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 $\mathbb{T}^d$ where 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.

1. INTRODUCTION

1.1. **Equidistribution of Kronecker sequences of $\mathbb{T}^d$.** It is 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: given a set $B \subset \mathbb{T}^d$ let

$$D_B(x, \alpha, N) = \sum_{n=0}^{N-1} \chi_B(x + k\alpha) - N\mu(B)$$

where $(x, \alpha) \in \mathbb{T}^d \times \mathbb{T}^d$, μ is the Haar measure on the torus and χ_B is the characteristic function of the set B. Uniform distribution of the sequence $x + k\alpha$ on $\mathbb{T}^d$ is equivalent to the fact that, for regular sets B, $D_B(x, \alpha, N)/N \to 0$ as $N \to \infty$. A step further in the description of the uniform distribution is the study of the rate of convergence to 0 of $D_B(x, \alpha, N)/N$.

Already with $d = 1$, it is clear that if $\alpha \in \mathbb{T} - \mathbb{Q}$ is fixed, the discrepancy $D_B(x, \alpha, N)$ 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 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

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

where the supremum is taken over all sets B in some natural class of sets $\mathcal{B}$, 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 porperties of α. In particular, following earlier advances of [8, 6, 15, 11, 19] and others, [1] proves that for arbitrary positive increasing function $\phi(n)$ the ratio

$$\frac{\overline{D}(\alpha, N)}{(\ln N)^d \phi(\ln \ln N)}$$

is bounded for almost every $\alpha \in \mathbb{T}^d$ iff $\sum_n \phi(n) < \infty$. In dimension $d = 1$ this result is the content of Khinchine theorems obtained in the early 1920's [11] and it follows from the almost independence of the partial quotients of α. The higher dimensional case is signifivantly more difficult and the cited bound was only obtained in the 1990s.

The bound on $\overline{D}(\alpha, N)$ focus on how bad can the discrepancy become along a subsequence of N for a fixed α in a full measure set. In a sense, they deal with the worst case scenario and do not capture the oscillations of the discrepancy.

On the other hand, the restriction on α is necessary, since given any $\varepsilon_n \to 0$ it is easy to see that for $\alpha \in \mathbb{T}$ *sufficiently Liouville*, the discrepancy (relative to intervals) can be as bad as $N\varepsilon_n$. 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 $\mathbb{T}^d$ ([16]). Here again, due to the use of continued fractions the latter conjecture can be easily verified in dimension 1 (cf. discussion in [1]).

In another direction, but still studying the discrepancy for a fixed α and along subsequences of N, [7] obtains a Central Limit Theorem in the one dimensional case of circle rotations. The results of [7] apply either for a set of α of zero measure (so called badly approximable numbers) and the set of times of large density, or for all α but for small set of times.

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 [9, 10] (see also [2]) 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 [9, 10]. *Let $D_N(x, \alpha) = \sum_{n=0}^{N-1} \chi_{[a,b]}(x + k\alpha) - N(b - a)$. There is a number $\rho = \rho(b - a)$ such that if (x, α) is uniformly distributed on $\mathbb{T}^2$ then $\frac{D_N}{\rho \ln N}$ converges to the standard Cauchy distribution, that is,*

$$\mathrm{mes}((x, \alpha) : \frac{D_N}{\rho \ln N} \le z) \to \frac{\tan^{-1} z}{\pi} + \frac{1}{2}.$$

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, and as in the case of other results related to discrepancies of Kronecker sequences, the main difficulty will come 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 [5] we considered the case where $\mathcal{C}$ is analytic and strictly convex and showed that $D_{\mathcal{C}}(x, \alpha, N)/N^{(d-1)/2d}$ has a limiting distribution (which however depends on $\mathcal{C}$).

Here we address the case where $\mathcal{C}$ is a box and show that $D_{\mathcal{C}}(x, \alpha, N)/(\ln N)^d$ converges to a Cauchy distribution. To avoid the irregular behavior of the limiting distribution as the function 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 have also to apply (arbitrarily small) random affine deformations on the boxes.

More precisely, for $u = (u_1, \ldots, u_d)$ with $0 < u_i < 1/2$ for every i, we define a cube on the d-torus by $C_u = [-u_1, u_1] \times \ldots [-u_d, u_d]$. Let $\eta > 0$ and MC_u be the image of C_u by a matrix $M \in \mathrm{SL}(d, \mathbb{R})$ such that

$$M = (a_{ij}) \in G_\eta = \{|a_{i,i} - 1|, \text{ for every } i \text{ and } |a_{i,j}| < \eta \text{ for every } j \neq i\}.$$

For a point $x \in \mathbb{T}^d$ and a translation frequency vector $\alpha \in \mathbb{T}^d$ we denote $\xi = (u, M, \alpha, x)$ and define the following discrepancy function

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

Fix d segements $[v_i, w_i]$ such that $0 < v_i < w_i < 1/2 \forall i = 1, \dots, d$. Let $X = (u, \alpha, x, (a_{i,j})) \in [v_1, w_1] \times \dots [v_d, w_d] \times \mathbb{T}^{2d} \times G_\eta$. We denote by λ the normalized restriction of the Lebesgue $\times$ Haar measure on X.

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

$$\lim_{N\to\infty} \lambda\{\xi \in X \ / \ \frac{D(\xi, N)}{(\ln N)^d} \le z\} = C(\boldsymbol{\rho} z) \tag{1}$$

where C is the standard Cauchy cumulative distribution function

$$C(z) = \frac{1}{\pi}\int_{-\infty}^{z} \frac{1}{1+a^2} da = \frac{\tan^{-1} z}{\pi} + \frac{1}{2}.$$

Our proof of Theorem 1 shows that for typical α a *quenched* limit (that is, with fixed α and x uniformly distributed on $\mathbb{T}^d$) of $D_{\mathcal{C}}(x, \alpha, N)$ does not exist even if we would allow the normalizing sequence U_N to depend on α. The reason is that the main contribution to discrepancy comes from a small set of so called *small denominators* and, at different scales, different small denominators become important. We note that the absence of quenched limits is often observed in zero entropy systems [3, 5, 14].

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 result of Section 3 is Theorem 5 that reduces the proof of the main theorem to establishing a Poisson limit theorem for the distribution of small denominators and the corresponding numerators.

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 in sections 4–6 we assume that $d = 2$ and use the notations (x, y), (u, v), (α, β) instead of $(x_1, x_2), (u_1, u_2), (\alpha_1, \alpha_2)$ and

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

In dimension two we need to prove the Poisson limit theorem for the set

$$\begin{aligned} &\{(\ln^2 N(a_1 k + b_1 l)(a_2 k + b_2 l)\|k\alpha + l\beta\|, \{N(k\alpha + l\beta)\}, \{(a_1 k + b_1 l)u\}, \\ &\{(a_2 k + b_2 l)v\}, \{kx + ly\})\} \end{aligned} \tag{$\star$}$$

when (k, l) range over the resonant frequencies for (α, β) that contribute to the discrepancy $D(\xi, N)$, namely

$$|a_1 k + b_1 l| > 1, \quad |a_2 k + b_2 l| > 1, \quad |(a_1 k + b_1 l)(a_2 k + b_2 l)| < N,$$

$$(a_1 k + b_1 l)(a_2 k + b_2 l)\|k\alpha + l\beta\| < \frac{1}{\varepsilon \ln^2 N}.$$

In section 4, we reduce the Poisson limit of the first coordinate of $(\star)$ to the Poisson limit theorem (Theorem 10) for the number of visits to a cusp by orbits of the Cartan action on $\mathbf{M} = SL_3(\mathbb{R})/SL_3(\mathbb{Z})$.

The proof of Theorem 10 is given in Section 6. Poisson limit theorems for dynamical systems is a popular subject. The most relevant for our purposes is paper [4] where the Poisson Limit Theorem is proven for partially hyperbolic systems assuming that the images of local unstable manifolds became 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 which k and l contribute to the resonances in ($\star$). However Rogers identities provide sufficiently strong control to handle this issue. Secondly, in the higher rank case (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) 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. At the present setting we are able to prove the Poisson limit theorem using the fact the long leaves of the Lyapunov foliations become uniformly distributed at a polynomial rate, except for a set of small measure.

The relevant equdistribution results for unipotent subgroups of $SL_3(\mathbb{R})$ acting on the space of three dimensional lattices are presented in Section 5.

Unfortunately, a possible existence of small exceptional sets, requires us to introduce additional parameters in the form of small affine deformations of the box, because we can only prove Poisson Limit Theorem if the initial lattice has a smooth density on $\mathbf{M}$ whereas if we work with the straight boxes we only get a positive codimension submanifold of $\mathbf{M}$.

To prove the Poisson limit for all components of ($\star$) we need to show that the other components are asymptotically independent of the first one. This requires an extra work but the argument is similar to the original analysis of Kesten.

In section 7 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 the main theorem 1. The case $\gamma = 1/d$ was studied in [13] where a limit distribution is 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.

Finally, in Section 8 we discuss the continuous time case, that is, we study the discrepancies corresponding to linear flows on the torus. We show that in case of boxes the discrepancy is bounded in probability since 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, our proof of Theorem 1 implies a Cauchy limit theorem for continuous discrepancies relative to balls, and this only in dimension $d = 3$. Indeed, the latter is in sharp contrast with the higher dimension case obtained in [5] that states that for $d \geq 4$ the continuous discrepancies relative to balls converge in distribution after normalization by a factor $T^{(d-3)/2(d-1)}$.

2. Preliminaries.

2.1. **Poisson process.** The proofs of the facts listed below can be found in monographs [17, 18].

Let (X, μ) be a measure space. Recall that a Poisson process associated to (X, μ) is a point process with values in X such that

(a) if $N(A)$ is the number of points in $A \subset X$ then $N(A)$ has Poisson distribution with parameter $\mu(A)$ and

(b) if $A_1, A_2 \dots A_k$ are disjoint subsets of X then $N(A_1), N(A_2) \dots N(A_k)$ are mutually independent.

If $X \subset \mathbb{R}^d$ and μ has a density f with respect to the Lebesgue measure we say that f is the intensity of the Poisson process.

Lemma 2.

