similar result to the case $\gamma = 0$ that corresponds to our main Theorem 1.1. The case $\gamma = 1/d$ was studied in [23] where a limit distribution was obtained without any normalization. As for the case $\gamma > 1/d$, it is vacuous since most orbits do not visit a ball of size $N^{-\gamma}$ before time N (by the Borel Cantelli Lemma).

In Section 10 we discuss the continuous time case, that is, we study the discrepancies corresponding to linear flows on the torus. We show that in the case of boxes the discrepancy is bounded in probability. Namely, the indicator function of a box is a coboundary with probability one. We actually get convergence in distribution of the discrepancies without any normalization. However, the method used to prove Theorem 1.1 gives a Cauchy limit theorem for continuous discrepancies relative to *balls,* and this only in dimension $d = 3$. Indeed, the latter is in sharp and curious contrast with the higher dimension case obtained in [8] that states that for $d \geq 4$ the continuous discrepancies relative to balls converge in distribution after normalization by a factor $T^{(d-3)/2(d-1)}$.

Finally, some technical estimates are collected in the appendix.

2. Preliminaries

2.1. **Poisson processes.** Recall that a random variable N has Poisson distribution with parameter λ if $\mathbb{P}(N=k)=e^{-\lambda}\frac{\lambda^k}{k!}$. Now easy combinatorics shows the following facts

(I) If $N_1, N_2 \dots N_m$ are independent random variables and each N_j has Poisson distribution with parameter λ_j, then $N=\sum_{j=1}^m N_j$ has Poisson distribution with parameter $\sum_{j=1}^m \lambda_j$.

(II) Conversely, take N points distributed according to a Poisson distribution with parameter λ and color each point independently with one of m colors where color j is chosen with probability p_j. Let N_j be the number of points of color j. Then N_j are independent and N_j has Poisson distribution with parameter $\lambda_j=p_j\lambda$.

Now let $(\mathfrak{X}, \mathfrak{m})$ be a measure space. By a Poisson process on this space we mean a random point process on $\mathfrak{X}$ such that if $\mathfrak{X}_1, \mathfrak{X}_2 \dots \mathfrak{X}_m$ are disjoint sets and N_j is the number of points in X_j then N_j are independent Poisson random variables with parameters $\mathfrak{m}(\mathfrak{X}_j)$ (note that this definition is consistent due to (I)). We will write $\{x_j\} \sim \mathfrak{P}(\mathfrak{X}, \mathfrak{m})$ to indicate that $\{x_j\}$ is a Poisson process with parameters $(\mathfrak{X}, \mathfrak{m})$. If $\mathfrak{X} \subset \mathbb{R}^d$ and $\mathfrak{m}$ has a density f with respect to the Lebesgue measure we say that f is the intensity of the Poisson process.

The following properties of the Poisson process are straightforward consequences of (I) and (II) above, and their proofs can be found in the monographs [19, 28].

Lemma 2.1. *(see* [19]*, Sections 2.3 and 5.2)*

(a) If $\{\Theta_j'\} \sim \mathfrak{P}(\mathfrak{X}, \mathfrak{m}')$ *and* $\{\Theta_j''\} \sim \mathfrak{P}(\mathfrak{X}, \mathfrak{m}'')$ *are independent then*

$$\{\Theta_j'\} \cup \{\Theta_j''\} \sim \mathfrak{P}(\mathfrak{X}, \mathfrak{m}'+\mathfrak{m}'').$$

(b) If $\{\Theta_j\} \sim \mathfrak{P}(\mathfrak{X}, \mathfrak{m})$ *and* $f : \mathfrak{X} \to \mathfrak{Y}$ *is a measurable map then* $\{f(\Theta_j)\} \sim \mathfrak{P}(\mathfrak{Y}, f^{-1}\mathfrak{m})$.

(c) Let $\mathfrak{X} = \mathfrak{Y} \times Z$, $\mathfrak{m} = \nu \times \lambda$ *where* λ *is a probability measure on* Z. *Then* $\{(\Theta_j, \Gamma_j)\} \sim \mathfrak{P}(\mathfrak{X}, \mathfrak{m})$ *iff* $\{\Theta_j\} \sim \mathfrak{P}(\mathfrak{Y}, \nu)$ *and* Γ_j *are random variables independent from* $\{\Theta_j\}$ *and each other and distributed according to* λ.

(d) If in (c) $\mathfrak{Y} = Z = \mathbb{R}$ *then* $\tilde{\Theta} = \{\Gamma_j \Theta_j\}$ *is a Poisson process. If* $\{\Theta_j\}$ *has intensity* f *then* $\tilde{\Theta}$ *has intensity*

$$\tilde{f}(\theta) = \mathbb{E}_\Gamma \left(f\left(\frac{\theta}{\Gamma}\right) \frac{1}{|\Gamma|} \right).$$

Next, recall [10, Chapter XVII] that the Cauchy distribution is the unique (up to scaling) symmetric distribution such that if Z, Z' and Z'' are independent random variables with that distribution then $Z'+Z''$ has the same distribution as $2Z$. This gives the following representation of the Cauchy distribution.

Lemma 2.2. *(see* [28]*, Theorem 1.4.2)* *(a) If $\{\Theta_j\}$ is a Poisson process on $\mathbb{R}$ with intensity $c\theta^{-2}$ then*

$$\lim_{\delta\to 0}\frac{1}{\rho}\sum_{\delta<|\Theta_j|}\Theta_j$$

has a standard Cauchy distribution, with $\rho = c\pi$.

(b) If $\{\Theta_j\}$ is a Poisson process on $\mathbb{R}$ with constant intensity c and if Γ_j are iid random variables having a symmetric distribution with compact support then

$$\lim_{M\to\infty}\frac{1}{\rho}\sum_{|\Theta_j|<M}\frac{\Gamma_j}{\Theta_j}$$

has a standard Cauchy distribution with $\rho = c\mathbb{E}(|\Gamma|)\pi$.

To see part (a) let $\{\Theta'_j\}, \{\Theta''_j\}$ and $\{\Theta_j\}$ be independent Poisson processes with intensity c. Then

$$\sum \frac{1}{\Theta'_j} + \sum \frac{1}{\Theta''_j} = \sum_{y\in\{\Theta'_j\}\cup\{\Theta''_j\}} \frac{1}{y}$$

and by Lemma 2.1 (a) and (b) both $\{\Theta'_j\}\cup\{\Theta''_j\}$ and $\{\frac{\Theta_j}{2}\}$ are Poisson processes with intensity $2c$.

The proof of Lemma 2.2 (b) follows from Lemma 2.2 (a) and parts (b),(c) and (d) of Lemma 2.1.

2.2. **Rogers identities.** For $d\in\mathbb{N}, d\geq 2$, denote the space of unimodular $d+1$-dimensional lattices by $\mathcal{M}_{d+1} = \mathrm{SL}_{d+1}(\mathbb{R})/\mathrm{SL}_{d+1}(\mathbb{Z})$ and let μ be the Haar measure on $\mathcal{M}_{d+1}$. Denote

$$\mathbf{c}_1 = \zeta(d+1)^{-1}, \quad \mathbf{c}_2 = \zeta(d+1)^{-2}, \quad \text{where } \zeta(d+1) = \sum_{n=1}^{\infty} n^{-(d+1)}$$

is the Riemann zeta function.

The following identities (see [23, 30]) play an important role in our argument. Let f, f_1, f_2 be piecewise smooth functions with compact support on $\mathbb{R}^{d+1}$. For a lattice $\mathcal{L}\subset\mathcal{M}_{d+1}$ let

$$F(\mathcal{L}) = \sum_{e\in\mathcal{L},\text{ prime}} f(e), \quad \bar{F}(\mathcal{L}) = \sum_{e_1\neq\pm e_2\in\mathcal{L},\text{ prime}} f_1(e_1)f_2(e_2).$$

F is called *Siegel transform of f* so we will sometimes denote F by $\mathcal{S}(f)$.

Lemma 2.3. *We have*

$$(a)\quad \int_{\mathcal{M}} F(\mathcal{L})d\mu(\mathcal{L}) = \mathbf{c}_1\int_{\mathbb{R}^{d+1}} f(x)dx,$$

$$(b)\quad \int_{\mathcal{M}} \bar{F}(\mathcal{L})d\mu(\mathcal{L}) = \mathbf{c}_2\int_{\mathbb{R}^{d+1}} f_1(x)dx\int_{\mathbb{R}^{d+1}} f_2(x)dx.$$

(c) Consequently

$$\int_{\mathcal{M}} F^2(\mathcal{L})d\mu(\mathcal{L}) = \mathbf{c}_1 \int_{\mathbb{R}^{d+1}} f^2(x)dx + \mathbf{c}_1 \int_{\mathbb{R}^{d+1}} f(x)f(-x)dx + \mathbf{c}_2 \left(\int_{\mathbb{R}^{d+1}} f(x)dx\right)^2.$$

3. Negligible contribution of non-resonant terms

As we already mentioned, the proof of the main Theorem 1.1 is obtained by applying the results of Section 2.1 to a sum of resonant terms in the Fourier series of $D(\vartheta, N)/(\ln N)^d$. But first we need to isolate the resonant terms that contribute to the limiting distribution. This will be done in the current section, the outcome of which is summarized in the Proposition 3.1 below. The rest of the section is devoted to the proof of Proposition 3.1. The proof is independent from the rest of the paper and can be skipped in a first reading.

3.1. Recall the definition

$$X = \left\{(\alpha, x, u, (a_{ij})) \in \mathbb{T}^d \times \mathbb{T}^d \times ([v_1, w_1] \times \dots [v_d, w_d]) \times G_\eta\right\}.$$

For $\vartheta \in X$ and $k \in \mathbb{Z}^d$, we use the notation

$$\bar{k}_i = a_{i,1}k_1 + \cdots + a_{i,d}k_d \tag{3.1}$$

Writing the Fourier series of the characteristic function of a box we get that

$$D(\vartheta, N) = \sum_{k \in \mathbb{Z}^d - \{0\}} U_k(\vartheta, N)$$

where

$$U_k(\vartheta, N) = A \prod_i \left(\frac{\sin\left(2\pi \bar{k}_i u_i\right)}{\bar{k}_i}\right) \frac{\sin(\pi N(k, \alpha))}{\sin(\pi(k, \alpha))} \cos(2\pi(k, x) + \varphi_{k,N,\alpha})$$

and $\varphi_{k,\alpha,N} = \pi(N-1)(k, \alpha)/2$, $A = \frac{1}{\pi^d}$, $(k, x) = \sum_{i=1}^d k_i x_i$.

Fix a small number $\bar{\epsilon} > 0$. For $y \in \mathbb{R}$ we use the notation $\|y\|$ for the closest distance of y to the integers. In all this section, we will use the notation C for a constant that does not depend on $\bar{\epsilon}$ or N but that may be different from line to line.

Define $W(\vartheta, N) = W(\alpha, (a_{i,j}), N)$ by

$$W(\vartheta, N) := \left\{ k \in \mathbb{Z}^d : \left|\prod_{i=1}^d \bar{k}_i\right| < N, \quad \forall i = 1, \dots, d, \quad |\bar{k}_i| \geq 1, \quad \left|\prod_{i=1}^d \bar{k}_i\right| \, \big\|(k, \alpha)\big\| \leq \frac{1}{\bar{\epsilon}(\ln N)^d} \right\}. \tag{3.2}$$

Next, we let

$$Z(\vartheta, N) = \big\{k \in W(\vartheta, N) : \bar{k}_1 > 0 \text{ and } \exists m \in \mathbb{Z} \text{ such that } k_1 \wedge \ldots \wedge k_d \wedge m = 1 \text{ and } \big\|(k,\alpha)\big\| = (k,\alpha) + m\big\}. \tag{3.3}$$

Then define

$$D_6(\vartheta, N) = \sum_{k \in Z} \frac{\Gamma_k(\vartheta, N)}{\Theta_k(\vartheta, N)} \tag{3.4}$$

where

$$\Theta_k(\vartheta, N) = \left(\prod_{i=1}^{d} \bar{k}_i\right) \big\|(k,\alpha)\big\| (\ln N)^d,$$

$$\Gamma_k(\vartheta, N) = \frac{2A}{\pi} \phi\left(\bar{k}_1 u_1, \ldots, \bar{k}_d u_d, N(k,\alpha), (k,x) + \varphi_{k,\alpha,N}\right),$$

and

$$\phi(\eta_1, \ldots, \eta_d, \eta_{d+1}, \eta_{d+2}) = \sum_{j=1}^{\infty} \frac{\left[\Pi_{i=1}^d \sin(2\pi j \eta_i)\right] \sin(\pi j \eta_{d+1}) \cos(2\pi j \eta_{d+2})}{j^{d+1}}. \tag{3.5}$$

The purpose of this section is show that $\dfrac{|D - D_6|}{(\ln N)^d}$ is small in probability.

Proposition 3.1. *For any $\upsilon > 0$, if we take $\bar{\epsilon} > 0$ sufficiently small and then N sufficiently large we have that*

$$\lambda\left(\left\{\vartheta \in X : \left|\frac{D_6(\vartheta, N)}{(\ln N)^d} - \frac{D(\vartheta, N)}{(\ln N)^d}\right| \geq \upsilon\right\}\right) \leq \upsilon. \tag{3.6}$$

3.2. Let

$$D_1(\vartheta, N) = \sum_{|k_i| \leq N,\ k \neq (0,\ldots,0)} U_k(\vartheta, N)$$

We claim that there exists a constant C such that

$$\|D - D_1\|_2^2 \leq C \tag{3.7}$$

where the L^2 norm in (3.7) and below in Section 3 is taken with respect to the variables $(\alpha, x) \in \mathbb{T}^{2d}$.

Proof of (3.7). Assume $\vartheta \in X$ given. Then for any $q \geq N$ and any $q_1, \ldots, q_{d-1} \in \mathbb{N}$ let $\hat{Y}(q, q_1, \ldots q_{d-1})$ be the set of vectors $k \in \mathbb{Z}^d$ such that for some permutation of the indices $1, \ldots, d$ we have $|\bar{k}_{i_d}| \in [q, q+1]$ and $|\bar{k}_{i_j}| \in [q_j, q_j+1]$ for every $j \in [1, d-1]$. Note that the cardinality of $\hat{Y}(q, q_1, \ldots q_{d-1})$ is uniformly bounded.

Since for any $\omega \in \mathbb{T}$ and any $m \neq 0$,

$$\left|\frac{\sin(2\pi m\omega)}{m}\right| < \min\left(2\pi|\omega|, \frac{1}{|m|}\right) < \frac{C}{|m|+1},$$

the contributions of high frequencies can be bounded as follows.

$$\|D-D_1\|_2^2$$

$$\leq C \sum_{q\geq N, q_1,\dots,q_{d-1}\geq 0} \frac{1}{q^2(q_1+1)^2\dots(q_{d-1}+1)^2}\int_{\mathbb{T}^d}\left(\frac{\sin(\pi N(k,\alpha)}{\sin(\pi(k,\alpha))}\right)^2 d\alpha$$

$$\leq C \sum_{q\geq N, q_1,\dots,q_{d-1}\geq 0} \frac{1}{q^2(q_1+1)^2\dots(q_{d-1}+1)^2} N \leq C. \qquad \square$$

3.3. Define $S(\vartheta,N)=S((a_{i,j}),N):=\{k\in\mathbb{Z}^d : |k_i|\leq N, |\bar{k}_i|\geq 1\}$. Then let

$$D_2(\vartheta,N)=\sum_{k\in S(\vartheta,N)} U_k(\vartheta,N).$$

We want to replace D_1 by D_2. For a fixed matrix $(a_{i,j})$, we want to bound the contributions of frequencies k such that $|\bar{k}_{i_d}|<1$ for at least one index $i_d\in[1,d]$. Observe first that since $(a_{i,j})$ is close to Identity then $|\bar{k}_i|\leq 2N$ for every i. Moreover, there exists C such that for every $(q_1,\dots,q_{d-1})\in[0,2N]^{d-1}$ there is at most C vectors $k\in[-N,N]^d$ such that $|\bar{k}_{i_d}|\leq 1$ and $|\bar{k}_{i_j}|\in[q_j,q_j+1]$ for every $j\in[1,d-1]$, where i_j is some permutation of the indices $1,\dots,d$. We call $\hat{Y}(q_1,\dots,q_{d-1})$ the latter set of k. We then exclude the translation vectors α for which there exists $(q_1,\dots,q_{d-1})\in[0,2N]^{d-1}$ with at least one $k\in\hat{Y}(q_1,\dots,q_{d-1})$ satisfying $|\Pi_{i=1}^{d-1}(q_i+1)|\big\|(k,\alpha)\big\|\leq\bar{\epsilon}/(\ln N)^{d-1}$. The excluded set $E_N((a_{i,j}))$ has Lebesgue measure of order $\bar{\epsilon}$.

