

CENTRAL LIMIT THEOREMS FOR SIMULTANEOUS DIOPHANTINE APPROXIMATIONS

DMITRY DOLGOPYAT, BASSAM FAYAD, AND ILYA VINOGRADOV

ABSTRACT. We study the distribution modulo 1 of the values taken on the integers of r linear forms in d variables with random coefficients. We obtain quenched and annealed central limit theorems for the number of simultaneous hits into shrinking targets of radii $n^{-\frac{r}{d}}$. By the Khintchine-Groshev theorem on Diophantine approximations, $\frac{r}{d}$ is the critical exponent for the infinite number of hits.

1. INTRODUCTION

1.1. Results. An important problem in Diophantine approximation is the study of the speed of approach to 0 of a possibly inhomogeneous linear form of several variables evaluated at integers points. Such a linear form is given by $l : \mathbb{T} \times \mathbb{T}^d \times \mathbb{Z}^d \rightarrow \mathbb{T}$

$$(1.1) \quad l_{x,\alpha}(k) = x + \sum_{i=1}^d k_i \alpha_i \pmod{1}$$

More generally, for $r \geq 1$, one can consider r linear forms $l_{x^j, \alpha^j}(k)$ for $j = 1 \dots r$ corresponding to $\mathbf{a} := (\alpha_i^j) \in \mathbb{T}^{d \times r}$ and $\mathbf{x} := (x^1, \dots, x^r) \in \mathbb{T}^r$, where each $\alpha^j, j = 1, \dots, r$, is a vector in \mathbb{T}^d .

Diophantine approximation theory classifies the matrices \mathbf{a} and vectors \mathbf{x} according to how “resonant” they are; i.e., how well the vector $(l_{x^j, \alpha^j}(k))_{j=1}^r$ approximates $\mathbf{0} := (0, \dots, 0) \in \mathbb{R}^r$ as k varies over a large ball in \mathbb{Z}^d . One can then fix a sequence of targets converging to 0, say intervals of radius r_n centered at 0 with $r_n \rightarrow 0$, and investigate the integers for which the target is *hit*, namely the integers k such that $l_{x^j, \alpha^j}^j(k) \in [-r_{|k|}, r_{|k|}]$ for every $j = 1, \dots, r$. An important class of targets is given by radii following a power law, $r_n = cn^{-\gamma}$ for some $\gamma, c > 0$ (see for example [20, 18, 32] or [5] for a nice discussion related to the Diophantine properties of linear forms).

Fix a norm $|\cdot|$ on \mathbb{R}^d , and let $\|\cdot\|$ denote the Euclidean norm on \mathbb{R}^r . For $c > 0$ and $\iota = 1$ or 2, we define sets

$$(1.2) \quad B_\iota(k, d, r, c) = [0, c|k|^{-\frac{d}{r}}]^r \subset \mathbb{R}^r$$

for $\iota = 1$ and

$$(1.3) \quad B_\iota(k, d, r, c) = \{a \in \mathbb{R}^r : \|a\| \leq c|k|^{-\frac{d}{r}}\},$$

for $\iota = 2$. We then introduce

$$(1.4) \quad V_{N,\iota}(\mathbf{a}, \mathbf{x}, c) = \#\{0 \leq |k| < N : (l_{x^j, \alpha^j}(k))_{j=1}^r \in B_\iota(k, d, r, c)\},$$

$$(1.5) \quad U_{N,\iota}(\mathbf{a}, c) = V_{N,\iota}(\mathbf{a}, \mathbf{0}, c).$$

A matrix $\mathbf{a} \in \mathbb{T}^{d \times r}$ is said to be *badly approximable* if for some $c > 0$, the sequence $U_{N,\iota}(\mathbf{a}, c)$ is bounded. By contrast, matrices \mathbf{a} for which $U_{N,\iota}^\varepsilon(\mathbf{a}, c)$ is unbounded for some $\varepsilon > 0$, where $U_{N,\iota}^\varepsilon(\mathbf{a}, c)$ is defined as $U_{N,\iota}$, but with radii $cn^{-\frac{d}{r}-\varepsilon}$ instead of $cn^{-\frac{d}{r}}$ are called *very well approximable* or VWA. The obvious direction of the Borel-Cantelli lemma implies that almost every $\mathbf{a} \in \mathbb{T}^{d \times r}$ is not very well approximable (cf. [7, Chap. VII]). The celebrated Khintchine-Groshev theorem on Diophantine approximation implies that badly approximable matrices are also of zero measure [19, 18, 13, 30, 6, 4]. Analogous definitions apply in the inhomogeneous case of $V_{N,\iota}(\mathbf{a}, \mathbf{x}, c)$, and similar results hold.

For targets given by a power law, the radii $cn^{-\frac{d}{r}}$ are thus the smallest ones to yield an infinite number of hits almost surely. A natural question is then to investigate statistics of these hits, which we call *resonances*. In the present paper we address in this context the behavior of the resonances on average over \mathbf{a} and \mathbf{x} , or on average over \mathbf{a} while \mathbf{x} is fixed at $\mathbf{0}$ or fixed at random. Let Vol denote the Euclidean volume and consider the “expected” number of hits

$$(1.6) \quad \widehat{V}_{N,\iota} = \text{Vol}(B_\iota(1, d, r, c)) \ln N$$

when \mathbf{x} and \mathbf{a} are uniformly distributed on corresponding tori. Let $\mathcal{N}(m, \sigma^2)$ denote the normal distribution with mean m and variance σ^2 .

THEOREM 1.1. *Suppose $(r, d) \neq (1, 1)$, and let \mathbf{a} be uniformly distributed on \mathbb{T}^{dr} . Then,*

$$(1.7) \quad \frac{U_{N,\iota}(\mathbf{a}, c) - \widehat{V}_{N,\iota}}{\sqrt{\ln N}}$$

converges in distribution to $\mathcal{N}(0, \sigma_i^2)$ as $N \rightarrow \infty$, where

$$(1.8) \quad \sigma_1^2 = 2c^r d \frac{\zeta(r+d-1)}{\zeta(r+d)} \text{Vol}(\mathcal{B}), \quad \sigma_2^2 = \frac{\pi^{r/2}}{\Gamma(\frac{r}{2} + 1)} \sigma_1^2.$$

where \mathcal{B} is the unit ball in $|\cdot|$ -norm.

REMARK 1.2. The restriction $(r, d) \neq (1, 1)$ above is necessary. In fact, it is shown in [27, 29] using continued fractions that in that case the Central Limit Theorem (CLT) still holds for U_N , but the correct normalization should be $\sqrt{\ln N \ln \ln N}$ rather than $\sqrt{\ln N}$.

THEOREM 1.3. *Let \mathbf{a} and \mathbf{x} be uniformly distributed in \mathbb{T}^{dr} . Then,*

$$(1.9) \quad \frac{V_{N,\iota}(\mathbf{a}, \mathbf{x}, c) - \widehat{V}_{N,\iota}}{\sqrt{\widehat{V}_{N,\iota}}}$$

converges in distribution to $\mathcal{N}(0, 1)$ as $N \rightarrow \infty$.

The preceding theorems give CLTs in the cases of \mathbf{x} fixed to be 0 or \mathbf{x} random. The CLT also holds for almost every fixed \mathbf{x} .

THEOREM 1.4. *Suppose $(r, d) \neq (1, 1)$. For almost every \mathbf{x} , if \mathbf{a} is uniformly distributed in \mathbb{T}^{dr} , then $\frac{V_{N,\iota}(\mathbf{a}, \mathbf{x}, c) - \widehat{V}_{N,\iota}}{\sqrt{\widehat{V}_{N,\iota}}}$ converges in distribution to a normal random variable with zero mean and variance one.*

1.2. Plan of the paper. Using a by now standard approach of Dani correspondence (cf. [9, 25, 26, 1, 2, 3, 23]) we deduce our results about Diophantine approximations from appropriate limit theorems for homogeneous flows. Namely we need to prove a CLT for Siegel transforms of piecewise smooth functions; these limit theorems are formulated in Section 2. The reduction of the theorems of Section 1 to those of Section 2 is given in Section 3. The CLTs in the space of lattices are in turn deduced from an abstract Central Limit Theorem (Theorem 4.2) for weakly dependent random variables which is formulated and proven in Section 4. In order to verify the conditions of Theorem 4.2 for the problem at hand we need several results about regularity of Siegel transforms which are formulated in Section 5 and proven in the appendix. In Section 6 we deduce our Central Limit Theorems for homogeneous flows from the abstract Theorem 4.2. Section 8 contains the proof of the formula (1.8) on the variances. Section 7 discusses some applications of Theorem 4.2 beyond the subject of Diophantine approximation.

Acknowledgements. D. D. was partially supported by the NSF. I.V. appreciates the support of Fondation Sciences Mathématiques de Paris during his stay in Paris.

2. CENTRAL LIMIT THEOREMS ON THE SPACE OF LATTICES

2.1. Notation. We let $\mathbb{G} = SL_{d+r}(\mathbb{R})$, $\tilde{\mathbb{G}} = SL_{d+r}(\mathbb{R}) \ltimes \mathbb{R}^{d+r}$. The multiplication rule in $\tilde{\mathbb{G}}$ takes form $(A, a)(B, b) = (AB, a + Ab)$. We regard \mathbb{G} as a subgroup of $\tilde{\mathbb{G}}$ consisting of elements of the form $(A, 0)$. We let \mathbb{L} be the abelian subgroup of \mathbb{G} consisting of matrices $\Lambda_{\mathbf{a}}$, and $\tilde{\mathbb{L}}$ be the abelian subgroup of $\tilde{\mathbb{G}}$ consisting of matrices $(\Lambda_{\mathbf{a}}, (0, y))$ where 0 is an origin in \mathbb{R}^d , y is an r -dimensional vector, \mathbf{a} is a $d \times r$ matrix and

$$\Lambda_{\mathbf{a}} = \begin{pmatrix} \text{Id}_d & 0 \\ \mathbf{a} & \text{Id}_r \end{pmatrix}.$$

Let \mathcal{M} be the space of $d + r$ dimensional unimodular lattices and $\tilde{\mathcal{M}}$ be the space of $d + r$ dimensional unimodular affine lattices. We identify \mathcal{M} and $\tilde{\mathcal{M}}$ respectively with $\mathbb{G}/SL_{d+r}(\mathbb{Z})$ and $\tilde{\mathbb{G}}/SL_{d+r}(\mathbb{Z}) \ltimes \mathbb{Z}^{d+r}$.

We will need spaces $C^{\mathbf{s}, \mathbf{r}}(\mathbb{R}^p)$, $C^{\mathbf{s}, \mathbf{r}}(\mathcal{M})$, and $C^{\mathbf{s}, \mathbf{r}}(\tilde{\mathcal{M}})$ of functions which can be well approximated by smooth functions, given $\mathbf{s}, \mathbf{r} \geq 0$. Recall first that the space $C^{\mathbf{s}}(\mathbb{R}^p)$ consists of functions $f: \mathbb{R}^p \rightarrow \mathbb{R}$ whose derivatives up to order \mathbf{s} are bounded. To define spaces $C^{\mathbf{s}}(\mathcal{M})$ and $C^{\mathbf{s}}(\tilde{\mathcal{M}})$, fix bases for $\text{Lie}(\mathbb{G})$ and $\text{Lie}(\tilde{\mathbb{G}})$; then, $C^{\mathbf{s}}(\mathcal{M})$ and $C^{\mathbf{s}}(\tilde{\mathcal{M}})$ consist of functions whose derivatives corresponding to monomials of order up to \mathbf{s} in the basis elements are bounded (see Appendix for precise definitions). Now we define $C^{\mathbf{s}, \mathbf{r}}$ -norm on a space equipped with a $C^{\mathbf{s}}$ -norm and an L^1 -norm by

$$(2.1) \quad \|f\|_{C^{\mathbf{s}, \mathbf{r}}} = \sup_{0 < \varepsilon < 1} \sup_{\substack{f^- \leq f \leq f^+ \\ \|f^+ - f^-\|_{L^1} < \varepsilon}} \varepsilon^{\mathbf{r}} (\|f^+\|_{C^{\mathbf{s}}} + \|f^-\|_{C^{\mathbf{s}}}).$$

(in the examples considered above $(\mathbb{R}^{d+r}, \mathcal{M}$ and $\tilde{\mathcal{M}})$ the L^1 -norm is taken with respect to the Haar measure).

