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

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

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 $x + n\alpha$ 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 [30, 20, 35, 26, 40] and others, [2] proves that for arbitrary positive increasing function $\phi(n)$

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

In dimension $d = 1$, this result is the content of Khinchine theorems obtained in the early 1920's [26], 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 andÊ (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 will be as bad as $(\ln N)^d$ along a suitable subsequence but not much is known better than the general lower bound $(\ln N)^{d/2}$ that

holds for every sequence on $\mathbb{T}^d$ ([38]). 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, [23, 9] obtain a Central Limit Theorem in the one dimensional case of circle rotations. The results of [23, 9] 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 both 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 [24, 25] (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 [24, 25]**.** *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 ν_2 *is the Lebesgue measure on* $\mathbb{T}^2$ *and*

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

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. 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 [12] 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 $\mathcal{C}$ and is not a standard stable law).

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

$$(1.3) \quad G_\eta = \{U = (a_{ij}) \in \mathrm{SL}_d(\mathbb{R}) : |a_{i,i} - 1| < \eta, \forall i \text{ and } |a_{i,j}| < \eta \; \forall j \neq i\}.$$

We denote by UC_u the image of C_u by a matrix $U \in G_\eta$. 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

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

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

$$(1.4) \quad D(\vartheta, N) = \#\{1 \leq m \leq N : (x + m\alpha) \bmod 1 \in UC_u\} - 2^d (\Pi_i u_i) N.$$

Theorem A. *For any $z \in \mathbb{R}$ we have*

$$(1.5) \qquad \lim_{N \to \infty} \lambda\{\vartheta \in X \; / \; \frac{D(\vartheta, N)}{(\ln N)^d} \leq z\} = \mathfrak{C}(\boldsymbol{\rho} z)$$

where $\mathfrak{C}$ is defined by the Cauchy distribution function (1.2) and
(1.6)
$$\boldsymbol{\rho} = \frac{1}{\zeta(d+1)d!} \left(\frac{2}{\pi}\right)^d \int \cdots \int \left| \sum_{j=1}^\infty \frac{[\Pi_{i=1}^d \sin(2\pi j \eta_i)] \sin(\pi j \eta_{d+1}) \cos(2\pi j \eta_{d+2})}{j^{d+1}} \right| d\eta_1 \ldots d\eta_{d+2}$$

where $\zeta(d+1) = \sum\limits_{n=1}^\infty \frac{1}{n^{d+1}}$ is the Riemann zeta function.

As it 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 rU with r smoothly distributed in a neighborhood of 1 and $U \in G_\eta$. 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 [13] for more results and open questions on this subject.

Our proof of Theorem A shows that for typical α, a *quenched* limit (that is, with fixed α, and x uniformly distributed on $\mathbb{T}^d$) of $D(\alpha, x, C, N)$ 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 refer the reader to [14, 15] for a more detailed version of this argument). We note that the absence of quenched limits is often observed in zero entropy systems [4, 5, 6, 12, 17, 19, 33]. To finish, we mention that the paper [13] can be consulted for an introduction to problems and results related to limit theorems for toral translations, an active domain of research in the recent years.

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 §2.1 we recall the representation of the Cauchy distribution in terms of a Poisson process. In §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 us to isolate the frequencies that make essential contributions to the discrepancy and to 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 A to that of Theorem B establishing a Poisson limit theorem for the distribution of the small denominators that appear in the (resonant) Fourier terms that contribute to the discrepancy.

Namely, for each $\bar{\epsilon} > 0$ we need to prove the Poisson limit theorem for the sequence

$$(\star) \quad \left\{ (\ln N)^d \prod_i \bar{k}_i \big\| \langle k, \alpha \rangle \big\|, N \langle k, \alpha \rangle \bmod 2, \{\bar{k}_1 u_1\}, \ldots, \{\bar{k}_d u_d\}, \{\langle k, x \rangle\} \right\}_{k \in Z(\vartheta, N)}$$

where $\langle \cdot, \cdot \rangle$, $\| \cdot \|$ and $\{\cdot\}$ denote respectively the Eucledian scalar product, the closest distance to integers and the fractional part, and where

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

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

and

$$W(\vartheta, N) := \left\{ k \in \mathbb{Z}^d : \left| \prod_{i=1}^{d} \bar{k}_i \right| < N, \right.$$

$$\left. \forall i = 1, \ldots, d, \quad |\bar{k}_i| \geq 1, \quad \left| \prod_{i=1}^{d} \bar{k}_i \right| \|\langle k, \alpha \rangle\| \leq \frac{1}{\bar{\epsilon}(\ln N)^d} \right\}.$$

In Section 5, we reduce the Poisson limit of the first two coordinates of ($\star$) to a Poisson limit theorem (Theorem 5.4) for the number of visits to a cusp by orbits of the Cartan action on the space of $d + 1$ dimensional unimodular lattices $\mathcal{M} = SL_{d+1}(\mathbb{R})/SL_{d+1}(\mathbb{Z})$. 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 5.3, where the argument is similar to the original analysis of Kesten.

The proof of Theorem 5.4 occupies Sections 6, 7, and 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, 8, 10, 11, 18, 21, 22, 36] and the references therein). The most relevant work for our purposes is the paper [11] 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 $\bar{k}_i$s that contribute to the resonances in ($\star$). However Rogers identities ([41, 37]) provide sufficiently strong control to handle this issue. Secondly, we need to consider the action of the full diagonal subgroup of $SL_{d+1}(\mathbb{R})$ because, for a typical resonance, $\bar{k}_1, \bar{k}_2 \ldots,$and $\bar{k}_d$ 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 fact that the we need to prove the Poisson Limit Theorem for higher rank subgroups constitutes the main novelty of Sections 6–8.

The relevant equidistribution results for unipotent subgroups of $SL_{d+1}(\mathbb{R})$ acting on $\mathcal{M}$ are presented in Section 7. To exploit these equidistribution results in the proof of ($\star$), we introduce additional parameters in the form of small affine deformations of the box. Indeed, 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 $\mathcal{M}$; while with the randomly slightly tilted boxes we have to establish a Poisson Limit Theorem for lattices having a smooth density on $\mathcal{M}$.

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 A. The case $\gamma = 1/d$ was studied in [32] where a limit distribution was obtained without any normalization. In the case $\gamma > 1/d$, the problem 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 A 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 [12] 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 appendices.

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 $\mathfrak{X}_j$ 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 [27, 39].

Lemma 2.1. *(see [27], §§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 from 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 measure $f(\theta)d\theta$ then $\tilde{\Theta}$ has measure $\tilde{f}(\theta)d\theta$ with

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

Next, recall [16, 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. *(a) If $\{\Theta_j\}$ is a Poisson process on $\mathbb{R}$ with measure $c\theta^{-2}d\theta$ 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

$$\lim_{\bar{\epsilon} \to 0} \frac{1}{\rho} \sum_{|\Theta_j| < \bar{\epsilon}^{-1}} \frac{\Gamma_j}{\Theta_j}$$

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

We provide a sketch of the proof to illustrate the idea of the argument. For more detailed presentation we refer the readers to [39, Theorem 1.4.2], or [15, Appendix B].
Sketch of proof. To see part (a), let $\{U_j'\}$, $\{U_j''\}$ and $\{U_j\}$ be independent Poisson processes with measure c. We want to show that $\sum \frac{1}{U_j'} + \sum \frac{1}{U_j''}$ have

the same distribution as $\sum \frac{2}{U_j}$. But

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

and we finish by observing, in light of Lemma 2.1 (a) and (b), that both $\{U_j'\} \cup \{U_j''\}$ and $\{\frac{U_j}{2}\}$ are Poisson processes with intensity $2c$.

The proof of part (b) follows the same idea as the proof of part (a), but now we use parts (b), (c) and (d) of Lemma 2.1.

2.2. **Siegel and 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

$$(2.1) \quad \mathbf{c}_1 = \zeta(d+1)^{-1}, \quad \mathbf{c}_2 = \zeta(d+1)^{-2}, \quad \text{where } \zeta(d+1) = \sum_{n=1}^{\infty} n^{-(d+1)}$$

is the Riemann zeta function.

The following identities (see [41, 37] as well as [32, 42]) 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 $\mathcal{L} \subset \mathcal{M}_{d+1}$, we say that a vector in $\mathcal{L}$ is prime if it is not an integer multiple of another vector in $\mathcal{L}$. Let

$$F(\mathcal{L}) = \sum_{\mathbf{v} \in \mathcal{L}, \text{ prime}} f(\mathbf{v}), \quad \bar{F}(\mathcal{L}) = \sum_{\mathbf{v}_1 \neq \pm \mathbf{v}_2 \in \mathcal{L}, \text{ prime}} f_1(\mathbf{v}_1) f_2(\mathbf{v}_2).$$

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 A is obtained by applying the results of §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 proof of Proposition 3.1 is independent from the rest of the paper and can be skipped in a first reading.

3.1.   Recall from §1.1 the definition

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

where G_η is given by (1.3). For $\vartheta \in X$ and $k \in \mathbb{Z}^d$, we use the notation

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

Writing the Fourier series of the characteristic function of a box we get for the discrepancy $D(\vartheta, N)$ defined in (1.4)

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

where

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

and $\varphi_{k,\alpha,N} = \dfrac{\pi(N-1)\langle k, \alpha \rangle}{2}$, $\mathfrak{a} = \dfrac{1}{\pi^d}$, $\langle k, x \rangle = \displaystyle\sum_{i=1}^{d} k_i x_i$.

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 $u = O(v)$, or equivalently $u \ll v$, when $|u| \le C|v|$ for some constant C that does not depend on $\bar{\epsilon}$ or N.

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

$$(3.2) \quad W(\vartheta, N) := \left\{ k \in \mathbb{Z}^d : \left| \prod_{i=1}^{d} \bar{k}_i \right| < N, \right.$$

$$\left. \forall i = 1, \ldots, d, \quad |\bar{k}_i| \ge 1, \quad \left| \prod_{i=1}^{d} \bar{k}_i \right| \|\langle k, \alpha \rangle\| \le \frac{1}{\bar{\epsilon}(\ln N)^d} \right\}.$$

Next, we let

$$(3.3) \quad Z(\vartheta, N) = \left\{ k \in W(\vartheta, N) : \bar{k}_1 > 0 \text{ and } \exists m \in \mathbb{Z} \text{ such that} \right.$$

$$\left. k_1 \wedge \ldots \wedge k_d \wedge m = 1 \text{ and } \left\| \langle k, \alpha \rangle \right\| = |\langle k, \alpha \rangle + m| \right\}.$$

Then define

$$(3.4) \qquad\qquad \bar{D}(\vartheta, N) = \sum_{k \in Z} \frac{\Gamma_k(\vartheta, N)}{\Omega_k(\vartheta, N)}$$

where

$$(3.5) \qquad\qquad \Omega_k(\vartheta, N) = \left(\prod_{i=1}^{d} \bar{k}_i \right) \left\| \langle k, \alpha \rangle \right\| (\ln N)^d,$$

$$(3.6) \qquad \Gamma_k(\vartheta, N) = \frac{2\mathfrak{a}}{\pi} \phi \left(\bar{k}_1 u_1, \dots, \bar{k}_d u_d, N\langle k, \alpha \rangle, \langle k, x \rangle + \varphi_{k,\alpha,N} \right),$$

and

$$(3.7) \qquad \phi(\eta_1, \dots, \eta_d, \eta_{d+1}, \eta_{d+2}) =$$
$$\sum_{j=1}^{\infty} \frac{\left[\Pi_{i=1}^d \sin(2\pi j \eta_i) \right] \sin(\pi j \eta_{d+1}) \cos(2\pi j \eta_{d+2})}{j^{d+1}}.$$

The purpose of this section is show that $\left| \frac{D}{(\ln N)^d} - \bar{D} \right|$ 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*

$$(3.8) \qquad \lambda \left(\left\{ \vartheta \in X : \left| \frac{D(\vartheta, N)}{(\ln N)^d} - \bar{D}(\vartheta, N) \right| \geq \upsilon \right\} \right) \leq \upsilon.$$

Remark 3.2. *Differences with the discrepancies relative to convex sets.* Proposition 3.1 identifies the normalization term $(\ln N)^d$ and the resonant terms in the Fourier series of the discrepancy function that contribute to its limiting distribution after normalization. These terms involve multiplicative small denominators of the form $\left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\|$ with $\left| \prod_i \bar{k}_i \right| \leq N$. The fact that frequency vectors k coming from many different scales in the set $\left| \prod_i \bar{k}_i \right| \leq N$ yield small denominators $\left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\|$ of comparable strength, as well as the asymptotic independence of the small denominators from different scales, are key in proving that the limiting distribution of the normalized discrepancy turns out to be a classical stable law (see Section 4 below). A similar analysis lies behind Kesten's proof in the case $d = 1$.

In contrast, the resonant terms in the Fourier series of the discrepancy function relative to a convex set involve small denominators of the form $|k|^{\frac{d+1}{2}} \|\langle k, \alpha \rangle\|$. The latter form of the small denominators limits the contributing terms to frequencies k of the same scale. Namely it is shown in [12] that the main contribution to the discrepancy comes from frequencies $|k|$ that are of the order of $N^{\frac{1}{d}}$, which leads to the normalization factor $N^{\frac{d-1}{2d}}$. Using Dani correspondence and mixing of expanding translates of horoshperical subgroups, it was shown in [12] that the limiting distribution of the normalized discrepancy function is given by a Siegel transform of a certain function on the space of marked lattices where the marked lattice is chosen at random. This is a special case of the distribution obtained by taking a certain function on an appropriate moduli space and evaluating them at a random point. Such distributions are not well studied in probabilistic literature, even though recently it has been shown that they appear in many problems related to distribution of ergodic averages of renormalizable

systems ([5, 33, 19]). The first steps towards creating the general theory of these distributions are made in [7, 34] but more work is needed in this direction.

We will need several lemmas to prove Proposition 3.1. The lemmas will involve L^2 estimates with respect to the variables $(\alpha, x) \in \mathbb{T}^{2d}$ as well as an exclusion of a small measure of frequencies α where the discrepancy may go completely out of control due to the very small denominators $\Pi_i \bar{k}_i \| \langle k, \alpha \rangle \|$.

3.2. Let
$$D_1(\vartheta, N) = \sum_{|k_i| \leq N, \ k \neq (0,\dots,0)} U_k(\vartheta, N).$$

Lemma 3.3. *We have*

(3.9) $$\|D - D_1\|_2^2 = \mathcal{O}(1).$$

The L^2 norm in (3.9) and below in Section 3 is taken with respect to the variables $(\alpha, x) \in \mathbb{T}^{2d}$.

Proof. Assume $\vartheta \in X$ given. Then for any $q \geq N$ and any $q_1, \dots, q_{d-1} \in \mathbb{N}$ let $\hat{Y}(q, q_1, \dots q_{d-1})$ be the set of vectors $k \in \mathbb{Z}^d$ such that for some permutation of the indices $1, \dots, d$ 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 $\hat{Y}(q, q_1, \dots q_{d-1})$ is uniformly bounded.

Since for any $\omega \in \mathbb{T}$ and any $m \neq 0$,
$$\left| \frac{\sin(2\pi m \omega)}{m} \right| < \min\left(2\pi |\omega|, \frac{1}{|m|} \right) = \mathcal{O}\left(\frac{1}{|m|+1} \right),$$

the contributions of high frequencies can be bounded as follows.
$$\|D - D_1\|_2^2$$

$$\ll \sum_{q \geq N, q_1, \dots, q_{d-1} \geq 0} \frac{1}{q^2 (q_1+1)^2 \dots (q_{d-1}+1)^2} \sum_{k \in \hat{Y}(q, q_1, \dots q_{d-1})} \int_{\mathbb{T}^d} \left(\frac{\sin(\pi N \langle k, \alpha \rangle)}{\sin(\pi \langle k, \alpha \rangle)} \right)^2 d\alpha$$

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

3.3. Define $S(\vartheta, N) = S(U, N) := \{k \in \mathbb{Z}^d : |k_i| \leq N, |\bar{k}_i| \geq 1\}$. 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 U, 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 U is close to Identity then $|\bar{k}_i| \leq 2N$ for every i. Moreover, there exists $C(d)$ such that for every $(q_1, \dots, q_{d-1}) \in [0, 2N]^{d-1}$ 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 $\hat{Y}(q_1, \ldots, q_{d-1})$ 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 $k \in \hat{Y}(q_1, \ldots, q_{d-1})$ satisfying $\left| \prod_{i=1}^{d-1} (q_i + 1) \right| \|\langle k, \alpha \rangle\| \leq \bar{\epsilon}/(\ln N)^{d-1}$. The excluded set $E_N(U)$ has Lebesgue measure of order $\bar{\epsilon}$.