We claim that

$$\|D_2-D_1\|^2_{L_2((\mathbb{T}^d-E_N)\times\mathbb{T}^d)}\leq C\frac{(\ln N)^{2(d-1)}}{\bar{\epsilon}}. \tag{3.8}$$

Proof of (3.8). Let

$$B_p\left((q_1,\dots,q_{d-1}),(a_{i,j})\right)=\{\alpha\in\mathbb{T}^d : \exists k\in\hat{Y}(q_1,\dots,q_{d-1})((a_{i,j})),$$
$$p\bar{\epsilon}/(\ln N)^{d-1}\leq\left(\prod_{i=1}^{d-1}(q_i+1)\right)\big\|(k,\alpha)\big\|\leq(p+1)\bar{\epsilon}/(\ln N)^{d-1}\}$$

then $\mathrm{Leb}\left(B_p\left((q_1,\dots,q_{d-1}),(a_{i,j})\right)\right)\leq C\frac{\bar{\epsilon}}{(q_1+1)\dots(q_{d-1}+1)(\ln N)^{d-1}}$. Hence

$$\|D_2-D_1\|^2_{L_2((\mathbb{T}^d-E_N)\times\mathbb{T}^d)}\leq$$
$$C\sum_{q_1,\dots,q_{d-1}\in[0,2N]^{d-1}}\sum_{p\geq 1}\frac{\bar{\epsilon}}{(q_1+1)\dots(q_{d-1}+1)(\ln N)^{d-1}}\frac{(\ln N)^{2(d-1)}}{\bar{\epsilon}^2p^2}$$
$$\leq C\frac{(\ln N)^{2(d-1)}}{\bar{\epsilon}}. \qquad \square$$

3.4. Denote $K(k) = \Pi_{i=1}^d \bar{k}_i$. Let

$$\bar{S}(\vartheta, N) := \{k \in \mathbb{Z}^d : |K(k)| \leq N, |\bar{k}_i| \geq 1\} \text{ and } D_3(\vartheta, N) = \sum_{k \in \bar{S}} U_k(\vartheta, N).$$

We claim that

$$\|D_3 - D_2\|_2^2 \leq C(\ln N)^{d-1}. \tag{3.9}$$

Proof of (3.9).

$$\|D_3 - D_2\|_2^2 \leq \sum_{k \in S, |K(k)| \geq N} \frac{1}{K(k)^2} \int_{\mathbb{T}^d} \left(\frac{\sin(\pi N(k,\alpha))}{\sin(\pi(k,\alpha))}\right)^2 d\alpha \leq \sum_{k \in S, |K(k)| \geq N} \frac{N}{K(k)^2}.$$

For $s \in \mathbb{N}$, let $A_s = \{k \in S : |K(k)| \in [e^s N, e^{s+1} N]\}$ and observe that $\mathrm{Card}(A_s) \leq C e^s N (\ln N + s)^{d-1}$. Thus

$$\|D_3 - D_2\|_2^2 \leq C \sum_{s=0}^{\infty} e^s N (\ln N + s)^{d-1} \frac{N}{(e^s N)^2} \leq C \ln N^{d-1}. \quad \square$$

3.5. Define $W(\vartheta, N) = W((a_{i,j}), \alpha, N)$ by

$$W(\vartheta, N) \quad := \quad \left\{k \in \bar{S}((a_{i,j}), N) : |\Pi_{i=1}^d \bar{k}_i| \|(k,\alpha)\| \leq \frac{1}{\bar{\epsilon}(\ln N)^d}\right\}$$

and let

$$D_4(\vartheta, N) = \sum_{k \in W(\vartheta, N)} U_k(\vartheta, N).$$

We claim that

$$\|D_4 - D_3\|_{L_2((\mathbb{T}^d - E_N) \times \mathbb{T}^d)} \leq C\sqrt{\bar{\epsilon}}\,(\ln N)^d. \tag{3.10}$$

Proof of (3.10). Since $k \in \bar{S}$ and $(a_{i,j})$ is close to Identity we have that $1 \leq |\bar{k}_i| \leq 2N$ for every i. Now, for every $q_1, \ldots, q_d \in [1, 2N]^d$ there are at most $C(d)$ vectors $k \in [-N, N]^d$ such that $|\bar{k}_i| \in [q_i, q_i + 1]$. We denote the latter set of vectors $Y(q_1, \ldots, q_d)$. We have that

$$\|D_4 - D_3\|^2_{L_2((\mathbb{T}^d - E_N) \times \mathbb{T}^d)} \leq C \sum_{(q_1, \ldots, q_d) \in [1, 2N]^d} A_{Y(q_1, \ldots, q_d)}$$

where

$$A_{Y(q_1, \ldots, q_d)} = \sum_{k \in Y(q_1, \ldots, q_d)} \int_{\mathbb{T}^d} \frac{1}{((\Pi_{i=1}^d q_i)\|(k,\alpha)\|)^2} \mathbb{1}_{\left((\Pi_{i=1}^d q_i)\,\|(k,\alpha)\|\right) \geq 1/\bar{\epsilon}(\ln N)^d}\, d\alpha.$$

Consider for each $k \in Y(q_1, \ldots, q_d)$ and $p \in \mathbb{N}$ the sets

$$B_{k,p} = \left\{\alpha \in \mathbb{T}^d : \frac{p}{\epsilon(\ln N)^d} \leq \left(\prod_{i=1}^d q_i\right) \|(k,\alpha)\| < \frac{p+1}{\epsilon(\ln N)^d}\right\}.$$

We have that

$$\mathrm{Leb}_{\mathbb{T}^d}(B_{k,p}) \leq \frac{1}{\bar{\epsilon}(\Pi_{i=1}^d q_i)(\ln N)^d}.$$

Thus

$$A_{Y(q_1,\dots,q_d)} \le C \frac{1}{\bar\epsilon(\Pi_{i=1}^d q_i)(\ln N)^d} \sum_{p=1}^{\infty} \frac{\bar\epsilon^2 (\ln N)^{2d}}{p^2} \le C\bar\epsilon \frac{(\ln N)^d}{\Pi_{i=1}^d q_i}$$

and the claim follows as we sum over $(q_1, \dots, q_d) \in [1, 2N]^d$. □

3.6. Since the terms in D_4 satisfy $\|(k,\alpha)\| \le \frac{1}{\bar\epsilon(\ln N)^d}$, we can replace U_k in the definition of Δ by

$$V_k(\vartheta, N) = A \prod_i \left(\frac{\sin\left(2\pi \bar k_i u_i\right)}{\bar k_i} \right) \frac{\sin(\pi N(k,\alpha))}{\pi \|(k,\alpha)\|} \cos(2\pi(k,x) + \varphi_{k,N,\alpha}))$$

and introduce

$$D_5(\vartheta, N) = 2 \sum_{k \in W(\vartheta,N), k_1 > 0} V_k.$$

Indeed, for $k \in W(\vartheta, N)$,

$$|U_k - V_k| \le C \frac{\|(k,\alpha)\|}{|\Pi_{i=1}^d \bar k_i|} \le \frac{C}{\bar\epsilon(\ln N)^d \left(|\Pi_{i=1}^d \bar k_i|\right)^2}.$$

Summing over $k \in W(\vartheta, N)$ we obtain

$$|D_5(\vartheta, N) - D_4(\vartheta, N)| \le \frac{C}{\bar\epsilon(\ln N)^d}. \tag{3.11}$$

Putting together (3.7)–(3.11), we see that $(D_5(\vartheta, N) - D(\vartheta, N))/(\ln N)^d$ satisfies (3.6) if $\bar\epsilon > 0$ is sufficiently small and then N is sufficiently large.

Recall the definition of $D_6(\vartheta, N)$ given in Section 3.1. The difference between D_6 and D_5 is that for $k \in Z(\vartheta, N)$, we comprise in D_6 all its multiples whereas in D_5 we take only multiples such that $pk \in W$. Thus

$$D_6(\vartheta, N) - D_5(\vartheta, N) = 2 \sum_{k \in Z(\vartheta,N)} \sum_{p \ge 1 : pk \notin W(\vartheta,N)} V_{pk}$$

Since we have already shown in (3.7)–(3.11) that the frequencies which are not in $W(\vartheta, N)$ make a negligible contribution after normalization by $(\ln N)^d$ as $\bar\epsilon \to 0$ and $N \to \infty$ it follows that for each υ

$$\lambda \left(\left\{ \vartheta \in X : \left| \frac{D_6(\vartheta, N)}{(\ln N)^d} - \frac{D_5(\vartheta, N)}{(\ln N)^d} \right| \ge \frac{\upsilon}{2} \right\} \right) \le \frac{\upsilon}{2}$$

provided that $\bar\epsilon$ is sufficiently small and N is sufficiently large. This finishes the proof of Proposition 3.1 and ends this section. □

4. Poisson distribution of small divisors

In this section we reduce the Cauchy limit of the discrepancies to a Poisson limit theorem (Theorem 4.1) for the small divisors $\Pi_i \bar k_i \|(k,\alpha)\|$.

4.1. The following is the bulk of our proof of Theorem 1.1.

Theorem 4.1. *Assume $\vartheta \in X$ is distributed according to the normalized Lebesgue measure λ. For any $\bar\epsilon > 0$, as $N \to \infty$, the process*

$$\left\{\left((\ln N)^d \left(\prod_i \bar k_i\right) \|(k,\alpha)\|, N(k,\alpha) \ mod\ (2), \{\bar k_1 u_1\}, \dots, \{\bar k_d u_d\},\right.\right.$$
$$\left.\left.\{(k,x)\}\right)\right\}_{k\in Z(\vartheta,N)}$$

converges to a Poisson process on $[-\frac{1}{\bar\epsilon}, \frac{1}{\bar\epsilon}] \times (\mathbb{R}/(2\mathbb{Z})) \times \mathbb{T}^{d+1}$ with intensity $2^{d-1}\mathbf{c}_1/d!$ where $\mathbf{c}_1 = 1/\zeta(d+1)$ is the constant from Lemma 2.3.

Remark. Observe that it does not change anything in the result nor in the proof to take in the last coordinate of the process $\{(k,x) + \varphi_{k,\alpha,N}\}$ instead of $\{(k,x)\}$ since the phase $\varphi_{k,\alpha,N} = \pi(N-1)(k,\alpha)/2$ is independent of the variable x. It is with the phase $\varphi_{k,\alpha,N}$ that Theorem 4.1 is used to prove Theorem 1.1.

Here and below when we consider the Poisson process on a real line times a torus the intensity is always computed with respect to the Lebesgue measure on the line times the Haar measure on the torus. This normalization is convenient since in Lemma 2.1(c) we need to have a probability measure on the second factor.

Sections 5–8 are dedicated to the proof of Theorem 4.1.

Note that by standard properties of weak convergence the result remains valid in the limit $\bar\epsilon = 0$. That is, we get the following result which is of independent interest.

Corollary 4.2. *Let $((a_{ij}, \alpha)$ be distributed according to a smooth density on $SL_d(\mathbb{R}) \times \mathbb{T}^d$. Then as $N \to \infty$ the point process*

$$\left\{\left((\ln N)^d \left(\prod_i \bar k_i\right) \|(k,\alpha)\|\right)\right\}_{k\in Z^*(\vartheta,N)}$$

where

$$Z^*(\vartheta, N) = \left\{k \in \mathbb{Z}^d : \left|\prod_i \bar k_i\right| < N, |\bar k_i| \geq 1, \bar k_1 > 0\right\}.$$

converges to a Poisson process on $\mathbb{R}$ with intensity $2^{d-1}\mathbf{c}_1/d!$.

Proof. By definition of the weak convergence it is sufficient to prove that for each $\bar\epsilon$ the point process restricted by the condition

$$\left|(\ln N)^d \left(\prod_i \bar k_i\right) \|(k,\alpha)\|\right| \leq \frac{1}{\bar\epsilon}$$

converges to the Poisson process on $\left[-\frac{1}{\bar\epsilon}, \frac{1}{\bar\epsilon}\right] \times (\mathbb{R}/(2\mathbb{Z})) \times \mathbb{T}^d$. Thus the corollary follows from Theorem 4.1 and the invariance of Poisson processes under projection (Lemma 2.1(b)). □

4.2. **Proof that Theorem 4.1 implies Theorem 1.1.** From Proposition 3.1 it suffices to show that as $\bar{\epsilon} \to 0$ and $N \to \infty$, $\frac{D_6}{\boldsymbol{\rho}(\ln N)^d}$ converges in distribution to the standard Cauchy law. Due to Theorem 4.1 and the remark that follows it, Lemma 2.1(c) yields that $\{\Theta_k(\vartheta, N)\}_{k\in Z}$ converges to a Poisson process on $[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}}]$ with intensity $c = 2^{d-1}\mathbf{c}_1/d!$, and that $\{\Gamma_k(\vartheta, N)\}_{k\in Z}$ are asymptotically iid independent from the Θs and have a symmetric distribution with compact support. Moreover, it follows from the definition of Γ_k in (3.5) that $\mathbb{E}(|\Gamma|) = \frac{2A}{\pi}\zeta(d+1)\left(\frac{2}{\pi}\right)^{d+2}$. Hence, in the limit $\bar{\epsilon} \to 0$, Lemma 2.2 applies and yields Theorem 1.1 with

$$\boldsymbol{\rho} = \pi c \mathbb{E}(\Gamma) = \frac{1}{d!}\left(\frac{2}{\pi}\right)^{2d+2}. \qquad \square$$

4.3. **The case** $d = 2$**. Notations.** Since the proof of Theorem 4.1 is the same for general d as for the case $d = 2$, we specify in the sequel to the latter case. In our opinion, this will improve the readability of the proof to which sections 5–8 are devoted.

Recall from the introduction the notations (α, β), (x, y), $(\bar{u}, \bar{v})$, instead of α, x, u and

$$M_{a_1b_1a_2b_2} \in G_\eta \ = \{\ |a_1 - 1|, |b_2 - 1|, |a_2|, |b_1| < \eta; a_1b_2 - b_1a_2 = 1\}.$$

Let

$$X = \{((\alpha, \beta), (x, y), (u, v), (a_{i,j})) \in \mathbb{T}^2 \times \mathbb{T}^2 \times [-\bar{u}, \bar{u}] \times [-\bar{v}, \bar{v}] \times G_\eta\}$$

We still denote by λ the normalized Lebesgue measure on X. In dimension 2, Theorem 4.1 reads as follows.

Theorem 4.3. *Assume* $\vartheta \in X$ *is distributed according to the normalized Lebesgue measure* λ*. For any* $\bar{\epsilon} > 0$*, as* $N \to \infty$*, the process*

$$\Big\{\Big(\ln^2 N(a_1k + b_1l)(a_2k + b_2l)\big\|k\alpha + l\beta\big\|, N(k\alpha + l\beta) mod(2), \{(a_1k + b_1l)u\},$$

$$\{(a_2k + b_2l)v\}, \{kx + ly\}\Big)\Big\}_{(k,l)\in Z(\vartheta,N)}$$

converges in probability to a Poisson process on $[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}}] \times \mathbb{T}^4$ *with intensity* $\mathbf{c}_1$ *where* $\mathbf{c}_1 = 1/\zeta(3)$ *is the constant from Lemma 2.3. Here*

$$\begin{aligned} Z(\vartheta, N) = \Big\{(k, l) \in \mathbb{Z}^2 : &|a_1k + b_1l| \geq 1, |a_2k + b_2l| \geq 1, \\ &|(a_1k + b_1l)(a_2k + b_2l)| < N, a_1k + b_1l > 0, \\ &|(a_1k + b_1l)(a_2k + b_2l)|\big\|k\alpha + l\beta\big\| \leq \frac{1}{\bar{\epsilon}(\ln N)^2}, \\ &\exists m \in \mathbb{Z} \text{ such that } k \wedge l \wedge m = 1 \textit{ and } \big\|k\alpha + l\beta\big\| = k\alpha + l\beta + m\Big\}. \end{aligned}$$

We now turn to the proof of Theorem 4.3 that will occupy sections 5–8.

5. Reduction to dynamics on the space of lattices

The goal of this section is to reduce the proof of Theorem 4.3 to Theorem 5.3 which is a Poisson limit theorem for the diagonal action on the space of lattices.

We will adopt the following notations. Denote the space of unimodular three dimensional lattices by $\mathcal{M} = \mathrm{SL}_3(\mathbb{R})/\mathrm{SL}_3(\mathbb{Z})$ and let μ be the Haar measure on $\mathcal{M}$.

In all the sequel we associate to N the integer $M = [\ln N]$.

Define

$$\mathcal{P} = \{\mathbf{t} : \mathbf{t}_1 > 0, \mathbf{t}_2 > 0, \mathbf{t}_1 + \mathbf{t}_2 < 1\} \tag{5.1}$$

$$\Pi_M = \{\mathbf{t} : \mathbf{t}_1 > 0, \mathbf{t}_2 > 0, \mathbf{t}_1 + \mathbf{t}_2 < M\} \tag{5.2}$$

Given $\bar{\epsilon} > 0$ let

$$I = (1, e], \quad J = [-e, -1) \cup (1, e], \quad K = \left[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}}\right].$$

For $\vartheta = (\alpha, \beta, x, y, a_1, b_1, a_2, b_2) \in X$, we define

$$\Lambda(\vartheta) = \begin{pmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ \alpha & \beta & 1 \end{pmatrix}. \tag{5.3}$$

Consider the Cartan subgroup

$$g_{t_1,t_2} = \begin{pmatrix} e^{-t_1} & 0 & 0 \\ 0 & e^{-t_2} & 0 \\ 0 & 0 & e^{t_1+t_2} \end{pmatrix}$$

5.1. **Reduction to a Poisson Limit Theorem for the Cartan action.** Define on the space $\mathcal{M}$ of unimodular lattices $\mathcal{L}$ the function

$$\Phi(\mathcal{L}) = \sum_{e \in \mathcal{L} \text{ prime}} \mathbb{1}_I(x(e))\mathbb{1}_J(y(e))\mathbb{1}_K(M^2 xyz(e)) \tag{5.4}$$

Note that Φ is the Siegel transform of

$$f(x, y, z) = \mathbb{1}_I(x)\mathbb{1}_J(y)\mathbb{1}_{K_M}(xyz) \text{ where } K_M = \left[-\frac{1}{M^2\bar{\epsilon}}, \frac{1}{M^2\bar{\epsilon}}\right].$$

Define an $\mathbb{R} \times (\mathbb{R}/2\mathbb{Z})$ valued function on $\mathcal{M} \times \mathbb{R}$

$$\Psi(\mathcal{L}, b) = (\Psi_1(\mathcal{L}), \Psi_2(\mathcal{L}, b)) = \sum_{e \in \mathcal{L} \text{ prime}} \mathbb{1}_I(x(e))\mathbb{1}_J(y(e))\mathbb{1}_K(M^2 xyz(e))(M^2 xyz(e), bz(e) \bmod (2)). \tag{5.5}$$

Given $\bar{\epsilon}$ and N, suppose that $\vartheta \in X$ is such

$$\forall t \in \Pi_M, \quad \Phi(g_{t_1,t_2}\Lambda(\vartheta)) \le 1 \tag{5.6}$$

From the definitions, for ϑ as in (5.6), the following are equivalent :

(i) $\Phi(g_{t_1,t_2}\Lambda(\vartheta)) = 1$

(ii) There exists a unique $(k,l)\in Z(\vartheta,N)$ such that

$$e^{t_1}<L_1\le e^{t_1+1}, e^{t_2}<|L_2|\le e^{t_2+1},\qquad L_1=a_1k+b_1l, L_2=a_2k+b_2l.$$

Note that if (i) or (ii) holds then

$$(M^2(a_1k+b_1l)(a_2k+b_2l)||\alpha k+\beta l||, N(k\alpha+l\beta) \bmod (2))= \Psi(g_t\Lambda(\vartheta), Ne^{-(t_1+t_2)}).$$

Thus, for ϑ satisfying (5.6), we have that the sequence

$$\{M^2(a_1k+b_1l)(a_2k+b_2l)||\alpha k+\beta l||, N(k\alpha+l\beta) \bmod (2)\}_{(k,l)\in Z}$$

is exactly

$$\{\Psi(g_t\Lambda(\vartheta), Ne^{-(t_1+t_2)})\}_{t\in\Pi_M,\Phi(g_t\Lambda(\vartheta))=1}.$$

Hence, to show that the distribution of

$$\{M^2(a_1k+b_1l)(a_2k+b_2l)||k\alpha+l\beta||, N(k\alpha+l\beta) \bmod (2)\}_{k,l\in Z}$$

converges as $N\to\infty$ to that of a Poisson process on $[-\frac{1}{\bar\epsilon},\frac{1}{\bar\epsilon}]\times\mathbb{R}/(2\mathbb{Z})$ with intensity $2\mathbf{c}_1$ it is sufficient to prove:

(a) that the set of ϑ that do not satisfy (5.6) is small;

(b) that the process $\{\Psi(g_t\Lambda(\vartheta), Ne^{-(t_1+t_2)})\}_{t\in\Pi_M,\Phi(g_t\Lambda(\vartheta))=1}$ converges in probability to a Poisson distribution.