Some properties of these spaces are discussed in the Appendix.

Given a function f on \mathbb{R}^{r+d} we consider its Siegel transforms $\mathcal{S}(f) : \mathcal{M} \rightarrow \mathbb{R}$ and $\tilde{\mathcal{S}}(f) : \tilde{\mathcal{M}} \rightarrow \mathbb{R}$ defined by

$$(2.2) \quad \mathcal{S}(f)(\mathcal{L}) = \sum_{e \in \mathcal{L}} f(e), \quad \tilde{\mathcal{S}}(f)(\tilde{\mathcal{L}}) = \sum_{e \in \tilde{\mathcal{L}}} f(e).$$

We emphasize that Siegel transforms of smooth functions are never bounded but the growth of their norms at infinity is well understood, see Subsection 5.3.

2.2. Results for the space of lattices. In this section we present general Central Limit Theorems for Siegel transforms. The reduction of Theorems 1.1, 1.3, 1.4 to the results stated here will be performed in Section 3.

Let $f \in C^{s,r}(\mathbb{R}^{d+r})$ be a non-negative function supported on a compact set which does not contain 0. (The assumption that f vanishes at zero is needed to simplify the formulas for the moments of its Siegel transform. See Proposition 5.1.) Denote $\bar{f} = \iint_{\mathbb{R}^{d+r}} f(x, y) dx dy$.

Given positive numbers K and α we say that a subset $S \subset \mathcal{M}$ is (K, α) -regular if S is a union of codimension 1 submanifolds and there is a one-parameter subgroup $h_u \subset \mathbb{L}$ such that

$$\mu(\mathcal{L} : h_{[-\varepsilon, \varepsilon]} \mathcal{L} \cap S \neq \emptyset) \leq K \varepsilon^\alpha$$

where μ denotes the Haar measure on \mathcal{M} . We say that a function $\rho : \mathcal{M} \rightarrow \mathbb{R}$ is (K, α) -regular if $\text{supp}(\rho)$ has a (K, α) -regular boundary and the restriction of ρ on $\text{supp}(\rho)$ belongs to C^α with

$$\|\rho\|_{C^\alpha(\text{supp}(\rho))} \leq K.$$

(K, α) -regular functions on $\tilde{\mathcal{M}}$ are defined similarly.

Let \mathbb{A} be subgroup of \mathbb{G} consisting of diagonal matrices. We use the notation da for Haar measure on \mathbb{A} . We say that ρ is K -centrally smoothable if there is a positive function ϕ supported in a unit neighborhood of the identity in \mathbb{A} such that $\int_{\mathbb{A}} \phi(a) da = 1$ and

$$\rho_\phi(\mathcal{L}) := \int_{\mathbb{A}} \rho(a\mathcal{L}) \phi(a) da$$

has L^∞ norm less than K . We say that a function ρ on $\tilde{\mathcal{M}}$ is K -centrally smoothable if

$$\rho^*(\mathcal{L}) = \sup_{\mathbf{x}} \rho(\mathcal{L}, \mathbf{x})$$

is a K -centrally smoothable function on \mathcal{M} . As before, we write $\mathcal{N}(m, \sigma^2)$ for the normal distribution with m and variance σ^2 and “ \implies ” stands for convergence in distribution.

For $p \in \mathbb{N}$ and $t \in \mathbb{R}$, we denote the $p \times p$ diagonal matrix

$$(2.3) \quad D_p(t) = \begin{pmatrix} 2^t & & \\ & \ddots & \\ & & 2^t \end{pmatrix}$$

and let

$$(2.4) \quad g = \begin{pmatrix} D_d(-1) & 0 \\ 0 & D_r(\frac{d}{r}) \end{pmatrix}.$$

THEOREM 2.1. *Suppose that $(r, d) \neq (1, 1)$.*

(a) *There is a constant σ such that if \mathcal{L} is uniformly distributed on \mathcal{M} then*

$$\frac{\sum_{n=0}^{N-1} \mathcal{S}(f)(g^n \mathcal{L}) - N\bar{f}}{\sigma\sqrt{N}} \implies \mathcal{N}(0, 1)$$

as $N \rightarrow \infty$.

(b) *Fix constants $C, u, \alpha, \varepsilon$ with $\varepsilon < \frac{1}{2}$. Suppose that \mathcal{L} is distributed according to a density ρ_N which is (CN^u, α) -regular and C -centrally smoothable. Then*

$$\frac{\sum_{n=N^\varepsilon}^{N-1} \mathcal{S}(f)(g^n \mathcal{L}) - N\bar{f}}{\sigma\sqrt{N}} \implies \mathcal{N}(0, 1)$$

as $N \rightarrow \infty$.

THEOREM 2.2. (a) *Let $\tilde{\mathcal{L}}$ be uniformly distributed on $\tilde{\mathcal{M}}$. Then there is a constant $\tilde{\sigma}$ such that*

$$\frac{\sum_{n=0}^{N-1} \tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) - N\bar{f}}{\tilde{\sigma}\sqrt{N}} \implies \mathcal{N}(0, 1)$$

as $N \rightarrow \infty$.

(b) *Fix constants $C, u, \alpha, \varepsilon$ with $\varepsilon < \frac{1}{2}$. Suppose that $\tilde{\mathcal{L}}$ is distributed according to a density ρ_N which is (CN^u, α) -regular and C -centrally smoothable. Then*

$$\frac{\sum_{n=N^\varepsilon}^{N-1} \tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) - N\bar{f}}{\tilde{\sigma}\sqrt{N}} \implies \mathcal{N}(0, 1)$$

as $N \rightarrow \infty$.

Let $\tilde{\mathcal{D}}$ be an unstable box, that is

$$\tilde{\mathcal{D}} = \{(\Lambda_{\mathbf{a}}, (0, \mathbf{x}))\tilde{\mathcal{L}}_0\}_{(\mathbf{a}, \mathbf{x}) \in \mathfrak{R}_1 \times \mathfrak{R}_2}$$

where $\tilde{\mathcal{L}}_0$ is a fixed affine lattice and \mathfrak{R}_1 and \mathfrak{R}_2 are boxes in $\mathbb{R}^{d \times r}$ and \mathbb{R}^r respectively. Consider a partition Π of $\tilde{\mathcal{D}}$ into \mathbb{L} -boxes. Thus elements of Π are of the form

$$\{(\Lambda_{\mathbf{a}}, (0, \mathbf{x}_0))\tilde{\mathcal{L}}_0\}_{\mathbf{a} \in \mathfrak{R}_1}$$

for some fixed \mathbf{x}_0 .

THEOREM 2.3. *Suppose that $(r, d) \neq (1, 1)$. Then for each unstable cube $\tilde{\mathcal{D}}$ and for almost every $\tilde{\mathcal{L}} \in \tilde{\mathcal{D}}$, if $\tilde{\mathcal{L}}$ is uniformly distributed in $\Pi(\tilde{\mathcal{L}})$, then*

$$\frac{\sum_{n=0}^{N-1} \tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) - N\bar{f}}{\tilde{\sigma}\sqrt{N}} \implies \mathcal{N}(0, 1)$$

as $N \rightarrow \infty$.

The explicit calculation of σ and $\tilde{\sigma}$ when f is an indicator functions (the case needed for Theorems 1.1-1.4) will be given in Section 8.

REMARK 2.4. Central Limit Theorems for partially hyperbolic translations on homogeneous spaces are proven in [10] (for bounded observables) and in [24] (for L^4 observables). (See also [33, 28] for important special cases). It seems possible to prove Theorem 2.1 for sufficiently large values of $d + r$ by verifying the conditions of [24].

Instead, we prefer to present in the next section an abstract result which will later be applied to derive Theorems 2.1, 2.2, and 2.3. We chose this approach for three reasons. First, this will make the paper self contained. Second, we replace the L^4 assumption of [24] by a weaker $L^{2+\delta}$ assumption which is important for small $d+r$. Third, our approach allows to give a unified proof for Theorems 2.1, 2.2, and 2.3.

3. DIAGONAL ACTIONS ON THE SPACE OF LATTICES AND DIOPHANTINE APPROXIMATIONS

In this section we reduce Theorem 1.1, 1.3, and 1.4 to Theorems 2.1–2.3.

To fix our notation we consider $U_{N,1}$ and $V_{N,1}$, the analysis of $U_{N,2}$ and $V_{N,2}$ being similar. We also drop the extra subscript and write $U_{N,1}$ and $V_{N,1}$ as U_N and V_N , respectively, until the end of this section.

In Subsections 3.1 and 3.2 we explain how to reduce Theorem 1.1 to Theorem 2.1. The reductions of Theorems 1.3 and 1.4 to Theorems 2.2 and 2.3 require only minor modifications which will be detailed in Section 3.3.

3.1. Dani correspondence. In this subsection we use the Dani correspondence principle to reduce the problem to a CLT for the action of diagonal elements on the space of lattices of the form $\Lambda_{\mathbf{a}}$ where \mathbf{a} is random.

Let \mathbf{a} be the matrix with rows $\alpha_i \in \mathbb{R}^d$, $i = 1, \dots, r$. Let

$$(3.1) \quad \Lambda_{\mathbf{a}} = \begin{pmatrix} \text{Id}_d & 0 \\ \mathbf{a} & \text{Id}_r \end{pmatrix}.$$

Let ϕ be the indicator of the set

$$(3.2) \quad E_c := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^r \mid |x| \in [1, 2], |x|^{d/r} y_j \in [0, c] \text{ for } j = 1, \dots, r\}$$

and consider its Siegel transform $\Phi = \mathcal{S}(\phi)$.

Now $n = (n_1, \dots, n_d)$ with $|n| \leq N$ contributes to $U_N(\mathbf{a}, c)$ (from (1.5)) precisely when there exists $(m_1, \dots, m_r) \in \mathbb{Z}^r$ such that

$$(3.3) \quad \left(\sum_{i=1}^d n_i \alpha_{1,i} + m_1, \dots, \sum_{i=1}^d n_i \alpha_{r,i} + m_r \right) \in B(n, d, r, c).$$

Clearly such a vector (m_1, \dots, m_r) is unique. It is elementary to see that (3.3) holds if and only if

$$(3.4) \quad g^t \Lambda_{\mathbf{a}}(n_1, \dots, n_d, m_1, \dots, m_r) \in E_c.$$

for some integer $t \leq 2^{\lceil \log_2 N \rceil}$ where g is the diagonal matrix defined in (2.4).

Below we will use the notion $A_N \ll B_N$ to mean that the ratio $\frac{|A_N|}{|B_N|}$ is bounded. From (3.3)–(3.4) we obtain

LEMMA 3.1. *For each $\varepsilon > 0$*

$$U_N(\mathbf{a}, c) = \sum_{t=\lceil (\log_2 N)^\varepsilon \rceil}^{\lceil \log_2 N \rceil} \Phi \circ g^t(\Lambda_{\mathbf{a}} \mathbb{Z}^{d+r}) + R_N,$$

with

$$\|R_N\|_{L^1(\mathbf{a})} \ll (\log_2 N)^\varepsilon$$

and $L^1(\mathbf{a})$ denoting the L^1 -norm with respect to the Lebesgue measure on the unit cube in $\mathbb{R}^{d \times r}$.

Proof. From (3.3)–(3.4), it follows that

$$R_N \leq \# \left\{ \begin{array}{l} 2^{\lfloor \log_2 N \rfloor} \leq |n| < N, \\ \text{or } |n| \leq 2^{(\log_2 N)^\varepsilon} \end{array} : (l_{0,\alpha^1}(n), \dots, l_{0,\alpha^r}(n)) \in B(n, d, r, c) \right\}.$$

Note that for a fixed $n \in \mathbb{Z}^d$ and $j \in 1, \dots, r$, the form $\{l_j(n, 0, \alpha_j)\}$ is uniformly distributed on the circle. Hence

$$\text{Leb} \{ \mathbf{a} \in \mathbb{T}^{rd} : (l_{0,\alpha^1}(n), \dots, l_{0,\alpha^r}(n)) \in B(n, d, r, c) \} \ll \frac{1}{|n|^d}$$

and so

$$\|R_N\|_{L^1(\mathbf{a})} \ll \sum_{2^{\lfloor \log_2 N \rfloor} \leq |n| \leq N} \frac{1}{|n|^d} + \sum_{|n| \leq 2^{((\log_2 N)^\varepsilon)}} \frac{1}{|n|^d} \ll (\log_2 N)^\varepsilon. \quad \square$$

Hence, to prove Theorem 1.1 we can replace $U_N(\mathbf{a}, c)$ by

$$\sum_{t=[(\log_2 N)^\varepsilon]}^{\lfloor \log_2 N \rfloor} \Phi \circ g^t(\Lambda_{\mathbf{a}} \mathbb{Z}^{d+r}).$$

3.2. Changing the measure. Note that the action of g^t on the space of lattices \mathcal{M} is partially hyperbolic and its unstable manifolds are orbits of the action $\Lambda_{\mathbf{a}}$ with $\mathbf{a} \in M(d, r)$ ranging in the set of $d \times r$ matrices. This will allow us to reduce the proof of CLTs to CLTs for the diagonal action on the space of lattices. A similar reduction is possible for the g^t -action on the space of affine lattices since in that case the unstable manifolds for the action of g^t are given by $(\Lambda_{\mathbf{a}}, (0, \mathbf{x}))$, with $\mathbf{a} \in M(d, r)$, $\mathbf{x} \in \mathbb{R}^r$.