(a) If $\{\Theta_j\}$ is a Poisson process on X and $\psi : X \to \tilde{X}$ is a measurable map then $\tilde{\Theta}_j = \psi(\Theta_j)$ is a Poisson process. If $X = \tilde{X} = \mathbb{R}$ and if $\{\Theta_j\}$ has intensity f and if ψ is invertible, then the intensity of $\tilde{\Theta}$ is

$$\tilde{f}(\theta) = f(\psi^{-1}(\theta)) \frac{1}{|\psi'(\psi^{-1}(\theta))|}.$$

(b) Let (Θ_j, Γ_j) be a point process on $X \times Z$ such that $\{\Theta_j\}$ is a Poisson process on X and $\{\Gamma_j\}$ are Z-valued random variables which are i.i.d. and independent of $\{\Theta_k\}$ then (Θ_j, Γ_j) is a Poisson process on $X \times Z$.

(c) Conversely if (Θ_j, Γ_j) is a Poisson process on $X \times Z$ with measure $\mu \times \nu$ where ν is a probability measure then Θ_j is a Poisson process with measure μ and Γ_j are iid independent of Θs and having distribution ν.

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

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

Lemma 3.

(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 $\rho = c\mathbb{E}(|\Gamma|)\pi$.

The proof of Lemma 3 (b) follows from Lemma 3 (a) and parts (a) and (d) of Lemma 2.

2.2. **Rogers identities.** The following identities (see [13, 20]) play an important role in our argument. 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.

Lemma 4. *Let f, f_1, f_2 be piecewise smooth functions with compact support on $\mathbb{R}^{d+1}$. For a lattice $\mathcal{L} \subset \mathbb{R}^{d+1}$ let*

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

Then

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

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

(c) Consequantly

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

3. Estimating the contribution of non-resonant terms.

The goal of this section is to reduce of the proof of the main Theorem 1 to proving a Poisson limit distribution of a point process related to the resonant terms to the discrepancy function (Theorem 5). The subsequent Sections 4–6 will be dedicated to the proof of Theorem 5.

3.1. Recall the definition

$$X = \{(u, \alpha, x, (a_{i,j})) \in [v_1, w_1] \times \dots [v_d, w_d] \times \mathbb{T}^{2d} \times G_\eta\}$$

Let $\Phi_m(\omega) := \frac{\sin(2\pi m\omega)}{m}$. For $\xi \in X$ and $k \in \mathbb{Z}^d$, we use the notation

$$\bar{k}_i = a_{i,1}k_1 + \dots + a_{i,d}k_d \tag{2}$$

Let

$$U_k(\xi, N) = A\Pi_i \Phi_{\bar{k}_i}(u_i) \frac{\sin(\pi N(k,\alpha))}{\sin(\pi(k,\alpha))} \cos(2\pi(k,x) + \varphi_{k,N,\alpha}))$$

where $\varphi_{k,\alpha,N} = \pi(N-1)(k,\alpha)/2$, $A = \frac{1}{\pi^d}$ and $(k,x) := \sum_{i=1}^d k_i x_i$. Writing the Fourier series of the characteristic function of a box we get that

$$D(\xi, N) = \sum_{k\in\mathbb{Z}^d-\{0\}} U_k(\xi, N) = 2 \sum_{k\in\mathbb{Z}^d-\{0\}, k_1>0} U_k(\xi, N)$$

3.2. Let

$$D_1(\xi, N) = \sum_{|k_i|\le N, k\ne(0,\dots,0)} U_k(\xi, N)$$

We claim that there exists a constant C such that

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

where the L^2 norm refers by default to functions of the variables $(\alpha, x) \in \mathbb{T}^{2d}$. As a consequence we can replace D by D_1 in (1).

Proof of the claim. Assume $\xi \in X$ given. Then for any $q \ge N$ and any $q_1, \dots, q_{d-1} \in \mathbb{N}$ there exists only finitely many $k \in \mathbb{Z}^d$ such that $k_{i_d} = q$ 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 $1, \dots, d$. Since for any ω, $|\Phi_m(\omega)| < \min(2\pi|\omega|, 1/m)$, the contributions of the latter frequencies can thus be bounded as follows

$$\begin{aligned}\|D - D_1\|_2^2 &\le C \sum_{q\ge N, q_1,\dots,q_{d-1}\ge 0} \frac{1}{q^2(q_1+1)^2\dots(q_{d-1}+1)^2} \int_{\mathbb{T}^d} \left(\frac{\sin(\pi N(k,\alpha))}{\sin(\pi(k,\alpha))}\right)^2 d\alpha \\ &\le C \sum_{q\ge N, q_1,\dots,q_{d-1}\ge 0} \frac{1}{q^2(q_1+1)^2\dots(q_{d-1}+1)^2} N \le C. \end{aligned}$$

□

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

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

We want to show that it is possible to replace the study of D_1 by that of 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 $C(d)$ such that for every $(q_1, \ldots, q_{d-1}) \in [0, 2N]^{d-1}$ there is at most $C(d)$ 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 $1, \ldots, d$. We call $K_{q_1,\ldots,q_{d-1}}$ the latter set of k. We then exclude the translation vectors α for which there exists $(q_1, \ldots, q_{d-1}) \in [0, 2N]^{d-1}$ with at least one $k \in K_{q_1,\ldots,q_{d-1}}$ satisfying $|\Pi_{i=1}^{d-1}(q_i + 1)| \|(k, \alpha)\| \leq \varepsilon/(\ln N)^{d-1}$. The excluded set $E_N((a_{i,j}))$ has Lebesgue measure of order ε.

We claim that

$$\|D_2 - D_1\|^2_{L_2((\mathbb{T}^d - E_N) \times \mathbb{T}^d)} = \mathcal{O}\left(\frac{(\ln N)^{2(d-1)}}{\varepsilon}\right).$$

Therefore we can replace D_1 by D_2 in (1).

Proof of the claim. Let

$$B_p\left((q_1, \ldots, q_{d-1}), (a_{i,j})\right) = \{\alpha \in \mathbb{T}^d : \exists k \in K_{q_1,\ldots,q_{d-1}}((a_{i,j})),$$
$$p\varepsilon/(\ln N)^{d-1} \leq |\Pi_{i=1}^{d-1}(q_i + 1)| \|(k, \alpha)\| \leq (p+1)\varepsilon/(\ln N)^{d-1}\}$$

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

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

□

3.4. Let $\bar{S}(\xi, N) := \{k \in \mathbb{Z}^d : \Pi_{i=1}^d |k_i| \leq N, |\bar{k}_i| \geq 1\}$. Then let

$$D_3(\xi, N) = \sum_{k \in \bar{S}} U_k(\xi, N).$$

We claim that there exists a constant $C > 0$ such that

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

Therefore we can replace D_2 by D_3 in (1).

Proof of the claim. Denote $K(k) = \Pi_{i=1}^d \bar{k}_i$. We have that

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

Let

$$A_s = \{k \in S : |K(k)| \in [2^s N, 2^{s+1} N]\}$$

and observe that $\#A_s \le C2^s N(\ln N + s)^{d-1}$. Thus

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

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

$$T(\xi, N) := \left\{ k \in \bar{S}((a_{i,j}), N) : |\Pi_{i=1}^d \bar{k}_i| \big\|(k, \alpha)\big\| \le \frac{1}{\varepsilon(\ln N)^d} \right\}$$

and let

$$D_4(\xi, N) = \sum_{k \in T} U_k(\xi, N).$$

We claim that

$$\|D_4 - D_3\|_{L_2((\mathbb{T}^d - E_N) \times \mathbb{T}^d)} \le C\sqrt{\varepsilon}(\ln N)^d$$

therefore to prove (1), we can study the limiting distribution as $N \to \infty$ of $D_4/(\ln N)^d$.

Proof of the claim. Since $k \in \bar{S}$ and $(a_{i,j})$ is close to Identity we have that $1 \le |\bar{k}_i| \le 2N$ for every i. Now, for every $q_1, \dots, q_d \in [1, 2N]^d$ 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 $K(q_1, \dots, q_d)$. We have that

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

where

$$A_{K(q_1, \dots, q_d)} = \sum_{k \in K(q_1, \dots, q_d)} \int_{\mathbb{T}^d} \frac{1}{(\Pi_{i=1}^d q_i \big\|(k, \alpha)\big\|)^2} \chi_{\Pi_{i=1}^d q_i \|(k,\alpha)\| \ge 1/\varepsilon(\ln N)^d} d\alpha.$$

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

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

We have that $\text{Leb}_{\mathbb{T}^d}(B_{k,p}) \le 1/(\varepsilon \Pi_{i=1}^d q_i (\ln N)^d)$. Thus

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

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

3.6. Let $\delta = \varepsilon^4$. Define $W(\xi, N) = W((a_{i,j}), \alpha, N)$ by

$$W(\xi, N) := \Big\{ k \in \mathbb{Z}^d : |\Pi_{i=1}^d \bar{k}_i| < N^{1-\delta}, \quad \forall i = 1, \dots, d, \quad |\bar{k}_i| \ge 1, |\Pi_{i=1}^d \bar{k}_i| \big\|(k, \alpha)\big\| \le \frac{1}{\varepsilon(\ln N)^d} \Big\}$$

and let

$$D_5(\xi, N) = \sum_{k \in W} U_k(\xi, N).$$

Let

$$F_N((a_{i,j})) = E_N \bigcup \Big\{ \alpha \in \mathbb{T}^d : \exists k \in T \text{ such that } |\Pi_{i=1}^d \bar{k}_i| \big\|(k, \alpha)\big\| < \frac{\varepsilon}{(\ln N)^d} \Big\}.$$

Since for any $q_1, \dots, q_d \in [1, 2N]^d$ there are at most $C(d)$ vectors $k \in \bar{S}$ such that $\bar{k}_i \in [q_i, q_i + 1]$ we get that

$$\mathrm{Leb}_{\mathbb{T}^d}(F_N) \le \mathrm{Leb}_{\mathbb{T}^d}(E_N) + \sum_{1 \le q_1, \dots, q_d \le N} \frac{C\varepsilon}{(\ln N)^d \Pi_{i=1}^d q_i} \le C\varepsilon.$$

We claim that

$$\|D_5 - D_4\|_{L_2((\mathbb{T}^d - F_N) \times \mathbb{T}^d)} \le C\sqrt{\varepsilon}(\ln N)^d.$$

Therefore to prove (1) we can study the limiting distribution as $N \to \infty$ of $D_5/(\ln N)^d$.