Lemma 3.4. $\|D_2 - D_1\|^2_{L_2((\mathbb{T}^d - E_N) \times \mathbb{T}^d)} \ll \dfrac{(\ln N)^{2(d-1)}}{\bar{\epsilon}}.$

Proof. Let

$$B_p\left((q_1, \ldots, q_{d-1}), U\right) = \{\alpha \in \mathbb{T}^d : \exists k \in \hat{Y}(q_1, \ldots, q_{d-1})(U),$$

$$p\bar{\epsilon}/(\ln N)^{d-1} \leq \left(\prod_{i=1}^{d-1} (q_i + 1)\right) \|\langle k, \alpha \rangle\| \leq (p+1)\bar{\epsilon}/(\ln N)^{d-1}\}.$$

Then $\mathrm{Leb}\left(B_p\left((q_1, \ldots, q_{d-1}), U\right)\right) \ll \frac{\bar{\epsilon}}{(q_1+1)\ldots(q_{d-1}+1)(\ln N)^{d-1}}$. Hence

$$\|D_2 - D_1\|^2_{L_2((\mathbb{T}^d - E_N) \times \mathbb{T}^d)} \ll$$

$$\sum_{q_1, \ldots, q_{d-1} \in [0, 2N]^{d-1}} \sum_{p \geq 1} \frac{\bar{\epsilon}}{(q_1 + 1) \ldots (q_{d-1} + 1)(\ln N)^{d-1}} \frac{(\ln N)^{2(d-1)}}{\bar{\epsilon}^2 p^2}$$

$$\ll \frac{(\ln N)^{2(d-1)}}{\bar{\epsilon}}. \qquad \square$$

3.4. For $k \in \mathbb{Z}^d$, denote $K(\vartheta, k) = \Pi_{i=1}^d \bar{k}_i$. Let

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

Lemma 3.5.

(3.10) $$\|D_3 - D_2\|^2_2 = \mathcal{O}\left((\ln N)^{d-1}\right).$$

Proof.

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

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

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

3.5.   Recall the definition (3.2) of $W(\vartheta, N) = W(U, \alpha, N)$

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

and let

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

Lemma 3.6. $\|D_4 - D_3\|_{L_2((\mathbb{T}^d - E_N) \times \mathbb{T}^d)} \ll \sqrt{\bar{\epsilon}} \, (\ln N)^d.$

Proof. Since $k \in \bar{S}$ and U 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

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

where

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

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

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

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

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

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

3.6.   Since the terms in D_4 satisfy $\big\| \langle k, \alpha \rangle \big\| \leq \frac{1}{\bar{\epsilon}(\ln N)^d}$, we can replace U_k defined in §3.1 by

$$V_k(\vartheta, N) = \mathfrak{a} \prod_i \left(\frac{\sin\left(2\pi \bar{k}_i u_i\right)}{\bar{k}_i} \right) \frac{\sin(\pi N \langle k, \alpha \rangle)}{\pi \big\| \langle k, \alpha \rangle \big\|} \cos(2\pi \langle k, x \rangle + \varphi_{k, N, \alpha}))$$

and introduce

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

Note that for small θ we have $|\sin \theta - \theta| \leq |\theta|^3$, so

$$\left| \frac{1}{\sin \theta} - \frac{1}{\theta} \right| \leq \left| \frac{\theta^3}{\theta \sin \theta} \right| \leq 2\theta.$$

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

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

Summing over $k \in W(\vartheta, N)$ we have thus obtained

Lemma 3.7.

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

3.7. *Proof of Proposition 3.1.* Putting together Lemmas 3.3–3.7, we see that $(D_5(\vartheta, N) - D(\vartheta, N))/(\ln N)^d$ satisfies (3.8) if $\bar{\epsilon} > 0$ is sufficiently small and then N is sufficiently large.

Recall the definition of $\bar{D}(\vartheta, N)$ given in (3.4). The difference between $(\ln N)^d \bar{D}(\vartheta, N)$ and D_5 is that for $k \in Z(\vartheta, N)$, we comprise in $(\ln N)^d \bar{D}(\vartheta, N)$ all its multiples whereas in D_5 we take only multiples such that $pk \in W$. Since we have already shown in Lemmas 3.3–3.7 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

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

provided that $\bar{\epsilon}$ is sufficiently small and N is sufficiently large. This finishes the proof of Proposition 3.1. ☐

4. POISSON DISTRIBUTION OF SMALL DIVISORS

In this section we reduce the Cauchy limit of the discrepancies to a Poisson limit theorem (Theorem B) for the small divisors $\left(\prod_i \bar{k}_i \right) \left\| \langle k, \alpha \rangle \right\|$.

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

Theorem B. *Assume that the distribution of $\vartheta \in X$ is absolutely continuous with respect to the Lebesgue measure. For any $\bar{\epsilon} > 0$, as $N \to \infty$, the process*

$$\left\{ (\ln N)^d \left(\prod_i \bar{k}_i \right) \left\| \langle k, \alpha \rangle \right\|, N \langle k, \alpha \rangle \bmod 2, \{\bar{k}_1 u_1\}, \ldots, \{\bar{k}_d u_d\}, \{\langle k, x \rangle\} \right\}_{k \in Z(\vartheta, N)}$$

where $Z(\vartheta, N)$ is defined by (3.3), converges to a Poisson process on

$$\left[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}} \right] \times (\mathbb{R}/(2\mathbb{Z})) \times \mathbb{T}^{d+1}$$

with intensity

(4.1) $$c = 2^{d-1} \mathbf{c}_1 / d!, \quad \mathbf{c}_1 = 1/\zeta(d+1).$$

Remark 4.1. It is sufficient to prove Theorem B when ϑ is distributed according to a smooth bounded density with respect to the normalized Lebesgue measure λ on X.

Indeed, let us suppose that Theorem B is known for smooth measures and assume now that ϑ is distributed according to an integrable density $\mathfrak{p}(\vartheta)$. Let $\mathcal{K}_1, \mathcal{K}_2, \ldots \mathcal{K}_r$ be any partition of the target space

$$\left[-\frac{1}{\epsilon}, \frac{1}{\epsilon} \right] \times \mathbb{R}/(2\mathbb{Z}) \times \mathbb{T}^{d+1}.$$

Let $\mathcal{N}_1(\vartheta, N), \ldots \mathcal{N}_r(\vartheta, N)$ be the number of points of our process inside $\mathcal{K}_1, \ldots \mathcal{K}_r$ respectively. We need to know that as $N \to \infty$, $\mathcal{N}_j(\mathcal{L}, N)$ are asymptotically independent Poisson random variables with means $\mathbf{c}\hat{\mathcal{K}}_j$ where $\hat{\mathcal{K}}_j$ is the volume of $\mathcal{K}_j$. Equivalently we need to show that for each r tuple $s_1, \ldots s_r$

$$\lim_{N \to \infty} \int \exp\left(i \sum_{j=1}^{r} s_j \mathcal{N}_j(\vartheta, N) \right) \mathfrak{p}(\vartheta) d\lambda(\vartheta) = \psi(s_1, \ldots s_r)$$

where ψ is characteristic function of multivariate Poisson, namely

$$\psi(s_1, \ldots s_r) = \exp\left(\mathbf{c} \sum_{j=1}^{r} \hat{\mathcal{K}}_j \left[e^{is_j} - 1 \right] \right)$$

(the precise form of ψ is not important for the argument below). Fix $\upsilon > 0$. Take a smooth density $\bar{\mathfrak{p}}$ on X such that

$$(4.2) \qquad ||\mathfrak{p} - \bar{\mathfrak{p}}||_{L^1(\lambda)} \leq \frac{\upsilon}{2}.$$

Since we assume that Theorem B holds for smooth densities, for large N we have

$$(4.3) \qquad \left| \int \exp\left(i \sum_{j=1}^{r} s_j \mathcal{N}_j(\vartheta, N) \right) \bar{\mathfrak{p}}(\vartheta) d\lambda(\vartheta) - \psi(s_1, \ldots s_r) \right| < \frac{\upsilon}{2}.$$

Combining (4.2) with (4.3) we obtain

$$\left| \int \exp\left(i \sum_{j=1}^{r} s_j \mathcal{N}_j(\vartheta, N) \right) \mathfrak{p}(\vartheta) d\mu(\vartheta) - \psi(s_1, \ldots s_r) \right| < \upsilon.$$

Since υ as well as the partition $\mathcal{K}_1, \ldots \mathcal{K}_r$ are arbitrary we obtain that Theorem 5.4 holds for absolutely continuous initial distributions.

Therefore we assume henceforth that the initial distribution of ϑ has smooth density on X.

Remark 4.2. Observe that it does not change anything in the result nor in the proof to take in the last coordinate of the process $\{\langle k, x \rangle + \varphi_{k,\alpha,N}\}$ instead of $\{\langle k, x \rangle\}$ since the phase $\varphi_{k,\alpha,N} = \pi(N-1)\langle k, \alpha \rangle/2$ is independent

of the variable x. It is with the phase $\varphi_{k,\alpha,N}$ that Theorem B is used to prove Theorem A.

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 B.

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 4.3. *Let (U, α) have absolutely continuous distribution on $SL_d(\mathbb{R}) \times \mathbb{T}^d$. Let*

$$(4.4) \quad \hat{Z}(\vartheta, N) := \left\{ k \in \mathbb{Z}^d : \left| \prod_{i=1}^{d} \bar{k}_i \right| < N, \quad \forall i = 1, \ldots, d, \quad |\bar{k}_i| \geq 1, \right.$$

$$\left. \bar{k}_1 > 0 \text{ and } \exists m \in \mathbb{Z} \text{ s. t. } k_1 \wedge \ldots \wedge k_d \wedge m = 1 \text{ and } \|\langle k, \alpha \rangle\| = |\langle k, \alpha \rangle + m| \right\}.$$

Then as $N \to \infty$ the point process

$$\left\{ (\ln N)^d \left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\| \right\}_{k \in \hat{Z}(\vartheta, N)}$$

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

Note that

$$Z(\vartheta, N) = \left\{ k \in \hat{Z}(\vartheta, N) : \left| (\ln N)^d \left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\| \right| \leq \frac{1}{\bar{\epsilon}} \right\}$$

(compare (4.4) with (3.3), (3.2)).

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

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

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 B and the invariance of Poisson processes under projection (Lemma 2.1(b)). $\square$

4.2. **Proof that Theorem B implies Theorem A.** Fix $z \in \mathbb{R}$ and $\eta > 0$. We want to show that for N sufficiently large we have

$$(4.5) \qquad \left| \lambda \{ \vartheta \in X : \frac{D(\vartheta, N)}{(\ln N)^d} \leq z \} - \mathfrak{C}(\boldsymbol{\rho} z) \right| < \eta.$$

We first use the approximation of $\frac{D(\vartheta, N)}{(\ln N)^d}$ by $\bar{D}(\vartheta, N)$ given by Proposition 3.1. Observe that unlike $D(\vartheta, N)$, the definition of $\bar{D}(\vartheta, N)$ depends implicitly on some $\bar{\epsilon} > 0$. Having fixed $z \in \mathbb{R}$, we know from Proposition 3.1, that if we fix $\bar{\epsilon} > 0$ sufficiently small, and then consider N sufficiently large the following holds

$$\lambda \left\{ \vartheta \in X : \bar{D}(\vartheta, N) \leq z - \frac{\eta}{4} \right\} - \frac{\eta}{2}$$

$$< \lambda \left\{ \vartheta \in X : \frac{D(\vartheta, N)}{(\ln N)^d} \leq z \right\} <$$

$$\lambda \left\{ \vartheta \in X : \bar{D}(\vartheta, N) \leq z + \frac{\eta}{4} \right\} + \frac{\eta}{2}.$$

Hence it suffices to prove that for $\bar{\epsilon}$ sufficiently small and N sufficiently large we have

$$(4.6) \qquad \left| \lambda \{ \vartheta \in X : \bar{D}(\vartheta, N) \leq z \} - \mathfrak{C}(\boldsymbol{\rho} z) \right| < \frac{\eta}{2}.$$

We now want to use the representation of the Cauchy distribution in terms of a Poisson process recalled in §2.1, and the Poisson limit of Theorem B. Recall that $\bar{D}(\vartheta, N) = \sum_{k \in Z} \Gamma_k(\vartheta, N)/\Omega_k(\vartheta, N)$ where $\Omega_k(\vartheta, N)$ and $\Gamma_k(\vartheta, N)$ are given by (3.5) and (3.6) respectively.

First of all, Theorem B asserts that the point process $\{\Omega_k(\vartheta, N)\}_{k \in Z(\vartheta, N)}$, that converges to a Poisson process on $[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}}]$ with constant intensity $c = 2^{d-1}\mathbf{c}_1/d!$. Secondly, Theorem B, with Remark 4.2, and Lemma 2.1 (c) tell us that $\{\Gamma_k(\vartheta, N)\}_{k \in Z(\vartheta, N)}$ behave asymptotically, as $N \to \infty$, like iid symmetric variables with compact support, that are independent of $\{\Omega_k(\vartheta, N)\}_{k \in Z(\vartheta, N)}$.

Hence, the limiting distribution, as $N \to \infty$, of $\sum_{k \in Z} \Gamma_k(\vartheta, N)/\Omega_k(\vartheta, N)$ is approached by that of $\sum \Theta_k / \bar{\Gamma}_k$ where $\{\Theta_k\}$ is a Poisson process on $[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}}]$ with constant intensity c and $\{\bar{\Gamma}_k\}$ consisting of symmetric variables with compact support, that are independent of $\{\Theta_k\}$.

Hence, if $\bar{\epsilon}$ was chosen sufficiently small, then N sufficiently large, it follows from Lemma 2.2 (b) that (4.6) holds with $\boldsymbol{\rho} = \pi c \mathbb{E}(|\bar{\Gamma}_k|)$, where the expectation is taken with respect to the distribution of the iid variables $\{\bar{\Gamma}_k\}$.

We finish by explicitly computing this limit expectation. Using the definition of Γ_k in (3.6) we get

$$(4.7) \qquad \mathbb{E}(|\bar{\Gamma}_k|) = \frac{2}{\pi^{d+1}} \int \cdots \int |\phi(\eta_1, \ldots \eta_{d+1})| \, d\eta_1 \ldots d\eta_{d+2}$$

where ϕ is given by (3.7).

The formula (1.6) for $\boldsymbol{\rho}$ follows from (2.1), (4.1), and (4.7). Theorem A is thus proved. $\square$

5.   Reduction to dynamics on the space of lattices

5.1. **Notation.** The goal of this section is to reduce the proof of Theorem B to Theorem 5.4 which is a Poisson limit theorem for the diagonal action on the space of lattices. For this we rely on Dani's correspondance principle relating Diophantine approximation problems to visits to a cusp by orbits of the Cartan action on the space of $d+1$ dimensional unimodular lattices $\mathcal{M}_{d+1} = SL_{d+1}(\mathbb{R})/SL_{d+1}(\mathbb{Z})$. For more details on the specific approach adopted here we refer the reader to the surveys [31] and [13] and references therein.

For simplicity we drop the subindex and refer to $\mathcal{M}_{d+1}$ as $\mathcal{M}$. We recall that μ denotes the Haar measure on $\mathcal{M}$.

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

Define

$$(5.1) \qquad \Pi_M = \{ \mathbf{t} \in \mathbb{N}^d : \sum_{j=1}^d \mathbf{t}_j \le M - d \}.$$

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 $\vartheta = (\alpha, x, U) \in X$, we define

$$(5.2) \qquad \Lambda(\vartheta) = \begin{pmatrix} U & 0 \\ \alpha & 1 \end{pmatrix}.$$

We recall that in all the sequel, it is assumed that the distribution of $\vartheta \in X$ has smooth density with respect to the Lebesgue measure.