Let $\eta = 1/k^{10d}$ where $k = \lfloor \log_2 N \rfloor$. For $i = 1, \dots, r+d-1$ let t_i be independent uniformly distributed in $[-\eta, \eta]$. Also introduce a random matrix $\mathbf{b} \in M(r, d)$ where all the entries of \mathbf{b} are independent uniformly distributed in $[-1, 1]$. Let

$$D_{\mathbf{t}} = \text{diag} \left(1 + t_1, \dots, 1 + t_{d+r-1}, \left(\prod_{l=1}^{r+d-1} (1 + t_l) \right)^{-1} \right)$$

and

$$\bar{\Lambda}_{\mathbf{b}} = \begin{pmatrix} \text{Id}_d & \mathbf{b} \\ 0 & \text{Id}_r \end{pmatrix}.$$

Let $\tilde{\Lambda}(\mathbf{a}, \mathbf{b}, \mathbf{t}) = D_{\mathbf{t}} \bar{\Lambda}_{\mathbf{b}} \Lambda_{\mathbf{a}}$.

It is clear that if \mathbf{a} is distributed uniformly in a unit cube, then $\tilde{\Lambda}(\mathbf{a}, \mathbf{b}, \mathbf{t})$ is distributed according to a $(Ck^{10d}, 1)$ -regular and C -centrally smoothable density for some constant C . Note that

$$g^t \tilde{\Lambda}(\mathbf{a}, \mathbf{b}, \mathbf{t}) = D_{\mathbf{t}} \bar{\Lambda}_{\mathbf{b}_t} g^t \Lambda_{\mathbf{a}} \text{ where } |\mathbf{b}_t| \leq e^{-t}.$$

Observe also that for $h \in \mathbb{G}$ and $E \subset \mathbb{R}^{d+r}$, we have

$$\mathcal{S}(\mathbb{1}_E)(h\mathcal{L}) = \mathcal{S}(\mathbb{1}_{h^{-1}E})(\mathcal{L}),$$

and hence if ε is sufficiently small then for $t \geq k^\varepsilon$ we have

$$\left| \mathcal{S}(\mathbb{1}_{E_c})(g^t \tilde{\Lambda}(\mathbf{a}, \mathbf{b}, \mathbf{t}) \mathbb{Z}^{d+r}) - \mathcal{S}(\mathbb{1}_{E_c})(g^t \Lambda_{\mathbf{a}} \mathbb{Z}^{d+r}) \right| \leq \mathcal{S}(\mathbb{1}_{\tilde{E}})(g^t \Lambda_{\mathbf{a}} \mathbb{Z}^{d+r})$$

where \tilde{E} denotes a Ck^{-10d} neighborhood of the boundary of E_c . Now the same argument as in Lemma 3.1 gives

LEMMA 3.2. *For each $\varepsilon > 0$,*

$$\sum_{t=[(\log_2 N)^\varepsilon]}^{[\log_2 N]} \Phi \circ g^t(\Lambda_{\mathbf{a}} \mathbb{Z}^{d+r}) = \sum_{t=[(\log_2 N)^\varepsilon]}^{[\log_2 N]} \Phi \circ g^t(\tilde{\Lambda}(\mathbf{a}, \mathbf{b}, \mathbf{t}) \mathbb{Z}^{d+r}) + \tilde{R}_N$$

with $\|\tilde{R}_N\|_{L^1(\mathbf{a}, \mathbf{b}, \mathbf{t})} \ll 1$.

Now Theorem 1.1 follows from Theorem 2.1 except for the formula for σ which is derived in Section 8.

3.3. Inhomogeneous case. The reduction of Theorems 1.3 and 1.4 to Theorems 2.2 and 2.3 requires only small changes compared to the preceding section. To wit, Lemmas 3.1 and 3.2 take the following form.

LEMMA 3.3. (a) *For each $\varepsilon > 0$,*

$$V_N(\mathbf{a}, \mathbf{x}, c) = \sum_{t=[(\log_2 N)^\varepsilon]}^{[\log_2 N]} \tilde{\mathcal{S}}(\mathbb{1}_{E_c})(g^t(\Lambda_{\mathbf{a}} \mathbb{Z}^{d+r} + (0, \mathbf{x}))) + R_N(\mathbf{a}, \mathbf{x}, c)$$

where $\|R_N\|_{L^1(\mathbf{a})} \ll (\log_2 N)^\varepsilon$.

(b) *Let \mathbf{b} and \mathbf{t} have the same distribution as in Subsection 3.2 and \mathbf{y} be uniformly distributed in $[-1, 1]^d$. Then for each $\varepsilon > 0$*

$$(3.5) \quad \begin{aligned} & \sum_{t=[(\log_2 N)^\varepsilon]}^{[\log_2 N]} \tilde{\mathcal{S}}(\mathbb{1}_{E_c})(g^t(\Lambda_{\mathbf{a}} \mathbb{Z}^{d+r} + (0, \mathbf{x}))) \\ &= \sum_{t=[(\log_2 N)^\varepsilon]}^{[\log_2 N]} \tilde{\mathcal{S}}(\mathbb{1}_{E_c})(g^t(\tilde{\Lambda}(\mathbf{a}, \mathbf{b}, \mathbf{t}) \mathbb{Z}^{d+r} + (\mathbf{y}, \mathbf{x}))) + \tilde{R}_N(\mathbf{a}, \mathbf{b}, \mathbf{t}, \mathbf{x}, \mathbf{y}, c) \end{aligned}$$

with $\|\tilde{R}_N\|_{L^1(\mathbf{a}, \mathbf{b}, \mathbf{t}, \mathbf{y})} \ll 1$.

Note that in part (a) the error has small $L^1(\mathbf{a})$ norm for each fixed \mathbf{x} . This follows from the fact that for each \mathbf{x} and k , $\mathbf{a}k + \mathbf{x}$ is uniformly distributed on \mathbb{T}^r . This is useful in the proof of Theorem 1.4 since we want to have a control for each (or at least, most) \mathbf{x} . We also note that part (b) is only needed for Theorem 1.3 since in Theorem 2.3 we start with initial conditions supported on a positive codimension submanifold of $\tilde{\mathcal{M}}$. (One of the steps in the proof of Theorem 2.3 consists of fattening the support of the initial measure, see Subsection 6.3, however Lemma 3.3(b) is not needed at the reduction stage).

4. AN ABSTRACT CLT THEOREM

In this section, we present an abstract Central Limit Theorem for weakly dependent random variables that is well adapted for variables coming from deterministic dynamical systems. It is well adapted for mixing flows on non compact manifolds for this that it allows the variables to be unbounded and because it takes into account the existence of small exceptional "bad sets" on which the variables are not controlled. In Section 6, we will use the abstract CLT Theorem 4.2 to prove Theorems 2.1, 2.2, 2.3. There, we will take the variables ξ_l to be $\xi_l(\mathcal{L}) = \Phi(g^l \mathcal{L})$, $\Phi = \mathcal{S}(f)$ (or $\tilde{\mathcal{S}}(f)$), where $f \in C^{s,r}(\mathbb{R}^{d+r})$ is a positive function supported on a compact set which does not contain 0. The functions Φ are not bounded but, since we excluded the cases of linear lattices with $r = d = 1$, they are in L^s for $s > 2$. The latter will be formulated in Section 5 on the regularity of Siegel transforms.

4.1. Bounded random variables. Before we state our CLT theorem for variables in L^s , $s > 2$, we give a simplified version of it in the case of bounded variables.

In what follows $C, u > 0, \theta \in (0, 1)$ are fixed constants. Let ξ_n be a sequence of random variables satisfying the following conditions. Write $\hat{\xi}_n = \xi_n - \mathbb{E}(\xi_n)$ for the corresponding centered random variable.

(H1) There exists filtration $\{\mathcal{F}_l\}_{l \geq 0}$ such that for every l, k there exists a bounded \mathcal{F}_{l+k} -measurable random variable $\xi_{l,l+k}$ with $\mathbb{E}\xi_{l,l+k} = \mathbb{E}\xi_l$ such that

$$\mathbb{P}(|\xi_l - \xi_{l,l+k}| \geq \theta^k) \leq C(l+1)^u \theta^k.$$

(H2) For l, k , there exists $G_{l,k}$ such that $\mathbb{P}(G_{l,k}^c) \leq C\theta^k$ and for $\omega \in G_{k,l}$

$$|\mathbb{E}(\hat{\xi}_{l+k} | \mathcal{F}_l)(\omega)| \leq C(l+1)^u \theta^k.$$

(H3) For $\omega \in G_{k,l}$ and $k' \geq k$

$$\left| \mathbb{E}(\hat{\xi}_{l+k} \hat{\xi}_{l+k'} | \mathcal{F}_l)(\omega) - b_{k'-k} \right| \leq C(l+1)^u e^{u(k'-k)} \theta^k.$$

THEOREM 4.1. *If ξ_l is a bounded sequence satisfying (H1)–(H3) then*

$$\frac{\sum_{l=0}^{n-1} (\xi_l - \mathbb{E}(\xi_l))}{\sqrt{n}}$$

converges as $n \rightarrow \infty$ to a normal distribution with zero mean and variance

$$\sigma^2 = b_0 + 2 \sum_{k=1}^{\infty} b_k.$$

The proof of Theorem 4.1 follows from that of the more general statement that we now formulate and prove.

4.2. The general statement. In what follows $C, u > 0, \theta \in (0, 1)$, and $s > 2$ are fixed constants. Let ξ_n be a sequence of random variables satisfying the following conditions. Write $\hat{\xi}_n = \xi_n - \mathbb{E}(\xi_n)$ for the corresponding centered random variable.

(H1) Given any K , there is a sequences ξ_n^K of random variables such that

(H1a) $|\xi_n^K| \leq K$ almost surely;

$$(H1b) \quad \mathbb{E}(|\xi_n^K - \xi_n|) = O(K^{1-s}), \quad \mathbb{E}((\xi_n^K - \xi_n)^2) = O(K^{2-s}).$$

(H2) There exists a filtration $\mathcal{F}_l = \mathcal{F}_{l,n}$ defined for $0 \leq l < n$ such that for every pair of nonnegative numbers l, k with $l + k < n$, there exists a variable $\xi_{l,l+k}^K$ that is \mathcal{F}_{l+k} -measurable, $|\xi_{l,l+k}^K| \leq K$ almost surely, $\mathbb{E}\xi_{l,l+k}^K = \mathbb{E}\xi_l^K$, and

$$\mathbb{P}(|\xi_l^K - \xi_{l,l+k}^K| \geq \theta^k) \leq CK^u(l+1)^u\theta^k.$$

(H3) For l, k , there exists an event $G_{k,l}$ such that $\mathbb{P}(G_{k,l}^c) \leq C\theta^k$ and for $\omega \in G_{k,l}$

$$|\mathbb{E}(\hat{\xi}_{l+k}^K | \mathcal{F}_l)(\omega)| \leq C(l+1)^u K^u \theta^k.$$

(H4) There exists a numerical sequence $b_{K,k}$ for $k \geq 0$ such that for $\omega \in G_{k,l}$ and $k' \geq k$,

$$\left| \mathbb{E}(\hat{\xi}_{l+k}^K \hat{\xi}_{l+k'}^K | \mathcal{F}_l)(\omega) - b_{K,k'-k} \right| \leq C(l+1)^u K^u e^{u(k'-k)} \theta^k.$$

In the following Theorem, part (a) will be sufficient to prove results as in Theorem 2.1 (a), where the distribution of the lattices is given by the Haar measure, while part (b) is tailored to adapt to the case of localized initial conditions such as in Theorem 2.1 (b).