This is the content of the following Theorem 5.1.

Theorem 5.1. *Assume $\vartheta\in X$ is distributed according to the normalized Lebesgue measure λ. Let Λ be the matrix $\Lambda(\vartheta)$ as defined in (5.3). Then, for any $\bar\epsilon>0$, we have*

(a) For any $t\in\mathcal{P}$, $\lambda(\Phi(g^t\Lambda)>1)=\mathcal{O}(M^{-4})$.

(b) $\{\Psi(g^t\Lambda, Ne^{-(t_1+t_2)})\}_{t\in\Pi,\Phi(g^t\Lambda)=1}$ converges as $N\to\infty$ to the Poisson process on $[-\frac{1}{\bar\epsilon},\frac{1}{\bar\epsilon}]\times\mathbb{R}/(2\mathbb{Z})$ with intensity $\mathbf{c}_1$.

The notation $X=\mathcal{O}(M^{-4})$ means that $|X|\le CM^{-4}$ where C may depend on other variables (such as $\bar\epsilon$) but not on M.

In order to get the full Poisson limit in Theorem 4.3 we will also need an additional effort to prove the independence and uniform distribution of the rest of the variables namely of

$$\{\{(a_1k+b_1l)u\},\{(a_2k+b_2l)v\},\{kx+ly\}\}_{(k,l)\in Z(\vartheta,N)}.$$

This issue is addressed below.

Definition 5.1. Let $A>0$. Consider a sequence $\{t^{(1)},\dots,t^{(s)}\}$ of points in Π_M where $t^{(j)}=(t_1^{(j)},t_2^{(j)})$. We say that this sequence is is A-split if for any pair i,j we have

$$|t_1^{(i)}-t_1^{(j)}|\ge A\quad |t_2^{(i)}-t_2^{(j)}|\ge A \text{ and } |\max(t_1^{(i)},t_2^{(i)})-\max(t_1^{(j)},t_2^{(j)})|\ge A,$$

and for any i we have $t_1^{(i)}>A$, $t_2^{(i)}>A$.

Proposition 5.2. *Let $R > 0$ be fixed. Given any $s \in \mathbb{N}$, let $(k_1, l_1), \dots, (k_s, l_s)$ be a sequence (that depends on N) of 2-tuples such that the sequence*

$$t^{(j)} = ([\ln |a_1 k_j + b_1 l_j|], [\ln |a_2 k_j + b_2 l_j|])$$

is $\sqrt{M}$-split ($[\cdot]$ denotes the integer part).

Suppose that (x, y, u, v) are distributed according to a density ρ_N such that

$$||\rho_N||_{C^1} \leq R. \tag{5.7}$$

Then the distribution of the s 3-tuples

$$\{(a_1k_1 + b_1l_1)u\}, \{(a_2k_1 + b_2l_1)v\}, \{k_1x + l_1y\} \dots,$$
$$\{(a_1k_s + b_1l_s)u\}, \{(a_2k_s + b_2l_s)v\}, \{k_sx + l_sy\}$$

converges to the uniform distribution on $\mathbb{T}^{3s}$ and the convergence is uniform with respect to N, $(a_1, a_2, b_1, b_2, \alpha, \beta)$, the choices of s 2-tuples satisfying the splitness condition, and ρ_N satisfying (5.7).

Proof of Proposition 5.2. We need to show that if $f_j : \mathbb{T}^3 \to \mathbb{C}$ are exponentials

$$f_j(\theta_1, \theta_2, \theta_3) = \exp\left(2\pi i \sum_{k=1}^{3} m_{jk}\theta_k\right)$$

and not all m_{jk} are equal to zero then

$$\int_{\mathbb{T}^4} \prod_{j=1}^{s} f_j\left((a_1k_j + b_1l_j)u, (a_2k_j + b_2l_j)v, (k_jx + l_jy + \phi_{l_j,k_j,\alpha,\beta,N})\right) \tag{5.8}$$
$$\rho_N(u, v, x, y)dudvdxdy \to 0.$$

uniformly in the parameters involved. Note that if not all m_{j1} are equal to zero then the coefficient in front of u in the above product is large since, due to splitness, it is dominated by the contribution of the largest of $|a_1k_j + b_1l_j|$ for which m_{j1} is non zero. In this case we show that the integral (5.8) is small by integrating by parts with respect to u. Similarly if not all m_{j2} are zero we can show that the integral (5.8) is small by integrating by parts with respect to v. Finally, suppose that all m_{j1} and all m_{j2} are zero. Let $\bar{j}$ be such that $\max(t_1^{(\bar{j})}, t_2^{(\bar{j})})$ is the largest among those indices for which $m_{j3} \neq 0$. Note that either $k_{\bar{j}}$ or $l_{\bar{j}}$ (or both) is of order $\exp(\max(t_1^{(\bar{j})}, t_2^{(\bar{j})}))$. In case $|k_{\bar{j}}| \geq |l_{\bar{j}}|$ we have, due to the splitness condition, that $\bar{k}_j$ dominates the coefficient in front of x and so we conclude that (5.8) is small by integrating by parts with respect to x. In the opposite case when $|k_{\bar{j}}| < |l_{\bar{j}}|$ we conclude that (5.8) is small by integrating by parts with respect to y. □

Proof that Theorem 5.1 implies Theorem 4.3. As demonstrated earlier, parts (a) and (b) of Theorem 5.1 imply that

$$\{M^2(a_1k + b_1l)(a_2k + b_2l)||\alpha k + \beta l||, N(k\alpha + l\beta) \bmod 2\}_{(k,l)\in Z}$$

converges as $N \to \infty$ to a Poisson process on $[-\frac{1}{\epsilon}, \frac{1}{\epsilon}] \times \mathbb{R}/(2\mathbb{Z})$ with intensity $\mathbf{c}_1$. Next, it follows from part (b) of Theorem 4.3 that if $t^{(1)}, t^{(2)}, \dots \in \Pi$ are the points such that $\Phi(g_t\Lambda)=1$, listed in any order; then for any $s\in\mathbb{N}$, we have that

$$\mathbb{P}(\{t^{(1)},t^{(2)},\dots\} \text{ is } \sqrt{M}-\text{split}) \to 1 \text{ as } N\to\infty. \tag{5.9}$$

Indeed, given $\bar\epsilon$, $\tilde\epsilon$ we can choose δ such that the probability that the Poisson process on $[-\frac{1}{\bar\epsilon},\frac{1}{\bar\epsilon}]\times\mathbb{R}/(2\mathbb{Z})$ with intensity $\mathbf{c}_1$ has two points within distance δ from each other is less than $\tilde\epsilon$. Since $M^{-1/2}<\delta$ for large M (5.9) follows. Therefore outside a set of small measure of $\vartheta\in X$, the sequence $(k,l)\in Z$ satisfies the hypothesis of Proposition 5.2.

Thus $\{(\{(a_1k+b_1l)u\},\{(a_2k+b_2l)v\},\{kx+ly\})\}_{(k,l)\in Z(\vartheta,N)}$ converge to uniformly distributed iid's on $\mathbb{T}^3$ independent of

$$\{M^2(a_1k+b_1l)(a_2k+b_2l)||\alpha k+\beta l||, N(k\alpha+l\beta) \bmod (2)\}_{(k,l)\in Z(\vartheta,N)}.$$

Lemma 2.1 hence yields the full Poisson limit of Theorem 4.3. □

5.2. **Modifying the initial distribution.** Before we close this section we use a last observation that allows us to complete the reduction of our problem to a clear cut dynamics problem on the space of lattices, namely the following.

Theorem 5.3. *Assume that $\mathcal{L}\in\mathcal{M}$ is distributed according to a probability measure $\tilde\mu$ that has a smooth density with respect to Haar measure on $\mathcal{M}$. Then*

(a) For any $t\in\mathcal{P}$, $\tilde\mu(\Phi(g^t\mathcal{L})>1)=\mathcal{O}(M^{-4})$.

(b) $\{\Psi(g^t\mathcal{L},Ne^{-(t_1+t_2)})\}_{\Phi(g^t\mathcal{L})=1,t\in\Pi_M}$ converges in probability as $N\to\infty$ to a Poisson process on $[-\frac{1}{\bar\epsilon},\frac{1}{\bar\epsilon}]\times\mathbb{R}/(2\mathbb{Z})$ with intensity $\mathbf{c}_1$.

Proof that Theorem 5.3 implies Theorem 5.1.

Let $\eta>0$ and define for an interval $A=[a,b]$ the intervals $A^+=[a(1-\eta),b(1+\eta)]$ and $A^-=[a(1+\eta),b(1-\eta)]$. Fix an interval $\bar K\subset K$. Let $\bar\Phi^\pm$ be defined as in (5.4) with the intervals $I^\pm,J^\pm,\bar K^\pm$ instead of I,J,K. Next, given $\Lambda=\Lambda(\vartheta)$ for some $\vartheta\in X$, define

$$\tilde\Lambda=\begin{pmatrix}(1+\sigma_1) & 0 & 0\\ 0 & (1+\sigma_2) & 0\\ 0 & 0 & (1+\sigma_1)^{-1}(1+\sigma_2)^{-1}\end{pmatrix}\begin{pmatrix}1&0&c_1\\0&1&c_2\\0&0&1\end{pmatrix}\Lambda$$

where σ_1,σ_2, c_1 and c_2 are distributed according to any smooth density on $[-\eta^2,\eta^2]^4$. This guarantees that when ϑ is distributed according to a smooth density on X, the lattice $\tilde\Lambda$ has a smooth distribution. The equivalence then between Theorem 5.1 and Theorem 5.3 stems from the straightforward observation that if M is sufficiently large, then for any $n\in\mathbb{N}$ it holds that

$$\bar\Phi^-(g_{t_1+\ln(1+\sigma_1),t_2+\ln(1+\sigma_2)}\tilde\Lambda)\ge n \implies \Phi(g_{t_1,t_2}\Lambda)\ge n$$
$$\implies \bar\Phi^+(g_{t_1+\ln(1+\sigma_1),t_2+\ln(1+\sigma_2)}\tilde\Lambda)\ge n. \quad \square$$

The rest of the paper will be devoted to the proof of Theorem 5.3.

6. Poisson limit for almost independent rare events

To prove Theorem 5.3 we will start with an abstract result that establishes a Poisson limit theorem for M^2 variables that behave similarly to iid variables with expectation of order $1/M^2$. The variables to which the abstract Poisson limit theorem must be applied to imply Theorem 5.3 will be defined precisely in Section 8 (see (8.1)–(8.4)). Essentially, there will be a counting variable ξ_t that takes integer values and that corresponds to $\Phi(g^t\mathcal{L})$ and two related variables ν_t and ζ_t that give the value of Ψ_1 and Ψ_2 in $\Psi(g^t\mathcal{L}, Ne^{-(t_1+t_2)})\}$ when $\xi_t = 1$.

6.1. **Setting and results.**

• Let $(\Omega, \mathbb{P})$ be a probability space. We denote by $\mathbb{E}$ the expectation with respect to $\mathbb{P}$.

• Let $\mathcal{P}$ be a bounded domain in $\mathbb{R}^2$ with piecewise smooth boundary (in order to establish Theorem 5.3 it is sufficient to consider the case when $\mathcal{P}$ is the triangle given by (5.1)). For $M \in \mathbb{N}$ we let

$$\Pi_M = \{t \in \mathbb{N}^2 : t/M \in \mathcal{P}\}.$$

• We are given a linear form on $\mathbb{R}^2$, $\lambda_1(x, y) = ax+by+c$ with $|a|+|b| > 0$.

• We let $(\mathfrak{X}, \mathfrak{m})$ and $(\tilde{\mathfrak{X}}, \tilde{\mathfrak{m}})$ be two measurable spaces. Let $\mathcal{Q}$ be a countable collection of finite partitions of $\mathfrak{X}$ and $\tilde{\mathcal{Q}}$ be a countable collection of finite partitions of $\tilde{\mathfrak{X}}$. We assume that $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$ converge to the point partitions of $(\mathfrak{X}, \mathfrak{m})$ and $(\tilde{\mathfrak{X}}, \tilde{\mathfrak{m}})$ respectively.

• For every M we consider a sequence $\{\xi_t^M\}_{t\in\Pi_M}$ of random variables taking values in non-negative integers and a sequence $\{\nu_t^M\}_{t\in\Pi_M}$ of $\mathfrak{X}$ valued random variables on Ω.

• For each fixed partition $Q = (K_1, \dots, K_P) \in \mathcal{Q}$ we suppose that ξ_t^M can be decomposed as $\xi_t^M = \sum_{p=1}^P \xi_{t,p}^M$ where $\xi_{t,p}^M$ take values in non-negative integers and on the set $\{\xi_t^M = 1\}$, it holds that $\xi_{t,p}^M = \mathbb{1}_{\nu_t^M \in K_p}$.

We define

$$\eta_t^M = \xi_t \mathbb{1}_{\xi_t^M=1}, \quad \eta_{t,p}^M = \xi_{t,p}^M \mathbb{1}_{\xi_{t,p}^M=\xi_t^M=1}.$$

(Note that, in fact, $\eta_t^M = \mathbb{1}_{\xi_t^M=1}$, and $\eta_{t,p}^M = \mathbb{1}_{\xi_{t,p}^M=\xi_t^M=1}$ but we use a more complicated definition above to emphasize that $\eta_t^M \approx \xi_t^M$, $\eta_{t,p}^M \approx \xi_{t,p}^M$.)

• Since all the variables depend on M, we will omit sometimes the superscript or subscript M and denote Π_M simply by Π, ξ_t^M by ξ_t etc.

• In all this section, C denotes a constant that may change from line to line and may depend on $Q, \tilde{Q}$ but that does not depend on the parameters M, t, or $\tilde{\delta}$ that will be introduced later. We also use the notation $Y = \mathcal{O}(X)$ if $|Y| \le CX$ for such a constant C.

• We assume that for every fixed M, a sequence of partitions F_t, $t \in \Pi$ of $(\Omega, \mathbb{P})$ is given. For $\omega \in \Omega$ we denote by $F_t(\omega)$ the element of F_t containing ω. We will denote by $\mathcal{F}_t$ the σ-algebra generated by F_t. We assume that the following hypotheses hold: there exists $R > 0$ (that does not depend on M) and a set E such that

$$\mathbb{P}(E^c) \le CM^{-100}$$

and

(h1) For any $t \in \Pi$,
$$\mathbb{E}(\xi_t) = \mathcal{O}(M^{-2})$$

(h2) For any $t \in \Pi$,
$$\mathbb{P}(\xi_t > 1) = \mathcal{O}(M^{-4})$$

(h3) For $t, t' \in \Pi$, $t \ne t'$,
$$\mathbb{P}(\xi_t \ge 1, \xi_{t'} \ge 1) = \mathcal{O}(M^{-4})$$

(h4) For $t, t' \in \Pi$ with $\lambda_1(t) \ge \lambda_1(t') + R \ln M$, for any $p \in [1, P]$ and for any $\omega \in E$

$$\text{(h4a)} \quad \mathbb{E}(\xi_t | \mathcal{F}_{t'})(\omega) = \frac{\mathbf{c}\mathfrak{m}(\mathfrak{X})}{M^2} + \mathcal{O}\left(M^{-4}\right)$$

$$\text{(h4b)} \quad \mathbb{E}(\xi_{t,p} | \mathcal{F}_{t'})(\omega) = \frac{\mathbf{c}\mathfrak{m}(K_p)}{M^2} + \mathcal{O}\left(M^{-4}\right)$$

$$\text{(h4c)} \quad \mathbb{E}(\eta_{t,p} | \mathcal{F}_{t'})(\omega) = \frac{\mathbf{c}\mathfrak{m}(K_p)}{M^2} + \mathcal{O}\left(M^{-4}\right)$$

(h5) For $t, \bar{t} \in \Pi$ with $\lambda_1(\bar{t}) > \lambda_1(t) + R \ln M$, for any $p \in [1, P]$, for any $\omega \in E$
$$\xi_{t,p} \text{ is constant on } F_{\bar{t}}(\omega)$$

(h6) The algebras $\mathcal{F}_t$ have a filtration like property in the sense that for $t, \bar{t} \in \Pi$ with $\lambda_1(\bar{t}) > \lambda_1(t) + R \ln M$, for any $\omega \in E$
$$F_{\bar{t}}(\omega) \subset F_t(\omega).$$

Theorem 6.1. *Under conditions (h1)–(h6), the sequence of point processes*

$$\left\{ \nu_t^M, \frac{t}{M} \right\}_{\xi_t^M = 1, t \in \Pi_M}$$

converges as $M \to \infty$ to a Poisson process with intensity $\mathbf{c}$ *on*

$$(\mathfrak{X} \times \mathcal{P}, \mathfrak{m} \times Leb).$$

Assume now that there is another form $\hat{\lambda}(x, y) = a'x + b'y + c'$ such that $|a'| + |b'| > 0$ and $\hat{\lambda}(t) > \lambda(t)$ on $\mathrm{Int}(\mathcal{P})$.

Suppose that for each M we have a sequence of ζ_t^M of $\tilde{\mathfrak{X}}$-valued random variables and assume that for any fixed element $\tilde{Q} = (\tilde{K}_1, \dots, \tilde{K}_J) \in \tilde{\mathfrak{Q}}$ there exists $\tilde{E}$ satisfying $\mathbb{P}(\tilde{E}_M^c) \le CM^{-10}$, such that the following conditions are satisfied.