Proof of the claim. For every $k \in \mathbb{Z}^d$, we have that

$$\mathrm{Leb}\left\{\alpha \in \mathbb{T}^d : |\Pi_{i=1}^d \bar{k}_i| \big\|(k, \alpha)\big\| < \frac{1}{\varepsilon(\ln N)^d}\right\} \le \frac{1}{|\Pi_{i=1}^d \bar{k}_i| \varepsilon (\ln N)^d}.$$

Hence the contribution of the $k \in T - W$ for $\alpha \in \mathbb{T}^d - F_N$ can be bounded by

$$\|D_5 - D_4\|^2_{L_2((\mathbb{T}^d - F_N) \times \mathbb{T}^d)} \le \sum_{k \in T, |\Pi_{i=1}^d \bar{k}_i| \ge N^{1-\delta}} \frac{(\ln N)^{2d}}{\varepsilon^2} \frac{1}{|\Pi_{i=1}^d \bar{k}_i| \varepsilon (\ln N)^d}.$$

For $s = 0, \dots, [\delta \ln N]$, define

$$P_s = \{k \in T : |\Pi_{i=1}^d \bar{k}_i| \in [2^s N^{1-\delta}, 2^{s+1} N^{1-\delta}]\}.$$

Then $\#(P_s) \le C 2^s N^{1-\delta} (\ln N)^{d-1}$. Thus the terms in P_s contribute to $\|D_5 - D_4\|^2_{L_2((\mathbb{T}^2 - F_N) \times \mathbb{T}^2)}$ with less than

$$\#(P_s) \frac{(\ln N)^d}{\varepsilon^3 2^s N^{1-\delta}} \le C \frac{(\ln N)^{2d-1}}{\varepsilon^3}.$$

Summing over $s = 0, \dots, [\delta \ln N]$ we get the required estimate. □

3.7. Observe that that given ε for each $\eta > 0$ there is a number $n(\eta)$ such that

$$\mathrm{mes}(\xi \text{ such that } \#(W(\xi, N)) > n(\eta)) < \eta$$

uniformly in N. Since the contributing terms in D_5 satisfy $\big\|(k, \alpha)\big\| \le \frac{1}{\varepsilon(\ln N)^d}$, we can replace U_k in the definition of Δ by

$$V_k(\xi, N) = A \Pi_i \Phi_{\bar{k}_i}(u_i) \frac{\sin(\pi N (k, \alpha))}{\big\|(k, \alpha)\big\|} \cos(2\pi(k, x) + \varphi_{k,N,\alpha}))$$

and consider instead of D_5 in (1)

$$D_6(\xi, N) = 2 \sum_{k \in W, k_1 > 0} V_k.$$

Next, we let

$$Z(\xi, N) = \{k \in W(\xi, N), k_1 > 0 : \exists m \in \mathbb{Z} \text{ such that } k_1 \wedge \ldots \wedge k_d \wedge m = 1 \quad \textit{and } \big\|(k, \alpha)\big\| = (k, \alpha) + m\}$$

and we can replace the study of $\frac{D_6}{(\ln N)^d}$ by

$$\frac{D_7(\xi, N)}{(\ln N)^d} = \frac{2A}{\pi} \sum_{k \in Z} \frac{\Gamma_k(\xi, N)}{\Theta_k(\xi, N)}$$

where

$$\Theta_k(\xi, N) = \Pi_{i=1}^d \bar{k}_i \big\|(k,\alpha)\big\| (\ln N)^d$$
$$\Gamma_k(\xi, N) = \phi\left(\bar{k}_1 u_1, \ldots, \bar{k}_d u_d, N(k,\alpha), (k,x) + \varphi_{k,\alpha,N}\right)$$

and

$$\phi(\eta_1, \ldots, \eta_d, \eta_{d+1}, \eta_{d+2}) = \sum_{j=1}^{\infty} \frac{\left[\Pi_{i=1}^d \sin(2\pi j \eta_i)\right] \sin(\pi j \eta_{d+1}) \cos(2\pi j \eta_{d+2})}{j^{d+1}}. \tag{3}$$

Note that $\|\partial_{\eta_i}\phi\| \leq C$, for any $i = 1, \ldots, d+2$. The difference between D_7 and D_6 is that for $k \in Z$, we comprise in D_7 all its multiples whereas in D_6 we take only multiples such that $jk \in W$. This does not make any difference in the limit because $\sum \frac{1}{j^{d+1}} < \infty$ and because we can of course add to D_6 the multiples of k such that $\|j(k,\alpha)\| \leq \frac{1}{\varepsilon'(\ln N)^d}$ with $\varepsilon' \ll \varepsilon$ which accounts for most of the sum in ϕ.

3.8. By the general facts about Poisson processes listed in section 2.1, Theorem 1 follows from the next result

Theorem 5. *For any $\varepsilon, \delta > 0$ we have that as $N \to \infty$ and $\xi \in X$ is distributed according to the normalized Lebesgue measure λ, the process*

$$\Big\{\Big((\ln N)^d \Pi_i \bar{k}_i \big\|(k,\alpha)\big\|, \{N(k,\alpha)\}, \{\bar{k}_1 u_1\}, \ldots, \{\bar{k}_d u_d\}, \{(k,x) + \varphi_{k,\alpha,N}\}\Big)\Big\}_{k \in Z(\xi,N)}$$

converges to a Poisson process on $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}] \times \mathbb{T}^{d+2}$ with intensity $2^{d-1}(1-\delta)^d \mathbf{c}_1/d!$ where $\mathbf{c}_1 = 1/\zeta(d+1)$ is the constant from Lemma 4.

Note that by standard properties of weak convergence the result remains valid for $\varepsilon = \delta = 0$. That is, we get the following result which is of independent interest

Corollary 6. *Let $\xi \in X$ be ditributed according to the normalized Lebesgue measure λ. Then as $N \to \infty$ the point process*

$$\Big\{\Big((\ln N)^d \Pi_i \bar{k}_i \big\|(k,\alpha)\big\|, \{N(k,\alpha)\}, \{\bar{k}_1 u_1\}, \ldots, \{\bar{k}_d u_d\}, \{(k,x) + \varphi_{k,\alpha,N}\}\Big)\Big\}_{k \in Z^*(\xi,N)}$$

where

$$Z^*(\xi, N) = \Big\{k \in \mathbb{Z}^d : |\bar{k}_i| \geq 1, |\Pi_i \bar{k}_i| < N, k_1 > 0, \quad |\Pi_i \bar{k}_i| \big\|(k,\alpha)\big\| \leq \frac{1}{\varepsilon(\ln N)^d}, \quad \exists m \in \mathbb{Z} \text{ such that } k_1 \wedge \ldots \wedge k_d \wedge m = 1 \textit{ and } \big\|(k,\alpha)\big\| = (k,\alpha) + m\Big\}$$

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

3.9. Proof of Theorem 1. We have that (a) and (c) of Lemma 2 imply that $\{\Theta_k(\xi,N)\}_{k \in Z}$ converges to a Poisson process on $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}]$ and that $\{\Gamma_k(\xi,N)\}_{k \in Z}$ are asymptotically iid independent from the Θs and have a symmetric distribution with compact support. Hence Lemma 3 applies and yields Theorem 1. Note that the limiting distribution of $D_7/(\ln N)^d$ has form $\frac{2}{\pi^{d+1}} \sum_j \frac{\Gamma_j}{\Theta_j}$. Observe that due to (3) we have $\mathbb{E}(\Gamma) = \zeta(d+1)\left(\frac{2}{\pi}\right)^{d+2}$ and hence

$$\boldsymbol{\rho} = \frac{2}{\pi^d} \frac{2^{d-1}\mathbf{c}_1}{d!} \mathbb{E}(\Gamma) = \frac{1}{d!}\left(\frac{2}{\pi}\right)^{2d+2}.$$

3.10. The case $d = 2$. Notations. Since the proof of Theorem 5 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 4–6 are devoted. There, we will prove the following version of Theorem 5 in the case $d = 2$.

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

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

Let

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

We denote by λ the normalized Lebesgue measure on X.

Theorem 7. *For each $\varepsilon, \delta > 0$ the following holds. As $N \to \infty$ and $\xi \in X$ is ditributed according to λ, the process*

$$\Big\{ \Big(\ln^2 N (a_1 k + b_1 l)(a_2 k + b_2 l) \big\| k\alpha + l\beta \big\|, \{N(k\alpha + l\beta)\}, \{(a_1 k + b_1 l)u\}, \\ \{(a_2 k + b_2 l)v\}, \{kx + ly + \varphi_{l,k,\alpha,\beta,N}\}\Big)\Big\}_{(k,l)\in Z(\xi,N)}$$

converges to a Poisson process on $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}] \times \mathbb{T}^4$ with intensity $(1-\delta)^2 \mathbf{c}_1$ where $\mathbf{c}_1 = 1/\zeta(3)$ is the constant from Lemma 4. Here

$$\begin{aligned} Z(\xi, N) = \Big\{ (k, l) \in \mathbb{Z}^2 : |a_1 k + b_1 l| \geq 1, |a_2 k + b_2 l| \geq 1, \\ |(a_1 k + b_1 l)(a_2 k + b_2 l)| < N^{1-\delta}, k > 0, \\ |(a_1 k + b_1 l)(a_2 k + b_2 l)| \big\| k\alpha + l\beta \big\| \leq \frac{1}{\varepsilon (\ln N)^2}, \\ \exists m \in \mathbb{Z} \text{ such that } k \wedge l \wedge m = 1 \textit{ and } \big\| k\alpha + l\beta \big\| = k\alpha + l\beta + m \Big\} \end{aligned}$$

4. Reduction to dynamics on the space of lattices.

Denote $M = [\ln N]$.

Introduce the following notations

$$I = (1, e], \quad J = [-e, -1) \cup (1, e], \quad K = \left[-\frac{1}{\varepsilon M^2}, \frac{1}{\varepsilon M^2} \right],$$

$$\Lambda = \begin{pmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ \alpha & \beta & 1 \end{pmatrix}, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = g_{t_1, t_2} \Lambda \begin{pmatrix} k \\ l \\ m \end{pmatrix}. \tag{4}$$

Define on the space $\mathbf{M}$ of unimodular lattices $\mathcal{L}$ the function

$$\Phi(\mathcal{L}) = \sum_{e \in \mathcal{L} \text{ prime}} 1_I(x(e)) 1_J(y(e)) 1_K(xyz(e)). \tag{5}$$

and on $\mathbf{M} \times \mathbb{R}$ define an $\mathbb{R} \times \mathbb{T}$ valued function

$$\Psi(\mathcal{L}, b) = (\Psi_1(\mathcal{L}), \Psi_2(\mathcal{L}, b)) = \\ \sum_{e \in \mathcal{L} \text{ prime}} 1_I(x(e)) 1_J(y(e)) 1_K(xyz(e)) (M^2 xyz(e), bz(e) \bmod (2)). \tag{6}$$

Given N, suppose that $\xi \in X$ is such that for every $t_1, t_2 \in [0, M]$ there exists at most one $(k, l) \in Z(\xi, N)$ such that

$$e^{t_1} < L_1 \leq e^{t_1 + 1}, \quad e^{t_2} < |L_2| \leq e^{t_2 + 1} \tag{7}$$

where $L_1 = a_1k + b_1l, L_2 = a_2k + b_2l$. Note that if (7) holds then $(k,l) \in Z$ iff $\Phi(g_{t_1,t_2}\Lambda(\xi)) = 1$. Thus, for such ξ we have that the sequence

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

is exactly

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

with

$$\mathbf{\Pi} = \{t = (t_1, t_2) \in \mathbb{N}^2 : t_1 + t_2 < (1-\delta)M\}.$$

Hence, to show that the distribution of $\{M^2(a_1k+b_1l)(a_2k+b_2l)||k\alpha+l\beta||, N\big\|k\alpha+l\beta\big\| \bmod (2)\}_{k,l\in Z}$ converges as $N \to \infty$ to that of a Poisson process on $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}] \times \mathbb{R}/(2\mathbb{Z})$ with intensity $2(1-\delta)^2\mathbf{c}_1$ it is sufficient to have (a) and (b) of the following theorem.