THEOREM 4.2. (a) *Under conditions (H1)-(H4) if ω is distributed according to \mathbb{P} then*

$$(4.1) \quad \frac{\sum_{l=0}^{n-1} \hat{\xi}_l}{\sqrt{n}} \implies \mathcal{N}(0, \sigma^2)$$

with

$$(4.2) \quad \sigma^2 = \sigma_0 + 2 \sum_{j=1}^{\infty} \sigma_j, \quad \sigma_j = \lim_{K \rightarrow \infty} b_{K,j}.$$

(b) *Suppose conditions (H1)-(H4) are satisfied. Fix $\varepsilon > 0$ such that $\frac{1+\varepsilon}{s} + \varepsilon < \frac{1}{2}$ and set $K_n = n^{\frac{1+\varepsilon}{s}}$. Suppose that ω is distributed according to a measure \mathbb{P}_n which has a density $\rho_n = \frac{d\mathbb{P}_n}{d\mathbb{P}}$ satisfying*

$$(D1) \quad \rho_n \leq Cn^u;$$

(D2) *for each k there is an \mathcal{F}_k -measurable density $\rho_{n,k}$ such that*

$$\mathbb{P}(|\rho_n - \rho_{n,k}| \geq \theta^k) \leq Cn^u \theta^k;$$

(D3) *For each $n^\varepsilon \leq l \leq n$,*

$$\mathbb{E}_n(|\xi_l - \xi_l^{K_n}|) \leq CK_n^{1-s}$$

where \mathbb{E}_n denotes the expectation with respect to \mathbb{P}_n , that is, $\mathbb{E}_n(\eta) = \mathbb{E}(\eta\rho_n)$.

Then,

$$(4.3) \quad \frac{\sum_{l=n^\varepsilon}^{n-1} \hat{\xi}_l}{\sqrt{n}} \implies \mathcal{N}(0, \sigma^2).$$

4.3. Limiting variance. Here we show that the normalized variance converges.

LEMMA 4.3. *Under conditions (H1)–(H4) we have that*

$$(4.4) \quad \lim_{n \rightarrow \infty} \frac{\mathbb{E}((\sum_{l=0}^{n-1} \hat{\xi}_l)^2)}{n} = \sigma^2$$

with σ as in (4.2).

Proof. First we record a property of cross-terms in the sum that lets us pass from ξ_n to the truncated sequence. We have

$$(4.5) \quad D := \mathbb{E}(\hat{\xi}_{m+j} \hat{\xi}_m) - \mathbb{E}(\hat{\xi}_{m+j}^K \hat{\xi}_m^K) = O(K^{2-s}).$$

Indeed, we have

$$\begin{aligned} D &= \mathbb{E}(\xi_{m+j} \xi_m) - \mathbb{E}(\xi_{m+j}^K \xi_m^K) \\ &\quad - \mathbb{E}(\xi_{m+j}) \mathbb{E}(\xi_m) + \mathbb{E}(\xi_{m+j}^K) \mathbb{E}(\xi_m^K) \\ &= \mathbb{E}(\xi_{m+j}^K (\xi_m - \xi_m^K)) + \mathbb{E}(\xi_m^K (\xi_{m+j} - \xi_{m+j}^K)) + \mathbb{E}((\xi_m - \xi_m^K)(\xi_{m+j} - \xi_{m+j}^K)) \\ &\quad - \mathbb{E}(\xi_{m+j}^K) \mathbb{E}(\xi_m - \xi_m^K) - \mathbb{E}(\xi_m^K) \mathbb{E}(\xi_{m+j} - \xi_{m+j}^K) - \mathbb{E}(\xi_m - \xi_m^K) \mathbb{E}(\xi_{m+j} - \xi_{m+j}^K). \end{aligned}$$

Now applying the Cauchy-Schwarz inequality followed by (H1a) and (H1b), we arrive at the bound (4.5) since $s > 2$.

Next, applying (H1), (H3), and (H4) with $l = 0$ gives

$$(4.6) \quad \mathbb{E}(\hat{\xi}_{m+j}^K \hat{\xi}_m^K) = \mathbb{E}(\hat{\xi}_{m+j}^K \hat{\xi}_m^K \mathbb{1}_{G_{m,0}}) + \mathbb{E}(\hat{\xi}_{m+j}^K \hat{\xi}_m^K \mathbb{1}_{G_{m,0}^c})$$

$$(4.7) \quad = b_{K,j} + O(K^u e^{uj} \theta^m) + O(\|\hat{\xi}_{m+j}^K\|_{L^\infty} \|\hat{\xi}_m^K\|_{L^\infty} \mathbb{P}(G_{m,0}^c))$$

$$(4.8) \quad = b_{K,j} + O(K^u e^{uj} \theta^m) + O(K^2 \theta^m).$$

Combining (4.5) and (4.8) we get

$$(4.9) \quad b_{K,j} = \mathbb{E}(\hat{\xi}_{m+j} \hat{\xi}_m) + O(K^{2-s}) + O(K^u e^{uj} \theta^m) + O(K^2 \theta^m).$$

Take a small number $\tilde{\varepsilon} > 0$ and assume that $j \leq \tilde{\varepsilon}K$, $K < \tilde{K} < 2K$. Taking $m = K/2$ we see that

$$b_{K,j} - b_{\tilde{K},j} = O(K^{2-s}).$$

Therefore for each j , the following limits exist,

$$\sigma_j := \lim_{K \rightarrow \infty} b_{K,j}.$$

and moreover

$$(4.10) \quad b_{K,j} = \sigma_j + O(K^{2-s}).$$

We give now the key estimates on the cross-terms in the variance according to different regimes of l and j .

SUBLEMMA 4.4. *We have the following estimates*

(a) *There exists $A > 0$ such that for all $(l, l+j) \in [0, n]^2$*

$$|\mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j})| \leq A$$

(b) Let $\eta = \min(\frac{s-2}{4u}, \frac{1}{10})$. If $j \geq \ln n^2$ and $l \leq n$, then

$$\mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j}) = O(\theta^{\eta j})$$

(c) If $j \leq \ln n^2$ and $l \geq \ln n^4$, then

$$\mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j}) - \sigma_j = O(n^{-2})$$

(d) There exists $C > 0$ and $\nu > 0$ such that for $j \geq \ln n^2$,

$$|\sigma_j| \leq C\theta^{\nu j}$$

Before we prove the sublemma, we show how it implies Lemma 4.4. We have

$$\begin{aligned} \frac{\mathbb{E}((\sum_{l=0}^{n-1} \hat{\xi}_l)^2)}{n} &= \frac{1}{n} \left[\sum_{l=0}^{n-1} \mathbb{E}(\hat{\xi}_l^2) + 2 \sum_{l \in [0, n], j \in [0, n-l]} \mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j}) \right] \\ &= \frac{1}{n} \left[\sum_{l=\ln n^4}^{n-1} \mathbb{E}(\hat{\xi}_l^2) + 2 \sum_{l \in [\ln n^4, n], j \in [0, \ln n^2]} \mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j}) \right] + o(1) \\ &= \sigma_0 + 2 \sum_{j=0}^{\ln n^2} \sigma_j + o(1) \\ &= \sigma_0 + 2 \sum_{j=0}^{\infty} \sigma_j + o(1) \end{aligned}$$

where the second equality follows from parts (a) and (b) of Sublemma (4.4), the third equality follows from part (c), and the fourth equality from part (d). \square

Proof of Sublemma 4.4. Part (a) follows from (4.5) and (H1a).

To prove part (b) assume $j \geq \ln n^2$ and let $K = \theta^{-\frac{j}{2u}}$. We claim that

$$|\mathbb{E}(\hat{\xi}_l^K \hat{\xi}_{l+j}^K)| \leq C\theta^{j/5}.$$

This is seen by writing

$$\mathbb{E}(\hat{\xi}_l^K \hat{\xi}_{l+j}^K) = \mathbb{E}(\hat{\xi}_{l,l+j/2}^K \mathbb{E}(\hat{\xi}_{l+j}^K | \mathcal{F}_{l+j/2})) + \mathbb{E}(\hat{\xi}_{l+j}^K (\hat{\xi}_{l,l+j/2}^K - \hat{\xi}_l^K)) = I + II.$$

and then bounding II using (H1a) and (H2), and bounding I using (H1a) and (H3).

Now (4.5) implies that

$$|\mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j})| \leq C\theta^{j/5} + C\theta^{\frac{s-2}{2u}j} = O(\theta^{\eta j})$$

as needed.

To prove part (c) we let $K = n^{-\frac{2}{2-s}}$. Then (4.5) yields

$$\mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j}) - \mathbb{E}(\hat{\xi}_l^K \hat{\xi}_{l+j}^K) = O(n^{-2}).$$

Now (H4) and (4.10) imply

$$\mathbb{E}(\hat{\xi}_l^K \hat{\xi}_{l+j}^K) - \sigma_j = O(n^{-2}) + O(\theta^{\ln n^3})$$

and (c) follows.

To prove part (d), fix j , let $K = \theta^{-\bar{\varepsilon}j}$ where $\bar{\varepsilon} \ll 1$, and take $l = j/\bar{\varepsilon}$. Then (H4) and (4.10) imply that

$$\mathbb{E}(\hat{\xi}_l^K \hat{\xi}_{l+j}^K) - \sigma_j = O(\theta^{\nu j})$$

with $\nu = \varepsilon(s-2)/2$. From (4.5) it follows that

$$\mathbb{E}(\hat{\xi}_l \hat{\xi}_{l+j}) - \sigma_j = O(\theta^{\nu j})$$

which together with (b) yields (d). \square

4.4. Proof of Theorem 4.2. In our proof, we shall use a standard Bernstein method based on “big block-small block” technique. That is, we divide the interval $[0, n]$ into big blocks of length n^ε alternating with small blocks of length n^{ε^2} . The big blocks will be almost “independent”, each from the preceding ones, due to the buffer zones that are the small blocks, while the contribution of these small blocks can be neglected.

From now on we fix $\varepsilon > 0$ so that $\frac{1+\varepsilon}{s} + \varepsilon < \frac{1}{2}$ and let $K = K_n = n^{\frac{1+\varepsilon}{s}}$. Let $l_m = m[n^\varepsilon]$, $m = 0, \dots, [n^{1-\varepsilon}] := m_n$. Denote

$$(4.11) \quad Z_m = \sum_{l=l_m+n^{\varepsilon^2}}^{l_{m+1}-1} \hat{\xi}_l^K, \quad \tilde{Z}_m = \sum_{l=l_m+n^{\varepsilon^2}}^{l_{m+1}-1} \hat{\xi}_l, \quad \tilde{\tilde{Z}}_m = \sum_{l=l_m}^{l_m+n^{\varepsilon^2}-1} \hat{\xi}_l, \quad \check{Z} = \sum_{l=l_{m_n}}^{n-1} \hat{\xi}_l,$$

so that

$$\sum_{l=0}^{n-1} \hat{\xi}_l = \sum_{m=0}^{m_n-1} \tilde{\tilde{Z}}_m + \sum_{m=0}^{m_n-1} \tilde{Z}_m + \check{Z}.$$

We claim that

$$(4.12) \quad \frac{\sum_{m=0}^{m_n-1} \tilde{\tilde{Z}}_m + \check{Z}}{\sqrt{n}} \quad \text{and} \quad \frac{\sum_{m=0}^{m_n-1} (Z_m - \tilde{Z}_m)}{\sqrt{n}}$$

converge to 0 in probability; it will therefore suffice for the proof of Theorem 4.2 (a), to prove the following

LEMMA 4.5.

$$(4.13) \quad \frac{\sum_{m=0}^{m_n} Z_m}{\sigma \sqrt{n}} \implies \mathcal{N}(0, 1).$$

We first prove the claim. To begin with, observe that following the proof of (4.4), we have

$$(4.14) \quad \mathbb{E} \left(\left(\sum_{m=0}^{m_n} \tilde{\tilde{Z}}_m + \check{Z} \right)^2 \right) = O(n^{1-\varepsilon+\varepsilon^2}) = o(n).$$

As for the second sum in (4.12), we write

$$Z_m - \tilde{Z}_m = \sum_{l=l_m+n^{\varepsilon^2}}^{l_{m+1}-1} [(\xi_l^K - \xi_l) - (\mathbb{E}(\xi_l^K) - \mathbb{E}(\xi_l))],$$

and condition (H1b) gives

$$\mathbb{E}|Z_m - \tilde{Z}_m| \leq 2 \sum_{l=l_m+n^{\varepsilon^2}}^{l_{m+1}-1} \mathbb{E}(|\xi_l^K - \xi_l|) \leq Cn^\varepsilon K^{1-s}.$$

Therefore,

$$(4.15) \quad \sum_{m=1}^{m_n} \frac{\mathbb{E}|Z_m - \tilde{Z}_m|}{\sqrt{n}} = O(K^{1-s}\sqrt{n}) = O\left(n^{\frac{1+\varepsilon}{s}-\frac{1}{2}}\right) = O(n^{-\varepsilon}) \rightarrow 0.$$

We now prepare for the proof of Lemma 4.5.