(h7) There exists a sequence $\upsilon_M \to 0$ as $M \to \infty$, and $R > 0$ such that if $t, t' \in \Pi_M$ satisfy $\hat{\lambda}(t) \geq \lambda_1(t') + R \ln M \geq \lambda_1(t) + 2R \ln M$, then for any $p \in [1, P]$ and any $j \in [1, J]$ and any $\omega \in \tilde{E}$ such that $\xi_t(\omega) = \xi_{t,p}(\omega) = 1$

$$|\mathbb{P}(\zeta_t^M \in \tilde{K}_j | \mathcal{F}_{t'})(\omega) - \tilde{\mathfrak{m}}(\tilde{K}_j)| \leq \upsilon_M.$$

(h8) For $t, \bar{t} \in \Pi$ with $\lambda_1(\bar{t}) > \hat{\lambda}(t) + R \ln M$, for any $j \in [1, J]$, for any $\omega \in \tilde{E}$ such that $\xi_t(\omega) = 1$

$$\mathbb{1}_{\zeta_t \in \tilde{K}_j} \text{ is constant on } F_{\bar{t}}(\omega)$$

Then we have the following strengthening of Theorem 6.1.

Theorem 6.2. *Under hypothesis (h1)–(h8), the sequence of point processes*

$$\left\{ \nu_t^M, \zeta_t^M, \frac{t}{M} \right\}_{\xi_t^M = 1, t \in \Pi_M}$$

converges as $M \to \infty$ to a Poisson process with intensity $\mathbf{c}$ *on*

$$(\mathfrak{X} \times \tilde{\mathfrak{X}} \times \mathcal{P}, \mathfrak{m} \times \tilde{\mathfrak{m}} \times Leb).$$

6.2. **Proof of Theorem 6.1.** Fix a small number $\bar{\delta}$. Divide Π into strips parallel to $\text{Ker}\lambda_1$, of the form $\lambda_1^{-1}[s_{j-1}, s_j]$, of width $\bar{\delta} M$. These strips have common boundaries. To create some independence between the strips we let $\bar{s}_j = s_{j-1} + \sqrt{M}$ and define the strips $\Pi_j = \lambda_1^{-1}[\bar{s}_j, s_j]$, for $j \geq 1$. Denote their width by $\tilde{\delta} M$ (where $\tilde{\delta} = \bar{\delta} - 1/\sqrt{M}$). Let L be the total number of the strips (observe that L is of order $\tilde{\delta}^{-1}$). Then, further subdivide these strips into disjoint squares $C_1, \ldots, C_H$ of side size $\tilde{\delta} M$. Doing so, we can make sure that the leftover part of Π, call it Π' contains less than $C\tilde{\delta} M^2$ points. Clearly, (h1) implies that

$$\mathbb{P}(\exists t \in \Pi' : \xi_t \geq 1) = \mathcal{O}(\tilde{\delta}) \tag{6.1}$$

and we can concentrate on the contribution of the points that are in the partition $C_1 \sqcup \ldots \sqcup C_H$.

Fix $k \in \mathbb{N}$. Pick k squares $S_1, S_2 \ldots S_k \subset \{C_1, \ldots, C_H\}$. We call the square configuration $\tilde{\delta}$-generic if the images of any two of them under λ_1 are distant by more than $3\tilde{\delta} M$. Also fix an index $i_q \in \{1 \ldots P\}$ for each $1 \leq q \leq k$.

In all the sequel we denote by $o_{\tilde{\delta}}(1)$ a quantity that goes to zero with $\tilde{\delta}$ independently of the parameters M and t.

For a set $B \subset \mathbb{R}^2$, we denote by $\hat{B}$ the area of B.

To obtain Theorem 6.1, we shall prove

Proposition 6.3.

(a1) $\mathbb{P}(\exists t \in \Pi_M : \xi_t > 1) = \mathcal{O}(M^{-2})$.

(a2) $\mathbb{P}(\exists t', t'' \in \Pi_M : \xi_{t'} \geq 1, \xi_{t''} \geq 1$ *and* $|\lambda_1(t') - \lambda_1(t'')| \leq 3\tilde{\delta} M) \leq C\tilde{\delta}$.

(b) If $S_1, S_2 \dots S_k$ is generic then

$$\mathbb{P}\Big(\exists(t^{(1)},\dots,t^{(k)})\in S_1\times\dots\times S_k:\\ \eta_{t^{(q)},i_q}=1,\forall q\in[1,k],\eta_t=0 \textit{ for } t\notin\{t^{(1)},\dots,t^{(k)}\}\Big)\\ =\mathbf{c}^k\tilde{\delta}^{2k}\left(\prod_{q=1}^{k}\mathfrak{m}(K_{i_q})\right)\exp(-\mathbf{c}\mathfrak{m}(\mathfrak{X})\hat{\mathcal{P}})(1+o_{\tilde{\delta}}(1)).$$

Proof that Proposition 6.3 impliesTheorem 6.1. Divide $\mathcal{P}$ into subsets $\mathcal{P}_1, \mathcal{P}_2, \dots \mathcal{P}_S$ of positive area. Let $\mathcal{A}$ be the event that that for each $(p,s)\in[1,\dots,P]\times[1,\dots,S]$,

$$\text{there are exactly } l_{p,s} \text{ points, satisfying } \frac{t}{M}\in\mathcal{P}_s,\quad \xi_t=\xi_{t,p}=1. \tag{6.2}$$

By parts (a1) and (a2) of Proposition 6.3

$$\mathbb{P}\left(\exists j\leq H:\quad \sum_{t\in C_j}\xi_t>1\right)=o_{\tilde{\delta}}(1).$$

Combining this with (6.1) we see that

$$\mathbb{P}(\mathcal{A})=\sum_{S_1,S_2\dots S_k}\mathbb{P}\Big(\exists(t^{(1)},\dots,t^{(k)})\in S_1\times\dots\times S_k:\\ \eta_{t^{(q)},i_q}=1,\forall q\in[1,k],\eta_t=0 \text{ for } t\notin\{t^{(1)},\dots,t^{(k)}\}\Big)+o_{\tilde{\delta}}(1)$$

where $k=\sum_{p=1,\dots,P;s=1,\dots,S} l_{p,s}$ and the sum is over all k-tuples $((S_1,i_1),\dots(S_k,i_k))$ satisfying (6.2).

Accordingly to compute the asymptotics of $\mathbb{P}(\mathcal{A})$ we apply Proposition 6.3. Namely, by Proposition 6.3(a2) the contribution of non-generic choices of k squares is negligible as $\tilde{\delta}\to 0$. On the other hand, since for each s, there are $n_s\approx\frac{(\hat{\mathcal{P}}_s)}{\tilde{\delta}^2}$ squares in $M\mathcal{P}_s$, Proposition 6.3(b) shows that generic choices contribute (recall that $\hat{\mathcal{P}}=\sum(\hat{\mathcal{P}}_s)$, and $\mathfrak{m}(\mathfrak{X})=\sum\mathfrak{m}(K_i)$)

$$\prod_{p,s}\left[\begin{pmatrix} n_s\\ l_{p,s}\end{pmatrix}\left(\mathbf{c}\mathfrak{m}(K_p)\tilde{\delta}^2\right)^{l_{p,s}}\right]\exp\left(-\mathbf{c}\mathfrak{m}(\mathfrak{X})\hat{\mathcal{P}}\right)(1+o_{\tilde{\delta}}(1))\\ =\prod_{p,s}\left[\frac{(\mathbf{c}\mathfrak{m}(K_p)(\hat{\mathcal{P}}_s))^{l_{p,s}}}{l_{p,s}!}\exp\left(-\mathbf{c}\mathfrak{m}(K_p)(\hat{\mathcal{P}}_s)\right)\right](1+o_{\tilde{\delta}}(1))$$

which is exactly the result required by Theorem 6.1. Since the partition $\mathcal{Q}$ could be chosen arbitrarily fine in the sequence $(\mathcal{Q}_n)$ that converges to the point partition the result follows. □

Observe next that parts (a1) and (a2) of Proposition 6.3 follow by direct summation of (h2) and (h3) respectively.

We still have to prove part (b) of the proposition.

6.3. **Proof of Proposition 6.3 (b).** We call a strip Π_j which contains a square from our collection $\{S_q\}$ a *type A* strip. The remaining strips (they are a majority) are called *type B* strips.

If Π_j is of type B we say that it is compatible if $\eta_t = 0$ for all $t \in \Pi_j$. If Π_j is of type A we say that it is compatible if for q such that $S_q \subset \Pi_j$, there exists $t \in S_q$ such that $\eta_{t,i_q} = 1$ and $\eta_{\bar{t}} = 0$ for $\bar{t} \in \Pi_j - \{t\}$. Denote $p_0 = 1$,

$$p_j = \mathbb{P}(\Pi_l \text{ are compatible for } l \leq j).$$

Our goal now is to show that (recall that L is the total number of strips)

$$p_L = \mathbf{c}^k \tilde{\delta}^{2k} \left(\prod_{q=1}^{k} \mathfrak{m}(K_{i_q}) \right) \exp(-\mathbf{c}\mathfrak{m}(\mathfrak{X})\hat{\mathcal{P}})(1 + o_{\tilde{\delta}}(1)).$$

The latter is derived immediately by an iterative application of (6.3) or (6.4) of the following lemma, according to whether a strip is of type A or B respectively.

Lemma 6.4. *If* Π_{j+1} *is of type A then*

$$p_{j+1} = \mathbf{c}\mathfrak{m}(K_{i_q})\tilde{\delta}^2 p_j (1 + o_{\tilde{\delta}}(1)) \tag{6.3}$$

and if Π_{j+1} *is of type B then*

$$p_{j+1} = p_j \left(1 - \mathbf{c}\mathfrak{m}(\mathfrak{X})(\hat{\mathbf{\Pi}}_{j+1})(1 + o_{\tilde{\delta}}(1))\right) \tag{6.4}$$

with $\mathbf{\Pi}_{j+1} = \Pi_{j+1}/M$.

Proof of Lemma 6.4. We shall prove (6.3), (6.4) is similar. So, we will estimate p_{j+1} when Π_{j+1} is supposed to be of type A. Let $\tilde{t}$ be such that

$$\min_{t \in \Pi_{j+1}} \lambda_1(t) - R \ln M \geq \lambda_1(\tilde{t}) \geq \max_{t \in \Pi_j} \lambda_1(t) + R \ln M. \tag{6.5}$$

We define $\mathcal{F}_j := \mathcal{F}_{\tilde{t}}$. Accordingly we let $F_j(\omega) := F_{\tilde{t}}(\omega)$. Also consider, for any $t \in \Pi_{j+1}$ satisfying the LHS of (6.5), the partition $\tilde{F}_t = F_t \wedge F_j$ and let $\tilde{\mathcal{F}}_t$ denote the σ-algebra generated by $\tilde{F}_t$. We need a preliminary estimate. Denote

$$V_j = \sum_{t,t' \in \Pi_j, t \neq t' : |\lambda_1(t) - \lambda_1(t')| \leq 3R \ln M} \eta_t \eta_{t'}.$$

Sublemma 6.5. *There is an* $\mathcal{F}_j$ *measurable set* $E_j \subset E$ *such that* $\mathbb{P}(E_j^c) \leq \frac{C \ln M}{\sqrt{M}}$ *and for* $\omega \in E_j$

$$\mathbb{E}(V_{j+1} | \mathcal{F}_j)(\omega) \leq 1/\sqrt{M} \tag{6.6}$$

and for $t, \bar{t} \in \Pi_{j+1}$ *such that* $\lambda_1(\bar{t}) \geq \lambda_1(t) + 3R \ln M$ *and* $\omega \in E_j$ *we have*

$$\mathbb{E}(\eta_t \eta_{\bar{t}} | \mathcal{F}_j)(\omega) \leq C M^{-4}. \tag{6.7}$$

Proof. We start with (6.6). Observe that

$$\mathbb{E}(\mathbb{E}(V_{j+1}|\mathcal{F}_j)) = \mathbb{E}(V_{j+1}) \leq \frac{C\ln M}{M}$$

where the last step relies on (h3). Hence, let

$$E_j' = \left\{\omega \in E : \mathbb{E}(V_{j+1}|\mathcal{F}_j) \leq \frac{1}{\sqrt{M}}\right\}.$$

Then $\mathbb{P}((E_j')^c) = \mathcal{O}(\ln M/\sqrt{M})$ by Markov inequality. Hence (6.6) holds for $\omega \in E_j'$.

Let now $\hat{t}$ be such that $\lambda_1(\bar{t}) - R\ln M \geq \lambda_1(\hat{t}) \geq \lambda_1(t) + R\ln M$. The idea in addressing (6.7) is essentially the following :

$$\begin{aligned}\mathbb{E}(\eta_t\eta_{\bar{t}}|\mathcal{F}_j) &\overset{\text{by (h6)}}{\sim} \mathbb{E}\left(\mathbb{E}(\eta_t\eta_{\bar{t}}|\tilde{\mathcal{F}}_{\hat{t}})|\mathcal{F}_j\right) \overset{\text{by (h5)}}{\sim} \mathbb{E}\left(\eta_t\mathbb{E}(\eta_{\bar{t}}|\tilde{\mathcal{F}}_{\hat{t}})|\mathcal{F}_j\right) \\ &\overset{\text{by (h4)}}{\sim} \frac{\mathbf{cm}(\mathfrak{X})}{M^2}\mathbb{E}(\eta_t|\mathcal{F}_j) \overset{\text{by (h4)}}{\sim} \frac{\mathbf{cm}(\mathfrak{X})}{M^2}\frac{\mathbf{cm}(\mathfrak{X})}{M^2}\end{aligned}$$

The fact that (h4)–(h6) only hold outside of some exceptional set, makes the argument above incomplete and we now complete it. Due to (h6), we have that $F_{\hat{t}}(\omega) \subset F_j(\omega)$ for $\omega \in E$. Hence, for $\omega \in E$ we have

$$\begin{aligned}\mathbb{E}(\eta_t\eta_{\bar{t}}|\mathcal{F}_j) &= \mathbb{E}\left(\mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right)\eta_{\bar{t}}|\mathcal{F}_j\right) + \mathcal{R}_1(t,\bar{t}) \\ &= \mathbb{E}\left(\mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right)\mathbb{E}\left(\eta_{\bar{t}}|\tilde{\mathcal{F}}_{\hat{t}}\right)|\mathcal{F}_j\right) + \mathcal{R}_1 \\ &= \mathbb{E}\left(\mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right)\mathbb{E}\left(\eta_{\bar{t}}|\mathcal{F}_{\hat{t}}\right)|\mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2 \\ &= \mathbb{E}\left(\eta_t\mathbb{E}\left(\eta_{\bar{t}}|\mathcal{F}_{\hat{t}}\right)|\mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 \\ &= \mathbb{E}\left(\eta_t\mathbb{1}_E\mathbb{E}\left(\eta_{\bar{t}}|\mathcal{F}_{\hat{t}}\right)|\mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4 \\ &\leq \frac{C}{M^2}\mathbb{E}\left(\eta_t|\mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4 \\ &\leq \frac{C}{M^4} + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4\end{aligned}$$

where the inequalities follow from (h4c) and for $i = 1, \ldots, 4$, $\mathcal{R}_i = \mathcal{R}_i(t,\bar{t})$ are given by

$$\begin{aligned}\mathcal{R}_1 &= \mathbb{E}\left(\left[\eta_t - \mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right)\right]\eta_{\bar{t}}|\mathcal{F}_j\right), \\ \mathcal{R}_2 &= \mathbb{E}\left(\mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right)\left[\mathbb{E}\left(\eta_{\bar{t}}|\tilde{\mathcal{F}}_{\hat{t}}\right) - \mathbb{E}\left(\eta_{\bar{t}}|\mathcal{F}_{\hat{t}}\right)\right]|\mathcal{F}_j\right), \\ \mathcal{R}_3 &= \mathbb{E}\left(\left[\mathbb{E}\left(\eta_t|\mathcal{F}_{\hat{t}}\right) - \eta_t\right]\mathbb{E}\left(\eta_{\bar{t}}|\mathcal{F}_{\hat{t}}\right)|\mathcal{F}_j\right), \\ \mathcal{R}_4 &= \mathbb{E}\left(\eta_t\mathbb{E}\left(\eta_{\bar{t}}|\mathcal{F}_{\hat{t}}\right)|\mathcal{F}_j\right) - \mathbb{E}\left(\eta_t\mathbb{1}_E\mathbb{E}\left(\eta_{\bar{t}}|\mathcal{F}_{\hat{t}}\right)|\mathcal{F}_j\right).\end{aligned}$$

Note that $\mathcal{R}_l$ are $\mathcal{F}_j$ measurable for $l \in \{1\ldots4\}$. We claim that all of them have L^1 norm of order $\mathcal{O}\left(M^{-100}\right)$. Indeed, first,

$$\mathbb{E}\left(|\mathcal{R}_4|\right) \leq \mathbb{P}(E^c) = \mathcal{O}\left(M^{-100}\right).$$

Next

$$\mathbb{E}\left(|\mathcal{R}_1|\right) \leq \mathbb{E}\left(\left|\eta_t - \mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right)\right|\right)$$

$$= \mathbb{E}\left(\left|\eta_t - \mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right)\right| \mathbb{1}_E\right) + \mathcal{O}\left(M^{-100}\right) = \mathcal{O}\left(M^{-100}\right)$$

since on E, $\tilde{F}_{\hat{t}}(\omega) = F_{\hat{t}}(\omega)$ due to (h6) and hence $\mathbb{E}\left(\eta_t|\tilde{\mathcal{F}}_{\hat{t}}\right) = \mathbb{E}\left(\eta_t|\mathcal{F}_{\hat{t}}\right) = \eta_t$ due to (h5). $\mathbb{E}\left(|\mathcal{R}_3|\right)$ and $\mathbb{E}\left(|\mathcal{R}_2|\right)$ are estimated similarly.

Let now

$$E_j'' = \{\omega \in E : \forall t, \bar{t} \in \Pi_j \quad \forall l \in \{1 \dots 4\} \quad |\mathcal{R}_l(t, \bar{t})| \leq M^{-4}\}.$$

Then by Markov inequality $\mathbb{P}((E_j'')^c) = \mathcal{O}\left(M^{-94}\right)$ and (6.7) holds for $\omega \in E_j''$.