Theorem 8. *Assume that $\xi \in X$ is distributed according to a probability measure with smooth density with respect to the Lebesgue measure. We will denote by Λ the matrix $\Lambda(\xi)$ as defined in (4). Then*

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

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

(c) Let $\tau(t) = \max(t_1, t_2)$ and let $\tau_1 < \tau_2 < \ldots < \tau_s \ldots$ be the set of points $\{\tau(t) : \Phi(g_t\Lambda) = 1, t \in \mathbf{\Pi}\}$ listed in the increasing order. Then for each s

$$\mathbb{P}(\tau_j - \tau_{j-1} > \sqrt{M} \text{ for each } j \leq s) \to 1 \text{ as } N \to \infty.$$

(d) Let $\tau'(t) = \min(t_1, t_2)$ and let $\tau'_1 > \tau'_2 > \ldots > \tau'_s \ldots$ be the set of points $\{\tau'(t) : \Phi(g_t\Lambda) = 1, t \in \mathbf{\Pi}\}$ listed in the decreasing order. Then for each s

$$\mathbb{P}(\tau'_s > \sqrt{M}) \to 1 \text{ as } N \to \infty.$$

In order to get the full Poissonean limit in Theorem 7 we will also need the following fact.

Theorem 9. *Let $K > 0$ be fixed. Given any $s \in \mathbb{N}$, let $(k_1^N, l_1^N), \ldots, (k_s^N, l_s^N)$ be a sequence of s 2-tuples such that*

$$|(k_j^N, l_j^N)| > |(k_{j-1}^N, l_{j-1}^N)|e^{\sqrt{\ln N}}. \tag{8}$$

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

$$||\rho_N||_{C^1} \leq K. \tag{9}$$

Then the distribution of the s 3-tuples

$$((a_1k_1 + b_1l_1)u, (a_2k_1 + b_2l_1)v, (k_1x + l_1y + \phi_{l_1,k_1,\alpha,\beta,N}) \ldots$$
$$(a_1k_s + b_1l_s)u, (a_2k_s + b_2l_s)v, (k_sx + l_sy + \phi_{l_s,k_s,\alpha,\beta,N}))$$

converges to the uniform distribution on $\mathbb{T}^{3s}$ and the convergence is uniform with respect to N and $(a_1, a_2, b_1, b_2, \alpha, \beta)$ and to the choices of s 2-tuples satisfying (8) and ρ_N satisfying (9).

Proof of Theorem 9. We need to show that given smooth functions $f_1, f_2 \ldots f_s$ on $\mathbb{T}^3$ with zero average we have

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

uniformly in the parameters involved but this follows easily by considering the Fourier series of f_j. □

Proof of Theorem 7. As mentioned before, (a) and (b) of Theorem 8 imply that $\{M^2(a_1k+b_1l)(a_2k+b_2l)||\alpha k+\beta l||, N\big\|k\alpha+l\beta\big\| \mod 2\}_{k,l\in Z}$ converges as $N\to\infty$ to a Poisson process on $[-\frac{1}{\varepsilon},\frac{1}{\varepsilon}]\times\mathbb{R}/(2\mathbb{Z})$ with intensity $2(1-\delta)^2\mathbf{c}_1$. Next, Theorem 9 and (c) of Theorem 8 imply that $\big\{\big(\{(a_1k+b_1l)u\},\{(a_2k+b_2l)v\},\{kx+ly+\varphi'_{l,k,\alpha,\beta,N}\}\big)\big\}_{(k,l)\in Z(\xi,N)}$ converge to uniformly distributed iid's on $\mathbb{T}^3$ independent of $\{M^2(a_1k+b_1l)(a_2k+b_2l)||\alpha k+\beta l||, N\big\|k\alpha+l\beta\big\| \mod (2)\}_{k,l\in Z(\xi,N)}$. Lemma 2 hence yields the full Poissonean limit of Theorem 7. □

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 10. *Assume that $\mathcal{L}$ has a smooth density on $\mathbf{M}$. Then*

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

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

(c) Let $\tau(t)=\max(t_1,t_2)$ and let $\tau_1<\tau_2<\dots<\tau_s\dots$ be the set of points $\{\tau(t):\Phi(g_t\mathcal{L})=1,t\in\mathbf{\Pi}\}$ listed in the increasing order. Then for each s

$$\mathbb{P}(\tau_j-\tau_{j-1}>\sqrt{M}\text{ for each } j\le s)\to 1 \text{ as } N\to\infty.$$

(d) Let $\tau'(t)=\min(t_1,t_2)$ and let $\tau'_1>\tau'_2>\dots>\tau'_s\dots$ be the set of points $\{\tau'(t):\Phi(g_t\mathcal{L})=1,t\in\mathbf{\Pi}\}$ listed in the decreasing order. Then for each s

$$\mathbb{P}(\tau'_s>\sqrt{M})\to 1 \text{ as } N\to\infty.$$

Proof that Theorem 10 implies Theorem 8.

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 some interval $\bar{K}\subset K$. Let $\bar{\Phi}^\pm$ be defined as in (5) with the intervals $I^\pm,J^\pm,\bar{K}^\pm$ instead of I,J,K. Next, given $\Lambda=\Lambda(\xi)$ for some $\xi\in X$, define

$$\tilde{\Lambda}=\begin{pmatrix}(1+\sigma)&0&0\\0&(1+\sigma)&0\\0&0&(1+\sigma)^{-2}\end{pmatrix}\begin{pmatrix}1&0&c_1\\0&1&c_2\\0&0&1\end{pmatrix}\Lambda$$

where σ, c_1 and c_2 are random variables independent of each other and of $(a_1,b_1,a_2,b_2,\alpha,\beta)$ and having uniform distribution on $[0,\eta^2]$. The equivalence between Theorem 8 and Theorem 10 stems from the straightforward observation that if M is sufficiently large, then for any $n\in\mathbb{N}$ it holds that

$$\bar{\Phi}^-(g_{t_1+\ln(1+\sigma),t_2+\ln(1+\sigma)}\tilde{\Lambda})\ge n \implies \Phi(g_{t_1,t_2}\Lambda)\ge n$$
$$\implies \bar{\Phi}^+(g_{t_1+\ln(1+\sigma),t_2+\ln(1+\sigma)}\tilde{\Lambda})\ge n$$

□

The rest of the paper will be devoted to the proof of Theorem 10.

5. Rate of equi-distirbution of unipotent flows.

Let $\mathbf{M}=SL_3(\mathbb{R})/SL_3(\mathbb{Z})$. Denote

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

We shall use the fact that the action of g_t on $\mathbf{M}$ is partially hyperbolic in the sense that

$$T\mathbf{M}=E_0+\sum_{q=1}^{3}E_q^+\oplus E_q^-$$

where E_0 is tangent to the orbit of g_t 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 $W_q = W_q^+$.

Definition 5.1. Given $\mathbf{s}, \mathbf{r} \geq 0$, we say that a function $A : M \to \mathbb{R}$ is in $H^{\mathbf{s},\mathbf{r}}$ with $||A||_{\mathbf{s},\mathbf{r}} = K$ if given $0 < \varepsilon \leq 1$ there are H^s-functions $A^- \leq A \leq A^+$ such that

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

Definition 5.2. Fix $\kappa_0 > 0$. Let $L > 0$ and $\mathcal{P}$ be a partition of 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

$$\mu\left(x : \left|\frac{1}{L_n}\int_{g^{k_n}\gamma(x)} A(s)ds - \widehat{A}\right| \geq \mathcal{K}_A L_n^{-\kappa_0}\right) \leq L_n^{-\kappa_0} \tag{10}$$

where $\widehat{A} = \int_M A(x)d\mu(x)$, $\mathcal{K}_A = ||A||_{\mathbf{s},\mathbf{r}} + 1$, and $L_n = Le^{\lambda_q(k_n)}$ is the length of the W_q curve $g^{k_n}\gamma(x)$ that goes through $g^{k_n}(x)$.

We call the points x such that

$$\forall n : \left|\frac{1}{L_n}\int_{g^{k_n}\gamma(x)} A(s)ds - \widehat{A}\right| \leq \mathcal{K}_A L_n^{-\kappa_0}$$

representative with respect to $(\mathcal{P}, \{k_n\}, A)$. Observe that if

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

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

The goal of this section is to show the following.

Proposition 11. *There exists $\mathbf{s}, \kappa_0, \varepsilon_0 > 0$ such that for any $0 \leq \mathbf{r} \leq \mathbf{s}$, $0 < \varepsilon \leq \varepsilon_0$, and any function $A \in H^{\mathbf{s},\mathbf{r}}$, and any L and a sequence $\{k_n\}$ satisfying*

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

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

The requirement that $\mathbf{r} \leq \mathbf{s}$ will only serve to maintain the exponent κ in the speed of equidistribution in (10) 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 $H^{\mathbf{s}_0,\mathbf{s}_0}$, since we will have to deal with characterisitc functions of nice sets (cf. section 6.3).

Proof. Without loss of generality we will work with functions A having zero average, that is $\widehat{A} = 0$. We will first prove proposition 11 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.