Consider the Cartan subgroup

$$(5.3) \qquad g^{\mathbf{t}} = \mathrm{diag}\left(e^{-\mathbf{t}_1}, \ldots e^{-\mathbf{t}_d}, e^{\sum_{j=1}^d \mathbf{t}_j} \right), \quad \mathbf{t} \in \mathbb{R}^d.$$

5.2. **Reduction to a Poisson Limit Theorem for the Cartan action.** Let $\mathcal{L}$ be a unimodular lattice in $\mathbb{R}^{d+1}$. We decompose elements of $\mathcal{L}$ as

$$\mathbf{v} = (x(\mathbf{v}), z(\mathbf{v})) \text{ where } x \in \mathbb{R}^d, \ z \in \mathbb{R}.$$

Let $\Pi(\mathbf{v}) = \left(\prod_{j=1}^d x_j \right) z$. Define

(5.4) $\mathcal{D} = \{ (x, z) : x_1 \in I, \ x_j \in J \text{ for } j = 2, \ldots, d \text{ and } M^d \Pi((x,z)) \in K \}.$

Let $\Phi : \mathcal{M} \to \mathbb{R}$ be the Siegel transform of $\mathbb{1}_{\mathcal{D}}$, as defined in §2.2, that is

$$(5.5) \qquad \Phi(\mathcal{L}) = \sum_{\mathbf{v} \in \mathcal{L} \text{ prime}} \mathbb{1}_{\mathcal{D}}(\mathbf{v})$$

The bulk of the derivation of the distribution of the small divisors of Theorem B from a Poisson limit theorem for the diagonal action on the space of lattices is encapsulated in the following simple observation. Recall the definition of $Z(\vartheta, N)$ from (3.3).

CLAIM. *For ϑ as in (5.7), the following are equivalent for $\mathbf{t} \in \Pi_M$:*
  (i) $\Phi(g^{\mathbf{t}}\Lambda(\vartheta)) = 1$
  (ii) *There exists a unique $(k, m) \in Z(\vartheta, N)$ such that $e^{\mathbf{t}_j} < |\bar{k}_j| \leq e^{\mathbf{t}_j + 1}$ for $j = 1, \ldots, d$.*

Proof. From the definition of Φ we have that

$$\Phi(g^{\mathbf{t}}\Lambda(\vartheta)) = \#\{(k, m) \in \mathbb{Z}^d \times \mathbb{Z}, k_1 \wedge \ldots \wedge k_d \wedge m = 1 : g^{\mathbf{t}}\Lambda(\vartheta)(k, m) \in \mathcal{D}\}.$$

Recall the notation (3.1): $\bar{k}_i = a_{i,1}k_1 + \cdots + a_{i,d}k_d$. Then observe that the definition (5.2) of $\Lambda(\vartheta)$ implies that

$$g^{\mathbf{t}}\Lambda(\vartheta)(k, m) = \left(e^{-\mathbf{t}_1}\bar{k}_1, \ldots, e^{-\mathbf{t}_d}\bar{k}_d, e^{\sum_{j=1}^d \mathbf{t}_j}(\langle k, \alpha \rangle + m) \right).$$

Hence, from the definitions of $\mathcal{D}$ in (5.4), of Π_M in (5.1) and the definition of $Z(\vartheta, N)$ in (3.3), we conclude that

$$\Phi(g^{\mathbf{t}}\Lambda(\vartheta)) = \#\{(k, m) \in Z(\vartheta, N) : e^{\mathbf{t}_j} < |\bar{k}_j| \leq e^{\mathbf{t}_j + 1}, j = 1, \ldots, d\}.$$

We have thus proved the equivalence between (i) and (ii) of the claim. $\quad\square$

Define also an $\mathbb{R} \times (\mathbb{R}/2\mathbb{Z})$ valued function on $\mathcal{M} \times \mathbb{R}$
(5.6)
$$\Psi(\mathcal{L}, b) = (\Psi_1(\mathcal{L}), \Psi_2(\mathcal{L}, b)) = \sum_{\mathbf{v} \in \mathcal{L} \text{ prime}} \mathbb{1}_{\mathcal{D}}(\mathbf{v})(M^d \Pi(\mathbf{v}), bz(\mathbf{v}) \bmod 2).$$

Given $\bar{\epsilon}$ and N, suppose that $\vartheta \in X$ is such
(5.7)
$$\forall \mathbf{t} \in \Pi_M, \quad \Phi(g^{\mathbf{t}}\Lambda(\vartheta)) \leq 1$$

Note that if (i) or (ii) of the Claim holds then

$$\left((\ln N)^d \left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\|, N\langle k, \alpha \rangle \bmod 2 \right) = \Psi\left(g^{\mathbf{t}}\Lambda(\vartheta), Ne^{-\sum_{j=1}^d \mathbf{t}_j} \right).$$

Thus, for ϑ satisfying (5.7), we have that the sequence

$$\left\{ \left((\ln N)^d \left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\|, N\langle k, \alpha \rangle \bmod 2 \right) \right\}_{k \in Z(\vartheta, N)}$$

is exactly

$$\left\{ \Psi\left(g^{\mathbf{t}}\Lambda(\vartheta), Ne^{-\sum_{j=1}^d \mathbf{t}_j} \right) \right\}_{\mathbf{t} \in \Pi_M, \Phi(g^{\mathbf{t}}\Lambda(\vartheta)) = 1}.$$

Hence, to show that the distribution of

$$\left\{ \left((\ln N)^d \left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\|, N\langle k, \alpha \rangle \bmod 2 \right) \right\}_{k \in Z(\vartheta, N)}$$

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 (5.7) is small;

(b) that the process $\left\{ \Psi \left(g^{\mathbf{t}} \Lambda(\vartheta), N e^{-\sum_{j=1}^{d} \mathbf{t}_j} \right) \right\}_{\mathbf{t} \in \Pi_M, \Phi(g^{\mathbf{t}} \Lambda(\vartheta)) = 1}$ converges in probability to a Poisson distribution.

This is the content of the following Theorem 5.1.

Theorem 5.1. *Assume that the distribution of $\vartheta \in X$ has smooth density with respect to the Lebesgue measure. Let Λ be the matrix $\Lambda(\vartheta)$ as defined in (5.2). Then, for any $\bar{\epsilon} > 0$, we have*

(a) *For any $\mathbf{t} \in \mathcal{P}$, $\lambda(\Phi(g^{\mathbf{t}} \Lambda) > 1) = \mathcal{O}(M^{-2d})$.*

(b) $\left\{ \left(\Psi \left(g^{\mathbf{t}} \Lambda(\vartheta), N e^{-\sum_{j=1}^{d} \mathbf{t}_j} \right), \frac{\mathbf{t}}{M} \right) \right\}_{\mathbf{t} \in \Pi_M, \Phi(g^{\mathbf{t}} \Lambda(\vartheta)) = 1}$ *converges as $N \to \infty$ to the Poisson process on $[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}}] \times \mathbb{R}/(2\mathbb{Z}) \times \mathcal{P}$ with intensity $\mathbf{c}_1$.*

The notation $X = \mathcal{O}(M^{-2d})$ means that $|X| \leq C M^{-2d}$ 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 B we will also need an additional effort to prove the independence and uniform distribution of the rest of the variables namely of

$$\left\{ \{\bar{k}_1 u_1\}, \ldots, \{\bar{k}_d u_d\}, \{\langle k, x \rangle\} \right\}_{k \in Z(\vartheta, N)}.$$

This issue is addressed below.

Definition 5.2. Let $L > 0$. Consider a sequence $\{t^{(1)}, \ldots, t^{(s)}\}$ of points in Π_M where $t^{(j)} = (t_1^{(j)}, \ldots t_d^{(j)})$. We say that this sequence is *L-split* if for any pair i, j we have

$$|t_p^{(i)} - t_p^{(j)}| \geq L \text{ for each } p \text{ and } |\max_p(t_p^{(i)}) - \max_p(t_p^{(j)})| \geq L,$$

and for any i we have $\min_p(t_p^{(i)}) > L$.

Proposition 5.3. *Let $R \in \mathbb{R}$ and $s \in \mathbb{N}$ be fixed. Let $k^{(1)}(N) \ldots, k^{(s)}(N)$, $k^{(j)}(N) \in \mathbb{R}^d$, be such that*

$$t^{(j)} = ([\ln |\bar{k}^{(j)_1}|], \ldots [\ln |\bar{k}_d^{(j)}|])$$

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

Suppose that $(u_1, \ldots u_d, x_1 \ldots x_d)$ are distributed according to a density ρ_N such that

(5.8) $$\|\rho_N\|_{C^1} \leq R.$$

Then the distribution of the numbers

$$\{\bar{k}_1^{(1)} u_1\}, \ldots, \{\bar{k}_d^{(1)} u_d\}, \{\langle k^{(1)}, x \rangle\} \ldots,$$

$$\{\bar{k}_1^{(s)} u_1\}, \ldots, \{\bar{k}_d^{(s)} u_d\}, \{\langle k^{(s)}, x \rangle\}$$

converges to the uniform distribution on $\mathbb{T}^{(d+1)s}$ and the convergence is uniform with respect to N, (U, α), the choices of s vectors satisfying the splitness condition, and ρ_N satisfying (5.8).

5.3. **Proof of Proposition 5.3.** By Weyl equidistribution criterion we need to show that if $f_j : \mathbb{T}^{d+1} \to \mathbb{C}$ are exponentials

$$f_j(\theta_1, \ldots, \theta_{d+1}) = \exp\left(2\pi i \sum_{p=1}^{d+1} m_{jp}\theta_p\right)$$

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

$$(5.9) \qquad \int_{\mathbb{T}^{2d}} \prod_{j=1}^{s} f_j\left(\bar{k}_1^{(j)} u_1, \ldots \bar{k}_d^{(j)} u_d, \langle k^{(j)}, x \rangle\right) \rho_N(u, x)dudx \to 0.$$

uniformly in the parameters involved. Suppose there exists some $p \leq d$ such that not all m_{jp} are equal to zero. Then the coefficient in front of u_p in the above product is large since, due to splitness, it is dominated by the contribution of the largest of $\bar{k}_p^{(j)}$ for which m_{jp} is non zero. In this case we show that the integral (5.9) is small by integrating by parts with respect to u_p. Next suppose that all m_{jp} with $j \leq s$ and $p \leq d$ are zero. Let $\bar{j}$ be such that $\max_p(t_p^{(\bar{j})}) =: t_{\bar{p}}^{(\bar{j})}$ is the largest among those indices for which $m_{j(d+1)} \neq 0$. Note that $k_{\bar{p}}^{(\bar{j})}$ is of order $\exp(t_{\bar{p}}^{(\bar{j})})$. Then, due to the splitness condition, $\bar{k}_{\bar{p}}^{(\bar{j})}$ dominates the coefficient in front of $x_{\bar{p}}$ and so we conclude that (5.9) is small by integrating by parts with respect to $x_{\bar{p}}$.   □

5.4. **Proof that Theorem 5.1 implies Theorem B.**

As demonstrated earlier, parts (a) and (b) of Theorem 5.1 imply that

$$\left\{\left((\ln N)^d \left(\prod_i \bar{k}_i\right) \|\langle k, \alpha \rangle\|, N\langle k, \alpha \rangle \mod 2\right)\right\}_{k \in Z(\vartheta, N)}$$

converges 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 $\mathbf{c}_1$. Next, it follows from part (b) of Theorem 5.1 that if $t^{(1)}, t^{(2)}, \ldots \in \Pi$ are the points such that $\Phi(g_t\Lambda) = 1$, listed in any order; then for any $s \in \mathbb{N}$, we have that

$$(5.10) \qquad \lambda(\{t^{(1)}, t^{(2)}, \ldots\} \text{ is } \sqrt{M} - \text{split}) \to 1 \text{ as } N \to \infty.$$

Indeed, given $\bar{\epsilon}$, $\tilde{\epsilon}$ we can choose δ such that the probability that the Poisson process on $[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}}] \times \mathbb{R}/(2\mathbb{Z}) \times \mathcal{P}$ with intensity $\mathbf{c}_1$ has two points within distance δ from each other in projection on the last coordinate is less than $\tilde{\epsilon}$. Since $M^{-1/2} < \delta$ for large M (5.10) follows. Therefore outside a set of small measure of $\vartheta \in X$, the set $Z(\vartheta, N)$ satisfies the hypothesis of Proposition 5.3.

Thus $\left\{ \left(\{\bar{k}_1 u_1\}, \ldots \{\bar{k}_d u_d\}, \{\langle k, x \rangle\} \right) \right\}_{k \in Z(\vartheta, N)}$ converge to uniformly distributed iid's on $\mathbb{T}^{d+1}$ independent of

$$\left\{ \left((\ln N)^d \left(\prod_i \bar{k}_i \right) \|\langle k, \alpha \rangle\|, N\langle k, \alpha \rangle \bmod 2 \right) \right\}_{k \in Z(\vartheta, N)}$$

Lemma 2.1 hence yields the full Poisson limit of Theorem B. $\qquad\square$

5.5. Modifying the initial distribution.
Before we close this section we make a last observation that allows us to complete the reduction of our problem to a clear cut ergodic theory problem on the space of lattices, namely the following.

Theorem 5.4. *Assume that $\mathcal{L} \in \mathcal{M}$ is distributed according to a probability measure $\tilde{\mu}$ that has a smooth bounded density with respect to the Haar measure on $\mathcal{M}$. Then*

(a) For any $\mathbf{t} \in \mathcal{P}$, $\tilde{\mu}(\Phi(g^{\mathbf{t}}\mathcal{L}) > 1) = \mathcal{O}(M^{-2d})$.

(b) $\left\{ \left(\Psi\left(g^{\mathbf{t}}\mathcal{L}, Ne^{-\sum_{j=1}^{d} \mathbf{t}_j} \right), \frac{\mathbf{t}}{M} \right) \right\}_{\Phi(g^{\mathbf{t}}\mathcal{L})=1, \, \mathbf{t} \in \Pi_M}$ converges in probability,

as $N \to \infty$, to a Poisson process on $[-\frac{1}{\epsilon}, \frac{1}{\epsilon}] \times \mathbb{R}/(2\mathbb{Z}) \times \mathcal{P}$ with intensity $\mathbf{c} = 2^{d-1}\mathbf{c}_1/d!$.

Remark 5.5. As in Remark 4.1, it is easy to see that the statement with a smooth bounded density of Theorem 5.4, implies the same results with a merely bounded density. Moreover, part 9b) of Theorem 5.4 holds assuming that the initial distribution of $\mathcal{L}$ is absolutely continuous with respect to the Haar measure.

Proof that Theorem 5.4 implies Theorem 5.1.
Let $\eta > 0$ and define for an interval $A = [a, b]$ the intervals

$$A^+ = [a(1 - \eta), b(1 + \eta)] \quad \text{and} \quad A^- = [a(1 + \eta), b(1 - \eta)].$$

Fix an interval $\bar{K} \subset K$. Let $\bar{\Phi}^{\pm}$ be defined as in (5.5) 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

$$\tilde{\Lambda} = \begin{pmatrix} (1+\sigma_1) & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & (1+\sigma_d) & 0 \\ 0 & \cdots & 0 & \left(\prod_{j=1}^{d}(1+\sigma_j)\right)^{-1} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & c_1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & c_d \\ 0 & \cdots & 0 & 1 \end{pmatrix} \Lambda$$

where $\sigma_1, \ldots, \sigma_d, c_1, \ldots, c_2$ are distributed according to any smooth density on $[-\eta^2, \eta^2]^{2d}$. This guarantees that when ϑ is distributed according to a smooth density on X, the lattice $\tilde{\Lambda}$ has a smooth bounded distribution with respect to the Haar measure. Thus, the implication of Theorem 5.1 from Theorem 5.4 stems from the straightforward observation that if M is

sufficiently large, then for any $n \in \mathbb{N}$ it holds that

$$\bar{\Phi}^-(g^{(\mathbf{t}_1+\ln(1+\sigma_1),\ldots,\mathbf{t}_d+\ln(1+\sigma_d))}\tilde{\Lambda}) \geq n \implies \Phi(g^{\mathbf{t}}\Lambda) \geq n$$
$$\implies \bar{\Phi}^+(g^{(\mathbf{t}_1+\ln(1+\sigma_1),\ldots,\mathbf{t}_d\ln(1+\sigma_d))}\tilde{\Lambda}) \geq n. \quad \square$$

The proof of Theorem 5.4 occupies Sections 6–8.