We start by defining an exceptional set \mathbf{G}_m^c on which we will not be able to exploit the almost independence of Z_{m+1} from (Z_1, \dots, Z_m) . The exact reasons for the definition of each condition on the “good set” \mathbf{G}_m will become evident in the course of the proof.

Let $\hat{l}_{m+1} = l_{m+1} + n^{\varepsilon^2/2}$. With sets $G_{l,k}$ as in hypotheses (H2)–(H4), we let

$$G_m^{(1)} = \bigcap_{k=n^{\varepsilon^2}}^{n^\varepsilon} \left(G_{l_{m+1},k} \cap G_{\hat{l}_{m+1},k} \right)$$

for $m \leq m_n$. Next, define for $l \geq l_{m+1} + n^{\varepsilon^2}$

$$\tilde{G}_l = \left\{ \omega : \mathbb{E} \left(\left| \xi_l^K - \xi_{l, l+n^{\varepsilon^2/2}}^K \right| \mid \mathcal{F}_{\hat{l}_{m+1}} \right) (\omega) \leq \theta^{n^{\varepsilon^2/10}} \right\}$$

and set

$$G_m^{(2)} = \bigcap_{k=n^{\varepsilon^2}}^{n^\varepsilon} \tilde{G}_{l_{m+1}+k}.$$

For $k, k' \in [n^{\varepsilon^2}, n^\varepsilon]$ with $k' - k \geq n^{\varepsilon^2/2}$ define

$$E_{m,k,k'} = \left\{ \omega' : \left| \mathbb{E}(\hat{\xi}_{l_{m+1}+k'}^K \mid \mathcal{F}_{l_{m+1}+k+n^{\varepsilon^2/2}})(\omega') \right| \geq \theta^{n^{\varepsilon^2/10}} \right\}$$

and let

$$\bar{G}_{m,k,k'} = \left\{ \omega : \mathbb{E} \left(\mathbb{1}_{E_{m,k,k'}}(\omega') \mid \mathcal{F}_{\hat{l}_{m+1}} \right) (\omega) \leq \theta^{n^{\varepsilon^2/10}} \right\}$$

and

$$G_m^{(3)} = \bigcap_{k,k' \in [n^{\varepsilon^2}, n^\varepsilon]; k'-k \geq n^{\varepsilon^2/2}} \bar{G}_{m,k,k'}.$$

Finally define “the good set”

$$\mathbf{G}_m = G_m^{(1)} \cap G_m^{(2)} \cap G_m^{(3)}.$$

Observe that (H2)–(H4) show that

$$(4.16) \quad \mathbb{P}(\mathbf{G}_m^c) \leq C\theta^{n^{\varepsilon^2/100}}.$$

The main step in the proof of Lemma 4.5, and thus of our CLT, is the following

SUBLEMMA 4.6. For $\omega \in \mathbf{G}_m$,

$$(4.17) \quad \ln \mathbb{E} \left(e^{i\lambda \frac{Z_{m+1}}{\sqrt{n}}} \middle| \mathcal{F}_{\hat{l}_{m+1}} \right) (\omega) = -\frac{n^\varepsilon}{n} \lambda^2 \sigma^2 (1 + o(1))$$

where $o(1)$ is uniform in $m = 1, \dots, m_n$.

Proof. To prove (4.17), we expand the exponential

$$(4.18) \quad \begin{aligned} \mathbb{E} \left(e^{i\lambda \frac{Z_{m+1}}{\sqrt{n}}} \middle| \mathcal{F}_{\hat{l}_{m+1}} \right) (\omega) &= 1 + i \frac{\lambda}{\sqrt{n}} \mathbb{E}(Z_{m+1} | \mathcal{F}_{\hat{l}_{m+1}})(\omega) - \frac{\lambda^2}{2n} \mathbb{E}(Z_{m+1}^2 | \mathcal{F}_{\hat{l}_{m+1}})(\omega) \\ &\quad + O \left(\frac{\lambda^3}{n^{\frac{3}{2}}} \mathbb{E}(Z_{m+1}^3 | \mathcal{F}_{\hat{l}_{m+1}})(\omega) \right) \\ &= 1 + i \frac{\lambda}{\sqrt{n}} \mathbb{E}(Z_{m+1} | \mathcal{F}_{\hat{l}_{m+1}})(\omega) - \frac{\lambda^2}{2n} \mathbb{E}(Z_{m+1}^2 | \mathcal{F}_{\hat{l}_{m+1}})(\omega) \\ &\quad + o \left(\frac{\lambda^2}{2n} \mathbb{E}(Z_{m+1}^2 | \mathcal{F}_{\hat{l}_{m+1}})(\omega) \right), \end{aligned}$$

where the last step uses that $|Z_{m+1}| \leq Kn^\varepsilon = o(n^{\frac{1}{2}})$.

Next, (H3) implies that on \mathbf{G}_m (which is contained in $G_{\hat{l}_{m+1},k}$ for each $k \in [n^{\varepsilon^2}, n^\varepsilon]$) we have

$$\mathbb{E}(Z_{m+1} | \mathcal{F}_{\hat{l}_{m+1}})(\omega) = \mathbb{E} \left(\sum_{j=n^{\varepsilon^2}}^{n^\varepsilon} \hat{\xi}_l^K \middle| \mathcal{F}_{\hat{l}_{m+1}} \right) (\omega) = o \left(\theta^{0.5n^{\varepsilon^2}} \right).$$

To complete the proof of (4.17) it suffices to show that for $\omega \in \mathbf{G}_m$,

$$(4.19) \quad |\mathbb{E}(Z_{m+1}^2 | \mathcal{F}_{\hat{l}_{m+1}})(\omega) - n^\varepsilon \sigma^2| = o(n^\varepsilon).$$

Note that

$$\mathbb{E}(Z_{m+1}^2 | \mathcal{F}_{\hat{l}_{m+1}})(\omega) = \sum_{k,k'=n^{\varepsilon^2}}^{n^\varepsilon} \mathbb{E}(\hat{\xi}_{l_{m+1}+k}^K \hat{\xi}_{l_{m+1}+k'}^K | \mathcal{F}_{\hat{l}_{m+1}})(\omega).$$

Let us estimate individual terms in this sum. Without loss of generality assume that $k' \geq k$. Let R be a large constant and consider two cases.

(a) $k > R(k' - k)$. In this case (H4) and (4.10) give

$$\mathbb{E}(\hat{\xi}_{l_{m+1}+k}^K \hat{\xi}_{l_{m+1}+k'}^K | \mathcal{F}_{\hat{l}_{m+1}})(\omega) + O(\tilde{\theta}^k) = b_{K,k'-k} = \sigma_{k'-k} + O(K^{2-s}) + O(\tilde{\theta}^k).$$

(b) $k \leq R(k' - k)$ and hence $k' - k > \frac{n^{\varepsilon^2}}{R}$. Then

(4.20)

$$(4.21) \quad \begin{aligned} \mathbb{E}(\hat{\xi}_{l_{m+1}+k}^K \hat{\xi}_{l_{m+1}+k'}^K | \mathcal{F}_{\hat{l}_{m+1}})(\omega) &= \mathbb{E}(\hat{\xi}_{l_{m+1}+k, l_{m+1}+k+n^{\varepsilon^2}/2}^K \hat{\xi}_{l_{m+1}+k'}^K | \mathcal{F}_{\hat{l}_{m+1}})(\omega) \\ &\quad + \mathbb{E} \left(\left(\hat{\xi}_{l_{m+1}+k}^K - \hat{\xi}_{l_{m+1}+k, l_{m+1}+k+n^{\varepsilon^2}/2}^K \right) \hat{\xi}_{l_{m+1}+k'}^K \middle| \mathcal{F}_{\hat{l}_{m+1}} \right) (\omega) \end{aligned}$$

(4.22)

$$= I + II$$

The second term is $O \left(\theta^{n^{\varepsilon^2}/10} K \right)$ since $\omega \in \tilde{G}_{l_{m+1}+k}$. For the first term use that $\omega \in \bar{G}_{m,k,k'}$ to obtain

$$\begin{aligned}
|I| &= \left| \mathbb{E}(\hat{\xi}_{l_{m+1}+k, l_{m+1}+k+n\varepsilon^2/2}^K | \mathcal{F}_{l_{m+1}+k+n\varepsilon^2/2}) | \mathcal{F}_{\hat{l}_{m+1}})(\omega) \right| \\
&\leq K^2 \mathbb{E}(\mathbb{1}_{E_{m,k,k'}}(\omega') | \mathcal{F}_{\hat{l}_{m+1}})(\omega) + K\theta^{n\varepsilon^2/10} \\
&\leq K^2 \theta^{n\varepsilon^2/10}
\end{aligned}$$

so both I and II are negligible. Combining the estimates of cases (a) and (b) we obtain (4.19) completing the proof of (4.17). \square

Proof of Lemma 4.5. It now remains to derive Lemma 4.5 from (4.17). For $j \leq m$ set

$$\hat{Z}_j = \sum_{l=l_j+n\varepsilon^2}^{l_{j+1}} \hat{\xi}_{l, n\varepsilon^2/2}^K.$$

Then

$$\mathbb{E} \left(\sum_{j=0}^m |\hat{Z}_j - Z_j| \right) = O \left(\theta^{n\varepsilon^2/10} \right)$$

and so

$$(4.23) \quad \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^{m+1} Z_j} \right) - \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} (\sum_{j=0}^m \hat{Z}_j) + i \frac{\lambda}{\sqrt{n}} Z_{m+1}} \right) = O \left(\theta^{n\varepsilon^2/10} \right),$$

$$(4.24) \quad \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^m Z_j} \right) - \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^m \hat{Z}_j} \right) = O \left(\theta^{n\varepsilon^2/10} \right).$$

Therefore,

$$(4.25) \quad \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^{m+1} Z_j} \right) \stackrel{(4.23)}{=} \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^m \hat{Z}_j} \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} Z_{m+1}} | \mathcal{F}_{\hat{l}_{m+1}} \right) \right) + O \left(\theta^{n\varepsilon^2/10} \right)$$

$$(4.26) \quad \stackrel{(4.16)}{=} \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^m \hat{Z}_j} \mathbb{1}_{\mathbf{G}_m} \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} Z_{m+1}} | \mathcal{F}_{\hat{l}_{m+1}} \right) \right) + O \left(\theta^{n\varepsilon^2/100} \right)$$

$$(4.27) \quad \stackrel{(4.17)}{=} e^{-\frac{\sigma^2 \lambda^2 n \varepsilon}{2n}} \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^m \hat{Z}_j} \right) + o \left(\frac{n^\varepsilon}{n} \right)$$

$$(4.28) \quad \stackrel{(4.24)}{=} e^{-\frac{\sigma^2 \lambda^2 n \varepsilon}{2n}} \mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^m Z_j} \right) + o \left(\frac{n^\varepsilon}{n} \right).$$

Iterating this recurrence relation m_n times we obtain

$$\mathbb{E} \left(e^{i \frac{\lambda}{\sqrt{n}} \sum_{j=0}^{m_n} Z_m} \right) = e^{-\frac{\sigma^2 \lambda^2}{2}} + o(1)$$

completing the proof of Lemma 4.5, and thus of part (a) of Theorem 4.2. \square

Part (b) of Theorem 4.2 can be established by a similar argument and we just briefly describe the necessary changes. To extend the proof of the Central Limit Theorem to the setting of part (b) we need to prove (4.12) and (4.17) with \mathbb{E}_n instead of \mathbb{E} . For (4.12) we need to prove the analogues of (4.14) and (4.15).