By [12], Theorem 2.4.5 there exists $\mathbf{s}$ and constants $C, c > 0$ 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

$$|\mu(A(x)B(gx)) - \mu(A)\mu(B)| \leq C||A||_{\mathbf{s}}||B||_{\mathbf{s}}e^{-c\max|t_j|}. \tag{11}$$

We claim that this implies that there exists $C > 0$ and $\kappa > 0$ such that

$$|\mu(A(x)B(h_3(u)x)) - \mu(A)\mu(B)| \leq C||A||_{\mathbf{s}}||B||_{\mathbf{s}}u^{-\kappa}. \tag{12}$$

Indeed let θ be such that $\tan\theta = e^{-2t}$ and let

$$U(t) = R(\theta)\begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{t} & 0 \\ 0 & 0 & 1 \end{pmatrix}R(-\theta)$$

where

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

A simple computation gives that $U(t) = h_3(u)$ with $u = e^{2t} + o(1)$, hence (12) follows from (11). Now, assuming that $\mu(A) = 0$, (12) implies that

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

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

$$\mu(S_L) = 0, \quad \mu(S_L^2) \leq CL^{-\kappa}\mathcal{K}_A^2.$$

This implies that

$$\mu(x : |S_L(x)| > \mathcal{K}_A L^{-\kappa/3}) \leq CL^{-\kappa/3}. \tag{13}$$

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 (13)

$$\bar{\mu}\left((x,u) : \frac{1}{L}\left|\int_{\gamma(x,u)} A(s)ds\right| \leq \mathcal{K}_A L^{-\kappa_0}\right) \leq CL^{-\kappa_0}$$

where $\bar{\mu}$ denotes the product of μ and the Lebesgue measure on $[0,1]$ and $\kappa_0 = \kappa/3$. Thus we can choose u so that $\hat{\mathcal{P}}^u$ satisfies

$$\mu\left(x : \frac{1}{L}\left|\int_{\gamma(x)} A(s)ds\right| \geq \mathcal{K}_A L^{-\kappa_0}\right) \leq CL^{-\kappa_0} \tag{14}$$

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

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

then there exists a partition $\mathcal{P}$ that is representative with respect to $(\{k_n\}, A)$ as in definition 5.2.

To extend (14) to functions in $H^{\mathbf{s},\mathbf{r}}$ (that eventually have infinite $H^{\mathbf{s}}$-norm), we use a standard approximation argument. Note first that (13) still holds for non zero mean $H^{\mathbf{s}}$-functions if we let

$$S_L^{\tilde{A}}(x) = \left(\frac{1}{L}\int_{\gamma(x)} \tilde{A}(h_3(u)x)du\right) - \mu(\tilde{A}).$$

On the other hand, we have that $0 \leq \widehat{A^+} \leq \varepsilon$, hence if $\mathcal{K}_A = ||A||_{\mathbf{s},\mathbf{r}} + 1$ we have that

$$\mu\left(S_L^A(x) \geq 2\mathcal{K}_A L^{-\tilde{\kappa}}\right) \leq \mu\left(S_L^{A^+}(x) \geq 2\mathcal{K}_A L^{-\tilde{\kappa}} - \varepsilon\right)$$

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

$$\mu\left(S_L^A(x) \geq 2\mathcal{K}_A L^{-\tilde{\kappa}}\right) \leq \mu\left(S_L^{A^+}(x) \geq ||A^+||_{\mathbf{s}} L^{-\kappa_0}\right) \leq L^{-\kappa_0}.$$

Using A^- to bound $\mu\left(S_L^A(x) \leq -2\mathcal{K}_A L^{-\tilde{\kappa}}\right)$ we see that (14) and thus the rest of the proof extends to $H^{\mathbf{s},\mathbf{r}}$ functions, provided the exponent κ_0 is reduced. □

6. Poisson Limit Theorem in the space of lattices.

6.1. Multiple solutions.

Lemma 12. *Assume that $\mathcal{L}$ has a smooth density on* $\mathbf{M}$*. Let Φ be defined as in (5). Denote $\Phi^t = \Phi \circ g_t$. Then we have as $M \to \infty$ and for any $t, t' \in \mathbb{Z}^2 - \{0,0\}$*

(*a*) $\mathbb{E}(\Phi^t) = \mathcal{O}(M^{-2})$;

(*a'*) $\mathbb{E}(\Phi^t) = \mathbf{c}_1 \frac{4}{\varepsilon} M^{-2} + \mathcal{O}(M^{-100})$ *if* $\min(t_1, t_2) \geq \sqrt{M}$

(*b*) $\mathbb{E}((\Phi^t)^2 - \Phi^t) = \mathcal{O}(M^{-4})$ *and hence* $\mathbb{P}(\Phi^t(\mathcal{L}) > 1) = \mathcal{O}(M^{-4})$;

(*c*) $\mathbb{P}(\Phi^t(\mathcal{L}) \neq 0$ *and* $\Phi^t(g_{t'}\mathcal{L}) \neq 0) = \mathcal{O}(M^{-4})$.

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 Roger's equalities of Lemma 4. Indeed, (a) of Lemma 4 would then imply that $\mathbb{E}(\Phi) = \mathbf{c}_1 \frac{4}{\varepsilon} M^{-2}$, since $\int_{\mathbb{R}^3} 1_I(x) 1_J(y) 1_K(xyz) dxdydz = \frac{4}{\varepsilon} M^{-2}$. On the other hand, if for $e = (x, y, z) \in \mathcal{L}$, we let $f(e) = 1_{I \times J \times K}(x, y, xyz)$, then since I is an interval of positive numbers, we have that

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

and the first estimate of part (b) follows by Lemma 4(b). The second estimate follows from the first by Markov inequality. As for (c) observe that if we define, for $e = (x, y, z) \in \mathcal{L}$, $g(e) = 1_{e^{-t'_1} I \times e^{-t'_2} J \times e^{t'_1 + t'_2} K}(x, y, xyz)$, then

$$\mathbb{E}(\Phi \Phi^{t'}) = \int_M \sum_{e_2 \neq \pm e_1 \in \mathcal{L} \text{ prime}} f(e_1) g(e_2) d\mu(\mathcal{L})$$

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 4(b) we get (*c*). Since in the case $\mathcal{L}$ is distributed according to the Haar measure we have that $\mathbb{E}(\Phi) = \mathbf{c}_1 \frac{4}{\varepsilon} M^{-2}$, then (a') follows for a smooth density by exponential mixing of the geodesic flow. □

Proof of Theorem 10(a). Part (*a*) of Theorem 10 is exactly the second part of (b) of Lemma 12. □

6.2. Poisson limit for the visits to the cusp. Let

$$\mathbf{\Pi} = \{\mathbf{t} : \mathbf{t}_1 > 0, \mathbf{t}_2 > 0, \mathbf{t}_1 + \mathbf{t}_2 < 1 - \delta\}$$
$$\Pi = \{\mathbf{t} : \mathbf{t}_1 > 0, \mathbf{t}_2 > 0, \mathbf{t}_1 + \mathbf{t}_2 < M(1 - \delta)\}$$

In this section we will prove the following result that will be extended in section 6.6 to yield Theorem 10.

Theorem 13. $\{\Psi_1(g_t\mathcal{L}), t_1/M, t_2/M\}_{\Phi(g_t\mathcal{L})=1}$ *converges to the Poisson process on* $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}] \times \mathbf{\Pi}$ *with intensity* $2\mathbf{c}_1$.

Proof. Similarly to the function Φ defined in (5), introduce for the rest of the paper the following shorthand notation. Divide K into a finite number of intervals $K_1, K_2 \dots K_{m_1}$ and let

$$\xi_t = \Phi(g_t\mathcal{L}), \quad \xi_{t,p} = \Phi_p(g_t\mathcal{L}) \tag{15}$$

where Φ_p is defined by (5) with K replaced by K_p.

Also let $\hat{\Phi}$ be defined by (5) with K replaced by $\hat{K} = \{z : d(z, \partial K) \le M^{-100}\}$. Last, consider the following collection of functions

$$\mathbf{\Phi} = \{\Phi_1 \dots \Phi_m, \hat{\Phi}\}.$$

Pick a small number $\tilde{\delta} \ll \delta$. Divide Π into H squares $C_1, \dots, C_H$ of sides $\tilde{\delta}M$ and one side parallel to $\text{Ker}\lambda_1$.

Fix $k \in \mathbb{Z}_+$. Pick k squares $S_1, S_2 \dots S_k \subset \{C_1, \dots, C_H\}$ with centers $t^{(q)}$. We call the square configuration $\tilde{\delta}$-generic if their images under λ_1 are distant by more than $3\tilde{\delta}M$. Also fix an index $i_q \in \{1 \dots m_1\}$ for each $1 \le q \le k$.

To obtain Theorem 13, we shall prove

Lemma 14.

(a) $\mathbb{P}(\exists t', t'' \in \Pi : \xi_{t'} = \xi_{t''} = 1$ *and* $|\lambda_1(t') - \lambda_1(t'')| \le 3\tilde{\delta}M) \to 0$ *as* $\tilde{\delta} \to 0$.

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

$$\mathbb{P}(\xi_t = 0 \text{ for } t \in \Pi \setminus \bigcup_q S_q \text{ and } \exists t^{(q)} \in S_q : \xi_{t^{(q)}, i_q} = 1 \text{ while } \xi_t = 0 \text{ for } t \in S_q - t^{(q)})$$

$$(2\mathbf{c}_1)^k \tilde{\delta}^{2k} \left(\prod_q |K_{i_q}|\right) \exp\left(-\frac{(1-\delta)^2}{\varepsilon}\right)(1 + o_{\tilde{\delta}\to 0}(1)).$$

Proof that Lemma 14 implies Theorem 13. Divide $\mathbf{\Pi}$ into subsets $\mathbf{\Pi}_1, \mathbf{\Pi}_2 \dots \mathbf{\Pi}_{m_2}$. Suppose that we want to find the probability that for each $(p, s) \in [1, \dots, m_1] \times [1, \dots, m_2]$, there are $l_{p,s}$ points, satisfying

$$\frac{t}{M} \in \mathbf{\Pi}_s, \Psi_1(g_t\mathcal{L}) \in K_p.$$

We will apply Lemma 14 with $k = \sum_{p=1,\dots,m_1; s=1,\dots,m_2} l_{p,s}$. For each s, there are $n_s \approx \frac{\text{Area}(\mathbf{\Pi}_s)}{\tilde{\delta}^2}$ squares in $M\mathbf{\Pi}_s$.