6. Poisson limit theorem for almost independent rare events

To prove Theorem 5.4 we will start with an abstract result that establishes a Poisson limit theorem for M^d variables that behave similarly to iid variables with expectation of order $1/M^d$. The variables to which the abstract Poisson limit theorem must be applied to imply Theorem 5.4 will be defined precisely in Section 8 (see (8.2)–(8.5)). Essentially, we will be considering a counting variable $\xi_{\mathbf{t}}$ that takes integer values and that corresponds to $\Phi(g^{\mathbf{t}}\mathcal{L})$ and two related variables $\nu_{\mathbf{t}}$ and $\zeta_{\mathbf{t}}$ that give the value of Ψ_1 and Ψ_2 in $\Psi\left(g^{\mathbf{t}}\mathcal{L}, Ne^{-\sum_j \mathbf{t}_j}\right)$ when $\xi_{\mathbf{t}} = 1$.

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 $\mathbb{R}^d$ with piecewise smooth boundary. For $M \in \mathbb{N}$ we let

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

• We are given an inhomogeneous non-constant linear form on $\mathbb{R}^d$,

$$\lambda_1(\mathbf{t}) = \sigma_0 + \sum_{j=1}^{d} \sigma_j \mathbf{t}_j.$$

• We let $(\mathfrak{X}, \mathfrak{m})$ and $(\tilde{\mathfrak{X}}, \tilde{\mathfrak{m}})$ be two probability 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 $\{\xi_{\mathbf{t}}^M\}_{\mathbf{t}\in\Pi_M}$ of random variables taking values in non-negative integers and a sequence $\{\nu_{\mathbf{t}}^M\}_{\mathbf{t}\in\Pi_M}$ of $\mathfrak{X}$ valued random variables on Ω.

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

We define

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

(Note that, in fact, $\eta_{\mathbf{t}}^M = \mathbb{1}_{\xi_{\mathbf{t}}^M=1}$, and $\eta_{\mathbf{t},p}^M = \mathbb{1}_{\xi_{\mathbf{t},p}^M=\xi_{\mathbf{t}}^M=1}$ but we use a more complicated definition above because condition (h2) below will ensure that with probabiilty close to 1 we have $\eta_{\mathbf{t}}^M = \xi_{\mathbf{t}}^M$, and $\eta_{\mathbf{t},p}^M = \xi_{\mathbf{t},p}^M$.)

● Since all the variables depend on M, we will omit sometimes the super-script or subscript M and denote Π_M simply by Π, $\xi_{\mathbf{t}}^M$ by $\xi_{\mathbf{t}}$ etc.

● In all this section, when we use the notation $Y = \mathcal{O}(X)$ or equivalently $Y \ll X$, it corresponds to $|Y| \leq C|X|$, where the implicit constant C is allowed to depend on $Q, \tilde{Q}$ but not on M, $\mathbf{t}$, and $\bar{\delta}$ that will be introduced later in the section.

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

$$\mathbb{P}(E^c) = \mathcal{O}(M^{-100d})$$

and

(h1) For any $\mathbf{t} \in \Pi$,
$$\mathbb{E}(\xi_{\mathbf{t}}) = \mathcal{O}(M^{-d});$$

(h2) For any $\mathbf{t} \in \Pi$,
$$\mathbb{P}(\xi_{\mathbf{t}} > 1) = \mathcal{O}(M^{-2d});$$

(h3) For $\mathbf{t}, \mathbf{t}' \in \Pi$, $\mathbf{t} \neq \mathbf{t}'$,
$$\mathbb{P}(\xi_{\mathbf{t}} \geq 1, \xi_{\mathbf{t}'} \geq 1) = \mathcal{O}(M^{-2d});$$

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

$$(\text{h4a}) \quad \mathbb{E}(\xi_{\mathbf{t}}|\mathcal{F}_{\mathbf{t}'})(\omega) = \frac{\mathbf{cm}(\mathfrak{X})}{M^d} + \mathcal{O}\left(M^{-2d}\right),$$

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

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

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

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

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

$$\left\{\nu_{\mathbf{t}}^M, \frac{\mathbf{t}}{M}\right\}_{\xi_{\mathbf{t}}^M = 1, \mathbf{t} \in \Pi_M}$$

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 $\hat{\lambda}(\mathbf{t}) = \hat{\sigma}_0 + \sum_{j=1}^{d} \hat{\sigma}_j \mathbf{t}_j$ such that $\hat{\lambda}(\mathbf{t}) > \lambda_1(\mathbf{t})$ on $\text{Int}(\mathcal{P})$. Assume that there is another domain $\widetilde{\mathcal{P}} \supset \mathcal{P}$ such that denoting $\tilde{\Pi}_M = \{\mathbf{t} \in \mathbb{N}^d : \mathbf{t}/M \in \widetilde{\mathcal{P}}\}$ we have that that (h1)–(h6) above are satisfied for $t \in \tilde{\Pi}_M$ and that $\max_{\widetilde{\mathcal{P}}} \lambda_1 > \max_{\mathcal{P}} \hat{\lambda}$.

Suppose that for each M we have a sequence of $\zeta_{\mathbf{t}}^M$ of $\tilde{\mathfrak{X}}$-valued random variables and assume that for any fixed element $\tilde{Q} = (\tilde{K}_1, \ldots, \tilde{K}_J) \in \tilde{\mathcal{Q}}$ the following conditions are satisfied.

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

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

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

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

Then we have the following strengthening of Theorem 6.1.

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

$$\left\{ \nu_{\mathbf{t}}^M, \zeta_{\mathbf{t}}^M, \frac{\mathbf{t}}{M} \right\}_{\xi_{\mathbf{t}}^M = 1, \mathbf{t} \in \Pi_M}$$

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.**
Divide $\mathcal{P}$ into arbitrarily small subsets $\mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_S$ of positive volume. Fix from the sequence $\mathcal{Q}$ a partition $Q = (K_1, \ldots, K_P)$ that is arbitrarily close to the point partition.

Given any double sequence $l_{p,s} \in \mathbb{N}, (p, s) \in [1, \ldots, P] \times [1, \ldots, S]$, we define the event

(6.1) $\mathcal{A} = \{\forall (p, s) \in [1, \ldots, P] \times [1, \ldots, S],$

there are exactly $l_{p,s}$ points $\mathbf{t} \in \Pi$ satisfying

$$\frac{\mathbf{t}}{M} \in \mathcal{P}_s \text{ and } \xi_{\mathbf{t}} = \xi_{\mathbf{t},p} = 1 \Bigg\}.$$

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

To prove Theorem 6.1 it suffices to see that

(6.2) $$\lim_{M \to \infty} \mathbb{P}(\mathcal{A}) = \prod_{p,s} \left[\frac{(\mathbf{cm}(K_p)(\hat{\mathcal{P}}_s))^{l_{p,s}}}{l_{p,s}!} \exp\left(-\mathbf{cm}(K_p)(\hat{\mathcal{P}}_s) \right) \right].$$

This section is devoted to the proof of (6.2). Fix an arbitrarily — small number $\bar{\delta} > 0$.

In the sequel we say that a function h which depends on $M, \bar{\delta}$, and maybe some other variables such as $\mathbf{t} \in \Pi_M$, is $o_{\bar{\delta}}(1)$ if the following holds : given $\nu > 0$, we can find $\hat{\delta}(\nu)$ such that if $\bar{\delta} < \hat{\delta}(\nu)$, then there exists $\bar{M}(\bar{\delta})$ such that for $M \geq \bar{M}(\bar{\delta})$, we have $|h| \leq \nu$ (uniformly in all the additional parameters). $O_{\bar{\delta}}(1)$ has a similar meaning.

Partition Π into cubes $C_1, \ldots, C_H$ of side size $\bar{\delta}M$ with one of the faces parallel to $\mathrm{Ker}(\lambda_1)$.

Definition 6.3. We say that a k-tuple $\{(S_l, i_l)\}$, $S_1, S_2, \ldots S_k \subset \{C_1, \ldots, C_H\}$, $i_1, \ldots, i_k \in \{1 \ldots P\}$, realizes the event $\mathcal{A}$ if the event $\mathcal{A}$ is realized, and if the $l_{p,s}$ points $\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(k)}$ of $\mathcal{A}$ are such that $\mathbf{t}^{(l)} \in S_l$ and $\xi_{\mathbf{t}^{(l)}} = \xi_{\mathbf{t}^{(l)}, i_l} = 1$.

Definition 6.4. We call a collection of cubes $\bar{\delta}$-generic if the images of any two of them under λ_1 are distant by more than $3\bar{\delta}M$. We say that a k-tuple $\{(S_l, i_l)\}$ is $\bar{\delta}$-generic if the cubes S_l are $\bar{\delta}$-generic.

To obtain (6.2), we shall need the following.

Proposition 6.5.
$$\mathbb{P}(\mathcal{A}) = \mathbb{P}(\mathcal{A} \text{ is realized by a } \bar{\delta}\text{-generic collection of cubes}) + o_{\bar{\delta}}(1).$$

Proposition 6.6. *Given a $\bar{\delta}$-generic k-tuple $\{(S_l, i_l)\}$ we have that*

$$\mathbb{P}(\mathcal{A} \text{ is realized by } \{(S_l, i_l)\}) = \mathbf{c}^k \bar{\delta}^{dk} \left(\prod_{l=1}^{k} \mathfrak{m}(K_{i_l}) \right) \exp(-\mathbf{cm}(\mathfrak{X})\hat{\mathcal{P}})(1 + o_{\bar{\delta}}(1)).$$

Proof of (6.2). By Proposition 6.5 we can restrict to the contribution of generic collections of squares.

Now, there are $n_s = \frac{(\hat{\mathcal{P}}_s)}{\bar{\delta}^d}(1 + \mathcal{O}(\bar{\delta}))$ cubes in $M\mathcal{P}_s$. The number of possible generic choices of k-tuples $\{(S_l, i_l)\}$ that realize $\mathcal{A}$ is thus

$$\prod_{p,s} \binom{n_s}{l_{p,s}} (1 + o_{\bar{\delta}}(1)).$$

Applying Proposition 6.6 we get that

$$\mathbb{P}(\mathcal{A}) = \prod_{p,s} \left[\binom{n_s}{l_{p,s}} \left(\mathbf{cm}(K_p)\bar{\delta}^d \right)^{l_{p,s}} \right] \exp\left(-\mathbf{cm}(\mathfrak{X})\hat{\mathcal{P}} \right) (1 + o_{\bar{\delta}}(1))$$

$$= \prod_{p,s} \left[\frac{(\mathbf{cm}(K_p)(\hat{\mathcal{P}}_s))^{l_{p,s}}}{l_{p,s}!} \exp\left(-\mathbf{cm}(K_p)(\hat{\mathcal{P}}_s) \right) \right] (1 + o_{\bar{\delta}}(1))$$

since $\hat{\mathcal{P}} = \sum(\hat{\mathcal{P}}_s)$, and $\mathfrak{m}(\mathfrak{X}) = \sum \mathfrak{m}(K_i)$. $\qquad \square$

We now turn to the proofs of Propositions 6.5 and 6.6. Proposition 6.5 is a direct consequence of Lemma 6.7(2) below.

Lemma 6.7. *We have*
(1) $\mathbb{P}(\exists \mathbf{t} \in \Pi_M : \xi_{\mathbf{t}} > 1) = \mathcal{O}(M^{-d})$.

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

Proof. Parts (1) and (2) follow by summation of (h2) and (h3) respectively. $\qquad\square$

Proof of Proposition 6.6. By Lemma 6.7 (1) we know that up to excluding a small probability set we have that each $\xi_{\mathbf{t}}, \mathbf{t} \in \Pi_M$ is either equal to 1 or 0. Hence, given a $\bar{\delta}$-generic k-tuple $\{(S_l, i_l)\}$ we have that

$$\mathbb{P}(\mathcal{A} \text{ is realized by } \{(S_l, i_l)\}) = \mathbb{P}\left(\exists (\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(k)}) \in S_1 \times \dots \times S_k : \right.$$

$$\left. \eta_{\mathbf{t}^{(l)}, i_l} = 1, \forall l \in [1, k], \eta_{\mathbf{t}} = 0 \text{ for } \mathbf{t} \notin \{\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(k)}\}\right)(1 + o_{\bar{\delta}}(1)).$$

We will thus finish if we prove the following

Lemma 6.8. *Given a $\bar{\delta}$-generic k-tuple $\{(S_l, i_l)\}$ we have that*

$$\mathbb{P}\left(\exists (\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(k)}) \in S_1 \times \dots \times S_k : \right.$$

$$\left. \eta_{\mathbf{t}^{(l)}, i_l} = 1, \forall l \in [1, k], \eta_{\mathbf{t}} = 0 \text{ for } \mathbf{t} \notin \{\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(k)}\}\right)$$

$$= \mathbf{c}^k \bar{\delta}^{dk} \left(\prod_{l=1}^{k} \mathfrak{m}(K_{i_l})\right) \exp(-\mathbf{c}\mathfrak{m}(\mathfrak{X})\hat{\mathcal{P}})(1 + o_{\bar{\delta}}(1)).$$

The proof of Lemma 6.8 will be given in the next subsection. $\qquad\square$

6.3. **Compatible strips. Proof of Lemma 6.8.** Divide $\Pi = \Pi_M$ into strips parallel to $\mathrm{Ker}\lambda_1$, 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 $j \geq 1$. We still denote their width by $\bar{\delta}M$ (up to changing the definition of $\bar{\delta}$ to $\bar{\delta} - 1/\sqrt{M}$). Let L be the total number of the strips (observe that L is of order $\bar{\delta}^{-1}$). We can naturally assume that the partition into cubes $C_1, \dots, C_H$ is such that every cube is completely included in a strip Π^j. Note that, similarly to the proof of Proposition 6.5, the probability that there is a point in a buffer zone is negligible.

Definition 6.9 (Type A and B strips). Given a $\bar{\delta}$-generic k-tuple $\{(S_l, i_l)\}$ we call a strip Π^j which contains a square S_l 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 $\eta_{\mathbf{t}} = 0$ 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 $\mathbf{t}^{(q)} \in S_q$ such that $\eta_{\mathbf{t}^{(q)}, i_q} = 1$ and $\eta_{\bar{\mathbf{t}}} = 0$ for $\bar{\mathbf{t}} \in \Pi^j - \{\mathbf{t}^{(q)}\}$.

Denote $p_0 = 1$, and for $j > 0$

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

Lemma 6.8 becomes thus equivalent to showing that (recall that L is the total number of strips)

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

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 6.10. *If Π^{j+1} is of type A, with $S_q \in \Pi^{j+1}$, then*

$$(6.3) \qquad\qquad p_{j+1} = \mathbf{cm}(K_{i_q})\bar{\delta}^d p_j (1 + o_{\bar{\delta}}(1))$$

and if Π^{j+1} is of type B then

$$(6.4) \qquad\qquad p_{j+1} = p_j \left(1 - \mathbf{cm}(\mathfrak{X})\hat{\mathbf{\Pi}}^{j+1}(1 + o_{\bar{\delta}}(1)) \right)$$

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

Proof of Lemma 6.10. We first prove (6.3). So, we assume Π^{j+1} is of type A, with $S_q \in \Pi^{j+1}$. Let $\tilde{\mathbf{t}}$ be such that

$$(6.5) \qquad \min_{\mathbf{t} \in \Pi^{j+1}} \lambda_1(\mathbf{t}) - R \ln M \geq \lambda_1(\tilde{\mathbf{t}}) \geq \max_{\mathbf{t} \in \Pi^j} \lambda_1(\mathbf{t}) + R \ln M.$$

We define $\mathcal{F}_j := \mathcal{F}_{\tilde{\mathbf{t}}}$.

We will need a control on the simultaneous occurrences of $\eta_{\mathbf{t}} = 1$ and $\eta_{\mathbf{t}'} = 1$ for pairs $(\mathbf{t}, \mathbf{t}')$. Denote

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

Sublemma 6.11. *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 $\omega \in E_j$*

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

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

$$(6.7) \qquad\qquad \mathbb{E}(\eta_{\mathbf{t}} \eta_{\bar{\mathbf{t}}}|\mathcal{F}_j)(\omega) \ll M^{-2d}.$$

The proof is technical and involves all the properties (h3)–(h6). We differ it to the Appendix B.

As a consequence we get the following.