We claim the following. First, (4.17) still holds with \mathbb{E}_n instead of \mathbb{E} . Second,

$$(4.29) \quad \mathbb{E}_n \left(\left(\sum_{m=1}^{m_n} \tilde{Z}_m + \tilde{Z} \right)^2 \right) = O(n^{1-\varepsilon+\varepsilon^2}) = o(n).$$

(Note that in contrast to (4.14) the sum here starts with $m = 1$, not $m = 0$.) Third,

$$(4.30) \quad \mathbb{P}_n \left(\sum_{l=n^\varepsilon}^{n-1} \xi_l \neq \sum_{l=n^\varepsilon}^{n-1} \xi_l^{K_n} \right) \rightarrow 0$$

where $K_n = n^{\frac{1+\varepsilon}{s}}$. To prove (4.30) note that

$$\mathbb{P}_n \left(\sum_{l=n^\varepsilon}^{n-1} \xi_l \neq \sum_{l=n^\varepsilon}^{n-1} \xi_l^{K_n} \right) \leq n \max_l \mathbb{P}_n(\xi_l \neq \xi_l^{K_n})$$

so (4.30) follows from (D3). Observe that once these three points of the claim are established, the rest of the proof of part (b) proceeds exactly as in part (a). To obtain the other two points of our claim we will need the following

SUBLEMMA 4.7. *There exists a set \bar{G}_m with $\mathbb{P}(\bar{G}_m^c) \leq C\theta^{n^\varepsilon/100}$ such that for $\omega \in \bar{G}_m$, for $l \geq n^\varepsilon$ and for η a bounded random variable we have that*

$$|\mathbb{E}_n(\eta|\mathcal{F}_l)(\omega) - \mathbb{E}(\eta|\mathcal{F}_l)(\omega)| \leq C\|\eta\|_\infty \theta^{n^\varepsilon/100}.$$

Proof. Let η be a bounded random variable, $l \geq n^\varepsilon$ and ω be such that

$$(4.31) \quad \rho_{n,l}(\omega) \geq \theta^{l/2}, \quad |\mathbb{E}(\tilde{\rho}_{n,l}|\mathcal{F}_l)(\omega)| \leq \theta^{2l/3} \quad \text{where} \quad \tilde{\rho}_{n,l} = \rho_n - \rho_{n,l}.$$

Then

$$\mathbb{E}_n(\eta|\mathcal{F}_l) = \frac{\mathbb{E}((\rho_{n,l} + \tilde{\rho}_{n,l})\eta|\mathcal{F}_l)}{\mathbb{E}((\rho_{n,l} + \tilde{\rho}_{n,l})|\mathcal{F}_l)} = \frac{\rho_{n,l}\mathbb{E}(\eta|\mathcal{F}_l) + O(\theta^{2l/3})}{\rho_{n,l} + O(\theta^{2l/3})} = \mathbb{E}(\eta|\mathcal{F}_l) + O(\theta^{l/6}).$$

We prove now that the set where (4.31) fails has measure that is exponentially small in l . The proof consists of two steps. First, it follows from (D1) and (D2) that

$$(4.32) \quad \mathbb{P}_n(|\tilde{\rho}_{n,l}| \geq \theta^l) \leq Cn^u \mathbb{P}(|\tilde{\rho}_{n,l}| \geq \theta^l) \leq Cn^{2u}\theta^l$$

so $\mathbb{P}_n(\{|\mathbb{E}(\tilde{\rho}_{n,l}|\mathcal{F}_l)(\omega)| \geq \theta^{2l/3}\})$ is exponentially small by Markov's inequality. Second,

$$\mathbb{P}_n(\rho_{n,l} < \theta^{l/2}, \tilde{\rho}_{n,l} < \theta^l) \leq \mathbb{P}_n(\rho_n < 2\theta^{l/2}) = \mathbb{E}(\rho_n 1_{\rho_n < 2\theta^{l/2}}) \leq 2\theta^{l/2}$$

and so $\mathbb{P}_n(\rho_{n,l} < \theta^{l/2})$ is exponentially small due to (4.32). \square

Sublemma 4.7, together with Sublemma 4.6, imply that (4.17) holds for \mathbb{E}_n instead of \mathbb{E} provided that we decrease slightly the set \mathbf{G}_m to $\bar{G}_m \cap \mathbf{G}_m$.

It remains to prove (4.29). Note that the arguments used to establish Sublemma 4.4 in fact give for $\omega \in \mathbf{G}_m$

$$\mathbb{E}(\xi_m \xi_{m+j} | \mathcal{F}_{n^{\varepsilon/2}})(\omega) = \begin{cases} \sigma_j + O(\theta^m) & \text{if } m > \frac{2uj}{|\ln \theta|} \\ O(\theta^j) & \text{otherwise.} \end{cases}$$

Hence Lemma 4.7 implies that for $\omega \in \bar{G}_m \cap \mathbf{G}_m$

$$\mathbb{E}_n(\xi_m \xi_{m+j} | \mathcal{F}_{n^{\varepsilon/2}})(\omega) = \begin{cases} \sigma_j + O(\theta^m) + O(\theta^{n^{\varepsilon/200}}) & \text{if } m > \frac{2uj}{|\ln \theta|} \\ O(\theta^j) + O(\theta^{n^{\varepsilon/200}}) & \text{otherwise.} \end{cases}$$

This estimate implies (4.29) by direct summation.

The proof of Theorem 4.2 is thus completed. \square

5. PRELIMINARIES ON DIAGONAL ACTIONS AND SIEGEL TRANSFORMS

In Section 6, we will use the abstract Theorem 4.2 to prove Theorem 2.1, 2.2, 2.3. For this, we just have to check (H1)–(H4) for the case where our probability space is \mathcal{M} equipped with the Haar measure and $\xi_l(\mathcal{L}) = \Phi(g^l \mathcal{L})$, $\Phi = \mathcal{S}(f)$, where $f \in C^{s,r}(\mathbb{R}^{d+r})$ is a positive function supported on a compact set which does not contain 0.

Before we construct the filtrations and prove (H1)–(H4) for the sequence $\xi_l(\mathcal{L}) = \Phi(g^l \mathcal{L})$, we recall and prove preliminary results about functions defined on the space of lattices, on Siegel transforms, and on the action of diagonal matrices. We will cover this in Sections 5.1, 5.3, and 5.2 respectively. Then we will prove Theorems 2.1, 2.2, 2.3 in Section 6. In Section 8, we compute the variances in the special case that interests us of f being the characteristic function of E_c given in (3.2). This will finish the proof of Theorems 1.1, 1.3, 1.4.

5.1. Siegel transforms and Rogers' identities.

PROPOSITION 5.1. [25, Theorems 3.15 and 3.16], [15, Appendix B] *Let $f: \mathbb{R}^{d+r} \rightarrow \mathbb{R}$ be a piecewise smooth function that is supported on a compact set which does not contain 0. Then*

$$\begin{aligned} (a) \quad & \int_{\mathcal{M}} \mathcal{S}(f)(\mathcal{L}) d\mu(\mathcal{L}) = \int_{\mathbb{R}^{d+r}} f(\mathbf{x}) d\mathbf{x}; \\ (b) \quad & \text{If } d+r > 2 \text{ then} \\ & \int_{\mathcal{M}} [\mathcal{S}(f)]^2(\mathcal{L}) d\mu(\mathcal{L}) = \left[\int_{\mathbb{R}^{d+r}} f(\mathbf{x}) d\mathbf{x} \right]^2 \\ & \quad + \sum_{\substack{(p,q) \in \mathbb{N}^2 \\ \gcd(p,q)=1}} \int_{\mathbb{R}^{d+r}} [f(p\mathbf{x})f(q\mathbf{x}) + f(p\mathbf{x})f(-q\mathbf{x})] d\mathbf{x}. \end{aligned}$$

Suppose that $f: \mathbb{R}^{d+r} \rightarrow \mathbb{R}$ is a piecewise smooth function of compact support. Then,

$$\begin{aligned} (c) \quad & \int_{\tilde{\mathcal{M}}} \tilde{\mathcal{S}}(f)(\tilde{\mathcal{L}}) d\mu(\tilde{\mathcal{L}}) = \int_{\mathbb{R}^{d+r}} f(\mathbf{x}) d\mathbf{x}; \\ (d) \quad & \int_{\tilde{\mathcal{M}}} [\tilde{\mathcal{S}}(f)]^2(\tilde{\mathcal{L}}) d\mu(\tilde{\mathcal{L}}) = \left[\int_{\mathbb{R}^{d+r}} f(\mathbf{x}) d\mathbf{x} \right]^2 + \int_{\mathbb{R}^{d+r}} f^2(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

5.2. Rate of equidistribution of unipotent flows and representative partitions.

Recall the notation of subsection 2.1. Let g be the matrix defined by (2.4) Let h_u be a one parameter subgroup of \mathbb{L} . For example one can take the matrices with ones on the diagonal, an arbitrary number in the upper left corner and zeros elsewhere. Note that

$$g^n h_u \mathcal{L} = h_{2^{d/r+1}u} g^n \mathcal{L}.$$

The filtrations for which we will prove (H2)–(H4) for the sequence $\xi_l(\mathcal{L}) = \Phi(g^l \mathcal{L})$, will consist of small arcs in the direction of the flow of h_u . The exponential mixing of the \mathbb{G} -action will underly the equidistribution and independence properties that are stated in (H1)–(H4).

We will need the notion of representative partitions that was already used in [12]. These will be partitions of \mathcal{M} whose elements are segments of h_u orbits, whose pushforwards by g^l will become rapidly equidistributed. To guarantee the filtration property, we would ideally consider an increasing sequence of such partitions with pieces of size 2^{-l} , $l = 0, \dots, n$. However, such partitions with fixed size pieces do not exist because h_u is weak mixing. We overcome this technical difficulty due to the following observations:

- (1) Rudolph's Theorem (see [8, Section 11.4]) shows that for each $\bar{\varepsilon}$ we can find a partition \mathcal{P} into h_u -orbits such that the length of each element is either L or $L/\sqrt{2}$ and if $\mathcal{P}_u = h_u(\mathcal{P})$ then

$$\mu(\mathcal{L} : \exists u \in [0, L] : \mathcal{P}_u(\mathcal{L}) \text{ has length } L/\sqrt{2}) \leq \bar{\varepsilon}.$$

- (2) Given $n \in \mathbb{N}$, it suffices to check the properties (H2)–(H4) away from a set of measure less than θ^n .

Having fixed n , we will therefore abuse notation and say that a partition is of size L if $\bar{\varepsilon}$ in 1) is less than θ^n . In light of this, let \mathcal{P} be a partition of size 1 and \mathcal{P}^l be its sub-partition of size 2^{-l} . Due to (1) and (2) we can assume without loss of generality that for every fixed $u \in [0, 1]$, the partitions \mathcal{P}_u^l form an increasing sequence and that as a consequence the sequence \mathcal{F}_l of σ -algebras generated by \mathcal{P}_u^l forms a filtration.