Letting $E_j = E_j' \cap E_j''$ we obtain the statement of the sublemma. □

With the set E_j from Sublemma 6.5 we have

$$p_{j+1} = \mathbb{P}(\Pi_1, \dots, \Pi_j \text{ are compatible and } \Pi_{j+1} \text{ is compatible}) \tag{6.8}$$

$$\mathbb{P}((\Pi_1, \dots, \Pi_j \text{ are compatible and } \Pi_{j+1} \text{ is compatible}) \cap E_j) + \mathcal{O}(\ln M/\sqrt{M})$$

$$= \mathbb{E}(\mathbb{1}_{\Pi_1,\dots,\Pi_j \text{ are compatible}} \mathbb{1}_{E_j} \mathbb{P}(\Pi_{j+1} \text{ is compatible } |\mathcal{F}_j)) + \mathcal{O}(\ln M/\sqrt{M})$$

where the last step relies on the fact that $E_j \subset E$ and so, by (h5), $\mathbb{1}_{\Pi_1,\dots,\Pi_j \text{ are compatible}}$ is constant on $F_j(\omega)$ for $\omega \in E_j$.

Hence (6.3) follows if we show that for $\omega \in E_j$

$$\mathfrak{p} := \mathbb{P}(\Pi_{j+1} \text{ is compatible } |\mathcal{F}_j) = \mathbf{c}\mathfrak{m}(K_{i_q})\tilde{\delta}^2(1 + o(1)). \tag{6.9}$$

Note that

$$\mathfrak{p} = \mathbb{P}(\exists t \in S_q \cap \Pi_{j+1} : \eta_{t,i_q} = 1, \text{ and } \eta_{\bar{t}} = 0 \text{ for } \bar{t} \neq t, \bar{t} \in \Pi_{j+1}|\mathcal{F}_j).$$

Since, for a fixed $t \in S_q$, Bonferroni inequalities imply that

$$\mathbb{E}\left(\eta_{t,i_q} - \sum_{t' \in \Pi_{j+1}, t' \neq t} \eta_t \eta_{t'} |\mathcal{F}_j\right) \leq$$

$$\mathbb{P}(\eta_{t,i_q} = 1, \text{ and } \eta_{t'} = 0 \text{ for } t' \neq t, t' \in \Pi_{j+1}|\mathcal{F}_j) \leq \mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j)$$

we have that for $\omega \in E_j$

$$\left|\mathfrak{p} - \sum_{t \in S_q \cap \Pi_{j+1}} \mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j)\right| \leq \sum_{t \in S_q \cap \Pi_{j+1}, t' \neq t \in \Pi_{j+1}} \mathbb{E}(\eta_t \eta_{t'}|\mathcal{F}_j) \leq \mathcal{O}(1/\sqrt{M}) + \mathcal{O}(\tilde{\delta}^2) \tag{6.10}$$

where we use (6.6) for the terms with $|\lambda_1(t) - \lambda_1(t')| \leq 3R \ln M$, and (6.7) for the terms with $|\lambda_1(t) - \lambda_1(t')| > 3R \ln M$.

On the other hand (h4c) implies that for $\omega \in E$

$$\sum_{t \in S_q \cap \Pi_{j+1}} \mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j) = \mathbf{c}\mathfrak{m}(K_{i_q})\tilde{\delta}^2(1 + o_{\tilde{\delta}}(1)).$$

proving (6.9).

This completes the proof of (6.3). The induction estimate (6.4) for type B strips follows likewise. The proof of Lemma 6.4 and hence of Proposition 6.3(b) is complete. □

6.4. **Proof of Theorem 6.2.** The proof of Theorem 6.2 is very similar to the proof of Theorem 6.1. Namely in addition to using (h4), (h5) and (h6) to control ν_t we use (h7) and (h8) to control ζ_t. Let us briefly describe the necessary modifications.

We need a stronger notion of genericity. Namely we call a configuration of squares $S_1, S_2 \dots S_k$ strongly $\tilde{\delta}$-generic if the distance of any two among $2k$ intervals

$$\lambda_1(S_1), \lambda_1(S_2), \dots, \lambda_1(S_k), \hat{\lambda}(S_1), \hat{\lambda}(S_2), \dots, \hat{\lambda}(S_k)$$

is at least $3\tilde{\delta}M$. The same argument as in the proof of Theorem 6.1 shows that the contribution of non strongly generic configurations becomes negligible as $\tilde{\delta} \to 0$.

For the proof of Theorem 6.2 we need the following generalization of Proposition 6.3 (b).

Proposition 6.6. *If $S_1, S_2 \dots S_k$ is strongly generic then for any choices of indices $\{i_q\}_{q=1}^k$, $\{j_q\}_{q=1}^k$ with $i_q \in \{1 \dots P\}$, $j_q \in \{1 \dots J\}$ we have*

$$\mathbb{P}\Big(\exists (t^{(1)}, \dots, t^{(k)}) \in S_1 \times \dots \times S_k :$$
$$\xi_{t^{(q)}} = \xi_{t^{(q)}, i_q} = 1, \zeta_{t^{(q)}} \in \tilde{K}_{j_q}, \forall q \in [1, k]; \xi_t = 0 \textit{ for } t \notin \{t^{(1)}, \dots, t^{(k)}\}\Big)$$
$$= \mathbf{c}^k \tilde{\delta}^{2k} \left(\prod_q \mathfrak{m}(K_{i_q}) \tilde{\mathfrak{m}}(\tilde{K}_{j_q}) \right) \exp(-\mathbf{c}\mathfrak{m}(\mathfrak{X})\hat{\mathcal{P}})(1 + o_{\tilde{\delta}}(1)).$$

The derivation of Theorem 6.2 from Proposition 6.6 is very similar to the derivation of Theorem 6.1 from Proposition 6.3.

The proof of Proposition 6.6 is similar to the proof of Proposition 6.3.

Namely, we divide all strips of our partition $\{\Pi_j\}$ into three types: A, B and C. As before, Π_j is of type A if it contains a square S_q from our configuration. We say that a strip Π_p is associated to S_q if $\lambda_1(\Pi_p) \cap \hat{\lambda}(S_q) \neq \emptyset$. We call the union of all strips associated to a given S_q a type C strips. Note that type C strips are wider than type A strips since they consist of finitely many strips from our original partition. The strips from our partition $\{\Pi_j\}$ which are neither type A nor belong to a type C strip will be called type B strips. (Note that this definition is slightly different from the definition used in Section 6.3 since some strips labeled B in Section 6.3 are now labeled C). Thus we have k type A strips, k type C strips and many type B strips.

The definition of compatibility for type A and type B strips remain the same as before. Type C strip associated to S_q is called compatible if $\zeta_{t^{(q)}} \in \tilde{K}_{j_q}$ where $t^{(q)} \in S_q$ is such that $\xi_{t^{(q)}} = 1$.

Now, as in the proof of Proposition 6.3, Proposition 6.6 follows if we show in addition to (6.3) and (6.4) that if Π_{j+1} is of type C then

$$p_{j+1} = p_j \tilde{\mathfrak{m}}(\tilde{K}_{j_q})(1 + o_{\tilde{\delta}}(1)) \tag{6.11}$$

where we use the same notation as before

$$p_j = \mathbb{P}(\Pi_l \text{ are compatible for } l \le j).$$

The proof of (6.3) and (6.4) is the same as before except that we need to appeal to both (h5) and (h8) instead of just (h5).

For type C strips we consider partition $F_j := F_{\tilde{t}}$ and σ-algebra $\mathcal{F}_j := \mathcal{F}_{\tilde{t}}$ with $\tilde{t}$ such that

$$\min_{t \in \Pi_{j+1}} \lambda_1(t) - R \ln M \ge \lambda_1(\tilde{t}) \ge \max_{t \in \Pi_j} \lambda_1(t) + R \ln M$$

Similarly to Section 6.3, (6.11) follows from the following estimate: if $\omega \in E \cap \tilde{E}$ and $\xi_{t^{(q)}} = \xi_{t^{(q)},p} = 1$ on $F_j(\omega)$ then

$$\mathbb{P}(\Pi_{j+1} \text{ is compatible } | \mathcal{F}_j) = \tilde{\mathfrak{m}}(\tilde{K}_{j_q})(1 + o(1)). \tag{6.12}$$

(6.12) follows directly from (h7) since $\lambda_1(\Pi_{j+1}) \ni \hat{\lambda}(t^{(q)}) > \lambda_1(\tilde{t}) + R \ln M$.
□

7. Rate of equi-distirbution of unipotent flows

7.1. **Notation.** In the proof of Theorem 5.3 we will need to show some independence between the moments of time such that $\Phi(g^t \mathcal{L}) \ge 1$ as well as randomness of the values taken then by $\Psi(g^t \mathcal{L}, N e^{-(t_1+t_2)})$. For this we will rely on the fact that the action of g_{t_1,t_2} on $\mathcal{M}$ is partially hyperbolic in the sense that

$$T\mathcal{M} = E_0 + \sum_{q=1}^{3} \left(E_q^+ \oplus E_q^- \right)$$

where E_0 is tangent to the orbit of g and $E_q^{\pm}$ are invariant one dimensional distributions. The corresponding Lyapunov exponents are $\pm\lambda_q$ where

$$\lambda_1 = 2t_1 + t_2, \quad \lambda_2 = t_1 + 2t_2, \quad \lambda_3 = t_1 - t_2.$$

$E_q^{\pm}$ are tangent to foliations $W_q^{\pm}$ which are orbit foliations for groups $h_q^{\pm}$ where

$$h_1^+(u) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ u & 0 & 1 \end{pmatrix}, \quad h_2^+(u) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & u & 1 \end{pmatrix}, \quad h_3^+(u) = \begin{pmatrix} 1 & 0 & 0 \\ u & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and $h_q^-(u)$ are transposes of $h_q^+(u)$. Below we shall abbreviate E_q^+, W_q^+, h_q^+ with E_q, W_q, h_q.

We also use the notation $\mu(A) = \int_{\mathcal{M}} A(x) d\mu(x)$, and for $g \in SL_3(\mathbb{R})$, we denote by $A(g\cdot)$ the function whose value at x is $A(gx)$.

The Sobolev norm of index $\mathbf{s}$ will be denoted by $|| \cdot ||_{\mathbf{s}}$. We will always assume that $\mathbf{s}$ is an integer.

7.2. **Mixing for smooth functions.** Here we recall the mixing properties of homogeneous flows.

By [20], Theorem 2.4.5 there exists $\mathbf{s}$ and constants $C, \kappa > 0$ such that if $A, B \in H^{\mathbf{s}}$ and if

$$g = \begin{pmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3} \end{pmatrix}$$

where $t_1 + t_2 + t_3 = 0$ then

$$|\mu(A(\cdot)B(g\cdot)) - \mu(A)\mu(B)| \leq C||A||_{\mathbf{s}}||B||_{\mathbf{s}}e^{-\kappa \max |t_j|} \tag{7.1}$$

We recall that this implies that there exists $C > 0$ such that

$$|\mu(A(\cdot)B(h_3(u)\cdot)) - \mu(A)\mu(B)| \leq C||A||_{\mathbf{s}}||B||_{\mathbf{s}}u^{-\kappa}. \tag{7.2}$$

Indeed let $\theta \in [0, \pi]$ be such that $\cos\theta = -e^{-t}$ and let

$$R(t) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad g_t = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{t} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

It is then immediate to observe that for any $r > 0$

$$M(t) := h_3(e^t)R(t)^{-1}g_{-t}$$

is uniformly bounded for $t > 0$. By invariance of μ we get

$$\mu(A(\cdot)B(h_3(e^t)\cdot)) = \mu(A(\cdot)B(M(t)g_tR(t)\cdot)) = \mu(A_t(\cdot)B_t(g_t\cdot))$$

with $A_t(\cdot) := A(R(t)^{-1}\cdot)$, $B_t(\cdot) := B(M_t\cdot)$. Note that $||A_t||_{\mathbf{s}} \leq ||A||_{\mathbf{s}}$, $||B_t||_{\mathbf{s}} \leq C||B||_{\mathbf{s}}$ for $t \geq 0$. Hence, (7.2) follows from (7.1) applied to A_t, B_t and g_t.

7.3. **Equidistribution.** The functions we are going to work with in Section 8 are not smooth but they can be well approximated by the smooth functions. This motivates the following definition.

Definition 7.1. Given $\mathbf{s}, \mathbf{r} \geq 0$, we say that a function $A : \mathcal{M} \to \mathbb{R}$ is in $H^{\mathbf{s},\mathbf{r}}$ with $||A||_{\mathbf{s},\mathbf{r}} = K$ if given $0 < \epsilon \leq 1$ there are H^s-functions $A^- \leq A \leq A^+$ such that

$$||A^+ - A^-||_{L^1(\mu)} \leq \epsilon \text{ and } ||A^\pm||_{\mathbf{s}} \leq K\epsilon^{-\mathbf{r}}$$

where μ is the Haar measure on $\mathcal{M}$ and $||\cdot||_{\mathbf{s}}$ denotes the Sobolev norm of index $\mathbf{s}$.

We say that γ is a W_q curve of size $L > 0$ if for some $y \in \mathcal{M}$ we have $\gamma = \{h_q(\tau)y : \tau \in [0, L]\}$. For a function $A : \mathcal{M} \to \mathbb{R}$ we use the notation

$$\int_\gamma A = \frac{1}{L}\int_0^L A(h_q(s)y)ds.$$

Definition 7.2. Fix $\kappa_0 > 0$. Let $L > 0$ and $\mathcal{P}$ be a partition of $\mathcal{M}$ into W_q-curves of length L and denote $\gamma(x)$ the element of $\mathcal{P}$ containing x. Given a finite or infinite sequence of integers (k_n) and a function $A \in H^{\mathbf{s},\mathbf{r}}$, we say that $\mathcal{P}$ is κ_0-*representative* with respect to $((k_n), A)$ if for any n

$$\mu\left(x \in \mathcal{M} : \left|\int_{g^{k_n}\gamma(x)} A - \mu(A)\right| \geq \mathcal{K}_A L_n^{-\kappa_0}\right) \leq L_n^{-\kappa_0} \tag{7.3}$$

where $\mathcal{K}_A = ||A||_{\mathbf{s},\mathbf{r}} + 1$, and $L_n = Le^{\lambda_q(k_n)}$ is the length of $g^{k_n}\gamma(x)$.

We call the points x such that

$$\text{for every } n : \left|\int_{g^{k_n}\gamma(x)} A - \mu(A)\right| \leq \mathcal{K}_A L_n^{-\kappa_0}$$

representative with respect to $(\mathcal{P}, (k_n), A)$. Observe that if $\mathcal{P}$ is κ_0-*representative* with respect to $((k_n), A)$ and

$$\sum_n (L_n)^{-\kappa_0} \leq \epsilon$$

then the set of representative points has measure larger than $1 - \epsilon$.

The goal of this section is to show the following.

Proposition 7.1. *There exists* $\mathbf{s}, \kappa_0, \epsilon_0 > 0$ *such that for any* $q \in \{1, 2, 3\}$ *for any* $0 \leq \mathbf{r} \leq \mathbf{s}$, $0 < \epsilon \leq \epsilon_0$, *any function* $A \in H^{\mathbf{s},\mathbf{r}}$, *any* L *and any sequence* $\{k_n\}$ *satisfying*

$$\sum_n \left(Le^{\lambda_q(k_n)}\right)^{-\kappa_0} \leq \epsilon,$$

there exists a partition $\mathcal{P}$ *of* M *into* W_q*-curves of length* L *that is* κ_0*-representative with respect to* $((k_n), A)$.

Remark. If $\mathcal{L} \in \mathcal{M}$ is distributed according to a probability measure $\tilde{\mu}$ that has a smooth density with respect to the Haar measure μ, then the same result of Proposition 7.1 holds with (7.3) in the definition of representative partitions replaced by

$$\tilde{\mu}\left(x \in \mathcal{M} : \left|\int_{g^{k_n}\gamma(x)} A - \mu(A)\right| \geq \mathcal{K}_A L_n^{-\kappa_0}\right) \leq CL_n^{-\kappa_0} \tag{7.4}$$

where C is the maximum of the density of $\tilde{\mu}$ with respect to μ.

Remark. The requirement that $\mathbf{r} \leq \mathbf{s}$ will only serve to maintain the exponent κ in the speed of equidistribution in (7.3) bounded from below. Any upper bound on $\mathbf{r}$ would yield a lower bound on κ but it will be sufficient for us in the sequel to consider functions in $H^{\mathbf{s},\mathbf{s}}$, since we will have to deal with characteristic functions of nice sets (cf. Section 8.2).

Proof. Without loss of generality we will work with functions A having zero average, that is $\mu(A) = 0$. We will first prove Proposition 7.1 for $A \in H^{\mathbf{s}}$ and then generalize it to $A \in H^{\mathbf{s},\mathbf{r}}$. Also, we will give the proof for the case $q = 3$ the other cases being similar.

Now, assuming that $\mu(A) = 0$, (7.2) implies that

$$|\mu(A(\cdot)A(h_3(u)\cdot))| \leq C\mathcal{K}_A^2 u^{-\kappa}$$

with $\mathcal{K}_A = ||A||_{\mathbf{s}}$, thus for $S_L(\cdot) = \frac{1}{L}\int_0^L A(h_3(u)\cdot)du$ we have

$$\mu(S_L^2) \leq CL^{-\kappa}\mathcal{K}_A^2.$$

This implies that for $\kappa_0 := \kappa/3$, we have

$$\mu(x \in \mathcal{M} : |S_L(x)| > \mathcal{K}_A L^{-\kappa_0}) \leq CL^{-\kappa_0}. \tag{7.5}$$

Next let $\hat{\mathcal{P}}$ be an arbitrary partition of M into W_3-curves of length L and let $\hat{\mathcal{P}}^u = h_3(Lu)\hat{\mathcal{P}}$. Then by (7.5)

$$\bar{\mu}\left((x,u) \in \mathcal{M}\times[0,1] : \left|\int_{\gamma(x,u)} A\right| > \mathcal{K}_A L^{-\kappa_0}\right) \leq CL^{-\kappa_0}$$

where $\gamma(x,u)$ denotes the piece of $\hat{\mathcal{P}}^u$ that goes through x and $\bar{\mu}$ denotes the product of μ and the Lebesgue measure on $[0,1]$. Thus, we can choose u so that $\hat{\mathcal{P}}^u$ satisfies

$$\mu\left(x \in \mathcal{M} : \left|\int_{\gamma(x)} A\right| > \mathcal{K}_A L^{-\kappa_0}\right) \leq CL^{-\kappa_0} \tag{7.6}$$

If L is large we can drop the constant C if we let κ_0 be slightly smaller than $\kappa/3$. Likewise, if (k_n) is a finite or infinite sequence with

$$\sum_n \left(Le^{\lambda_3(k_n)}\right)^{-\kappa_0} \leq \epsilon$$

then there exists a partition $\mathcal{P}$ that is representative with respect to $((k_n), A)$ as in definition 7.2.