By Lemma 14(a) the contribution of non-generic choices of k squares is negligible as $\tilde{\delta} \to 0$. On the other hand by Lemma 14(b) generic choices contribute (recall that $\text{Area}(\mathbf{\Pi}) = (1-\delta)^2$)

$$\prod_{p,s} \left[\binom{n_s}{l_{p,s}} \left(2\mathbf{c}_1 |K_p| \tilde{\delta}^2\right)^{l_{p,s}}\right] \exp\left(-\frac{\text{Area}(\mathbf{\Pi})}{\varepsilon}\right)$$

$$\approx \prod_{p,s} \left[\frac{(2\mathbf{c}_1 |K_p| \text{Area}(\mathbf{\Pi}_s))^{l_{p,s}}}{l_{p,s}!} \exp\left(-|K_p| \text{Area}(\mathbf{\Pi}_s)\right)\right]$$

which is exactly the result required by Theorem 13. ∎

Proof of Lemma 14(a). By Bonferroni inequality

$$\mathbb{P}(\exists t', t'' \in \Pi : \xi_{t'} = \xi_{t''} = 1 \text{ and } |\lambda_1(t') - \lambda_1(t'')| \leq 3\tilde{\delta}M)$$
$$\leq \sum_{|\lambda_1(t')-\lambda_1(t'')|\leq 3\tilde{\delta}M} \mathbb{P}(\xi_{t'} = \xi_{t''} = 1) = \mathcal{O}(\tilde{\delta})$$

where the last inequality follows from Lemma 12. □

Remark. Define $\gamma_M(t) = M + t_1$. The same argument as above actually yields that

$$\mathbb{P}(\exists t', t'' \in \Pi : \xi_{t'} = \xi_{t''} = 1 \text{ and } |v' - v''| \leq 3\tilde{\delta}M) \to 0 \text{ as } \tilde{\delta} \to 0$$

where $v' \in \{\lambda_1(t'), \gamma_M(t')\}$ and $v'' \in \{\lambda_1(t''), \gamma_M(t'')\}$. This will be useful in the proof of Theorem 10 that will be given in section 6.6.

Before we prove Lemma 14 in Section 6.4 we prove first a standard estimate on the $H_{s,s}$ norms of Φ, Φ_p and $\hat{\Phi}$.

6.3. Estimates of norms.

Lemma 15. *For any $s \geq 0$ we have that*

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

(b) $\mathbb{E}(\Phi_i) = 2\mathbf{c}_1|K_i|, \quad \mathbb{E}(\hat{\Phi}) = \mathcal{O}(M^{-100})$

Proof. (a) We shall prove the bound for Φ, the other estimates being similar. Let ϕ be a C^∞ functions 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

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

Then $\Phi^-_\varepsilon \leq \Phi \leq \Phi^+_\varepsilon$ where Φ^+_ε is defined by (5) with I_I, I_J and I_K replaced by $I^+_{I,\varepsilon}, I^+_{J,\varepsilon}$ and $I^+_{K,\varepsilon}$, and Φ^- is defined by (5) with I_I, I_J and I_K replaced by $I^-_{I,\varepsilon}, I^-_{J,\varepsilon}$ and $I^-_{K,\varepsilon}$. By Lemma 4(a) $||\Phi^+_\varepsilon - \Phi^-_\varepsilon||_{L^1} \leq C\varepsilon$ and $||\Phi^\pm_\varepsilon||_{\mathbf{s}} \leq C\varepsilon^{-\mathbf{s}}$.

(b) The result follows directly from Lemma 4(a). □

6.4. Proof of Lemma 14(b).

Proof. Let $\{S_q\}_{q=1,\dots,k}$ be a fixed $\tilde{\delta}$-generic configuration of squares and let $\{i_q\}$ be a sequence of indices with values in $\{1, \dots, m_1\}$. Projecting our squares into the real line by λ_1 we obtain $2k$ points which divide $[0, 2(1-\delta)M]$ into $2k+1$ segments. The segments which are projections of S_q have length $\tilde{\delta}M$ while complimentary segments are longer. Subdividing the complimentary segments into segments of length $\tilde{\delta}M$ we obtain a partition of the segment[1] $[0, 2(1-\delta)M]$ by points

$$0 = \zeta_0 < \zeta_1 < \dots < \zeta_n = 2(1-\delta)M.$$

We call segments which are projections of one of the squares *type A* segments and the remaining segments *type B* segments. Let $\Pi_j = \lambda_1^{-1}[\zeta_{j-1}, \zeta_j]$. These strips have common boundaries. To

[1]We can always adjust $\tilde{\delta}$ to make sure there are no rests and that the squares are completely contained in the strips Π_j.

create some independence we let $\bar{\zeta}_j = \zeta_{j-1} + \sqrt{M}$ and let $\bar{\Pi}_j = \lambda_1^{-1}[\bar{\zeta}_j, \zeta_j]$. Lemma 12(a) implies that $\mathbb{P}(\xi_t \geq 1) = \mathcal{O}(M^{-2})$, hence

$$\mathbb{P}(\exists t \in \Pi \setminus \bigcup_j \bar{\Pi}_j : \xi_t \geq 1) = \mathcal{O}(1/\sqrt{M}).$$

Accordingly we can concentrate on the contributions of $t \in \bigcup_j \bar{\Pi}_j$. Similarly, we may assume that $\min(t_1, t_2) \geq \sqrt{M}$.

We say that $\bar{\Pi}_j$ is of type A if $[\zeta_{j-1}, \zeta_j]$ is of type A and that $\bar{\Pi}_j$ is of type B if $[\zeta_{j-1}, \zeta_j]$ is of type B. If $\bar{\Pi}_j$ is of type B we say that it is compatible if $\xi_t = 0$ for all $t \in \bar{\Pi}_j$. If $\bar{\Pi}_j$ is of type A we say that it is compatible if for q such that $S_q \subset \bar{\Pi}_j$, there exists $t \in S_q$ such that $\xi_{t,i_q} = 1$ and $\xi_{\bar{t}} = 0$ for $\bar{t} \in \bar{\Pi}_j - \{t\}$. Denote

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

We shall show that if $\bar{\Pi}_{j+1}$ is of type A then

$$p_{j+1} = 2\mathbf{c}_1 |K_{i_q}| \tilde{\delta}^2 p_j (1 + o(1)) + \mathcal{O}\left(\frac{1}{\ln M}\right) \tag{16}$$

and if $\bar{\Pi}_{j+1}$ is of type B then

$$p_{j+1} = p_j \left(1 - \frac{2\mathbf{c}_1}{\varepsilon} \frac{\text{Area}(\bar{\Pi}_{j+1})}{\text{Area}(\Pi)}(1 + o(1))\right) + \mathcal{O}\left(\frac{1}{\ln M}\right). \tag{17}$$

Combining (16) and (17) for all j we obtain part (b) of Lemma 14. We shall prove (16), (17) is similar.

Let $\mathcal{P}_j$ be a sequence of increasing partitions of size $L = (e^{\zeta_j} M^{100})^{-1}$ such that $\mathcal{P}_j$ is κ_0-representative with respect to $(t, \mathbf{\Phi})$ for every $t \in \bar{\Pi}_{j+1}$. This is possible by Proposition 11 since $\mathbf{\Phi} \subset H^{\mathbf{s},\mathbf{s}}$ and

$$\sum_{t \in \bar{\Pi}_{j+1}} \left(L e^{\lambda_1(t)}\right)^{-\kappa_0} \ll 1.$$

Given $t \in \bar{\Pi}_{j+1}$ we also take partitions $\mathcal{P}_j^t$ of size $L_t = (e^{\lambda_1(t)} M^{100})^{-1}$ which are representative with respect to $(\{\bar{t} \in \Pi_{j+1} : \lambda_1(\bar{t}) > \lambda_1(t) + R \ln M\}, \mathbf{\Phi})$. We can assume that $\mathcal{P}_j^t$ refines $\mathcal{P}_j$ by adding the endpoints of $\mathcal{P}_j$ to $\mathcal{P}_j^t$.

Observe that $\sum_{t,\bar{t} \in \Pi_{j+1}, \lambda_1(\bar{t}) > \lambda_1(t) + R \ln M} \left(L_t e^{\lambda_1(\bar{t})}\right)^{-\kappa_0} = \mathcal{O}(M^{-1000})$. Thus, if we let E_j be the set of $\mathcal{L}$ that are representative for $(t, \mathcal{P}_j, \mathbf{\Phi})$ for every $t \in \bar{\Pi}_j$ and are representative for $(\bar{t}, \mathcal{P}_j^t, \mathbf{\Phi})$ for every couple $t, \bar{t} \in \bar{\Pi}_{j+1}$ such that $\lambda_1(\bar{t}) > \lambda_1(t) + R \ln M$, then $\mathbb{P}(E_j^c) = \mathcal{O}(M^{-1000})$ by Section 5.

Let $\mathcal{F}_j$ denote the σ algebra generated by $\mathcal{P}_j$. Note that the set $\{\bar{\Pi}_1 \dots \bar{\Pi}_j \text{ are compatible}\}$ can be modified on a set of measure $\mathcal{O}\left(M^{-98}\right)$ so that the new set is $\mathcal{F}_j$ measurable. (Indeed, if $1_{\bar{\Pi}_1 \dots \bar{\Pi}_j \text{ are compatible}}$ is not constant on the element of $\mathcal{P}_j$ containing $\mathcal{L}$ then $g_t(\mathcal{L})$ passes in the $\mathcal{O}(M^{-100})$ neighborhood of the boundary of the set defining $\mathbf{\Phi}$ for some $t \in \bigcup_{\bar{j} \leq j} \bar{\Pi}_{\bar{j}}$.) Hence

$$\begin{aligned} p_{j+1} &= \mathbb{P}(\bar{\Pi}_1 \dots \bar{\Pi}_j \text{ are compatible and } \bar{\Pi}_{j+1} \text{ is compatible}) \\ &= \mathbb{E}(1_{\bar{\Pi}_1 \dots \bar{\Pi}_j \text{ are compatible}} \mathbb{P}(\bar{\Pi}_{j+1} \text{ is compatible } |\mathcal{F}_j)) + o(M^{-98}). \end{aligned} \tag{18}$$

Our goal for the rest of this section is to prove that outside of a set of negligible measure we have

$$\mathbb{P}(\bar{\Pi}_{j+1} \text{ is compatible } |\mathcal{F}_j)) = 2\mathbf{c}_1 |K_{i_q}| \tilde{\delta}^2 (1 + o(1)). \tag{19}$$

We let

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

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

We then get the following

Claim. *For $\mathcal{L} \in E_j$ and $t, t', \bar{t} \in \Pi_{j+1}$ with $\lambda_1(\bar{t}) > \lambda_1(t) + R \ln M$*

$$\mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j) = \frac{2\mathbf{c}_1|K_{i_q}|}{M^2}(1+o(1)) \tag{20}$$

$$\mathbb{E}(\xi_t^2 - \xi_t|\mathcal{F}_j) = \mathcal{O}(M^{-4}) \tag{21}$$

$$\mathbb{E}(\eta_t\eta_{t'}|\mathcal{F}_j) = \mathcal{O}(M^{-2}) \tag{22}$$

$$\mathbb{E}(\eta_t\eta_{\bar{t}}|\mathcal{F}_j) = \mathcal{O}(M^{-4}) \tag{23}$$