Sublemma 6.12. *We have for Π^{j+1} of type A, with $S_q \subset \Pi^{j+1}$, and $\omega \in E_j$*

$$(6.8) \qquad \mathfrak{p} := \mathbb{P}(\Pi^{j+1} \text{ is compatible } |\mathcal{F}_j)(\omega) = \mathbf{cm}(K_{i_q})\bar{\delta}^d(1 + o_{\bar{\delta}}(1)).$$

Proof. Recall that by definition,

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

We omit in the rest of this proof the mention of ω, that is assumed to be fixed in E_j. Since, for a fixed $\mathbf{t} \in S_q$, Bonferroni inequalities imply that

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

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

Hence, because $\omega \in E_j$,

$$\left|\mathfrak{p} - \sum_{\mathbf{t}\in S_q}\mathbb{E}(\eta_{\mathbf{t},i_q}|\mathcal{F}_j)\right| \leq \sum_{\mathbf{t},\in S_q,\,\mathbf{t}'\neq\mathbf{t}\in\Pi^{j+1}} \mathbb{E}(\eta_{\mathbf{t}}\eta_{\mathbf{t}'}|\mathcal{F}_j) \leq \mathcal{O}(1/\sqrt{M}) + \mathcal{O}\left(\bar{\delta}^{d+1}\right)$$

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

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

$$\sum_{t\in S_q}\mathbb{E}(\eta_{t,i_q}|\mathcal{F}_j) = \mathbf{cm}(K_{i_q})\bar{\delta}^d(1 + o_{\bar{\delta}}(1)).$$

proving (6.8).   □

Proof of (6.3). With the set E_j from Sublemma 6.11 we have

$$p_{j+1} = \mathbb{P}((\Pi^1,\ldots,\Pi^{j+1}\ \text{are compatible}) \cap E_j) + \mathcal{O}(\ln M/\sqrt{M})$$

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

where the last step relies on the fact that $E_j \subset E$ and so, by (h5), the event $\mathbb{1}_{\Pi^1,\ldots,\Pi^j\ \text{are compatible}}$ is constant on $F_j(\omega)$ for $\omega \in E_j$. We finish using (6.8) and one more time the fact that $\mathbb{P}(E_j^c) \leq \frac{C\ln M}{\sqrt{M}}$.   □

Proof of (6.4). (6.4) follows from Sublemma 6.13 below exactly in the same way as (6.3) follows from Sublemma 6.12.

Sublemma 6.13. *We have for Π^{j+1} of type B and $\omega \in E_j$*

(6.10) $\mathfrak{p}' := \mathbb{P}(\Pi^{j+1}\ \text{is compatible}\ |\mathcal{F}_j)(\omega) = 1 - \mathbf{cm}(\mathfrak{X})\hat{\mathbf{\Pi}}_{j+1}(1 + o_{\bar{\delta}}(1)).$

Proof. We omit in the rest of this proof the mention of ω, that is assumed to be fixed in E_j. Observe that

$$\mathfrak{p}' = \mathbb{P}\left(\eta_{\mathbf{t}} = 0\ \text{for all}\ \mathbf{t} \in \Pi^{j+1}|\mathcal{F}_j\right) = 1 - \bar{\mathfrak{p}} - \hat{\mathfrak{p}}$$

with

$$\bar{\mathfrak{p}} = \mathbb{P}\left(\text{there exists a unique}\ \mathbf{t} \in \Pi^{j+1}\ \text{such that}\ \eta_{\mathbf{t}} = 1|\mathcal{F}_j\right)$$

$$\hat{\mathfrak{p}} = \mathbb{P}\left(\text{there exists at least a pair}\ (\mathbf{t},\mathbf{t}') \in \Pi^{j+1}\ \text{such that}\ \eta_{\mathbf{t}} = \eta_{\mathbf{t}'} = 1|\mathcal{F}_j\right).$$

From the proof of (6.3) (taking the sum over all the squares of Π^{j+1} and $i_q \in [1, P]$ in (6.8) and (6.9)) we have that

$$\bar{\mathfrak{p}} = \mathbf{cm}(\mathfrak{X})\hat{\mathbf{\Pi}}_{j+1}(1 + o_{\bar{\delta}}(1)).$$

On the other hand

$$\hat{\mathfrak{p}} \leq \sum_{(\mathbf{t},\mathbf{t}')\in\Pi^{j+1},\mathbf{t}'\neq\mathbf{t}} \mathbb{E}(\eta_{\mathbf{t}}\eta_{\mathbf{t}'}|\mathcal{F}_j) \leq \mathcal{O}(1/\sqrt{M}) + \mathcal{O}(\bar{\delta}^2)$$

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

Since $\hat{\mathfrak{p}} = o_{\bar{\delta}}(\bar{\mathfrak{p}})$, (6.10) follows. □

The proof of Lemma 6.10, and therefore of Lemma 6.8 and Theorem 6.1 is now complete. □

6.4. **Proof of Theorem 6.2.** The proof of Theorem 6.2 is similar to the proof of Theorem 6.1. In addition to using (h4), (h5) and (h6) to control $\nu_{\mathbf{t}}$ we use (h7) and (h8) to control $\zeta_{\mathbf{t}}$. Let us briefly describe the necessary modifications.

We need a stronger notion of generic collections of squares than that of Definition 6.4.

Definition 6.14. We call a collection of cubes $S_1, \ldots, S_k$ strongly $\bar{\delta}$-generic if the distance of any two among $2k$ intervals

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

is at least $3\bar{\delta}M$. We say that a k-tuple $\{(S_l, i_l, j_l)\}$, where $i_l \in [1, P], j_l \in [1, J]$, is strongly $\bar{\delta}$-generic if $S_1, \ldots, S_k$ is strongly $\bar{\delta}$-generic.

The same argument as in the proof of Theorem 6.1 shows that the contribution of non strongly generic configurations becomes negligible as $\bar{\delta} \to 0$.

For the proof of Theorem 6.2 we need the following generalization of Lemma 6.8.

Lemma 6.15. *Given a strongly $\bar{\delta}$-generic k-tuple $\{(S_l, i_l, j_l)\}$, we have that*

$$\mathbb{P}\left(\exists(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(k)}) \in S_1 \times \ldots \times S_k : \right.$$

$$\left. \eta_{\mathbf{t}^{(l)},i_l} = 1, \zeta_{\mathbf{t}^{(l)}} \in \tilde{K}_{j_l} \forall l \in [1, k], \eta_{\mathbf{t}} = 0 \ for \ \mathbf{t} \notin \{\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(k)}\}\right)$$

$$= \mathbf{c}^k \bar{\delta}^{dk}\left(\prod_{l=1}^{k} \mathfrak{m}(K_{i_l})\tilde{\mathfrak{m}}(\tilde{K}_{j_l})\right)\exp(-\mathbf{cm}(\mathfrak{X})\hat{\mathcal{P}})(1 + o_{\bar{\delta}}(1)).$$

The derivation of Theorem 6.2 from Lemma 6.15 is exactly the same as the derivation of Theorem 6.1 from Lemma 6.8.

We thus focus on explaining the difference in the proof of Lemma 6.15 from that of Lemma 6.8.

We need to replace Definition 6.9 of type A and B strips by the following definition that takes into account a third kind of strips. We keep the division $\{\Pi^j\}_{j=1,\dots,L}$ of $\Pi = \Pi_M$ into the parallel strips of width $\bar{\delta}M$.

Definition 6.16 (Type A, B and C strips)**.** For $j \in [1, L]$, we say that Π^j is of type A if it contains a cube S_q from our configuration. For such a pair (Π^j, S_q) we say that a strip Π^p is associated to S_q if $\lambda_1(\Pi^p) \cap \hat{\lambda}(S_q) \neq \emptyset$. Note that since $\hat{\lambda}(\mathbf{t}) > \lambda_1(\mathbf{t})$ on $\mathrm{Int}(\mathcal{P})$, we have that $p > j$. We call the union of all strips associated to a given S_q a type C strip. Note that type C strips are a union of a uniformly bounded number of consecutive strips Π^i, so their width is still $\mathcal{O}(\bar{\delta})$. The strips from the partition $\{\Pi^j\}$ which are neither type A nor belong to a type C strip will be called type B strips. For convenience, after slightly decreasing the total number of strips to $L' \leq L$, we still call the collection of strips that we get $\{\Pi^j\}_{j=1,\dots,L'}$ (the fact that the type C strips are slightly larger than the others will have no consequence on the subsequent proofs).

Observe that there may exists $\bar{k} < k$ such that for $q \in \{\bar{k} + 1, \dots, k\}$ it holds that $\lambda_1(\Pi_M) \cap \hat{\lambda}(S_q) = \emptyset$. If not, we just take $\bar{k} = k$. Thus, we have k type A strips, $\bar{k}$ type C strips and $L - k - \bar{k}$ type B strips. It will be convenient for the proof below to add $\Pi^{L'+1}, \dots, \Pi^{L'+k-\bar{k}}$ strips in $\tilde{\Pi}_M$ where for every $j \in [1, k - \bar{k}]$

$$\Pi^{L'+j} = \{\mathbf{t} \in \tilde{\Pi}_M : \exists \mathbf{t}' \in S_{\bar{k}+j} : |\lambda_1(\mathbf{t}) - \hat{\lambda}(\mathbf{t}')| \leq R \ln M\}.$$

We call $\Pi^{L'+j}$ type C strips associated to $S_{\bar{k}+j}$. The fact that these additional strips are well defined is due to our assumption that $\tilde{\mathcal{P}}$ is sufficiently large so that $\max_{\tilde{\mathcal{P}}} \lambda_1 > \max_{\mathcal{P}} \hat{\lambda}$.

Denote the total number of strips $L = L' + k - \bar{k}$.

The definition of compatibility for type A and type B strips remains the same as in Definition 6.9. Given a cube S_l of our collection, and a strip Π^j of type A such that $S_l \subset \Pi^j$ and suppose that Π^j is compatible. This gives $\mathbf{t}^{(l)} \in S_l$ satisfying the conditions of Definition 6.9 ($\eta_{\mathbf{t}^{(l)}, i_l} = 1$ and $\eta_{\mathbf{t}} = 0$ for all $\mathbf{t} \in \Pi^j - \{\mathbf{t}^{(l)}\}$). Assume now that Π^p, $p > j$ is a type C strip associated to S_l. Then, Π^p is called compatible if $\zeta_{\mathbf{t}^{(l)}} \in \tilde{K}_{j_l}$ and $\eta_{\mathbf{t}} = 0$ for all $\mathbf{t} \in \Pi^p \cap \Pi_M$. In particular, for type C strips which are disjoint from Π_M the only requirement that we have is that $\zeta_{\mathbf{t}^{(l)}} \in \tilde{K}_{j_l}$.

Let $p_j = \mathbb{P}(\Pi^l$ are compatible for $l \leq j)$. Then Lemma 6.15 becomes thus equivalent to showing that

$$p_L = \mathbf{c}^k \bar{\delta}^{dk} \left(\prod_{l=1}^k \mathfrak{m}(K_{i_l}) \tilde{\mathfrak{m}}(\tilde{K}_{j_l}) \right) \exp(-\mathbf{c}\mathfrak{m}(\mathfrak{X})\hat{\mathcal{P}})(1 + o_{\bar{\delta}}(1)).$$

Then, as in the proof of Lemma 6.8, Lemma 6.15 follows inductively from

Lemma 6.17. *If Π^{j+1} is of type A, with $S_q \in \Pi^{j+1}$, then*

(6.11) $$p_{j+1} = \mathbf{c}\mathfrak{m}(K_{i_q})\bar{\delta}^d p_j(1 + o_{\bar{\delta}}(1));$$

if Π^{j+1} is of type B then

(6.12) $$p_{j+1} = p_j \left(1 - \mathbf{cm}(\mathfrak{X})\hat{\Pi}^{j+1}(1 + o_{\bar{\delta}}(1)) \right);$$

and if Π^{j+1} is a type C strip associated to a square S_q then

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

Proof. We consider the same σ-algebra $\mathcal{F}_j$ as in the proof of Lemma 6.10, that is $\mathcal{F}_j := \mathcal{F}_{\tilde{\mathbf{t}}}$ with $\tilde{\mathbf{t}}$ such that

$$\min_{\mathbf{t} \in \Pi^{j+1}} \lambda_1(\mathbf{t}) - R \ln M \geq \lambda_1(\tilde{\mathbf{t}}) \geq \max_{\mathbf{t} \in \Pi^j} \lambda_1(\mathbf{t}) + R \ln M.$$

We then have from (h5) and (h8) that, regardless of the type of the strips, for $\omega \in E$

$$\mathbb{1}_{\Pi^1,\dots,\Pi^j \text{ are compatible}} \text{ is constant on } F_j(\omega).$$

Indeed, events of the form $\mathbb{1}_{\eta_{\mathbf{t}^{(l)}} \in K_{i_l}}$, with $\lambda_1(\mathbf{t}^{(l)}) \in \Pi^1 \cup \dots \cup \Pi^j$ and $\mathbb{1}_{\zeta_{\mathbf{t}^{(l)}} \in \tilde{K}_{j_l}}$ with $\hat{\lambda}(\mathbf{t}^{(l)}) \in \lambda_1(\Pi^1 \cup \dots \cup \Pi^j)$ are constant on $F_j(\omega)$ by (h5) and (h8).

Observe that (6.11) and (6.12) as well as their proofs are then identical to (6.3) and (6.4) of Lemma 6.10. As for (6.13), it follows from

Sublemma 6.18. *If Π^{j+1} is a type C strip associated to a square S_q, and $\omega \in E$ is such that on ω, Π^l are compatible for $l \leq j$, we have*

(6.14) $$\mathbb{P}(\Pi^{j+1} \text{ is compatible } |\mathcal{F}_j)(\omega) = \tilde{\mathfrak{m}}(\tilde{K}_{j_q})(1 + o_{\bar{\delta}}(1)).$$

Proof. By definition there exists some $\mathbf{t}^{(q)} \in S_q$ such that $\hat{\lambda}(\mathbf{t}^{(q)}) \in \lambda_1(\Pi^{j+1})$. By the definition of $\tilde{\mathbf{t}}$, this implies that $\hat{\lambda}(\mathbf{t}^{(q)}) > \lambda_1(\tilde{\mathbf{t}}) + R \ln M$. Next

$$\mathbb{P}(\zeta^M_{\mathbf{t}^{(q)}} \in \tilde{K}_{j_q}|\mathcal{F}_j)(\omega) - \mathbb{P}(\exists \mathbf{t} \in \Pi^{j+1} : \eta_{\mathbf{t}} \neq 0|\mathcal{F}_j)(\omega)$$

$$\leq \mathbb{P}(\Pi^{j+1} \text{ is compatible } |\mathcal{F}_j)(\omega) \leq \mathbb{P}(\zeta^M_{\mathbf{t}^{(q)}} \in \tilde{K}_{j_q}|\mathcal{F}_j)(\omega)$$

Since $\omega \in E$, $\mathbf{t}^{(q)}$ is constant on $\mathcal{F}_j(\omega)$. Hence (h7) gives

$$\mathbb{P}(\zeta^M_{\mathbf{t}^{(q)}} \in \tilde{K}_{j_q}|\mathcal{F}_j)(\omega) = \tilde{\mathfrak{m}}(\tilde{K}_{j_q}) + o_{\bar{\delta}}(1)$$

while (h4c) gives

$$\mathbb{P}(\exists \mathbf{t} \in \Pi^{j+1} : \eta_{\mathbf{t}} \neq 0|\mathcal{F}_j)(\omega) = \mathcal{O}(\bar{\delta}).$$

Combining the last three estimates we obtain (6.14). □

We have completed the proof of Lemma 6.17 and thus of Theorem 6.2. □

7. Rate of equi-distribution of unipotent flows

7.1. **Notation.** Recall that $\{g^{\mathbf{t}}\}_{\mathbf{t} \in \mathbb{R}^d}$ is the d parameter subgroup given by (5.3). We let

$$(7.1) \qquad T(\mathbf{t}) = \sum_{j=1}^{d} \mathbf{t}_j.$$

In the proof of Theorem 5.4 we will need to show asymptotic independence between the moments of time such that $\Phi(g^{\mathbf{t}}\mathcal{L}) \geq 1$ as well as randomness of the values taken then by $\Psi(g^{\mathbf{t}}\mathcal{L}, Ne^{-T(\mathbf{t})})$. For this we will rely on the fact that the action of $g^{\mathbf{t}}$ on $\mathcal{M}$ is partially hyperbolic in the sense that

$$T\mathcal{M} = E_0 \oplus \sum_{q<p}^{d+1} \left(E_{qp}^+ \oplus E_{qp}^- \right)$$

where E_0 is tangent to the orbit of g and $E_q^{\pm}$ are invariant one dimensional distributions. The corresponding Lyapunov exponents are $\pm\lambda_{qp}$ with $q < p \leq d$ and $\pm\lambda_q$ (corresponding to $p = d+1$) where

$$\lambda_{qp} = \mathbf{t}_q - \mathbf{t}_p, \quad \lambda_q = T(\mathbf{t}) + \mathbf{t}_q.$$

$E_{qp}^{\pm}$ are tangent to foliations $W_{qp}^{\pm}$ which are orbit foliations for groups $h_{qp}^{\pm}$ where $h_{qp}^+(u)$ is the matrix with ones on the main diagonal, u in the p-th column of the q-th row, and zero in all other places, while $h_{qp}^-(u)$ are transposes of $h_{qp}^+(u)$. Below we shall abbreviate $E_{1(d+1)}^+, W_{1(d+1)}^+, h_{1(d+1)}^+$ with E_1, W_1, h_1.