Fix a small constant $\kappa > 0$. Given a collection $\Psi \subset C^{\text{s,r}}(\mathcal{M})$, a set of natural numbers $\{k_n\}_{n \in \mathbb{N}}$, and a number L , we call a partition \mathcal{P} of size L *representative* with respect to $(\{k_n\}, \Psi)$ if for each $A \in \Psi$ and for each $n \in \mathbb{N}$,

$$\mu \left(\mathcal{L} : \left| \int_{g^{k_n} \mathcal{P}(\mathcal{L})} A - \mu(A) \right| \geq \|A\|_{C^{\text{s,r}}(\mathcal{M})} (2^{k_n} L)^{-\kappa} \right) \leq \|A\|_{C^{\text{s,r}}(\mathcal{M})} (2^{k_n} L)^{-\kappa}.$$

The curve $g^{k_n} \mathcal{P}(\mathcal{L})$ is of the form $h_u \bar{\mathcal{L}}$ with $u \in [0, 2^{k_n} L]$, and we use the notation $\int_\gamma A$ for the normalized integral $\frac{1}{2^{k_n} L} \int_0^{2^{k_n} L} A(h_u \bar{\mathcal{L}}) du$. Given a finite collection $\Psi \subset C^{\text{s,r}}(\mathcal{M})$, $\{k_n\}_{n \in \mathbb{N}}$, and $L > 0$, we let

$$\delta(\{k_n\}, \Psi, L) := \sum_{A \in \Psi} \|A\|_{C^{\text{s,r}}(\mathcal{M})} \sum_n (2^{k_n} L)^{-\frac{\kappa}{2}}.$$

Let δ be a small number. Then we have as in [12, Proposition 7.1]

PROPOSITION 5.2. *Let $\mathcal{R}(\{k_n\}, \Psi) \subset [0, 1]$ be the set of u such that \mathcal{P}_u is representative with respect to $(\{k_n\}, \Psi)$. Then $\text{Leb}(\mathcal{R}(\{k_n\}, \Psi)) \geq 1 - \delta(\{k_n\}, \Psi, L)$.*

Proof. We quickly recall how Proposition 5.2 can be deduced, exactly as in [12, Proposition 7.1], from the polynomial mixing of the unipotent flow h_u . Indeed, assuming that $\mu(A) = 0$, polynomial mixing implies that

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

with $\mathcal{K}_A = \|A\|_{\mathbf{s},\mathbf{r}}$. Thus for curves $\gamma(\bar{\mathcal{L}})$ of the form $h_u\bar{\mathcal{L}}$ with $u \in [0, L]$ we get that

$$\mu\left(\bar{\mathcal{L}} : \left| \int_{\gamma(\bar{\mathcal{L}})} A \right| \geq \mathcal{K}_A L^{-\kappa_0} \right) \leq CL^{-\kappa_0}$$

with $\kappa_0 := \kappa/3$. This implies that if we consider a partition \mathcal{P} of size L and its corresponding shifted partitions \mathcal{P}_u , $u \in [0, L]$ we get for the measure $\bar{\mu} = \mu \times \text{Leb}_{[0,1]}$ that

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

where $\mathcal{P}_u(\mathcal{L})$ denotes the piece of \mathcal{P}_u that goes through \mathcal{L} . The claim of Proposition 5.2 then follows by Markov's inequality. \square

REMARK 5.3. Proposition 5.2 will be used in the next section to obtain a partition of \mathcal{M} into pieces of h_u orbits satisfying the condition of Theorem 4.2. We could also use a partition into whole \mathbb{L} -orbits. The proof of Proposition 5.2 in that case would be simpler since we could use effective equidistribution of horospherical subgroups [21]. We prefer to use h_u orbits instead since it allows us to give unified proofs of Theorems 2.1, 2.2, and 2.3 (as well as Theorem 7.1 in Section 7).

5.3. Truncation of Siegel transforms. In this Section we give some useful results on truncations of a Siegel transform of a compactly supported function $f \in C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^p)$. These bounds are essential to control the truncated ξ_n^K that appear in the abstract CLT of Section 4. We will leave all the proofs and constructions to Appendix A. In particular, we will define $\mathfrak{h}_{2,K} : \mathcal{M}(\text{or } \tilde{\mathcal{M}}) \rightarrow \mathbb{R}$, with the properties described in Lemma 5.4 below. We will always use the following notation for $\Phi = \mathcal{S}(f)$ (or $\tilde{\mathcal{S}}(f)$) : $\Phi^K = \Phi \mathfrak{h}_{2,K}$. In the sequel we will consider $\xi_n^K := \Phi^K \circ g^n$.

LEMMA 5.4. [12] *There exists a constant $Q > 1$ such that for each pair of integers \mathbf{s}, \mathbf{r} and each R there is a constant $C = C(R, \mathbf{s}, \mathbf{r})$ such that the following holds. Let f be supported on $B(0, R)$ in \mathbb{R}^p .*

(a) *If $f \in C^{\mathbf{s}}(\mathbb{R}^p)$ then*

$$\|\Phi^K\|_{C^{\mathbf{s}}(\mathcal{M})} \leq CK \|f\|_{C^{\mathbf{s}}(\mathbb{R}^p)}.$$

(b) *If $f \in C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^p)$ then*

$$\|\Phi^K\|_{C^{\mathbf{s},\mathbf{r}}(\mathcal{M})} \leq CK \|f\|_{C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^p)}.$$

(c) *If $f \in C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^p)$ then*

$$\|\Phi^K \cdot (\Phi^K \circ g^j)\|_{C^{\mathbf{s},2\mathbf{r}}(\mathcal{M})} \leq CK^2 \|f\|_{C^{\mathbf{s},\mathbf{r}}(\mathbb{R}^p)}^2 Q^j.$$

LEMMA 5.5. *For every r, d there exists $C > 0$ such that $\Phi = \mathcal{S}(f)$ or $\Phi = \tilde{\mathcal{S}}(f)$ satisfy*

$$(5.1) \quad \mathbb{E}(\Phi - \Phi^K) \leq \frac{C}{K^{d+r-1}}.$$

$$(5.2) \quad \mathbb{E}((\Phi - \Phi^K)^2) \leq \frac{C}{K^{d+r-2}}.$$

If $r = d = 1$, then $\Phi = \tilde{\mathcal{S}}(f)$ satisfies

$$(5.3) \quad \mathbb{E}(\Phi - \Phi^K) \leq \frac{C}{K^2}.$$

$$(5.4) \quad \mathbb{E}((\Phi - \Phi^K)^2) \leq \frac{C}{K}.$$

In addition, the same inequalities (5.1)–(5.4) hold if the expectation is considered with respect to a measure that has a C -centrally smoothable density.

Recall that an \mathbb{L} -box is a set of the form $\Pi(\mathfrak{R}, \tilde{\mathcal{L}}) = \{\Lambda_{\mathbf{a}}\tilde{\mathcal{L}}\}$ where \mathbf{a} belongs to the box \mathfrak{R} in \mathbb{R}^{dr} . Also define $\mathbf{a}: \mathcal{M} \rightarrow \mathbb{R}$ by

$$\mathbf{a}(\mathcal{L}) = \max\{(\text{covol}(\bar{\mathcal{L}}))^{-1} : \bar{\mathcal{L}} \leq \mathcal{L}\}.$$

LEMMA 5.6. *For each $\bar{\varepsilon}, L$ there exists a constant $C > 0$ such that for any box \mathfrak{R} whose sides are longer than $\bar{\varepsilon}$ and which is contained in $[-L, L]^{dr}$ and for any $\tilde{\mathcal{L}}$ with $\mathbf{a}(\tilde{\mathcal{L}}) \leq L$, $\Pi = \Pi(\mathfrak{R}, \tilde{\mathcal{L}})$ satisfies*

$$(5.5) \quad \mathbb{P}_{\Pi}(|\Phi \circ g^l| \geq K) \leq \frac{C}{K^{d+r}},$$

$$(5.6) \quad |\mathbb{E}_{\Pi}(\Phi \circ g^l - \Phi^K \circ g^l)| \leq \frac{C}{K^{d+r-1}},$$

$$(5.7) \quad \mathbb{E}_{\Pi}((\Phi \circ g^l - \Phi^K \circ g^l)^2) \leq \frac{C}{K^{d+r-2}},$$

where \mathbb{P}_{Π} is a restriction of the Haar measure on $\mu_{\mathbb{L}}$ to Π and \mathbb{E}_{Π} is expectation with respect to \mathbb{P}_{Π} .

6. PROOF OF THE CLT FOR DIAGONAL ACTIONS

We are ready now to prove Theorem 2.1, 2.2, 2.3 using Theorem 4.2.

6.1. CLT for lattices. Proof of Theorem 2.1. For f as in the statement of Theorem 2.1 (that is, $f \in C^{s,r}(\mathbb{R}^{d+r})$ non-negative and supported on a compact set which does not contain $0 \in \mathbb{R}^{d+1}$), recall that we defined the Siegel transform $\Phi = \mathcal{S}(f)$ and $\xi_l(\mathcal{L}) = \Phi(g^l\mathcal{L})$. Recall also the notation $\Phi^K = \Phi \mathfrak{h}_{2,K}$ and $\xi_l^K := \Phi^K \circ g^l$. We will now prove (H1)–(H4) for the sequence $\{\xi_n\}$.

6.1.1. Property (H1). Fix any $s \in (2, d+r)$. Property (H1) follows from inequalities (5.1) and (5.2) of Lemma 5.5. The fact that (H1) fails to hold when $d = r = 1$ is the reason why the CLT does not hold in this case.

6.1.2. *Constructing filtrations.* We will use the notion of representative partitions of Section 5.2 to construct the desired filtrations.

First of all, note that to prove (H4) we need to deal with function of the form $\Phi^{K_n} \cdot \Phi^{K_n} \circ g^j$. Therefore we define for every $j \leq n$ the collection of functions

$$\Phi^{(j)} := \{\Phi, \Phi^{K_n}, \Phi^{K_n} \cdot \Phi^{K_n} \circ g^j\}$$

Fix constants $R_1 \gg R_2 \gg R_3 \gg 1$, and define for every $l \leq n$ the following collection of functions and sequences of integers

$$(6.1) \quad \bigcup_{j \leq n} (\{k + l\}_{k \geq R_3(\log_2 K_n + j)}, \Phi^{(j)})$$

Next let \mathcal{P} be a partition of size 1 and \mathcal{P}^l be its subpartition of size 2^{-l} . By Proposition 5.2 and 5.4 there is u such that for each $0 \leq l \leq n$, \mathcal{P}_u^l is representative with respect to the collections of integers and functions in (6.1). Let \mathcal{F}_l be the filtration of σ -algebras generated by \mathcal{P}_u^l . Denote $\xi_{l,l+k}^K = \mathbb{E}(\xi_l^K | \mathcal{F}_{k+l})$.

We claim that $(\xi_l^K, \{\mathcal{F}_l\})$ satisfies (H1)–(H4) with $u = 2s$ provided that θ is sufficiently close to 1. Since (H1) has been checked above it remains to verify (H2)–(H4).

6.1.3. *Property (H2).* If $k \leq C \log_2 K$ then (H2) holds if we take u sufficiently large. By Lemma 5.4 there are functions Φ^\pm such that

$$\Phi^- \leq \Phi^K \leq \Phi^+, \quad \|\Phi^+ - \Phi^-\|_{L^1} \leq 2^{-\varepsilon k}, \quad \|\Phi^\pm\|_{C^1} \leq CK2^{\varepsilon r k}.$$

Then

$$\xi_l^- - \xi_{l,k}^+ \leq \xi_l^K - \xi_{l,k}^K \leq \xi_l^+ - \xi_{l,k}^-$$

where ξ_l^\pm and $\xi_{l,k}^\pm$ are defined analogously to ξ_l^K and $\xi_{l,k}^K$ with Φ^K replaced by Φ^\pm . Since Φ^\pm are Lipschitz, we have

$$|\xi_l^\pm - \xi_{l,k}^\pm| \leq CK2^{(\varepsilon r - 1)k}.$$

So if $2^{\varepsilon r - 1} \leq \theta^2$ and C is sufficiently large then $|\xi_l^K - \xi_{l,k}^K| \geq \theta^k$ implies $\xi_l^+ - \xi_l^- \geq \frac{\theta^k}{3}$. Hence Markov's inequality gives

$$\mathbb{P}(|\xi_l^K - \xi_{l,k}^K| \geq \theta^k) \leq C \left(\frac{2^{-\varepsilon}}{\theta} \right)^k.$$

This proves (H2) provided that u is large enough and

$$2^{(\varepsilon r - 1)/2} < \theta < 2^{-\varepsilon}.$$

6.1.4. *Properties (H3) and (H4).* (H3) follows from the definition of representative partition if $k \geq R_3 \log_2 K_n$ while for $k < R_3 \log_2 K_n$,

$$\mathbb{E}(\xi_{k+l}^K | \mathcal{F}_l) \leq K \leq K^u \theta^k$$

provided that θ is sufficiently close to 1.

Likewise, if $k > R_1(\log_2 K_n + k' - k)$ then (H4) holds by the definition of the representative partition with

$$b_{K,k} = \mathbb{E}(\xi_k^K \xi_0^K) - (\mathbb{E}(\xi_0^K))^2.$$

If $k \leq R_1(\ln K_n + k' - k)$ we consider two cases

(a) $k' - k \leq R_2 \log_2 K_n$ and so $k < 2R_1^2 \log_2 K_n$. In this case (H4) trivially holds similarly to (H3).