Proof of the claim. Because $\mathcal{L}$ is representative for $(t, \mathcal{P}_j, \mathbf{\Phi})$, (20) and (21) follow from parts (a) and (b) of Lemma 12. Now since since $\eta_t\eta_{t'} \le \xi_t$, (22) also follows from Lemma 12 (a). Equality (23) needs a little more work. Since $\eta_t \le 1$ and $\eta_{\bar{t}} \le 1$ we have

$$\begin{aligned}\mathbb{E}(\eta_t\eta_{\bar{t}}|\mathcal{F}_j) &= \mathbb{E}(\mathbb{E}(\eta_t\eta_{\bar{t}}|\mathcal{F}_j^t)|\mathcal{F}_j)\\ &\le \mathbb{E}(\mathbb{E}(1_{\mathbb{E}(\eta_t|\mathcal{F}_j^t)=1}\eta_{\bar{t}}|\mathcal{F}_j^t)|\mathcal{F}_j) + \mathbb{P}(0 < \mathbb{E}(\eta_t|\mathcal{F}_j^t) < 1|\mathcal{F}_j).\end{aligned}$$

Next

$$\mathbb{E}(\mathbb{E}(1_{\mathbb{E}(\eta_t|\mathcal{F}_j^t)=1}\eta_{\bar{t}}|\mathcal{F}_j^t)|\mathcal{F}_j) = \mathbb{E}(1_{\mathbb{E}(\eta_t|\mathcal{F}_j^t)=1}\mathbb{E}(\eta_{\bar{t}}|\mathcal{F}_j^t)|\mathcal{F}_j).$$

Recall that if $\mathcal{L}$ is representative for $(t, \mathcal{F}_j^t)$ then

$$\mathbb{E}(\eta_{\bar{t}}|\mathcal{F}_j^t) \le \mathbb{E}(\xi_{\bar{t}}|\mathcal{F}_j^{\bar{t}}) = \mathbb{E}(\xi_{\bar{t}}) + \mathcal{O}(M^{-100}).$$

Since $\mathbb{E}(\xi_{\bar{t}}) = \mathcal{O}(M^{-2})$, the last expression is bounded by

$$\begin{aligned}&\mathcal{O}(M^{-2})\mathbb{E}(1_{\mathbb{E}(\eta_{\bar{t}}|\mathcal{F}_j^{\bar{t}})=1}|\mathcal{F}_j) + \mathcal{O}(M^{-100})\\ &\qquad \le \mathcal{O}(M^{-2})\mathbb{E}(\eta_{\bar{t}}|\mathcal{F}_j) + \mathcal{O}(M^{-100}) = \mathcal{O}(M^{-4}).\end{aligned}$$

On the other hand if $0 < \mathbb{E}(\eta_{\bar{t}}|\mathcal{F}_j^{\bar{t}}) < 1$ then $z(g_{\bar{t}}\mathcal{L})$ is M^{-100} close to the boundary of K, that is $\mathbb{P}(0 < \mathbb{E}(\eta_t|\mathcal{F}_j^t) < 1|\mathcal{F}_j) \le \mathbb{E}(\hat{\phi} \circ g^t|\mathcal{F}_j) = \mathcal{O}(M^{-100})$. The claim is proved. □

Back to the proof of (16), we have that

$$\mathbb{P}(\exists t \in \bar{\Pi}_{j+1} : \xi_t > 1|\mathcal{F}_j) \le \sum_{t\in\bar{\Pi}_{j+1}} \mathbb{E}(\xi_t^2 - \xi_t|\mathcal{F}_j).$$

Let

$$\begin{aligned}I &= \mathbb{P}(\exists t \in S_q \cap \bar{\Pi}_{j+1} : \xi_{t,i_q} = 1,\ \text{and } \xi_{\bar{t}} = 0 \text{ for } \bar{t} \ne t, \bar{t} \in \bar{\Pi}_{j+1}|\mathcal{F}_j)\\ \mathit{II} &= \mathbb{P}(\exists t \in S_q \cap \bar{\Pi}_{j+1} : \eta_{t,i_q} = 1,\ \text{and } \eta_{\bar{t}} = 0 \text{ for } \bar{t} \ne t, \bar{t} \in \bar{\Pi}_{j+1}|\mathcal{F}_j)\end{aligned}$$

Then, due to (21),

$$|I - \mathit{II}| \le \sum_{t\in\bar{\Pi}_{j+1}} \mathbb{E}(\xi_t^2 - \xi_t|\mathcal{F}_j) = \mathcal{O}(M^{-2}). \tag{24}$$

Next, since for a fixed $t \in S_q$ we have that

$$\begin{aligned}&\mathbb{E}\left(\eta_{t,i_q} - \sum_{t'\in\bar{\Pi}_{j+1}, t'\ne t} \eta_t\eta_{t'}|\mathcal{F}_j\right) \le\\ &\qquad \mathbb{P}(\eta_{t,i_q} = 1,\ \text{and } \eta_{t'} = 0 \text{ for } t' \ne t, t' \in \bar{\Pi}_{j+1}|\mathcal{F}_j) \le \mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j)\end{aligned}$$

then

$$(25)\qquad \left| I\!I - \sum_{t\in S_q\cap\bar\Pi_{j+1}} \mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j)\right| \le \sum_{t\in S_q\cap\bar\Pi_{j+1}, t'\neq t\in\bar\Pi_{j+1}} \mathbb{E}(\eta_t\eta_{t'}|\mathcal{F}_j) \le \mathcal{O}(\tilde\delta^3)+\mathcal{O}(M^{-1}\ln M)$$

using (22) for bounding the terms with $|t-t'|\le R\ln M$, and (23) for $|t-t'|>R\ln M$.

Due to (20) we get that

$$(26)\qquad \sum_{t\in S_q\cap\bar\Pi_{j+1}} \mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j) = 2\mathbf{c}_1|K_{i_q}|\tilde\delta^2(1+o(1)).$$

Now, (19) follows for $\mathcal{L}\in E_j$ from (24), (25) and (26). Finally,

$$\begin{aligned} p_{j+1} &= \mathbb{P}(\bar\Pi_1\dots\bar\Pi_j \text{ are compatible and } \bar\Pi_{j+1} \text{ is compatible}) \\ &= \mathbb{P}(\bar\Pi_1\dots\bar\Pi_j \text{ are compatible and} \\ &\qquad \bar\Pi_{j+1} \text{ is compatible and } \mathcal{L}\in E_j) + \mathcal{O}\left(M^{-1000}\right) \\ &= p_j 2\mathbf{c}_1|K_{i_q}|\tilde\delta^2(1+o(1)). \end{aligned}$$

This completes the proof of (16). (17) follows lickewise. Part (b) of Lemma 14 is thus completed. □

6.5. Proof of Theorem 10 (c) and (d). Note that for Poisson process on $[0,\frac{1}{\varepsilon}]\times\mathbf{\Pi}$ given η we can find $\bar\delta$ such that

$$\mathbb{P}(\min_{i,j}|\tau(\mathbf{t}^{(i)})-\tau(\mathbf{t}^{(j)})|<\bar\delta)<\eta.$$

Therefore (c) of Theorem 10 follows from Theorem 13. Theorem 10(d) follows likewise. □

6.6. Proof of Theorem 10(b). In the proof of Poisson limit for $\{(\Psi_1(g_t\mathcal{L}), t/M\}$ an important role was played by foliation of $\mathbf{M}$ into the leaves of W_1. Indeed, in the typical situation where $t_1,\dots,t_s\in Z$ are sufficiently separated (our genericity condition) the independence between the $\Psi_1(g_{t_i}\cdot)$ is due to the fact that $\Psi_1(g_{t_i}\cdot)$ is determined on a scale $e^{-\lambda_1(t_i)}$ of W_1 leafs, a scale on which the successive $\Psi_1(g_{t_j}\cdot)$, $j>i$ are completely free.

The key to the proof of part (b) of Theorem 10 is that Ψ_1 and Ψ_2 are determined at different scales along the W_1 leafs. More precisely $\Psi_2(g_{t_i}\cdot)$ is determined on a scale $1/(Ne^{\lambda_1(t_i)})$, which means that $\Psi_2(g_{t_i}\mathcal{L})$ is uniformly distributed on $[0,2]$ if $\mathcal{L}$ moves along a W_1 leaf of size much larger than $1/(Ne^{\lambda_1(t_i)})$. As a consequence, since the scales $\{e^{-\lambda_1(t_1)},\dots,e^{-\lambda_1(t_s)},1/(Ne^{\lambda_1(t_1)}),\dots,1/(Ne^{\lambda_1(t_s)})\}$ (that have to be rearranged in an increasing order) are typically sufficiently split (see the generectiy condition below) then the independence of the quantites $\Psi_1(g_{t_1}\mathcal{L}),\dots,\Psi_1(g_{t_s}\mathcal{L}),\Psi_2(g_{t_1}\mathcal{L}),\dots,\Psi_2(g_{t_s}\mathcal{L})$ is insured.

Thus, the proof of Theorem 10(b) proceeds along the lines of the proof of Theorem 13 with the modifications described below.

Now in addition to choosing $\{S_q\}_{q\le k}$ and $\{i_q\}_{q\le k}$ as in Lemma 14, we also divide $\mathbb{T}^1$ into segments $\mathbf{T}_1\dots\mathbf{T}_{m_2}$ and choose a sequence $\{l_q\}$ with values in $\{1\dots m_2\}$.

We shall say that $\bar t$ is $\tilde\delta$-resonant with respect to t if

$$(27)\qquad |\lambda_1(\bar t)-\gamma_M(t)|\le 3\tilde\delta M$$

where $\gamma_M(t)=M+t_1$.

Let R_q be the set of points which are $\tilde\delta$-resonant with respect to some point in $t\in\{S_q\}_{q\le k}$. We modify the definition of $\tilde\delta$-genericity by requiring that the images of $\{S_q\}_{q=1}^k$ and $\{R_q\}_{q=1}^k$ by λ_1 are $\tilde\delta$-disjoint. We call the projections of R_q by λ_1 *type C intervals.*

As observed in the remark following the proof of Part (a) of Lemma 14 the following is valid :

$$\mathbb{P}(\exists t', t'' \in \Pi : \xi_{t'} = \xi_{t''} = 1 \text{ and } |v' - v''| \le 3\tilde{\delta} M) \to 0 \text{ as } \tilde{\delta} \to 0.$$

with $v' \in \{\lambda_1(t'), \gamma_M(t')\}$, $v'' \in \{\lambda_1(t''), \gamma_M(t'')\}$.