We also use the notation $\mu(A) = \int_{\mathcal{M}} A(x)d\mu(x)$, and for $g \in SL_{d+1}(\mathbb{R})$, 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. To fix the notation we discuss only the subgroup $h_1(u)$ (which is the only subgroup used in the proof of Theorem 5.4 given in Section 8), however similar results holds for all other unipotent subgroups including $h_{pq}^{\pm}$.

By [28], Theorem 2.4.5 there exists $\mathbf{s}$ and constants $C, \kappa > 0$ such that if $A, B \in H^{\mathbf{s}}$ then

$$(7.2) \qquad |\mu(A(\cdot)B(g^{\mathbf{t}}\cdot)) - \mu(A)\mu(B)| \leq C||A||_{\mathbf{s}}||B||_{\mathbf{s}} e^{-\kappa \max |\mathbf{t}_j|}$$

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

$$(7.3) \qquad |\mu(A(\cdot)B(h_1(u)\cdot)) - \mu(A)\mu(B)| \leq C||A||_{\mathbf{s}}||B||_{\mathbf{s}} u^{-\kappa}.$$

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

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

$$
\mathfrak{g}_t = \operatorname{diag}(e^{-t}, 0, \ldots 0, e^t).
$$

It is then immediate to observe that

$$
M(t) := h_1(e^t)R(t)^{-1}\mathfrak{g}_{-t}
$$

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

$$
\mu(A(\cdot)B(h_1(e^t)\cdot)) = \mu(A(\cdot)B(M(t)\mathfrak{g}_t R(t)\cdot)) = \mu(A_t(\cdot)B_t(\mathfrak{g}_t\cdot))
$$

with $A_t(\cdot) := A(R(t)^{-1}\cdot)$, $B_t(\cdot) := B(M_t\cdot)$. Note that $||A_t||_{\mathbf{s}} \leq ||A||_{\mathbf{s}}$, $||B_t||_{\mathbf{s}} \ll ||B||_{\mathbf{s}}$ uniformly in $t \geq 0$ Hence, (7.3) follows from (7.2) applied to A_t, B_t and $\mathfrak{g}_t$ with $t = \ln u$.

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 $H^{\mathbf{s}}$-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_1 curve of size $L > 0$ if for some $y \in \mathcal{M}$ we have $\gamma = \{h_1(\tau)y : \tau \in [0, L]\}$. For a function $A : \mathcal{M} \to \mathbb{R}$ we use the notation

$$
\int_{\gamma} A = \frac{1}{L} \int_0^L A(h_1(s)y)ds.
$$

Definition 7.2. Fix $\kappa_0 > 0$. Let $L > 0$ and $\mathcal{P}$ be a partition of $\mathcal{M}$ into W_1-curves of length L and denote by $\gamma(x)$ 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

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

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 7.3. *There exists* $\mathbf{s}, \kappa_0, \epsilon_0 > 0$ *such that 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_1-*curves of length* L *that is* κ_0-*representative with respect to* $((k_n), A)$.

(b) If $\mathcal{L} \in \mathcal{M}$ *is distributed according to a probability measure* $\tilde{\mu}$ *that has a bounded density with respect to the Haar measure* μ, *then the result of part (a) holds with (7.4) in the definition of representative partitions replaced by*

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

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

Remark 7.4. The requirement that $\mathbf{r} \leq \mathbf{s}$ will only serve to maintain the exponent κ in the speed of equidistribution in (7.4) 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. §8.3).

Proof. It suffices to prove part (a). Part (b) follows from (a) since for any set Ω we have $\tilde{\mu}(\Omega) \leq \tilde{C}\mu(\Omega)$.

Without loss of generality we will work with functions A having zero average, that is $\mu(A) = 0$. We will first prove Proposition 7.3 for $A \in H^{\mathbf{s}}$ and then generalize it to $A \in H^{\mathbf{s},\mathbf{r}}$.

Now, assuming that $\mu(A) = 0$, (7.3) implies that

$$|\mu(A(\cdot)A(h_1(u)\cdot))| \leq C\mathcal{K}_A^2 u^{-\kappa}$$

with $\mathcal{K}_A = ||A||_{\mathbf{s}}$, thus for $S_L(\cdot) = \frac{1}{L}\int_0^L A(h_1(u)\cdot)du$ we have

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

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

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

Next let $\hat{\mathcal{P}}$ be an arbitrary partition of M into W_1-curves of length L and for $u \in [0,1]$ let $\hat{\mathcal{P}}^u = h_1(Lu)\hat{\mathcal{P}}$. 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

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

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

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

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

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

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. $\qquad\square$

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.** §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 7.5. *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 (7.2).

Proof. Without the loss of generality we may assume that

$$(7.8) \qquad\qquad \mu(B) = 0.$$

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

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

Assume first that A is positive. Then

$$(7.9) \qquad \mu(A(\cdot)B_-(g\cdot)) \leq \mu(A(\cdot)B(g\cdot)) \leq \mu(A(\cdot)B_+(g\cdot)).$$

Next, by (7.2)

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

Note that due to (7.8)

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

Hence

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

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

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

Likewise

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

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

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

proving the lemma in the general case. □

8. Poisson limit theorem for the diagonal action

8.1. **Overview of the proof of Theorem 5.4.** The goal of this section is to prove Theorem 5.4 using the abstract Theorem 6.2 and the polynomial rate of uniform distribution of long pieces of horocycles given in Section 7.

We fix the probability space $(\Omega, \mathbb{P})$ to be the space $(\mathcal{M}, \tilde{\mu})$ where $\tilde{\mu}$ is the measure from Theorem 5.4 that is assumed to have a smooth bounded density with respect to the Haar measure.

In all this section, the expectation with respect to $\tilde{\mu}$ of a variable X will be denoted $\mathbb{E}(X)$.

Let

$$(8.1) \qquad\qquad \mathcal{P} = \{\mathbf{t} \in \mathbb{R}^d : \mathbf{t}_j > 0, \, T(\mathbf{t}) \leq 1\},$$

where $T(\mathbf{t}) = \sum_{j=1}^{d} \mathbf{t}_j$. In this section we will denote by $\Pi_M = M\mathcal{P}$. (Note that, in fact, (5.1) gives $\Pi_M = (M - d)\mathcal{P}$ but we ignore "$-d$" in order to simplify the notation. Since the result holds for arbitrary M, this does not cause any loss of genelity). Let

$$\widetilde{\mathcal{P}} = \{\mathbf{t} \in \mathbb{R}_+^d : T(\mathbf{t}) < 10\} \text{ so that } \tilde{\Pi}_M = \{\mathbf{t} \in \mathbb{R}_+^d : T(\mathbf{t}) < 10M\}.$$

Recall the definition of $\lambda_1(t) = \mathbf{T}(\lambda) + t_1$, 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 $T(\mathbf{t}) = M$).

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 $K_1, K_2 \ldots K_P$ and let Q be the partition of $(\mathfrak{X}, \mathfrak{m})$ into the intervals $K_1, K_2 \ldots K_P$.

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

Recall the definitions of Φ, Ψ given in (5.5) (5.6) and introduce Φ_p, $p \in [1, P]$ that are defined by formula (5.5) with K_p in place of K. Let

$$(8.2) \qquad\qquad \Phi^{\mathbf{t}} = \Phi(g^{\mathbf{t}}\mathcal{L})$$

$$(8.3) \qquad\qquad \Phi_p^{\mathbf{t}} = \Phi_p(g^{\mathbf{t}}\mathcal{L})$$

$$(8.4) \qquad\qquad \nu_{\mathbf{t}} = \Psi_1(g^{\mathbf{t}}\mathcal{L})$$

$$(8.5) \qquad\qquad \zeta_{\mathbf{t}} = \Psi_2\left(g^{\mathbf{t}}\mathcal{L}, Ne^{-T(\mathbf{t})}\right)$$

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 $\Phi^{\mathbf{t}}, \{\Phi_p^{\mathbf{t}}\}_{p \leq P}, \zeta_{\mathbf{t}}, \nu_{\mathbf{t}}$, for $\mathbf{t} \in \Pi_m$ satisfy (h1)–(h8) then we get the Poisson limit for

$$\{\Psi(g^{\mathbf{t}}\Lambda(\xi), Ne^{-T(\mathbf{T})})\}_{\mathbf{t} \in \Pi_M, \Phi(g^{\mathbf{t}}\Lambda(\xi))=1}$$

required in Theorem 5.4.

Proposition 8.1, proven in §8.2, shows that $\Phi^{\mathbf{t}}$ satisfies (h1)–(h3). The proof relies on Rogers' identities given in Lemma 2.3. §8.3 contains estimates of $\|\cdot\|_{\mathbf{s},\mathbf{s}}$ norms of the functions Φ_i and Ψ_j. Then in §8.4 we show, using Proposition 7.3, the existence of the partitions $F_{\mathbf{t}}$ and the set E such that (h4)–(h8) hold.

8.2. Multiple solutions. The following proposition asserts that $\Phi^{\mathbf{t}}$ satisfies (h1)–(h3).

Proposition 8.1. *Recall the notations* $\Phi^{\mathbf{t}} = \Phi \circ g^{\mathbf{t}}$, $\Phi_p^{\mathbf{t}} = \Phi_p \circ g^{\mathbf{t}}$. *Then uniformly in* $\mathbf{t} \in \mathbb{Z}^d$, $\mathbf{t}' \in \mathbb{Z}^d - \{\mathbf{0}\}$ *we have*

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

(a') $\qquad \tilde{\mu}(\Phi_p^{\mathbf{t}}) = 2^{d-1}\mathbf{c}_1|K_p|M^{-d} + \mathcal{O}(M^{-100d})$

provided that $\min_j(\mathbf{t}_j) \geq R \ln M$ *and* R *is sufficiently large;*

(b) $\qquad \tilde{\mu}\left(\left(\Phi^{\mathbf{t}}\right)^2 - \Phi^{\mathbf{t}}\right) = \mathcal{O}\left(M^{-2d}\right)$;

(c) $\qquad \tilde{\mu}(\{\mathcal{L} \in \mathcal{M} : \Phi^{\mathbf{t}}(\mathcal{L}) \neq 0 \text{ and } \Phi^{\mathbf{t}}(g_{\mathbf{t}'}\mathcal{L}) \neq 0\}) = \mathcal{O}\left(M^{-2d}\right)$.

Note that (b) implies via Markov inequality that

$$(8.6) \qquad \tilde{\mu}(\{\mathcal{L} \in \mathcal{M} : \Phi^{\mathbf{t}}(\mathcal{L}) > 1\}) = \mathcal{O}\left(M^{-2d}\right).$$

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. Namely, part (a) of Lemma 2.3 implies that $\mu(\Phi) = \mathbf{c}_1\text{Vol}(\mathcal{D}) = 2^{d-1}\mathbf{c}_1|K|M^{-d}$, where $\mathcal{D}$ is the set defined by (5.4). Indeed

$$\text{Vol}(\mathcal{D}) = \frac{|K|}{M^d} \int_{\mathbb{R}^d} \mathbb{1}_I(x_1(\mathbf{v})) \left(\prod_{j=2}^d \mathbb{1}_J(x_j(\mathbf{v}))\right) \frac{dx}{\prod_{j=1}^d |x_j|} = 2^{d-1}|K|M^{-d}.$$

On the other hand, letting $f = \mathbb{1}_{\mathcal{D}}$ we get, since I is an interval of positive numbers, that

(8.7)
$$\Phi^2(\mathcal{L}) - \Phi(\mathcal{L}) = \sum_{\mathbf{v}_1 \neq \mathbf{v}_2 \in L \text{ prime}} f(\mathbf{v}_1)f(\mathbf{v}_2) = \sum_{\mathbf{v}_1 \neq \pm\mathbf{v}_2 \in L \text{ prime}} f(\mathbf{v}_1)f(\mathbf{v}_2)$$

so part (b) follows by Lemma 2.3 (b). As for (c) observe that if we define, for $\mathbf{v} = (x, z) \in \mathcal{L}$,

$$\tilde{f}(\mathbf{v}) = \mathbb{1}_{e^{-\mathbf{t}'_1}I}(x_1(\mathbf{v})) \left(\prod_{j=2}^d \mathbb{1}_{e^{-\mathbf{t}'_j}J}(x_j(\mathbf{v}))\right) \mathbb{1}_{e^{T(\mathbf{t}')}K}(M^d\Pi(\mathbf{v})),$$

$$\mu(\Phi\Phi^{\mathbf{t}'}) = \int_{\mathcal{M}} \sum_{\mathbf{v}_2 \neq \pm\mathbf{v}_1 \in \mathcal{L} \text{ prime}} f(\mathbf{v}_1)\tilde{f}(\mathbf{v}_2) d\mu(\mathcal{L})$$

where the contribution of $\mathbf{v}_2 = -\mathbf{v}_1$ vanishes because both I and $e^{-\mathbf{t}'_1}I$ are positive intervals, while the contribution of $\mathbf{v}_2 = \mathbf{v}_1$ vanishes since either I and $e^{-\mathbf{t}'_1}I$ are disjoint or J and $e^{-\mathbf{t}'_j}J$ for some $j = 2, \ldots d$ are disjoint. Applying Lemma 2.3(b) we get (c).

Since $\mu(\Phi_p) = 2^{d-1}\mathbf{c}_1|K_p|M^{-d}$, (a') follows by exponential mixing of the geodesic flow (Lemma 7.5) and Lemma 8.2 from § 8.3. $\qquad \square$

8.3. **Estimates of norms.** Before we construct the partition $\mathcal{F}_\mathbf{t}$, 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 Lemma 8.2 below are proven in §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 from Appendix A. We set $\mathfrak{h}_{2,\delta} = 1 - \mathfrak{h}_{1,\delta}$.

Let $\hat{K} = \{z : d(z, \partial K) \leq M^{-1000d}\}$ and set $\hat{\Phi} = \mathcal{S}(1_{\hat{\mathcal{D}}})$ where $\hat{\mathcal{D}}$ is defined similarly to $\mathcal{D}$ (see equation (5.4)) with K replaced by $\hat{K}$. Define similarly $\hat{K}_p$, $\mathcal{D}_p$, 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 $\delta > 0$, $||(\Phi^2 - \Phi)\mathfrak{h}_{2,\delta}||_{\mathbf{s},\mathbf{s}} = \mathcal{O}(\delta^{-2(d+1)})$.
(d) $\mu(\Phi_i) = 2^{d-1}\mathbf{c}_1|K_i|$, $\mu(\hat{\Phi}_i) = \mathcal{O}(M^{-1000d})$.
(e) $\mu(\Phi \mathfrak{h}_{1,\delta}) = \mathcal{O}\left(\delta^{(d+1)/2}\right)$.
(f) $\mu((\Phi^2 - \Phi)\mathfrak{h}_{2,\delta}) = \mathcal{O}\left(M^{-2d}\right)$.

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

Consider the following collection of functions

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

Let $F_\mathbf{t}$ be a partition of $\mathcal{M}$ into W_1-curves of size $L_\mathbf{t} = (e^{\lambda_1(t)}M^{1000d})^{-1}$, which is κ_0-representative with respect to $(\Pi^+(\mathbf{t}), \boldsymbol{\Phi})$ (that is, representative for all $\bar{\mathbf{t}} \in \Pi^+(\mathbf{t})$). Such a partition exists due to Proposition 7.3, Lemma 8.2 and the fact that

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

if R is sufficiently large. Moreover, if we let E_1 be the set of $\mathcal{L}$ such that for any $\mathbf{t} \in \Pi$, $\mathcal{L}$ is representative with respect to $(\mathcal{F}_\mathbf{t}, \Pi^+(\mathbf{t}), \boldsymbol{\Phi})$ then

$$\mu(E_1^c) = \mathcal{O}(M^{-100d}).$$

Proposition 8.3. *There exist sets E with*

$$\tilde{\mu}(E^c) \ll M^{-100d},$$

such that the variables $\Phi^\mathbf{t}, \eta_t, \Phi_p^\mathbf{t}, \eta_{\mathbf{t},p}$ and the partitions $F_\mathbf{t}$ satisfy the properties (h4)–(h8) of Section 6.