To extend (7.6) to functions in $H^{\mathbf{s},\mathbf{r}}$ (that may have infinite $H^{\mathbf{s}}$-norm), we use a standard approximation argument. Note first that (7.5) still holds for non zero mean $H^{\mathbf{s}}$-functions if we replace S_L^A by $\bar{S}_L^A(\cdot) = S_L^A - \mu(A)$.

Now for $\epsilon > 0$ let A, A^+, A^- be as in Definition 7.1 where we assume that $\mu(A) = 0$. Let $\mathcal{K}_A = ||A||_{\mathbf{s},\mathbf{r}} + 1$. Since $0 \leq \mu(A^+) \leq \epsilon$, we have that

$$\mu\left(x : S_L^A(x) > 2\mathcal{K}_A L^{-\tilde{\kappa}}\right) \leq \mu\left(x : \bar{S}_L^{A^+}(x) > 2\mathcal{K}_A L^{-\tilde{\kappa}} - \epsilon\right). \tag{7.7}$$

So, if we choose ϵ and $\tilde{\kappa}$ such that $\epsilon = \mathcal{K}_A L^{-\tilde{\kappa}} \sim \mathcal{K}_A \epsilon^{-\mathbf{r}} L^{-\kappa_0}$, that is $\epsilon \sim L^{-\tilde{\kappa}}$ and $\tilde{\kappa} = \kappa_0/(r+1)$ we get from (7.7) using (7.5) that

$$\mu\left(x : S_L^A(x) > 2\mathcal{K}_A L^{-\tilde{\kappa}}\right) \leq \mu\left(x : \bar{S}_L^{A^+}(x) > ||A||_{\mathbf{s},\mathbf{r}} L^{-\kappa_0}\right) \leq L^{-\kappa_0}.$$

Using A^- to bound $\mu\left(x : S_L^A(x) \leq -2\mathcal{K}_A L^{-\tilde{\kappa}}\right)$ we see that (7.6) and thus the rest of the proof extends to $H^{\mathbf{s},\mathbf{r}}$ functions, provided the exponent κ_0 is reduced. □

If $\mathcal{A}$ is a finite collection of functions we say that $\mathcal{P}$ is representative with respect to $((k_n), \mathcal{A})$ if for each $A \in \mathcal{A}$, $\mathcal{P}$ is representative with respect to $((k_n), A)$.

7.4. **Mixing for approximately smooth functions.** Section 7.3 controls the deviations of ergodic sums for $H^{\mathbf{s},\mathbf{r}}$-functions. We also need a bound on the rate of mixing for diagonal flows. Namely, let A be a bounded $H^{\mathbf{s}}$-function.

Lemma 7.2. *There is a constant C such that for any $H^{\mathbf{s},\mathbf{r}}$ function B we have*

$$|\mu(A(\cdot)B(g\cdot)) - \mu(A)\mu(B)| \leq C\left(||A||_{\mathbf{s}} + ||A||_{L^\infty}\right)||B||_{\mathbf{s},\mathbf{r}} e^{-\frac{\kappa \max|t_j|}{r+1}}$$

where κ is the constant from (7.1).

Proof. Without the loss of generality we may assume that

$$\mu(B) = 0. \tag{7.8}$$

Given ϵ, let $B_\pm$ be the functions such that

$$B_- \leq B \leq B_+, \quad \mu(B_+ - B_-) \leq \epsilon \text{ and } ||B_\pm||_{\mathbf{s}} \leq ||B||_{\mathbf{s},\mathbf{r}}\epsilon^{-r}.$$

Assume first that A is positive. Then

$$\mu(A(\cdot)B_-(g\cdot)) \leq \mu(A(\cdot)B(g\cdot)) \leq \mu(A(\cdot)B_+(g\cdot)). \tag{7.9}$$

Next, by (7.1)

$$|\mu(A(\cdot)B_+(g\cdot))| \leq \mu(A)\mu(B_+) + C||A||_{\mathbf{s}}\,||B||_{\mathbf{s},\mathbf{r}}\epsilon^{-r}e^{-\kappa\max|t_j|}.$$

Note that due to (7.8)

$$\mu(B_+) = \mu(B_+ - B) \leq \mu(B_+ - B_-) \leq \epsilon.$$

Hence

$$|\mu(A(\cdot)B_+(g\cdot)| \leq C||A||_{\mathbf{s}}\left[\epsilon + ||B||_{\mathbf{s},\mathbf{r}}\epsilon^{-r}e^{-\kappa\max|t_j|}\right].$$

Choosing ϵ so that $\epsilon^{-(r+1)}e^{-\kappa\max|t_j|} = 1$ we get

$$|\mu(A(\cdot)B_+(g\cdot))| \leq C||A||_{\mathbf{s}}\,||B||_{\mathbf{s},\mathbf{r}}e^{-\frac{\kappa\max|t_j|}{r+1}}.$$

Likewise

$$|\mu(A(\cdot)B_-(g\cdot))| \leq C||A||_{\mathbf{s}}\,||B||_{\mathbf{s},\mathbf{r}}e^{-\frac{\kappa\max|t_j|}{r+1}}.$$

The last two inequalities together with (7.9) prove the lemma for non negative A.

In the general case decompose $A = A_1 - A_2$ where

$$A_1 = 2||A||_{L^\infty}, \quad A_2 = A_1 - A.$$

Since both A_1 and A_2 are non-negative we have

$$|\mu(A_j(\cdot)B(g\cdot)) - \mu(A_j)\mu(B)| \leq C\left(||A||_{\mathbf{s}} + ||A||_{L^\infty}\right)||B||_{\mathbf{s},\mathbf{r}}e^{-\frac{\kappa \max|t_j|}{r+1}}$$

proving the lemma in the general case. □

8. Poisson limit for the number of visits to the cusp of the Cartan action

The goal of this section is to prove Theorem 5.3 using the abstract Theorem 6.2 and the polynomial rate of uniform distribution of long pieces of horocycles given in Section 7.

First of all we fix the probability space $(\Omega, \mathbb{P})$ to be the space $(\mathcal{M}, \tilde{\mu})$ where $\tilde{\mu}$ denotes any given probability measure on $\mathcal{M}$ with a smooth density with respect to Haar measure. In all this section, the expectation with respect to $\tilde{\mu}$ of a variable X will be denoted $\mathbb{E}(X)$.

Recall the definitions (5.1) and (5.2) of $\mathcal{P} = \{(t_1, t_2) \in \mathbb{R}^2_+ : t_1 + t_2 < 1\}$ and thus $\Pi_M = \{(t_1, t_2) \in \mathbb{R}^2_+ : t_1 + t_2 < M\}$. Recall the definition of $\lambda_1(t) = 2t_1 + t_2$, and let $\hat{\lambda}(t) = t_1 + M$. Observe that $\hat{\lambda}(t) > \lambda_1(t)$ on Π_M (even though $\hat{\lambda}(t)$ can be equal to $\lambda_1(t)$ on the boundary of Π_M, that is, if $t_1 + t_2 = M$).

We take $(\mathfrak{X}, \mathfrak{m})$ and $(\tilde{\mathfrak{X}}, \tilde{\mathfrak{m}})$ to be the spaces $K = \left[-\frac{1}{\epsilon}, \frac{1}{\epsilon}\right]$ and $\mathbb{R}/2\mathbb{Z}$ equipped with their normalized Lebesgue measures.

Fix any $P \geq 1$ and divide K into a finite number of intervals $K_1, K_2 \dots K_P$ and let Q be the partition of $(\mathfrak{X}, \mathfrak{m})$ into the intervals $K_1, K_2 \dots K_P$.

Similarly, fix any $J \geq 1$ and divide $[0, 2)$ into a finite number of intervals $\tilde{K}_1, \dots, \tilde{K}_J$, and let $\tilde{Q}$ be the partition of $(\tilde{\mathfrak{X}}, \tilde{\mathfrak{m}})$ into the intervals $\tilde{K}_1, \dots, \tilde{K}_J$.

Recall the definitions of Φ, Ψ given in (5.4) (5.5) of Section 5 and introduce Φ_p, $p \in [1, P]$ that are defined by formula (5.4) with K_p in place of K. Let

$$\xi_t = \Phi(g_t\mathcal{L}) \tag{8.1}$$

$$\xi_{t,p} = \Phi_p(g_t\mathcal{L}) \tag{8.2}$$

$$\nu_t = \Psi_1(g_t\mathcal{L}) \tag{8.3}$$

$$\zeta_t = \Psi_2(g_t\mathcal{L}, Ne^{-(t_1+t_2)}) \tag{8.4}$$

Fix R to be a large number (the precise conditions on R are described later in this section).

By Theorem 6.2, if we prove that $\xi_t, \{\xi_{t,p}\}_{p\leq P}, \{\zeta_t\}_{t\leq P}$ satisfy (h1)–(h8) then we get the Poisson limit for

$$\{\Psi(g_t\Lambda(\xi), Ne^{-(t_1+t_2)})\}_{t\in\Pi_M, \Phi(g_t\Lambda(\xi))=1}$$

required in Theorem 5.3.

Proposition 8.1, proven in Section 8.1, shows that ξ_t satisfies (h1)–(h3). The proof relies on Rogers' identities given in Lemma 2.3. Section 8.2 contains estimates of $||\cdot||_{\mathbf{s},\mathbf{s}}$ norms of the functions Φ_i and Ψ_j. Then in Section

8.3 we show, using Proposition 7.1 of Section 7, the existence of the partitions F_t and the sets E and $\tilde{E}$ such that (h4)–(h8) hold.

8.1. **Multiple solutions.** The following proposition asserts that ξ_t satisfies (h1)–(h3).

Proposition 8.1. *Denote $\Phi^t = \Phi \circ g_t$, $\Phi_p^t = \Phi_p \circ g_t$. Then uniformly in $t, t' \in \mathbb{Z}^2 - \{0,0\}$ we have*

(a) $\tilde{\mu}(\Phi^t) = \mathcal{O}(M^{-2})$;

(a′) $\tilde{\mu}(\Phi_p^t) = 2\mathbf{c}_1|K_p|M^{-2} + \mathcal{O}(M^{-100})$
if $\min(t_1, t_2) \geq R \ln M$ *and R is sufficiently large;*

(b) $\tilde{\mu}((\Phi^t)^2 - \Phi^t) = \mathcal{O}(M^{-4})$;

(c) $\tilde{\mu}(\{\mathcal{L} \in \mathcal{M} : \Phi^t(\mathcal{L}) \neq 0 \text{ and } \Phi^t(g_{t'}\mathcal{L}) \neq 0\}) = \mathcal{O}(M^{-4})$.

Note that (b) implies that $\tilde{\mu}(\{\mathcal{L} \in \mathcal{M} : \Phi^t(\mathcal{L}) > 1\}) = \mathcal{O}(M^{-4})$.

Proof. Without loss of generality, we can assume in the proof of the inequalities (a), (b), (c), that $\mathcal{L}$ is distributed according to the Haar measure on M, and by invariance of the Haar measure take $t = 0$. The inequalities then follow from Rogers' equalities of Lemma 2.3. Indeed, part (a) of Lemma 2.3 implies that $\mu(\Phi) = 2\mathbf{c}_1|K|M^{-2}$, since

$$\int_{\mathbb{R}^3} \mathbb{1}_I(x)\mathbb{1}_J(y)\mathbb{1}_K(M^2xyz)dxdydz = |K|M^{-2}.$$

On the other hand, if for $e = (x, y, z) \in \mathcal{L}$, we let $f(e) = \mathbb{1}_{I\times J\times K}(x, y, M^2xyz)$, then since I is an interval of positive numbers, we have that

$$\Phi^2(\mathcal{L}) - \Phi(\mathcal{L}) = \sum_{e_1 \neq e_2 \in L \text{ prime}} f(e_1)f(e_2) = \sum_{e_1 \neq \pm e_2 \in L \text{ prime}} f(e_1)f(e_2) \tag{8.5}$$

and the first estimate of part (b) follows by Lemma 2.3 (b). The second estimate follows from the first by Markov inequality. As for (c) observe that if we define, for $e = (x, y, z) \in \mathcal{L}$,

$$\tilde{f}(e) = \mathbb{1}_{e^{-t_1'}I\times e^{-t_2'}J\times e^{t_1'+t_2'}K}(x, y, M^2xyz),$$

then

$$\mu(\Phi\Phi^{t'}) = \int_{\mathcal{M}} \sum_{e_2 \neq \pm e_1 \in \mathcal{L} \text{ prime}} f(e_1)\tilde{f}(e_2)d\mu(\mathcal{L})$$

where the contribution of $e_2 = -e_1$ vanishes because both I and $e^{-t_1'}I$ are positive intervals, while the contribution of $e_2 = e_1$ vanishes since either I and $e^{-t_1'}I$ or J and $e^{-t_2'}J$ are disjoint. Applying Lemma 2.3(b) we get (c).

Since $\mu(\Phi_p) = \mathbf{c}|K_p|M^{-2}$, (a') follows by exponential mixing of the geodesic flow (Lemma 7.2) and Lemma 8.2 proven in Section 8.2. □

8.2. **Estimates of norms.** Before we construct the partition $\mathcal{F}_t$, we first state estimates on the $H^{\mathbf{s},\mathbf{s}}$ norms of Φ and Φ_p. We also obtain an estimate on the norm of Φ^2 after making an appropriate cutoff. The results stated in this section 8.2 are proven in Section A.2.

Let $\mathfrak{h}_{1,\delta}$ be a smooth cutoff function supported on the set of lattices with a short vector of size $\mathcal{O}(\delta)$. The existence of such function is guaranteed by Lemma A.1. We set $\mathfrak{h}_{2,\delta} = 1 - \mathfrak{h}_{1,\delta}$.

Let $\hat{K} = \{z : d(z, \partial K) \leq M^{-1000}\}$ and set $\hat{\Phi} = \mathcal{S}(\mathbb{1}_I(x)\mathbb{1}_J(y)\mathbb{1}_{\hat{K}}(M^2xyz))$. We define similarly $\hat{K}_p$ and $\hat{\Phi}_p$ for $p \in [1, P]$.

Lemma 8.2. *For any $s \geq 0$ we have that*

(a) $||\Phi||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(1)$, $||\hat{\Phi}||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(1)$. *Also for each $p \in [1, P]$*

$$||\Phi_p||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(1), \quad ||\hat{\Phi}_p||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(1).$$

(b) For each $\delta > 0$, $||\Phi\mathfrak{h}_{1,\delta}||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(1)$.

(c) For each $\delta > 0$, $||(\Phi^2 - \Phi)\mathfrak{h}_{2,\delta}||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(\delta^{-6})$.

(d) $\mu(\Phi_i) = 2\mathbf{c}|K_i|, \quad \mu(\hat{\Phi}_i) = \mathcal{O}(M^{-1000})$.

(e) $\mu(\Phi\mathfrak{h}_{1,\delta}) = \mathcal{O}\left(\delta^{3/2}\right)$.

(f) $\mu((\Phi^2 - \Phi)\mathfrak{h}_{2,\delta}) = \mathcal{O}\left(M^{-4}\right)$.

8.3. **The partition $\mathcal{F}_t$ and the proof of (h4)–(h8).** Given $t \in \Pi$ we denote by $\Pi^+(t)$ the set of $\bar{t} \in \Pi$ such that $\lambda_1(\bar{t}) > \lambda_1(t) + R \ln M$.

Consider the following collection of functions

$$\mathbf{\Phi} = \{\Phi, \Phi_1 \dots \Phi_P, \hat{\Phi}, \hat{\Phi}_1, \dots, \hat{\Phi}_P, \Phi\mathfrak{h}_{1,M^{-1000}}, (\Phi^2 - \Phi)\mathfrak{h}_{2,M^{-1000}}\}.$$

Let F_t be a partition of $\mathcal{M}$ into W_1-curves of size $L_t = (e^{\lambda_1(t)}M^{1000})^{-1}$, which is κ_0-representative with respect to $(\Pi^+(t), \mathbf{\Phi})$ (that is, representative for all $\bar{t} \in \Pi^+(t)$). This is possible to do due to Proposition 7.1, Lemma 8.2 and the fact that

$$\sum_{t\in\Pi,\ \bar{t}\in\Pi^+(t)} \left(L_t e^{\lambda_1(\bar{t})}\right)^{-\kappa_0} = \mathcal{O}(M^{-10^{100}})$$

if R is sufficiently large. Moreover, if we let E_1 be the set of $\mathcal{L}$ such that for any $t \in \Pi$, $\bar{t} \in \Pi^+(t)$, $\mathcal{L}$ is representative with respect to $(\mathcal{F}_t, \bar{t}, \mathbf{\Phi})$ then we have that $\mu(E_1^c) = \mathcal{O}(M^{-100})$.

Proposition 8.3. *There exist sets $E, \tilde{E}$ with*

$$\tilde{\mu}(E^c) \leq CM^{-100}, \quad \tilde{\mu}(\tilde{E}^c) \leq CM^{-100},$$

such that the variables $\xi_t, \eta_t, \xi_{t,p}, \eta_{t,p}$ and the partitions F_t satisfy the properties (h4)–(h8) of Section 6.

Proof.
Property (h4). We prove that any $\mathcal{L} \in E_1$ satisfies (h4).

Properties (h4a) and (h4b) follow from parts (a) and (a') of Proposition 8.1 and the definition of representative points.