Part (b) of Lemma 14 has to be modified as follows:

If $S_1, S_2 \dots S_k$ is generic then

$$\mathbb{P}\Bigg(\xi_t = 0 \textit{ for } t \in \Pi \setminus \bigcup_q S_q \textit{ and } \exists t^{(q)} \in S_q : \\ \xi_{t^{(q)}, i_q} = 1, \Psi_2\left(\frac{N}{e^{t_1^{(q)} + t_2^{(q)}}}, g_{t^{(q)}}\mathcal{L}\right) \in T_{l_q} \textit{ while } \xi_t = 0 \textit{ for } t \in S_q - t^{(q)}\Bigg) \\ = (2\mathbf{c}_1)^k \tilde{\delta}^{2k} \left(\prod_q |K_{i_q}||\mathbf{T}_{l_q}|\right) \exp\left(-\frac{(1-\delta)^2}{\varepsilon}\right)(1 + o_{\tilde{\delta}\to 0}(1)).$$

As in the proof of Lemma 14 (b), the latter probability is established inductively. The definition of compatibility for type A and B strips remains the same as in the proof of Lemma 14.

Let $\bar{\Pi}_{j+1}$ be a strip of type C so that Π_{j+1} contains R_q. Note that $j + 1 > i$ for the i such that Π_i contains S_q (because in Π, $t_1, t_2 \ge 0$ and $t_1 + t_2 < M(1-\delta)$). Hence in the definition of compatibility of Π_{j+1} we assume given the value of $t^{(q)} \in S_q \subset \Pi_i$ such that $\xi_{t^{(q)}, i_q} = 1$.

We then say that $\bar{\Pi}_{j+1}$ is compatible if

$$\xi_t = 0 \text{ for } t \in \bar{\Pi}_{j+1} \text{ and } \Psi_2\left(\frac{N}{e^{t_1^{(q)} + t_2^{(q)}}}, g_{t^{(q)}}\mathcal{L}\right) \in T_{l_q} \tag{28}$$

Recall the definition $p_j = \mathbb{P}(\bar{\Pi}_l$ are compatible for $l \le j)$. Then the proof of Theorem 10(b) is the same as the proof of Theorem 13 if (16) and (17) that determine the inductive relation on the probabilities p_j are supplemented by the following equation when Π_{j+1} is of type C :

$$p_{j+1} = p_j |\mathbf{T}_{l_q}|(1 + o(1)).$$

We just need to show that

$$\mathbb{P}(\bar{\Pi}_{j+1} \text{ is compatible } | \mathcal{F}_j) = |\mathbf{T}_{l_q}|(1 + o(1)). \tag{29}$$

Proof of (29). By the definition (27) of resonance between R_q and S_q, we have that $Ne^{t_1^{(q)}} \ge N^{\tilde{\delta}} e^{\zeta_j}$ where ζ_j is the maximal value of λ_1 on Π_j. We claim that this implies

$$\mathbb{P}\left(\Psi_2\left(\frac{N}{e^{t_1^{(q)} + t_2^{(q)}}}, g_{t^{(q)}}\mathcal{L}\right) \in T_{l_q} \big| \mathcal{F}_j\right) = |\mathbf{T}_{l_q}|(1 + o(1)).$$

Indeed, observe that if $\Phi_{i_q}(g_{t^{(q)}}\mathcal{L}) = 1$ with (x, y, z) the vector of $g_{t^{(q)}}\mathcal{L}$ such that $(x, y, xyz) \in I \times J \times K_q$, then we can assume that $\Phi_{i_q}(g_{t^{(q)}} h^1_\tau \mathcal{L}) = 1$ for $\tau \in [0, e^{-\zeta_j}]$ due to the vector

$$(x, y, xe^{2t_1^{(q)} + t_2^{(q)}}\tau + z).$$

Since

$$\Psi_2\left(\frac{N}{e^{t_1^{(q)} + t_2^{(q)}}}, g_{t^{(q)}} h^1_\tau \mathcal{L}\right) = \frac{N}{e^{t_1^{(q)} + t_2^{(q)}}}(xe^{2t_1^{(q)} + t_2^{(q)}}\tau + z) \bmod (1),$$

the equidistribution follows from $Ne^{t_1^{(q)}} \ge N^{\tilde{\delta}} e^{\zeta_j}$.

On the other hand, the first condition in (28) is not very restrictive since it is violated with probability $o(1)$ due to (17). (29) is thus established finishing the proof of Theorem 10(b). $\square$

7. Small boxes.

One can ask what happens if we 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 while case $\gamma = 1/d$ was studied in [13]. For $\gamma > 1/d$ most orbits do not visit $\mathcal{C}_N$ so we consider the remaining case $0 < \gamma < \frac{1}{d}$.

Theorem 16. *Under the assumptions of Theorem 1* $\frac{D_{\mathcal{C}_N}(x,\alpha,N)}{\boldsymbol{\rho}((1-d\gamma)\ln N)^d}$ *converges to the standard Cauchy distribution.*

The proof of Theorem 16 is the same as the proof of Theorem 1 except that now we can neglect the contribution of ks where $|\bar{k}_j| < N^\gamma$ for some j (cf. Section 3.3). Accordingly in Thereom 10 $\mathbf{\Pi}$ has to be replaced by $\mathbf{\Pi}_\gamma = \{t : t_j > \gamma M, \sum_j t_j < M\}$ which decreases the intensity of the limiting Poisson process by factor $(1 - d\gamma)^d$.

8. 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 the continuous time discrepancy function as

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

where $S_v^t = x + vt$.

In the case of balls, it was shown in [5] 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 cubes and gives rise to a Cauchy distribution after normalization by $\ln T$. This will be proved in Section 8.2 below.

It was also shown in [5] that for balls in dimension $d = 2$ the continuous time discrepancy converges, without any normalization, in distribution. In the next Section 8.1 we will show that this is also the case in any dimension $d \geq 2$ for the continuous time discrepancy for boxes.

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

Proof. We have

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

where $\Phi_m(u) = \sin(2\pi m u)/m$ and $\bar{k}_j$ is given by (2). We claim that for almost all A, v there exist a constant $C(A, v)$ such that

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

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

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

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 $\sum_k \Gamma_k(A, v) < \infty$ where

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

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

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

Then

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

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

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

provided that $|k|$ is large enough.

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

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

Hence

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

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

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

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 $\mathbf{D}_{\mathcal{C}}(v, x, T)$ is close to the distribution of

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

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

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

Theorem 18. *Let $\mathcal{C} = \prod_j (0, u_j)$. Suppose that the pair (x, v) has a smooth distribution of compact support. Then $\mathbf{D}_{\mathcal{C}}(v, x, T)$ converges in distribution as $T \to \infty$.*

The proof of Theorem 18 is similar to the proof of Theorem 17 with the additional simplifications since there is only one small denominator (k, v). So we leave the proof to the reader.

8.2. 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 and r is uniformly distributed on some segment $[a, b]$. Let σ denote the product of the distribution of v, the distribution of r and the Haar measure on $\mathbb{T}^3$.

Theorem 19. *There exists a constant* $\tilde{\rho}$ *such that* $\frac{\mathbf{D}_{B(0,r)}(v,x,T)}{\tilde{\rho} r \ln T}$ *converges to the standard Cauchy distribution.*

Proof. The proof is similar to the proof of Theorem 1 so we just outline the main steps. We have

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

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 $\frac{\varepsilon}{\ln T} < |(k, v)| k^2 < \frac{1}{\varepsilon \ln T}$ and $|k| < T$. Therefore the key step in proving Theorem 19 is the following.

Proposition 20. *The set*

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

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

The proof of Proposition 20 is similar to the proof of Theorem 7 and consists of the following steps.

(a) We prove the Poisson limit for $\{k^2(k, v) \ln T\}$ using the argument of Section 6.2.

(b) We prove that $(v, T) \bmod 2$ is asymptotically independent of $k^2(k, v) \ln T$ using the fact that their values are determined at different scales (cf. Section 6.6).

(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. Theorem 9). □

References

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

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

[3] A. Bufetov, G. Forni *Limit Theorems for Horocycle Flows,* arXiv preprint 1104.4502.

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

[5] D. Dolgopyat, B. Fayad, Deviations of ergodic sums for toral translations-I: Convex bodies, preprint.

[6] 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.

[7] F. Huveneers *Subdiffusive behavior generated by irrational rotations,* Erg Th. Dyn. Sys. **29** (2009) 1217–1233.

[8] J. F. Koksma, *Diophantische Approximationen*, Ergebnisse der Mathematik, Band IV, Heft 4, Berlin, 1936.

[9] H. Kesten *Uniform distribution* mod 1*,* Ann. of Math. **71** (1960) 445–471.

[10] H. Kesten *Uniform distribution* mod1*, II,* Acta Arith. **7** (1961/1962) 355–380.

[11] A. Khintchine *Ein Satz ber Kettenbrche, mit arithmetischen Anwendungen,* Mathematische Zeitschrift **18** (1923) 289–306.

[12] D. Kleinbock, G. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces,* in *Sinai's Moscow Seminar on Dynamical Systems* AMS. Transl. **171** (1996) 141–172

[13] J. Marklof *The n-point correlations between values of a linear form,* Erg. Th., Dynam. Sys. **20** (2000) 1127–1172.

[14] J. Marklof *Almost modular functions and the distribution of* n^2x *modulo one,* Int. Math. Res. Not. **39** (2003) 2131–2151.

[15] A. Ostrowski *Bemerkungen zur Theorie der Diophantischen Approximationen,* Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg **1** (1922) 77 − 98.
[16] K. F. Roth, *On irregularities of distribution,* Mathematika 1 (1954), 73-79
[17] S. Resnick *Extreme Values, Regular Variation And Point Processes*, Springer, 2007, 320 pp.
[18] G. Samorodnitsky and M. Taqqu *Stable non-Gaussian random processes,* Chapman & Hall/CRC, 1994, 640 pp.
[19] W. Schmidt, *Metrical theorems on fractional parts of sequences*, Trans. Amer. Math. Soc. **110** (1964) 493–518.
[20] I. Vinogradov *Limiting distribution of visits of several rotations to shrinking intervals,* preprint.