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

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 $\Phi_p^{\mathbf{t}}$ are integer valued and so

$$\Phi_p^{\mathbf{t}} - \eta_{\mathbf{t},p} \leq \Phi_p^{\mathbf{t}\,2} - \Phi_p^{\mathbf{t}} \leq \Phi^{\mathbf{t}\,2} - \Phi^{\mathbf{t}}.$$

Since also $0 \leq \xi_{\mathbf{t},p} - \eta_{\mathbf{t},p} \leq \Phi_p^{\mathbf{t}} \leq \Phi^{\mathbf{t}}$ we get

$$0 \leq \Phi_p^{\mathbf{t}} - \eta_{\mathbf{t},p} \leq \hat{\xi}_{\mathbf{t}}$$

where

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

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

$$0 \leq \mathbb{E}(\Phi_p^{\mathbf{t}} - \eta_{\mathbf{t},p} | \mathcal{F}_{\mathbf{t}'}) \leq \mathbb{E}(\hat{\xi}_{\mathbf{t}} | \mathcal{F}_{\mathbf{t}'}) \leq \frac{C}{M^{2d}}$$

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 $(F_{\mathbf{t}'}, \mathbf{t}, \mathbf{\Phi})$. The last display implies that

$$\mathbb{E}(\Phi_p^{\mathbf{t}} | \mathcal{F}_{\mathbf{t}'}) - \frac{C}{M^{2d}} \leq \mathbb{E}(\eta_{\mathbf{t},p} | \mathcal{F}_{\mathbf{t}'}) \leq \mathbb{E}(\Phi_p^{\mathbf{t}} | \mathcal{F}_{\mathbf{t}'}).$$

Hence (h4c) follows from (h4b).

Property (h5). Consider $\mathbf{t} \in \Pi$, $\bar{\mathbf{t}} \in \Pi^+(\mathbf{t})$, and $\gamma_{\bar{\mathbf{t}}} \in F_{\bar{\mathbf{t}}}$. If $\Phi^{\mathbf{t}} \equiv 0$ on $\gamma_{\bar{\mathbf{t}}}$ then (h5) clearly holds on $\gamma_{\bar{\mathbf{t}}}$. On the other hand suppose that there exists $\bar{\mathcal{L}} \in \gamma_{\bar{\mathbf{t}}}$ and p such that $\xi_{p,\mathbf{t}}(\bar{\mathcal{L}}) \neq 0$. If follows that $g^{\mathbf{t}}(\bar{\mathcal{L}})$ contains a vector $(\bar{x}, \bar{z}) \in I \times J^{d-1} \times K_p$. Note that

$$g^{\mathbf{t}} \gamma = \{h_\tau g^{\mathbf{t}}(\bar{\mathcal{L}})\}_{\tau \in \mathcal{I}}$$

where $\mathcal{I}$ is an interval containing zero of length which less than $M^{-10^9 d}$. Observe that $h_\tau g^{\mathbf{t}}(\bar{\mathcal{L}})$ contains a vector (x_τ, z_τ) where $x_\tau = \bar{x}$, $z_\tau = \bar{z} + \tau x_1$. In particular $|z_\tau - z| < 10^{-1000d}$. Hence if $g^{\mathbf{t}} \gamma_{\bar{\mathbf{t}}}$ does not intersect

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

then it is completely contained in K_p. The measure of $\mathcal{L}$ such that $g^{\mathbf{t}} \gamma_{\bar{\mathbf{t}}}$ intersects $\hat{K}_p$ for some p is thus bounded by $\mathcal{O}(M^{-1000d})$ from Lemma 8.2(d). Taking the complement to the union of all these exceptional $\mathcal{L}$ for all $\mathbf{t} \in \Pi, \bar{\mathbf{t}} \in \Pi^+(\mathbf{t})$ we get a set E_2 such that $\tilde{\mu}(E_2^c) = \mathcal{O}(M^{-999d})$ and (h5) holds for $\mathcal{L} \in E_2$.

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

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

then we have $\tilde{\mu}(E_3^c) = \mathcal{O}(M^{-100d})$.

We let $E_4 := E_1 \cap E_2 \cap E_3$ and observe that $\tilde{\mu}(E_4^c) = O(M^{-100d})$ and (h4), (h5), and (h6) hold on E_4.

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 $\mathcal{L} \in E_4$ satisfies (h7). Then we further refine E_4 to ensure (h8).

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

$$(8.8) \qquad \tilde{\mu}\left(\Psi_2\left(g^{\mathbf{t}}\mathcal{L}, \frac{N}{e^{T(\mathbf{t})}} \right) \in \tilde{K}_j \big| \mathcal{F}_{\mathbf{t}'} \right)(\mathcal{L}) = |\tilde{K}_j|(1 + o(1)).$$

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

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

for some $\bar{\mathcal{L}} \in \mathcal{M}$. By property (h5) $\Phi_p^{\mathbf{t}} = 1$ on $\gamma_{\mathbf{t}'}$. In particular, $\Phi_p(g^{\mathbf{t}}\bar{\mathcal{L}}) = 1$, that is $g^{\mathbf{t}}\bar{\mathcal{L}}$ contains a vector $(x, z) \in \mathcal{D}_p$. Since $g^{\mathbf{t}}h_\tau = h_{e^{\lambda_1(\mathbf{t})}\tau} g^{\mathbf{t}}$ it follows that $g^{\mathbf{t}}h_\tau \bar{\mathcal{L}}$ contains the vector $(x_\tau, z_\tau) \in \mathcal{D}_p$. Namely, $x_\tau = x$, $z_\tau = z + e^{\lambda_1(\mathbf{t})}\tau x_1$. Now

$$(8.9) \qquad \Psi_2\left(g^{\mathbf{t}}h_\tau^1 \bar{\mathcal{L}}, \frac{N}{e^{T(\mathbf{t})}} \right) = \frac{N}{e^{T(\mathbf{t})}} z_\tau \bmod 2 = \frac{N}{e^{T(\mathbf{t})}} z + e^{\hat{\lambda}(\mathbf{t})}\tau x_1 \bmod 2.$$

and since τ varies on an interval of length $(e^{\lambda_1(\mathbf{t}')} M^{1000})^{-1}$, the uniform distribution (8.8) follows from the fact that $\hat{\lambda}(\mathbf{t}) \geq \lambda_1(\mathbf{t}') + R \ln M$ provided that R is sufficiently large.

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

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

for some $\bar{\mathcal{L}} \in \mathcal{M}$. By (h5), $\Phi^{\mathbf{t}} = 1$ on $F_{\bar{\mathbf{t}}}(\omega)$. (8.9) shows that if $\mathbb{1}_{\zeta_{\mathbf{t}} \in \tilde{K}_j}$ is not constant on $F_{\bar{\mathbf{t}}}(\mathcal{L})$ then $\Psi_2\left(g^{\mathbf{t}}h_\tau^1 \bar{\mathcal{L}}, \frac{N}{e^{T(\mathbf{t})}} \right)$ lies in a $\mathcal{O}(M^{-1000d})$ neighborhood of $\partial \tilde{K}_j$. Let

$$E = \left\{ \mathcal{L} \in E_4 : \forall \mathbf{t}, \bar{\mathbf{t}} \in \Pi \text{ with } \lambda_1(\bar{\mathbf{t}}) \geq \hat{\lambda}(\mathbf{t}) + R \ln M \text{ it holds that :} \right.$$

$$\left. \forall j \in [1, J], \quad \mathbb{1}_{\zeta_{\mathbf{t}} \in \tilde{K}_j} \text{ is constant on } F_{\bar{\mathbf{t}}}(\mathcal{L}) \right\}.$$

Then (h8) holds on E. On the other hand, the same argument as for property (h5) shows that $\tilde{\mu}(E - E_4) = \mathcal{O}(M^{-999d})$ as needed.

We have thus checked (h4)–(h8) for $\Phi^{\mathbf{t}}, \{\Phi_p^{\mathbf{t}}\}, \{\zeta_{\mathbf{t}}\}$ for $\mathbf{t} \in \Pi$ and $p \in [1, P]$, which completes the proof of Proposition 8.3. $\qquad\square$

Proof of Theorem 5.4. Theorem 5.4(a) is exactly property (h4a) that we proved in Proposition 8.3.

To prove Theorem 5.4(b), we note that it follows from Theorem 6.2 and from properties (h1)–(h8) that we proved in Propositions 8.1 and 8.3, that the process

$$\left\{ \left(\Psi \left(g^{\mathbf{t}} \mathcal{L}, N e^{-\sum_{j=1}^{d} \mathbf{t}_j} \right), \frac{\mathbf{t}}{M} \right) \right\}_{\Phi(g^{\mathbf{t}} \mathcal{L})=1, \, \mathbf{t} \in \Pi_M}$$

converges in probability, as $N \to \infty$, to a Poisson process on

$$\left[-\frac{1}{\bar{\epsilon}}, \frac{1}{\bar{\epsilon}} \right] \times \mathbb{R}/(2\mathbb{Z}) \times \mathcal{P}$$

with intensity $2^{d-1}\mathbf{c}_1$. Since $\hat{\mathcal{P}} = \frac{1}{d!}$ the claim of Theorem 5.4(b) follows from the invariance of Poisson processes by projection given by Lemma 2.1(b). $\qquad\square$

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 A while the case $\gamma = 1/d$ was studied in [32]. For $\gamma > 1/d$ most orbits do not visit $\mathcal{C}_N$ so we consider the remaining case $0 < \gamma < \frac{1}{d}$. Recall (1.6).

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

The proof of Theorem 9.1 is very similar to the proof of Theorem A so we just describe the necessary changes.

Similarly to Theorem A the proof consists of two parts: showing that non-resonant terms are negligible and establishing the Poisson limit for the resonant terms.

To describe the first part let

$$Z(\vartheta, N, \gamma) = \{k \in Z(\vartheta, N) : |\bar{k}_i| > N^\gamma\},$$

$$\bar{D}(\vartheta, N, \gamma) = \sum_{k \in Z(\vartheta, N, \gamma)} \frac{\Gamma_k(\vartheta, N, \gamma)}{\Omega_k(\vartheta, N)}$$

where $\bar{k}_i$ are defined by (3.1),

$$\Gamma_k(\vartheta, N, \gamma) = \frac{2A}{\pi} \phi(\bar{k}_1 u_1 N^{-\gamma}, \dots, \bar{k}_d u_d N^{-\gamma}, N(k, \alpha), \langle k, x \rangle + \varphi_{k,\alpha,N})$$

and Ω_k and ϕ are defined by (3.5) and (3.7) respectively. We have the following analogue of Proposition 3.1.

Proposition 9.2. *For any $\upsilon > 0$, if we take $\bar{\epsilon} > 0$ sufficiently small and N sufficiently large we have that*

$$\lambda \left(\left\{ \vartheta \in X : \left| \frac{\bar{D}(\vartheta, N, \gamma)}{(\ln N)^d} - \frac{D(x, \alpha, \mathcal{C}_N, N)}{(\ln N)^d} \right| \geq \upsilon \right\} \right) \leq \upsilon.$$

Proposition 9.2 follows from the estimates of Section 3 with the exception that §3.3 has to be modified to take into account that now we remove more frequencies (namely, we now discuard the frequncies with $|\bar{k}_i| < N^\gamma$).

Define $U_k(\vartheta, N, \gamma)$ similarly to $U_k(\vartheta, N)$ with u_j replaced by $u_j N^{-\gamma}$ and let

$$D_1(\vartheta, N, \gamma) = \sum_{0 < |\bar{k}_i| < N} U_k(\vartheta, N, \gamma), \quad D_2(\vartheta, N, \gamma) = \sum_{N^\gamma < |\bar{k}_i| < N} U_k(\vartheta, N, \gamma).$$

We claim that

(9.1) $\dfrac{D_1 - D_2}{(\ln N)^d}$ converges to 0 in probability as $N \to \infty$.

To this end fix a small $\tilde{\epsilon} > 0$ and let

$$\tilde{Y}(q_1, \ldots q_d) = \{k \in \mathbb{Z}^d : |\bar{k}_j| \in [q_j, q_j{+}1]\}, \quad Q(q_1, \ldots q_d) = \prod_{j=1}^{d} \max(q_j, N^\gamma),$$

$$\mathcal{Q} = \{(q_1, \ldots, q_d) \in \mathbb{Z}^d : \min(|q_j|) < N^\gamma, \max(|q_j|) < 2N\},$$

$$\tilde{E} = \{\alpha : \exists (q_1, \ldots, q_d) \in \mathcal{Q} \text{ and } k \in \tilde{Y}(q_1, \ldots q_d) : ||\langle k, \alpha \rangle|| \le \tilde{\epsilon} Q(q_1, \ldots q_d)\}.$$

Then

(9.2) $\mathrm{mes}(\tilde{E}) \le \tilde{\epsilon}.$

On the other hand, since for $k \in \tilde{Y}(q_1, \ldots q_d)$ we have

$$|U_k(\vartheta, N, \gamma)| \le \frac{|\cos(2\pi \langle k, x \rangle + \phi_{k,\alpha,N})|}{Q(q_1, \ldots q_d)||\langle k, \alpha \rangle||}$$

we get, repeating the arguments of Subsection 3.3, that

(9.3) $||D_2 - D||_{L^2((\mathbb{T}^d - \tilde{E}) \times \mathbb{T}^d)} \le C \dfrac{(\ln N)^{d-1}}{\sqrt{\tilde{\epsilon}}}.$

Combining (9.2) and (9.3) we obtain (9.1). Combining (9.1) with the estimates of Section 3 we obtain Proposition 9.2.

Next, Theorem B has to be modified as follows.

Proposition 9.3. *Assume $\vartheta \in X$ is distributed according to the normalized Lebesgue measure λ. For any $\bar{\epsilon} > 0$, as $N \to \infty$, the process*

$$\left\{\left((\ln N)^d \left(\prod_i \bar{k}_i\right) ||\langle k, \alpha \rangle||, (N\langle k, \alpha \rangle \ mod \ 2), \{\bar{k}_1 u_1 N^{-\gamma}\}, \ldots, \{\bar{k}_d u_d N^{-\gamma}\}, \right.\right.$$

$$\left.\left.\{\langle k, x \rangle\}\right)\right\}_{k \in Z(\vartheta, N, \gamma)}$$

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 $2^{d-1} \mathbf{c}_1 (1 - \gamma d)^d / d!$ where $\mathbf{c}_1 = 1/\zeta(d+1)$ is the constant from Lemma 2.3.

The Poisson limit for the pair

$$(9.4) \qquad \left\{ (\ln N)^d \left(\prod_i \bar{k}_i \right) \| \langle k, \alpha \rangle \|, (N \langle k, \alpha \rangle \bmod 2) \right\}$$

follows from the abstract Theorem 6.2 with $\mathcal{P} = \{t : t_j > 0, \sum_j t_j < M\}$ replaced by $\mathcal{P}_\gamma = \{t : t_j > \gamma M, \sum_j t_j < M\}$. Note that the change of variables $\bar{t}_j = t_j - \gamma M$ transforms $\mathcal{P}_\gamma$ to the simplex

$$\{\bar{t} : \bar{t}_j > 0, \sum_j \bar{t}_j < (1 - \gamma d)M\}$$

so that $\mathrm{Vol}(\mathcal{P}_\gamma) = (1 - d\gamma)^d \, \mathrm{Vol}(\mathcal{P})$. This explains the extra factor $(1 - \gamma d)^d$ in Theorem 9.1.

The asymptotic independence of the remaining coordinates in Proposition 9.3 from the pair (9.4) is a consequence of the following modification of Propoition 5.3.