To check (h4c) note that $\xi_{t,p}$ are integer valued and so

$$\xi_{t,p} - \eta_{t,p} \leq \xi_{t,p}^2 - \xi_{t,p} \leq \xi_t^2 - \xi_t.$$

Since also $0 \leq \xi_{t,p} - \eta_{t,p} \leq \xi_{t,p} \leq \xi_t$ we get

$$0 \leq \xi_{t,p} - \eta_{t,p} \leq \hat{\xi}_t$$

where

$$\hat{\xi}_t = \left[(\Phi^2 - \Phi)\mathfrak{h}_{2,M^{-1000}} + \Phi\mathfrak{h}_{1,M^{-1000}}\right] \circ g_t.$$

Accordingly for $\mathcal{L} \in E_1$

$$0 \leq \mathbb{E}(\xi_{t,p} - \eta_{t,p}|\mathcal{F}_{t'}) \leq \mathbb{E}(\hat{\xi}_t|\mathcal{F}_{t'}) \leq \frac{C}{M^4}$$

where the last inequality relies on parts (e) and (f) of Lemma 8.2 and the fact that $\mathcal{L}$ is representative with respect to $(F_{t'}, t, \mathbf{\Phi})$. The last display implies that

$$\mathbb{E}(\xi_{t,p}|\mathcal{F}_{t'}) - \frac{C}{M^4} \leq \mathbb{E}(\eta_{t,p}|\mathcal{F}_{t'}) \leq \mathbb{E}(\xi_{t,p}|\mathcal{F}_{t'}).$$

Hence (h4c) follows from (h4b).

Property (h5). For $t \in \Pi$ and $\bar{t} \in \Pi^+(t)$, and $\gamma_{\bar{t}} \in F_{\bar{t}}$ it holds that $g^t\gamma_{\bar{t}}$ has length less than M^{-10^9}. Hence if $g^t\gamma_{\bar{t}}$ does not intersect

$$\hat{K}_p = \{z : d(z, \partial K_p) \leq M^{-10000}\}$$

then it is completely contained in K_p or K_p^c. The measure of $\mathcal{L}$ such that $g_t\gamma_{\bar{t}}$ intersects $\hat{K}_p$ for some p is thus bounded by $\mathcal{O}(M^{-1000})$ from Lemma 8.2(b). Taking the complement to the union of all these exceptional $\mathcal{L}$ for all $t \in \Pi, \bar{t} \in \Pi^+(t)$ we get a set E_2 such that $\tilde{\mu}(E_2^c) = \mathcal{O}(M^{-996})$ and (h5) holds for $\mathcal{L} \in E_2$.

Property (h6). Since the size of the pieces of F_t is $L_t = (e^{\lambda_1(t)}M^{1000})^{-1}$ and the size of the pieces of $F_{\bar{t}}$ is $L_{\bar{t}} = (e^{\lambda_1(\bar{t})}M^{1000})^{-1}$ if we let

$$E_3 = \{\mathcal{L} : F_{\bar{t}}(\mathcal{L}) \subset F_t(\mathcal{L}) \text{ for all } t \in \Pi, \bar{t} \in \Pi^+(t)\}$$

then we have $\tilde{\mu}(E_3^c) = \mathcal{O}(M^{-1000})$.

We let $E := E_1 \cap E_2 \cap E_3$ and observe that $\tilde{\mu}(E^c) = O(M^{-100})$ and (h4), (h5), and (h6) hold on E.

Property (h7). Let $\tilde{K}_1, \ldots, \tilde{K}_J$ be a partition of $\mathbb{R}/2\mathbb{Z}$ with J intervals. We will show that $\mathcal{L} \in E$ satisfies (h7). Then we further refine E to ensure (h8).

Assume t, t' are such that $\hat{\lambda}(t) \geq \lambda_1(t') + R\ln M \geq \lambda_1(t) + 2R\ln M$. We need to show that if $\mathcal{L} \in E$ is such that $\xi_t = \xi_{t,p} = 1$ then for any $j \in [1, J]$

$$\tilde{\mu}\left(\Psi_2\left(g_t\mathcal{L}, \frac{N}{e^{t_1+t_2}}\right) \in \tilde{K}_j \middle| \mathcal{F}_{t'}\right)(\mathcal{L}) = |\tilde{K}_j|(1 + o(1)). \tag{8.6}$$

Let $\gamma_{t'} = F_{t'}(\mathcal{L})$. Then $\gamma_{t'}$ is of the form

$$\gamma_{t'} = \{h_\tau^1\bar{\mathcal{L}}\}_{0\leq\tau\leq(e^{\lambda_1(t')}M^{1000})^{-1}}$$

for some $\bar{\mathcal{L}} \in \mathcal{M}$. By property (h5) $\xi_{t,p} = 1$ on $\gamma_{t'}$. In particular, $\Phi_p(g_t\bar{\mathcal{L}}) = 1$, that is $g_t\bar{\mathcal{L}}$ contains a vector (x, y, z) such that

$$(x, y, xyz) \in I \times J \times K_p.$$

Since $g_t h_\tau = h_{e^{\lambda_1(t)}\tau} g_t$ it follows that $g_t h_\tau \bar{\mathcal{L}}$ contains the vector (x_τ, y_τ, z_τ) such that $(x_\tau, y_\tau, x_\tau y_\tau z_\tau) \in I \times J \times K_p$, namely,

$$x_\tau = x, \quad y_\tau = y, \quad z_\tau = z + e^{\lambda_1(t)}\tau x.$$

Now

$$\begin{aligned}\Psi_2\left(g_t h^1_\tau \bar{\mathcal{L}}, \frac{N}{e^{t_1+t_2}}\right) &= \frac{N}{e^{t_1+t_2}} z_\tau \bmod (2) \\ &= \frac{N}{e^{t_1+t_2}} z + e^{\hat{\lambda}(t)}\tau x \bmod (2)\end{aligned}$$

and since τ varies on an interval of length $(e^{\lambda_1(t')}M^{1000})^{-1}$, the uniform distribution (8.6) follows from the fact that $\hat{\lambda}(t) \geq \lambda_1(t') + R\ln M$ provided that R is sufficiently large.

Property (h8). $F_{\bar{t}}(\mathcal{L})$ is of the form

$$\{h^1_\tau \bar{\mathcal{L}}\}_{0 \leq \tau \leq (e^{\lambda_1(\bar{t})}M^{1000})^{-1}}$$

for some $\bar{\mathcal{L}} \in \mathcal{M}$. By (h5), $\xi_t = 1$ on $F_{\bar{t}}(\omega)$. The computation from property (h7) gives

$$\Psi_2\left(g_t h^1_\tau \bar{\mathcal{L}}, \frac{N}{e^{t_1+t_2}}\right) = \frac{N}{e^{t_1+t_2}} z + e^{\hat{\lambda}(t)}\tau x \bmod (2).$$

Accordingly if $\mathbb{1}_{\zeta_t \in \tilde{K}_j}$ is not constant on $F_{\bar{t}}(\mathcal{L})$ then $\Psi_2\left(g_t h^1_\tau \bar{\mathcal{L}}, \frac{N}{e^{t_1+t_2}}\right)$ lies in a $\mathcal{O}(M^{-1000})$ neighborhood of $\partial\tilde{K}_j$. Let

$$\begin{aligned}\tilde{E} = \Big\{\mathcal{L} \in E : \forall t, \bar{t} \in \Pi \text{ with } \lambda_1(t) \geq \hat{\lambda}(t) + R\ln M \text{ it holds that :} \\ \forall j \in [1, J], \quad \mathbb{1}_{\zeta_t \in \tilde{K}_j} \text{ is constant on } F_{\bar{t}}(\mathcal{L})\Big\}.\end{aligned}$$

Then (h8) holds on $\tilde{E}$. On the other hand, the same argument as for property (h5) shows that $\tilde{\mu}(E - \tilde{E}) = \mathcal{O}(M^{-996})$ as needed.

We have thus checked (h4)–(h8) for ξ_t, $\{\xi_{t,p}\}$, $\{\zeta_t\}$ for $t \in \Pi$ and $p \in [1, P]$, which completes the proof of Proposition 8.3. □

Proof of Theorem 5.3. Note that part (a) of Theorem 5.3 is exactly property (h4a) that we proved in Proposition 8.3. Part (b) of Theorem 5.3 follows from Theorem 6.2 and from properties (h1)–(h8) that we proved in Propositions 8.1 and 8.3. □

9. Small boxes

One can also consider the visits to small boxes $\mathcal{C}_N = \prod_j \left[-\frac{u_j}{N^\gamma}, \frac{u_j}{N^\gamma}\right]$. The case $\gamma = 0$ is treated in Theorem 1.1 while the case $\gamma = 1/d$ was studied in [23]. For $\gamma > 1/d$ most orbits do not visit $\mathcal{C}_N$ so we consider the remaining case $0 < \gamma < \frac{1}{d}$.

Theorem 9.1. *Under the assumptions of Theorem 1.1,* $\frac{D(x,\alpha,\mathcal{C}_N,N)}{\boldsymbol{\rho}((1-d\gamma)\ln N)^d}$ *converges to the standard Cauchy distribution.*

The proof of Theorem 9.1 is the same as the proof of Theorem 1.1 except that now, and with similar computations as in Section 3.3, we can neglect in the Fourier series of the discrepancy the frequencies k such that $|\bar{k}_j| < N^\gamma$ for some j. Accordingly, in Theorem 5.3, $\mathcal{P}$ has to be replaced by

$$\mathcal{P}_\gamma = \{t : t_j > \gamma M, \sum_j t_j < M\}$$

which decreases the intensity of the limiting Poisson process by a factor $(1 - d\gamma)^d$.

10. Continuous time

In this section we discuss briefly the behavior of the discrepancy function in the case of linear flows on the torus. Given a set $\mathcal{C}$ we consider the continuous time discrepancy function as

$$\mathbf{D}(v, x, \mathcal{C}, T) = \int_0^T \mathbb{1}_{\mathcal{C}}(S_v^t x)dt - T\text{Vol}(\mathcal{C})$$

where $S_v^t = x + vt$.

In the case of balls, it was shown in [8] that for $d \geq 4$, the continuous time discrepancy function has a similar behavior as the discrete time discrepancy, namely it converges in distribution after normalization by a factor $T^{(d-3)/2(d-1)}$.

Curiously, for balls in dimension $d = 3$, the continuous time discrepancy behaves similarly to the discrete discrepancy of boxes and gives rise to a Cauchy distribution after normalization by $\ln T$. This will be proved in Section 10.2 below.

It was also shown in [8] that for balls in dimension $d = 2$ the continuous time discrepancy converges, without any normalization, in distribution. In the next Section 10.1 we will show that this is also the case in any dimension $d \geq 2$ for the continuous time discrepancy for boxes.

10.1. **Boxes.** Let $\mathcal{C} = A(\prod_j (0, u_j))$. We assume that the triple (A, x, v) is distributed according to a smooth density of compact support and that $A \in \mathrm{SL_d}(\mathbb{R})$ is such that $||A - I|| \leq \eta$ where η is sufficiently small.

Theorem 10.1. *As* $T \to \infty$, $\mathbf{D}(v, x, \mathcal{C}, T)$ *converges in distribution.*

Proof. We have

$$\mathbf{D}(v,x,\mathcal{C},T)=4^d\sum_k\left[\prod_j\left(\frac{\sin\left(2\pi\bar{k}_j u_j\right)}{\bar{k}_j}\right)\right]\frac{\sin(\pi(k,vT))}{\pi(k,v)}\cos(2\pi(k,x)+\phi_{k,T,v}).$$

where $\bar{k}_j$ is given by (3.1). We claim that for almost all A, v there exist a constant $C(A,v)$ such that

$$||\mathbf{D}(v,x,\mathcal{C},T)||_{L^2_x}\leq C(A,v)$$

and moreover for each ϵ there exists $N=N(A,v)$ such that

$$\left\|\sum_{|k|>N}\left[\prod_j\left(\frac{\sin\left(2\pi\bar{k}_j u_j\right)}{\bar{k}_j}\right)\right]\frac{\sin(\pi(k,vT))}{\pi(k,v)}\cos(2\pi(k,x)+\phi_{k,T,v})\right\|_{L^2_x}\leq\epsilon.$$

To this end it suffices to demonstrate that for almost every (A,v)

$$\sum_k\left(\left(\prod_j\bar{k}_j\right)(k,v)\right)^{-2}<\infty.$$

Since $\det(A)\neq 0$ there exists $\delta(A)$ such that for each k there is $l\in\{1\dots d\}$ such that $|\bar{k}_l|>\delta|k|$. Accordingly it suffices to check that for each l

$$\sum_k\Gamma_k(A,v)<\infty\quad\text{where}\quad\Gamma_k(A,v)=\left(\left(\prod_{j\neq l}\bar{k}_j\right)(k,v)|k|\right)^{-2}.$$

All sums have the same form so we consider the case $l=d$. Given numbers $s_1,\dots s_{d-1},s_d$ and $\epsilon>0$ denote $\Omega(k,s_1\dots s_d)=$

$\{(A,v):|\bar{k}_j|\in[|k|^{s_j},|k|^{s_j+\epsilon}]$ for $j=1,\dots,d-1$ and $|(v,k)|\in[|k|^{s_d},|k|^{s_d+\epsilon}]\}$.

Then

$$\mathbb{P}(\Omega(k,s_1\dots s_d))\leq C|k|^{s+d\epsilon-d}$$

where $s=\sum_{j=1}^d s_j$. We draw two conclusions from this estimate. First, for almost all (A,v) we have

$$\left|\left(\prod_{j=1}^{d-1}\bar{k}_j\right)(k,v)\right|>|k|^{-2d\epsilon}$$

provided that $|k|$ is sufficiently large.

Second, for $s\geq-2d\epsilon$ we have

$$\mathbb{E}(\mathbb{1}_{\Omega(k,s_1\dots s_d)}(A,v)\Gamma_k(A,v))\leq C|k|^{d\epsilon-[(d+2)+s]}.$$

Hence

$$\mathbb{E}\left(\sum_k\mathbb{1}_{\Omega(k,s_1\dots s_d)}(A,v)\Gamma_k(A,v)\right)<\infty.$$

Summing over all d-tuples $(s_1 \dots s_d) \in (\epsilon\mathbb{Z})^d$ such that

$$s_j \leq 1, \quad s = \sum_{j=1}^{d} s_j > -2d\epsilon$$

we get $\mathbb{E}\left(\sum_k \Gamma_k(A, v)\right) < \infty$ proving our claim.

The claim implies that for large N the distribution of $\mathbf{D}(v, x, \mathcal{C}, T)$ is close to the distribution of

$$\mathbf{D}_N^-(v, x, \mathcal{C}, T) = 4^d \sum_{|k|\leq N} \prod_j \left(\frac{\sin\left(2\pi \bar{k}_j u_j\right)}{\bar{k}_j} \right) \frac{\sin(\pi(k, vT)}{\pi(k, v)} \cos(2\pi(k, x) + \phi_{k,T,v}).$$

Hence it remains to prove that $\mathbf{D}_N^-(v, x, \mathcal{C}, T)$ converges in distribution as $T \to \infty$. This convergence follows easily from the fact that as $T \to \infty$ $\{vT\}$ becomes uniformly distributed on $(\mathbb{R}/2\mathbb{Z})^d$. □

A similar argument shows that randomness in $\mathcal{C}$ is not necessary. Namely, we have the following result.

Theorem 10.2. *Let $\mathcal{C} = \prod_j (0, u_j)$. Suppose that the pair (x, v) has a smooth distribution of compact support. Then $\mathbf{D}(v, x, \mathcal{C}, T)$ converges in distribution as $T \to \infty$.*

The proof of Theorem 10.2 is similar to the proof of Theorem 10.1 with the additional simplifications since now $|\bar{k}_j| \geq 1$ and so only (k, v) may possibly be small. Therefore we leave the proof to the reader.

10.2. **Balls.** In this section, $\mathcal{C}$ is assumed to be a ball of radius r in $\mathbb{T}^3$. We suppose that v is chosen according to a smooth density p whose support is compact and does not contain the origin, r is uniformly distributed on some segment $[a, b]$, x is uniformly distributed on $\mathbb{T}^3$ and v, r and x are independent.

Theorem 10.3. *There exists a constant $\tilde{\rho}$ such that $\frac{\mathbf{D}(v,x,B(0,r),T)}{\tilde{\rho} r \ln T}$ converges as $T \to \infty$ to the standard Cauchy distribution.*

Proof. The proof is similar to the proof of Theorem 1.1 so we just outline the main steps. We have

$$\mathbf{D}(v, x, B(0, r), T) = \sum_{k\in\mathbb{Z}^3} f_k(r, v, x, T) = \sum_{k\in\mathbb{Z}^3, k \text{ prime}} g_k$$

where $f_k = c_k \frac{\cos[2\pi(k,x)+\pi(k,Tv)] \sin(\pi(k,Tv))}{\pi(k,v)}$, $g_k = \sum_{p=1}^{\infty} f_{kp}$ and

$$c_k \sim \frac{r}{\pi|k|^2} \sin(2\pi r|k|).$$

Similarly to Section 3 we show that the main contribution to the discrepancy comes from the harmonics where $\frac{\epsilon}{\ln T} < |(k, v)||k|^2 < \frac{1}{\epsilon \ln T}$ and $|k| < T$. Therefore the key step in proving Theorem 10.3 is the following.

Proposition 10.4. *The point process*

$$\left\{|k|^2(k,v)\ln T, (k,Tv) \bmod 2, \{(k,x)\}, \{r|k|\}\right\}_{|k|\leq T, \epsilon k^2|(k,v)|\ln T<1, k\ prime}$$

converges as $T\to\infty$ *to a Poisson process on* $[-\frac{1}{\epsilon},\frac{1}{\epsilon}]\times(\mathbb{R}/2\mathbb{Z})\times(\mathbb{R}/\mathbb{Z})^2$ *with constant intensity.*

The proof of Proposition 10.4 is similar to the proof of Theorem 4.3 and consists of the following steps.

(a) We prove the Poisson limit for $\{|k|^2(k,v)\ln T\}$ using the argument of Section 8. We first normalize one of the coordinates, say v_3, of the vector v to 1, which reduces the study of the Poisson limit for $\{|k|^2(k,v)\ln T\}$ to the study of the visits to the cusp in $\mathcal{M}=\mathrm{SL}_3(\mathbb{R})/\mathrm{SL}_3(\mathbb{Z})$ of $g_t\Lambda$ with

$$g_t=\begin{pmatrix} e^t & 0 & 0\\ 0 & e^t & 0\\ 0 & 0 & e^{-2t}\end{pmatrix} \text{ and } \Lambda(v_1,v_2)=\begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ v_1 & v_2 & 1\end{pmatrix}.$$

More precisely, the relevant neighborhood in the cusp is defined *via* the function

$$f(x,y,z)=\mathbb{1}_I(x^2+y^2)\mathbb{1}_K((x^2+y^2)z)$$

where $I=[1,e)$, $K=[-\frac{1}{\epsilon\ln T},\frac{1}{\epsilon\ln T}]$. We then define $\Phi(\mathcal{L})=\mathcal{S}(f)$.