Given $s \in \mathbb{N}$ consider a sequence of s-tuples $(k^{(1,N)}, \ldots k^{(s,N)})$ where $k^{(j,N)} \in \mathbb{Z}^d$. Let $\bar{k}^{(j,N)}$ be vector with components

$$\bar{k}_p^{(j,N)} = a_{p,1} k_1^{(j,N)} + \cdots + a_{p,d} k_d^{(j,N)}.$$

Proposition 9.4. *Suppose that*

$$(9.5) \qquad \forall j \in \{1, \ldots, s\} \quad |\bar{k}^{(j,N)}| > N^\gamma e^{\sqrt{\ln N}},$$

and

$$(9.6) \qquad \forall j' \neq j'' \in \{1, \ldots, s\} \quad \left| \ln |\bar{k}^{(j',N)}| - \ln |\bar{k}^{(j'',N)}| \right| > e^{\sqrt{\ln N}}$$

Let (x, u) be distributed according to a density ρ_N on $\mathbb{T}^d \times T^d$ such that

$$(9.7) \qquad \|\rho_N\|_{C^1} \leq R.$$

Then the distribution of the s $(d+1)$-tuples

$$\left(\{(\bar{k}^{(1,N)}, u_1 N^{-\gamma})\}, \ldots \{(\bar{k}^{(1,N)}, u_d N^{-\gamma})\}, \{(k^{(1,N)}, x)\} , \right.$$

$$\ldots$$

$$\left. \{(\bar{k}^{(d,N)}, u_1 N^{-\gamma})\}, \ldots \{(\bar{k}^{(d,N)}, u_d N^{-\gamma})\}, \{(k^{(d,N)}, x)\} \right)$$

converges to the uniform distribution on $\mathbb{T}^{(d+1)s}$ as $N \to \infty$ and the convergence is uniform with respect to the matrix (a_{pq}) the choices of $\{k^{(j,N)}\}_{j=1}^s$ satisfying (9.5) and (9.6), and ρ_N satisfying (9.7).

The proof of Proposition 9.4 is similar to the proof of Proposition 5.3 so we omit it.

Apart from the modifications described above the proof of Proposition 9.3 is identical to the proof of Theorem B.

Also, the derivation of Theorem 9.1 from Propositions 9.2 and 9.3 is the same as the derivation of Theorem A from Proposition 3.1 and Theorem B.

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

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

where $S_v^t = x + vt$.

In the case of balls, it was shown in [12] 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 §10.2 below.

It was also shown in [12] that for balls in dimension $d = 2$ the continuous time discrepancy converges, without any normalization, in distribution. In §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 $\mathcal{C} = U(\prod_j (0, u_j))$. We assume that the triple (U, x, v) is distributed according to a smooth density of compact support and that $U \in \mathrm{SL}_{\mathrm{d}}(\mathbb{R})$ is such that $||U - I|| \leq \eta$ 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

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

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

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

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

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

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

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

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

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

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

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

Then

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

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

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

provided that $|k|$ is sufficiently large.

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

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

Hence

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

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

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

we get $\mathbb{E} \left(\sum_k \Gamma_k(U, v) \right) < \infty$ 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

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

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 $(\mathbb{R}/2\mathbb{Z})^d$. $\qquad\square$

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

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

Similarly to Section 3 (see also [12, Section 3 and §6.4]) we show that the main contribution to the discrepancy comes from the harmonics where

$$\frac{\epsilon}{\ln T} < |\langle k, v \rangle||k|^2 < \frac{1}{\epsilon \ln T} \quad \text{and} \quad |k| < T.$$

Therefore the key step in proving Theorem 10.3 is the following.

Proposition 10.4. *The point process*

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

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 B and consists of the following steps.

(a) We prove the Poisson limit for $\{|k|^2 \langle k, v \rangle \ln T\}$ 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 $\{|k|^2 \langle k, v \rangle \ln T\}$ 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

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

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

The Poisson limit of $\{|k|^2 \langle k, v \rangle \ln T\}$ 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 $\langle k, Tv \rangle \bmod 2$ is asymptotically independent of $|k|^2 \langle k, v \rangle \ln T$ using the fact that their values are determined at different scales (cf. proof of (h7) in Section 8).

(c) We show that $\langle k, x \rangle$ and $\{r|k|\}$ are independent of the previous data using the superlacunarity of the sequence of small denominators (cf. Proposition 5.3). $\qquad\square$

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

$$\mathfrak{U}_1, \mathfrak{U}_2, \ldots, \mathfrak{U}_{(d+1)^2-1}$$

be a basis in the space of left invariant vectorfields on $\mathcal{M}_{d+1}$. We let

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

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

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

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 [29, §4.2]. For completeness, we reproduce the formula from [29]. 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 $\Phi, \Psi \in C^{\mathbf{s}}(\mathcal{M}_{d+1})$

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

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

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

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

(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

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

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

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

(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 $SL_{d+1}(\mathbb{R})$ 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 $SL_{d+1}(\mathbb{R})$ 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

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

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 8.3.**

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 $\mathbb{R}^{d+1}$

$$f_\epsilon^\pm(x, z) = \phi_{I,\epsilon}^\pm(x_1)\left(\prod_{j=2}^d \phi_{J,\epsilon}^\pm(x_j)\right)\phi_{K_M,\epsilon}^\pm(\Pi(\mathbf{v})),$$

where $K_M = [-\frac{1}{\bar{\epsilon}M^2}, \frac{1}{\bar{\epsilon}M^2}]$, and define as in (5.5) the Siegel transforms $\Phi_\epsilon^\pm$ of $f_\epsilon^\pm$ instead of

$$f = \mathbb{1}_I(x_1)\left(\prod_{j=2}^d \mathbb{1}_J(x_j)\right)\mathbb{1}_{K_M}(\Pi(\mathbf{v})).$$

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, (8.7) 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

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

$$\leq \mu\left(\left[(\Phi_\epsilon^+)^2 - \Phi_\epsilon^+\right] - \left[(\Phi_\epsilon^-)^2 - \Phi_\epsilon^-\right]\right) = \mathcal{O}(\epsilon).$$

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

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

$$||((\Phi_\epsilon^\pm)^2 - \Phi_\epsilon^\pm)h_{2,\delta}||_{s,s} = \mathcal{O}\left(\delta^{-2(d+1)}\right)$$

proving part (c).

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

To prove part (e) we note that

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

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

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

and Lemma 2.3(c).

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

APPENDIX B.   INDEPENDENCE

Proof of Sublemma 6.11. We start with (6.6). Observe that

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

where the last step comes directly from (h3) and the fact that there is $\mathcal{O}(M^{2d-1}\ln M)$ terms in V_{j+1}. Let

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

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

We turn now to (6.7). Let $\mathbf{t}, \bar{\mathbf{t}} \in \Pi^{j+1}$ such that $\lambda_1(\bar{\mathbf{t}}) \ge \lambda_1(\mathbf{t}) + 3R\ln M$. Let now $\hat{\mathbf{t}}$ be such that

(B.2) $$\lambda_1(\bar{\mathbf{t}}) - R\ln M \ge \lambda_1(\hat{\mathbf{t}}) \ge \lambda_1(\mathbf{t}) + R\ln M.$$

Recall that $\tilde{\mathbf{t}}$ is such that

$$\min_{\mathbf{t}\in\Pi^{j+1}} \lambda_1(\mathbf{t}) - R\ln M \ge \lambda_1(\tilde{\mathbf{t}}) \ge \max_{\mathbf{t}\in\Pi^j} \lambda_1(\mathbf{t}) + R\ln M.$$

and that $\mathcal{F}_j = \mathcal{F}_{\tilde{\mathbf{t}}}$. We let $F_j(\omega) := F_{\tilde{\mathbf{t}}}(\omega)$. We then consider the partition $\tilde{F}_{\mathbf{t}} = F_{\mathbf{t}} \wedge F_j$ and let $\tilde{\mathcal{F}}_{\mathbf{t}}$ denote the σ-algebra generated by $\tilde{F}_{\mathbf{t}}$.

The idea in addressing (6.7) is essentially the following:

$$\mathbb{E}(\eta_{\mathbf{t}}\eta_{\bar{\mathbf{t}}}|\mathcal{F}_j) \overset{\text{by (h6)}}{\sim} \quad \mathbb{E}\left(\mathbb{E}(\eta_{\mathbf{t}}\eta_{\bar{\mathbf{t}}}|\tilde{\mathcal{F}}_{\hat{\mathbf{t}}})|\mathcal{F}_j\right) \overset{\text{by (h5)}}{\sim} \quad \mathbb{E}\left(\eta_{\mathbf{t}}\mathbb{E}(\eta_{\bar{\mathbf{t}}}|\tilde{\mathcal{F}}_{\hat{\mathbf{t}}})|\mathcal{F}_j\right)$$

$$\overset{\text{by (h4)}}{\sim} \quad \frac{\mathbf{cm}(\mathfrak{X})}{M^d}\mathbb{E}(\eta_{\mathbf{t}}|\mathcal{F}_j) \overset{\text{by (h4)}}{\sim} \quad \frac{\mathbf{cm}(\mathfrak{X})}{M^d}\frac{\mathbf{cm}(\mathfrak{X})}{M^d}.$$

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 $F_{\hat{\mathbf{t}}}(\omega) \subset F_j(\omega)$ for $\omega \in E$. Hence, for $\omega \in E$ we have

$$\mathbb{E}(\eta_{\mathbf{t}} \eta_{\bar{\mathbf{t}}} | \mathcal{F}_j) = \mathbb{E}\left(\mathbb{E}\left(\eta_{\mathbf{t}} | \tilde{\mathcal{F}}_{\hat{\mathbf{t}}}\right) \eta_{\bar{\mathbf{t}}} | \mathcal{F}_j\right) + \mathcal{R}_1(\mathbf{t}, \bar{\mathbf{t}})$$

$$= \mathbb{E}\left(\mathbb{E}\left(\eta_{\mathbf{t}} | \tilde{\mathcal{F}}_{\hat{\mathbf{t}}}\right) \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \tilde{\mathcal{F}}_{\hat{\mathbf{t}}}\right) | \mathcal{F}_j\right) + \mathcal{R}_1$$

$$= \mathbb{E}\left(\mathbb{E}\left(\eta_{\mathbf{t}} | \tilde{\mathcal{F}}_{\hat{\mathbf{t}}}\right) \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \mathcal{F}_{\hat{\mathbf{t}}}\right) | \mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2$$

$$= \mathbb{E}\left(\eta_{\mathbf{t}} \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \mathcal{F}_{\hat{\mathbf{t}}}\right) | \mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3$$

$$= \mathbb{E}\left(\eta_{\mathbf{t}} \mathbb{1}_E \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \mathcal{F}_{\hat{\mathbf{t}}}\right) | \mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4$$

$$\leq \frac{C}{M^d} \mathbb{E}\left(\eta_{\mathbf{t}} | \mathcal{F}_j\right) + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4$$

$$\leq \frac{C}{M^{2d}} + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4$$

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

$$\mathcal{R}_1 = \mathbb{E}\left(\left[\eta_{\mathbf{t}} - \mathbb{E}\left(\eta_{\mathbf{t}} | \tilde{\mathcal{F}}_{\hat{\mathbf{t}}}\right)\right] \eta_{\bar{\mathbf{t}}} \Big| \mathcal{F}_j\right),$$

$$\mathcal{R}_2 = \mathbb{E}\left(\mathbb{E}\left(\eta_{\mathbf{t}} \Big| \tilde{\mathcal{F}}_{\hat{\mathbf{t}}}\right) \left[\mathbb{E}\left(\eta_{\bar{\mathbf{t}}} \Big| \tilde{\mathcal{F}}_{\hat{\mathbf{t}}}\right) - \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \mathcal{F}_{\hat{\mathbf{t}}}\right)\right] \Big| \mathcal{F}_j\right),$$

$$\mathcal{R}_3 = \mathbb{E}\left(\left[\mathbb{E}\left(\eta_{\mathbf{t}} \Big| \mathcal{F}_{\hat{\mathbf{t}}}\right) - \eta_{\mathbf{t}}\right] \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \mathcal{F}_{\hat{\mathbf{t}}}\right) | \mathcal{F}_j\right),$$

$$\mathcal{R}_4 = \mathbb{E}\left(\eta_{\mathbf{t}} \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \mathcal{F}_{\hat{\mathbf{t}}}\right) | \mathcal{F}_j\right) - \mathbb{E}\left(\eta_{\mathbf{t}} \mathbb{1}_E \mathbb{E}\left(\eta_{\bar{\mathbf{t}}} | \mathcal{F}_{\hat{\mathbf{t}}}\right) \Big| \mathcal{F}_j\right).$$

Note that $\mathcal{R}_l$ are $\mathcal{F}_j$ measurable for $l \in \{1 \ldots 4\}$. We claim that all of them have L^1 norm of order $\mathcal{O}\left(M^{-100d}\right)$. Indeed, first,

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

Next

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

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

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

Let now

$$E_j'' = \{\omega \in E : \forall \mathbf{t}, \bar{\mathbf{t}} \in \Pi^{j+1} \text{ and } \hat{\mathbf{t}} \text{ as in (B.2)} \ \forall l \in \{1 \ldots 4\} \ |\mathcal{R}_l(\mathbf{t}, \bar{\mathbf{t}})| \leq M^{-2d}\}.$$

Then by Markov inequality $\mathbb{P}((E_j'')^c) = \mathcal{O}\left(M^{-90d}\right)$ and (6.7) holds for $\omega \in E_j''$.

Letting $E_j = E_j' \cap E_j''$ where E_j' is given by (B.1), we finish the proof of Sublemma 6.11. $\qquad \square$

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. Math. **140** (1994) 449–502.

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

[4] A. I. Bufetov, *Limit theorems for translation flows,* Ann. Math. **179** (2014) 431–499.

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

[6] A. I. Bufetov, B. Solomyak *Limit theorems for self-similar tilings,* Comm. Math. Phys. **319** (2013) 761–789.

[7] F. Cellarosi, J. Marklof *Quadratic Weyl sums, automorphic functions and invariance principles,* Proc. LMS **113** (2016) 775–828.

[8] 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.

[9] J.-P. Conze, S. Isola, S. Le Borgne *Diffusive behavior of ergodic sums over rotations,* Stoch. Dyn. **19** (2019) paper 1950016, 26 pp.

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

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

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

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

[14] D. Dolgopyat, I. Goldsheid, *Quenched limit theorems for nearest neighbour random walks in 1D random environment,* Comm. Math. Phys. **315** (2012) 241–277.

[15] D. Dolgopyat, O. Sarig *Quenched and annealed temporal limit theorems for circle rotations,* Asterisque **415** (2020) 57–85.

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

[17] G. Forni, A. Kanigowski *Time-Changes of Heisenberg nilflows,* preprint arXiv: 1711.05543.

[18] 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.

[19] J. Griffin, J. Marklof *Limit theorems for skew translations,* J. Mod. Dyn. **8** (2014) 177–189.

[20] G. H. Hardy, J. E. Littlewood *Some Problems of Diophantine Approximation: The Lattice–Points of a Right–Angled Triangle,* Proc. LMS **3** (1922) 15-36.

[21] M. Hirata *Poisson law for Axiom A diffeomorphisms,* Erg. Th. Dyn. Sys. **13** (1993) 533–556.

[22] 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.

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

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

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

[26] A. Khintchine *Ein Satz über Kettenbrüche, mit arithmetischen Anwendungen,* Mathematische Zeitschrift **18** (1923) 289–306.

[27] J. Kingman *Poisson processes,* Oxford Studies in Probability **3** (1993) Oxford Univ. Press, New York 104 pp.

[28] 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.

[29] D. Y. Kleinbock, G. A. Margulis *Logarithm laws for flows on homogeneous spaces,* Invent. Math. **138** (1999) 451–494.

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

[31] Marklof J. *Kinetic limits of dynamical systems,* Proc. Symp. Pure Math. **87** (2015) 195–223.

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

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

[34] J. Marklof, I. Vinogradov *Spherical averages in the space of marked lattices,* Geom. Dedicata **186** (2017) 75–102.

[35] A. Ostrowski *Bemerkungen zur Theorie der Diophantischen Approximationen,* Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg **1** (1922) 77 – 98.

[36] B. S. Pitskel *Poisson Limit Law for Markov chains,* Erg. Th. Dyn. Sys. **11** (1991) 501–513.

[37] C. A. Rogers *The number of lattice points in a set,* Proc. LMS **6** (1956) 305–320.

[38] K. F. Roth, *On irregularities of distribution,* Mathematika 1 (1954), 73-79

[39] G. Samorodnitsky and M. Taqqu *Stable non-Gaussian random processes,* Chapman & Hall/CRC, 1994, 640 pp.

[40] W. Schmidt, *Metrical theorems on fractional parts of sequences*, Trans. Amer. Math. Soc. **110** (1964) 493–518.

[41] C. L. Siegel, *A mean value theorem in geometry of numbers,* Ann. Math. **46** (1945) 340–347.

[42] I. Vinogradov *Limiting distribution of visits of several rotations to shrinking intervals,* preprint.