The Poisson limit of $\{|k|^2(k,v)\ln T\}$ is obtained from a Poisson limit for

$$\{\Phi(g_t\Lambda)\}_{t\in[0,\ln T],}.$$

In this setting, the manifold determined by $\Lambda(v_1,v_2)$ consists of the full strong unstable foliation of g_t and there is no need for extra parameters to establish the Poisson limit.

(b) We prove that $(k,Tv) \bmod 2$ is asymptotically independent of $|k|^2(k,v)\ln T$ using the fact that their values are determined at different scales (cf. proof of (h7) in Section 8).

(c) We show that (k,x) and $\{r|k|\}$ are independent of the previous data using the superlacunarity of the sequence of small denominators (cf. Proposition 5.2). □

Appendix A. Norms.

A.1. **Preliminaries.** It is well known that the fluctuation of ergodic integrals depends strongly on the regularity properties of the observables. To gauge such regularity we will need several norms on the space of lattices.

Let $C^{\mathbf{s}}(\mathcal{M}_{d+1})$ denote the space of smooth functions on $\mathcal{M}_{d+1}$. Let $\mathfrak{U}_1,\mathfrak{U}_2,\dots,\mathfrak{U}_{(d+1)^2-1}$ be a basis in the space of left invariant vectorfields on $\mathcal{M}_{d+1}$. We let

$$||\Phi||_{C^{\mathbf{s}}}=\max_{0\leq k\leq \mathbf{s}}\max_{i_1,i_2\dots i_k}\max_{\mathcal{L}\in\mathcal{M}_{d+1}}\left|\partial_{\mathfrak{U}_{i_1}}\partial_{\mathfrak{U}_{i_2}}\dots\partial_{\mathfrak{U}_{i_k}}\Phi(\mathcal{L})\right|.$$

Let $H^{\mathbf{s}}$ denote the Sobolev space of index $\mathbf{s}$. It is equipped with the norm

$$||\Phi||_s^2=\sum_{0\leq k\leq\mathbf{s}}\sum_{i_1,i_2\dots i_k}\int\left|\partial_{\mathfrak{U}_{i_1}}\partial_{\mathfrak{U}_{i_2}}\dots\partial_{\mathfrak{U}_{i_k}}\Phi(\mathcal{L})\right|^2 d\mu(\mathcal{L}).$$

Let $\mathbf{a}(\mathcal{L})$ denote the length of the shortest nonzero vector in $\mathcal{L}$.

Lemma A.1. *For each* $\mathbf{s}$ *there are constants* C_1, C_2 *such that for each* $\delta \leq 1$ *there is a function* $\mathfrak{h}_{1,\delta} : \mathcal{M} \to \mathbb{R}$ *such that*

- $0 \leq \mathfrak{h}_{1,\delta} \leq 1$,
- $\mathfrak{h}_{1,\delta}(\mathcal{L}) = 1$ *if* $\mathbf{a}(\mathcal{L}) \leq \delta$,
- $\mathfrak{h}_{1,\delta}(\mathcal{L}) = 0$ *if* $\mathbf{a}(\mathcal{L}) \geq C_1\delta$,
- $||\mathfrak{h}_{1,\delta}||_{C^{\mathbf{s}}(\mathcal{M})} \leq C_2$.

Proof. This lemma is a special case of [21, Section 4.2]. For completeness, we reproduce the formula from [21]. Let Υ be a nonnegative function on $SL_{d+1}(\mathbb{R})$ with integral one supported on the set

$$||g||^2 \leq C_1, \quad ||g||^{-2} \leq C_1.$$

Then one can set

$$\mathfrak{h}_{1,\delta}(\mathcal{L}) = \int_{SL_{d+1}(\mathbb{R})} \Upsilon(g)\ \mathbb{1}_{\mathbf{a}(g\mathcal{L}) \leq C_1\delta}\ d\mu(g). \quad \square$$

We also need a space $H^{\mathbf{s},\mathbf{r}}$ of functions on $\mathcal{M}_{d+1}$ which can be well approximated by $H^{\mathbf{s}}$ functions (see Definition 7.1). Similar norms can be introduced on the Euclidean space $\mathbb{R}^{d+1}$. We note the following inequalities for $\Phi, \Psi \in C^{\mathbf{s}}(\mathcal{M}_{d+1})$

$$||\Phi||_{\mathbf{s}} \leq C_3||\Phi||_{C^{\mathbf{s}}}, \tag{A.1}$$

$$|||\Psi\Phi||_{\mathbf{s}} \leq C_4||\Psi||_{C^{\mathbf{s}}}||\Phi||_{\mathbf{s}} \tag{A.2}$$

where the constants C_3 and C_4 depend on $\mathbf{s}$. Accordingly if $\Psi \in C^{\mathbf{s}}(\mathcal{M}_{d+1})$ is positive and $\Phi \in H^{\mathbf{s},\mathbf{r}}$ we get

$$|||\Psi\Phi||_{\mathbf{s},\mathbf{r}} \leq C_5||\Psi||_{C^{\mathbf{s}}}||\Phi||_{\mathbf{s},\mathbf{r}} \tag{A.3}$$

(In fact, (A.3) holds for arbitrary smooth Ψ since Ψ can be represented as a difference of two smooth positive functions, but we will only need (A.3) for positive Ψ.)

We also need a space $C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^{d+1})$ which is defined similarly to $H^{\mathbf{s},\mathbf{r}}$.

Namely, given $\mathbf{s}, \mathbf{r} \geq 0$, we say that a function $f : \mathbb{R}^{d+1} \to \mathbb{R}$ is in $C^{\mathbf{s},\mathbf{r}}$ with $||f||_{C^{\mathbf{s},\mathbf{r}}} = K$ if given $0 < \epsilon \leq 1$ there are $C^{\mathbf{s}}$-functions $f^- \leq f \leq f^+$ such that

$$||f^+ - f^-||_{L^1(\mathbb{R}^{d+1})} \leq \epsilon \text{ and } ||f^{\pm}||_{C^{\mathbf{s}}(\mathbb{R}^{d+1})} \leq K\epsilon^{-\mathbf{r}}.$$

Lemma A.2. *For each integer* $\mathbf{s}$ *and each* R *there is a constant* $C = C(R, \mathbf{s})$ *such that:*

(a) If f *is a* $C^{\mathbf{s}}(\mathbb{R}^{d+1})$ *function supported in the ball of radius* R *about the origin then*

$$||\mathcal{S}(f)||_{H^{\mathbf{s}}(\mathcal{M}_{d+1})} \leq C||f||_{C^{\mathbf{s}}(\mathbb{R}^{d+1})}. \tag{A.4}$$

and if f is a $C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^{d+1})$ function supported in the ball of radius R about the origin then

$$||\mathcal{S}(f)||_{H^{\mathbf{s},\mathbf{r}}(\mathcal{M}_{d+1})} \leq C||f||_{C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^{d+1})}; \tag{A.5}$$

(b) Let $\mathfrak{h}_{2,\delta} = 1 - \mathfrak{h}_{1,\delta}$ where $\mathfrak{h}_{1,\delta}$ is a function from Lemma A.1. Let f be a $C^{\mathbf{s}}(\mathbb{R}^{d+1})$ function supported in the ball of radius R about the origin. Then

$$||\mathcal{S}(f)h_{2,\delta}||_{C^{\mathbf{s}}(\mathcal{M}_{d+1})} \leq C||f||_{C^{\mathbf{s}}(\mathbb{R}^{d+1})}\ \delta^{-(d+1)}.$$

Proof. Given a left invariant vectorfield $\mathfrak{U}$ on $\mathrm{SL}_{d+1}(\mathbb{R})$ let $\bar{\mathfrak{U}}$ be the corresponding left invariant vectorfield on $\mathbb{R}^{d+1}$. That is

$$(\partial_{\bar{\mathfrak{U}}}f)(x) = \frac{d}{dt}\Big|_{t=0} f(g(t)x)$$

where $g(t)$ is a one parameter subgroup of $\mathrm{SL}_{d+1}(\mathbb{R})$ such that $g'(0) = \mathfrak{U}$.

Since $\partial_{\mathfrak{U}}\mathcal{S}(f) = \mathcal{S}(\partial_{\bar{\mathfrak{U}}}f)$, (A.4) follows from Lemma 2.3(c).

(A.5) follows from (A.4) since $f^- \leq f \leq f^+$ implies

$$\mathcal{S}(f^-) \leq \mathcal{S}(f) \leq \mathcal{S}(f^+).$$

Next

$$|\mathcal{S}(f)(\mathcal{L})\mathfrak{h}_{2,\delta}(\mathcal{L})| \leq \mathbb{1}_{\mathbf{a}(\mathcal{L})\geq\delta} \sum_{e\in\mathcal{L},\ \text{prime}} |f(e)|.$$

However if the shortest vector in $\mathcal{L}$ is longer than δ there are at most $\mathcal{O}(\delta^{-(d+1)})$ terms contributing to this sum. Accordingly

$$||\mathcal{S}(f)h_{2,\delta}||_{C^0(\mathcal{M}_{d+1})} \leq C||f||_{C^0(\mathbb{R}^{d+1})}\ \delta^{-(d+1)}.$$

The higher derivatives are estimated similarly. □

A.2. **Proof of results of Section 8.2.**

Proof of Lemma 8.2. (a) Let ϕ be a C^∞ function such that $\phi(z) = 1$ for $z \leq 0$, $\phi(z) = 0$ for $z \geq 1$ and $0 \leq \phi(z) \leq 1$ for $0 \leq z \leq 1$. Given an interval $K = [k_1, k_2]$ let

$$\phi^+_{K,\epsilon}(z) = \frac{1}{2}\left[\phi\left(\frac{z-k_2}{\epsilon}\right) - \phi\left(\frac{z-k_1+\epsilon}{\epsilon}\right)\right]$$

$$\phi^-_{K,\epsilon}(z) = \frac{1}{2}\left[\phi\left(\frac{z-k_2+\epsilon}{\epsilon}\right) - \phi\left(\frac{z-k_1}{\epsilon}\right)\right].$$

Consider the following functions on $\mathbb{R}^3$

$$f^\pm_\epsilon(x,y,z) = \phi^\pm_{I,\epsilon}(x)\phi^\pm_{J,\epsilon}(y)\phi^\pm_{K_M,\epsilon}(xyz),$$

where $K_M = [-\frac{1}{\epsilon M^2}, \frac{1}{\epsilon M^2}]$, and define as in 5.4 the Siegel transforms $\Phi^\pm_\epsilon$ of $f^\pm_\epsilon$ instead of $f = \mathbb{1}_I(x)\mathbb{1}_J(y)\mathbb{1}_{K_M}(xyz)$. Since $f^-_\epsilon \leq f \leq f^+_\epsilon$ we conclude that $||f||_{C^{\mathbf{s},\mathbf{s}}(\mathbb{R}^3)} = \mathcal{O}(1)$. Now (A.5) shows that $||\Phi||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(1)$. The norms of $\hat{\Phi}, \Phi_p$ and $\hat{\Phi}_p$ are estimated similarly.

Part (b) follows from part (a) and (A.3).

Next, (8.5) shows that

$$(\Phi_\epsilon^-)^2 - \Phi_\epsilon^- \leq \Phi^2 - \Phi \leq (\Phi_\epsilon^+)^2 - \Phi_\epsilon^+.$$

Thus

$$\mathfrak{h}_{2,\delta}\left[(\Phi_\epsilon^-)^2 - \Phi_\epsilon^-\right] \leq \mathfrak{h}_{2,\delta}\left[\Phi^2 - \Phi\right] \leq \mathfrak{h}_{2,\delta}\left[(\Phi_\epsilon^+)^2 - \Phi_\epsilon^+\right].$$

We have

$$\begin{aligned}&\mu\left(\mathfrak{h}_{2,\delta}\left(\left[(\Phi_\epsilon^+)^2 - \Phi_\epsilon^+\right] - \left[(\Phi_\epsilon^-)^2 - \Phi_\epsilon^-\right]\right)\right)\\ &\quad\leq \mu\left(\left[(\Phi_\epsilon^+)^2 - \Phi_\epsilon^+\right] - \left[(\Phi_\epsilon^-)^2 - \Phi_\epsilon^-\right]\right) = \mathcal{O}(\epsilon).\end{aligned}$$

where the last step relies on Lemma 2.3(c).

Next, similarly to the proof of Lemma A.2 (b) we get

$$||((\Phi_\epsilon^\pm)^2 - \Phi_\epsilon^\pm)h_{2,\delta}||_{s,s} = \mathcal{O}\left(\delta^{-6}\right)$$

proving part (c).

Part (d) follows directly from Lemma 2.3(d).

To prove part (e) we note that

$$\mu(\Phi\mathfrak{h}_{1,\delta}) \leq \sqrt{\mu(\Phi^2)\mu(\mathfrak{h}_{1,\delta}^2)} \leq C\sqrt{\mu(\mathfrak{h}_{1,\delta}^2)} \leq \bar{C}\delta^{3/2}$$

where the estimate of $\mu(\Phi^2)$ follows from Lemma 2.3(c) and the estimate of $\mu(\mathfrak{h}_{1,\delta}^2)$ follows from the fact that

$$\mathfrak{h}_{1,\delta} \leq \mathcal{S}(\mathbb{1}_{x^2+y^2+z^2\leq(C_1\delta)^2})$$

and Lemma 2.3(c).

Finally part (f) follows from Proposition 8.1(b) since $\mathfrak{h}_{2,\delta} \leq 1$. □

References

[1] M. Abadi, N. Vergne *Sharp error for point-wise Poisson approximations in mixing processes,* Nonlinearity **21** (2008) 2871–2885.

[2] J. Beck, *Probabilistic Diophantine approximation. I. Kronecker sequences*, Ann. of Math. (2) **140** (1994), no. 2, 449–502.

[3] R. Bellman, *Research problem # 6*, Bull. Am. Math. Soc. **64** (1958) 60.

[4] A. Bufetov, G. Forni *Limit Theorems for Horocycle Flows,* Ann. ENS **47** (2014) 851–903.

[5] J.-R. Chazottes, P. Collet *Poisson approximation for the number of visits to balls in non-uniformly hyperbolic dynamical systems,* Ergodic Theory Dynam. Systems **33** (2013) 49–80.

[6] M. Denker, M. Gordin, A. Sharova. *A Poisson limit theorem for toral automorphisms,* Illinois J. Math. 48 (2004) 1–20.

[7] D. Dolgopyat *Limit Theorems for partially hyperbolic systems,* Trans. AMS **356** (2004) 1637–1689.

[8] D. Dolgopyat, B. Fayad, *Deviations of ergodic sums for toral translations-I: Convex bodies,* GAFA **24** (2014) 85–115.

[9] D. Dolgopyat, B. Fayad, *Limit Theorems for toral translations,* Proc. Symp. Pure Math. **89** (2015) 227–277.

[10] Feller W. *An introduction to probability theory and its applications,* Vol. **II** 2d ed. John Wiley & Sons, New York-London-Sydney, 1971, 669 pp.

[11] J. M. Freitas, N. Haydn, M. Nicol *Convergence of rare event point processes to the Poisson process for planar billiards,* Nonlinearity **27** (2014) 1669–1687.

[12] G. H. Hardy, J. E. Littlewood *Some Problems of Diophantine Approximation: The Lattice–Points of a Right–Angled Triangle,* Proc. London Math. Soc. **3** (1922) 15-36.
[13] M. Hirata *Poisson law for Axiom A diffeomorphisms,* Ergod. Th. Dynam. Sys. **13** (1993) 533–556.
[14] M. Hirata *Poisson law for the dynamical systems with self-mixing conditions,* In *Dynamical systems and chaos,* Ed. N. Aoki, K. Shiraiwa & Y. Takahashi, (1995) v. **1** 87–96. World Scientific, Singapore.
[15] F. Huveneers *Subdiffusive behavior generated by irrational rotations,* Erg Th. Dyn. Sys. **29** (2009) 1217–1233.
[16] H. Kesten *Uniform distribution* mod 1*,* Ann. of Math. **71** (1960) 445–471.
[17] H. Kesten *Uniform distribution* mod1*, II,* Acta Arith. **7** (1961/1962) 355–380.
[18] A. Khintchine *Ein Satz ber Kettenbrche, mit arithmetischen Anwendungen,* Mathematische Zeitschrift **18** (1923) 289–306.
[19] J. Kingman *Poisson processes,* Oxford Studies in Probability **3** (1993) Oxford Univ. Press, New York 104 pp.
[20] D. Y. Kleinbock, G. A. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces,* in *Sinai's Moscow Seminar on Dynamical Systems* AMS. Transl. **171** (1996) 141–172.
[21] D. Y. Kleinbock, G. A. Margulis *Logarithm laws for flows on homogeneous spaces,* Invent. Math. **138** (1999) 451–494.
[22] J. F. Koksma, *Diophantische Approximationen*, Ergebnisse der Mathematik, Band IV, Heft 4, Berlin, 1936.
[23] J. Marklof *The n-point correlations between values of a linear form,* Erg. Th., Dynam. Sys. **20** (2000) 1127–1172.
[24] J. Marklof *Almost modular functions and the distribution of n^2x modulo one,* Int. Math. Res. Not. **39** (2003) 2131–2151.
[25] A. Ostrowski *Bemerkungen zur Theorie der Diophantischen Approximationen,* Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg **1** (1922) 77 – 98.
[26] B. S. Pitskel *Poisson Limit Law for Markov chains,* Erg. Th. Dyn. Sys. **11** (1991) 501–513.
[27] K. F. Roth, *On irregularities of distribution,* Mathematika 1 (1954), 73-79
[28] G. Samorodnitsky and M. Taqqu *Stable non-Gaussian random processes,* Chapman & Hall/CRC, 1994, 640 pp.
[29] W. Schmidt, *Metrical theorems on fractional parts of sequences*, Trans. Amer. Math. Soc. **110** (1964) 493–518.
[30] I. Vinogradov *Limiting distribution of visits of several rotations to shrinking intervals,* preprint.