(b) $k' - k \geq R_2 \log_2 K_n$ and so $k < 2R_1(k' - k)$. Accordingly to establish (H4) with $b_{K,k} = 0$ it suffices to show that there is a constant $\tilde{\theta} < 1$ such that

$$(6.2) \quad \mathbb{P} \left(\left| \mathbb{E}(\hat{\xi}_{l+k}^K \hat{\xi}_{l+k'}^K | \mathcal{F}_l)(\omega) \right| \geq \tilde{\theta}^j \right) \leq \tilde{\theta}^j.$$

We are going to show that (6.2) follows from already established (H1)–(H3). The argument is similar to the proof of (?). Namely, denoting by $j = k' - k$ we get

$$\begin{aligned} \mathbb{E}(\hat{\xi}_{l+k'}^K \hat{\xi}_{l+k}^K | \mathcal{F}_l) &= \mathbb{E}(\hat{\xi}_{l+k+j}^K \hat{\xi}_{l+k, l+k+j/2}^K | \mathcal{F}_l) + \mathbb{E}(\hat{\xi}_{l+k+j}^K (\hat{\xi}_{l+k, l+k+j/2}^K - \hat{\xi}_{l+k}^K) | \mathcal{F}_l) \\ &= I + II. \end{aligned}$$

(H1a) and (H2) imply that $\mathbb{P}(|II| \geq K\theta^j) \leq \theta^j$. Next,

$$|I| = \left| \mathbb{E} \left(\left[\hat{\xi}_{l+k, l+k+j/2}^K \mathbb{E}(\hat{\xi}_{l+k+j}^K | \mathcal{F}_{l+k+j/2}) \right] \middle| \mathcal{F}_l \right) \right|$$

and (H3) shows that the expected value of the RHS is $O(K^{2u}\theta^{j/2})$. Now Markov's inequality shows that $\mathbb{P}(|I| \geq K^{2u}\theta^{j/4}) \leq \theta^{j/4}$. Combining the estimates of I and II we obtain (6.2).

Having checked (H1)–(H4), we have established Theorem 2.1(a) via Theorem 4.2(a).

6.1.5. *Starting from localized initial conditions.* To prove Theorem 2.1(b) we just need to check condition (D1)–(D3) of Theorem 4.2(b) for ρ_N .

Property (D1) follows from (CN^u, α) -regularity since $\|\rho_N\|_{L^\infty} \leq \|\rho_N\|_{C^\alpha(\text{supp}(\rho))}$.

To check (D2) let $\rho_{N,l} = \mathbb{E}(\rho_N | \mathcal{F}_l)$. Then if

$$(6.3) \quad h_{[-2^{-l}, 2^{-l}]} \mathcal{L} \cap \partial(\text{supp}(\rho)) = \emptyset$$

then ρ_N is Holder on the element of \mathcal{P}_u^l containing \mathcal{L} so

$$|\rho_N - \rho_{N,l}| \leq CN^u 2^{-\alpha l}.$$

so the exceptional set for (D2) consists of points violating (6.3). This set has a small measure since $\partial(\text{supp}(\rho))$ is (CN^u, α) -regular.

Finally (D3) follows from inequalities (5.1) and (5.2) of Lemma 5.5 applied to the centrally smoothable density ρ_N .

The proof of Theorem 2.1 is thus complete. \square

6.2. **CLT for affine lattices. Proof of Theorem 2.2.** For f as in the statement of Theorem 2.2 we define $\Phi = \tilde{\mathcal{S}}(f)$ and $\xi_l(\tilde{\mathcal{L}}) = \Phi(g^l \tilde{\mathcal{L}})$, with $\tilde{\mathcal{L}} \in \tilde{\mathcal{M}}$ distributed according to Haar measure. We also use $\Phi^K = \Phi \mathfrak{h}_{2,K}$ and $\xi_l^K := \Phi^K \circ g^l$.

If $(r, d) \neq (1, 1)$ the analysis is exactly the same as in the case of linear lattices.

If $(r, d) = (1, 1)$, (5.1) and (5.2) of Lemma 5.5 are not sufficient anymore to prove Property (H1), and we replace them by (5.3) and (5.4). The proof of properties (H2)–(H4) proceeds exactly as in the case of linear lattices.

Theorem 2.2(a) thus follows from Theorem 4.2(a).

The changes needed for Theorem 2.2(b) are the same as for for Theorem 2.1(b). Observe that (D3) and (H1) hold since (5.1)–(5.4) are valid if Haar measure is replaced with measures having centrally-smoothable densities. \square

6.3. Fixed x . Proof of Theorem 2.3. Here we deduce Theorem 2.3 from a refined version of Theorem 2.2. Using (5.5) we conclude that it is sufficient to prove the Central Limit Theorem for sums with a shorter range of summation,

$$\frac{\sum_{n=N^\varepsilon}^{N-1} \tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) - N\bar{f}}{\tilde{\sigma}\sqrt{N}}.$$

Next, take a large constant β and for $\tilde{\mathcal{L}} \in \tilde{\mathcal{D}}$ let

$$\mathcal{V}(\tilde{\mathcal{L}}) = \{(D_t \bar{\Lambda}_{\mathbf{b}}, (\mathbf{y}, 0))\tilde{\mathcal{L}}\}_{|t| \leq N^{-\beta}, |\mathbf{b}| \leq N^{-\beta}, |\mathbf{y}| \leq N^{-\beta}}.$$

Let $\hat{\mathcal{D}} = \bigcup_{\tilde{\mathcal{L}} \in \tilde{\mathcal{D}}} \mathcal{V}(\tilde{\mathcal{L}})$. We have a partition $\hat{\Pi}$ of $\hat{\mathcal{D}}$ so that each element of the partition is of the form $\hat{\Pi}(\tilde{\mathcal{L}}^*)$ for some $\tilde{\mathcal{L}}^* \in \tilde{\mathcal{M}}$, where $\hat{\Pi}(\tilde{\mathcal{L}}^*) := \bigcup_{\tilde{\mathcal{L}} \in \Pi(\tilde{\mathcal{L}}^*)} \mathcal{V}(\tilde{\mathcal{L}})$.

LEMMA 6.1. *If β is sufficiently large then there is a constant $\bar{\delta} > 0$ such that*

$$\mathbb{P}_{\hat{\Pi}(\tilde{\mathcal{L}}^*)} \left(\sum_{n=N^\varepsilon}^{N-1} \left| \tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) - \tilde{\mathcal{S}}(f)(g^n (D_t \bar{\Lambda}_{\mathbf{b}}, (\mathbf{y}, 0))\tilde{\mathcal{L}}) \right| \geq 1 \right) \leq N^{-(1+\bar{\delta})}$$

except possibly for a set of $\tilde{\mathcal{L}}^*$ of measure $O(N^{-10})$.

Proof. In accordance with notation of Theorem 4.2(b) we will use notation $K_N = N^{\frac{1+\varepsilon}{r+d}}$. Also denote $\hat{\mathcal{L}} = (D_t \bar{\Lambda}_{\mathbf{b}}, (\mathbf{y}, 0))\tilde{\mathcal{L}}$. First we replace $\tilde{\mathcal{S}}(f)$ by Φ^{K_N} . This can be done in view of the following estimate those proof will be given in the appendix

LEMMA 6.2.

$$\mathbb{P}_{\hat{\Pi}(\tilde{\mathcal{L}}^*)} \left(\tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) \neq \Phi^{K_N}(g^n \tilde{\mathcal{L}}) \quad \text{or} \quad \tilde{\mathcal{S}}(f)(g^n \hat{\mathcal{L}}) \neq \Phi^{K_N}(g^n \hat{\mathcal{L}}) \right) \leq \frac{C}{K^{d+r}}.$$

Hence denoting $\varepsilon_N := N^{-20}$ we get functions Φ^\pm such that

$$\Phi^- \leq \Phi^{K_N} \leq \Phi^+, \quad \|\Phi^+ - \Phi^-\| \leq \varepsilon_N, \quad \text{and} \quad \|\Phi^\pm\|_{C^s} \leq C\varepsilon_N^{-r}.$$

We claim that

$$\left| \Phi^{K_N}(g^n \tilde{\mathcal{L}}) - \Phi^{K_N}(g^n \hat{\mathcal{L}}) \right| \leq \left| \Phi^+(g^n \hat{\mathcal{L}}) - \Phi^-(g^n \hat{\mathcal{L}}) \right| + CK_N \varepsilon_N^{-r} N^{-\beta} = I_n + II_n.$$

Consider for example the case where $\Phi^{K_N}(g^n \hat{\mathcal{L}}) \geq \Phi^{K_N}(g^n \tilde{\mathcal{L}})$, the opposite case being similar. Then

$$\begin{aligned} 0 &\leq \Phi^{K_N}(g^n \hat{\mathcal{L}}) - \Phi^{K_N}(g^n \tilde{\mathcal{L}}) \leq \Phi^+(g^n \hat{\mathcal{L}}) - \Phi^-(g^n \tilde{\mathcal{L}}) \\ &\leq \Phi^+(g^n \hat{\mathcal{L}}) - \Phi^-(g^n \hat{\mathcal{L}}) + |\Phi^-(g^n \hat{\mathcal{L}}) - \Phi^-(g^n \tilde{\mathcal{L}})|. \end{aligned}$$

The second term can be estimated by

$$\|\Phi^-\|_{C^1} d(g^n \hat{\mathcal{L}}, g^n \tilde{\mathcal{L}}) \leq CK_N \varepsilon_N^{-r} N^{-\beta}$$

proving our claim. Next if β is sufficiently large then $|\sum_n II_n| \leq \frac{1}{2}$ while

$$\left\| \sum_n I_n \right\|_{L^1} \leq C\varepsilon_N N.$$

Now the lemma follows from Markov's inequality. \square

Lemma 6.1 allows to reduce Theorem 2.3 to the following result.

THEOREM 6.3. *Suppose that $(r, d) \neq (1, 1)$. For each $r > 0$ and each $\varepsilon > 0$ the following holds. If $\mathbb{P}_{\tilde{\mathcal{L}}^*}$ denotes the uniform distribution on $\hat{\Pi}(\tilde{\mathcal{L}}^*)$ then*

$$\mathbb{P} \left(\tilde{\mathcal{L}}^* : \sup_z \left| \mathbb{P}_{\tilde{\mathcal{L}}^*} \left(\frac{\sum_{n=N^\varepsilon}^{N-1} \tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) - N\bar{f}}{\tilde{\sigma}\sqrt{N}} \leq z \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-s^2/2} ds \right| \geq \varepsilon \right) = O(N^{-r}).$$

The proof of Theorem 6.3 is also very similar to the proof of Theorem 2.1. Let us describe the necessary modifications.

The property (H1) follows from Lemma 5.6 instead of (5.1) and (5.2).

To define the required filtration of (H2)–(H4) we need to adapt Proposition 5.2 as follows. Take δ_N going to 0 sufficiently slowly, for example, $\delta_N = 1/N$. We let \mathcal{P} be a partition into segments of h_u orbits of size δ_N and \mathcal{P}^l the corresponding subpartitions of pieces with length $\delta_N 2^{-l}$. We let \mathcal{P}_u^l be the translates by h_u of these partitions and denote by $\mathcal{P}_u^l(\tilde{\mathcal{L}}^*)$ the collection of pieces of \mathcal{P}_u^l which are contained in $\hat{\Pi}(\tilde{\mathcal{L}}^*)$. We say that $\tilde{\mathcal{L}}^*$ is N -good if there exists $u \in [0, 1/N]$ such that for each $N^\varepsilon \leq l \leq N$, $\mathcal{P}_u^l(\tilde{\mathcal{L}}^*)$ is representative with respect to the families (6.1). The proof of Proposition 5.2 also shows the following.

LEMMA 6.4. *Given $r \in \mathbb{N}$, if we take R_3 in (6.1) sufficiently large then*

$$\mathbb{P} \left(\tilde{\mathcal{L}}^* \text{ is not } N\text{-good} \right) \leq \frac{C}{N^r}.$$

On the other hand if $\tilde{\mathcal{L}}^*$ is N -good then the filtration generated by the partitions $\mathcal{P}_u^l(\tilde{\mathcal{L}}^*)$ satisfies (H2)–(H4). Theorem 6.3 thus follows from Theorem 4.2. \square

REMARK 6.5. The argument given above does not tell us for which \mathbf{x} Theorem 1.4 holds. Of course rational \mathbf{x} have to be excluded due to Theorem 1.1. Now a simple Baire category argument shows that Theorem 1.4 also fails for very Liouvilian \mathbf{x} . It is of interest to provide explicit Diophantine conditions which are sufficient for Theorem 1.4. The papers [14, 34] provide tools which may be useful in attacking this question.

7. RELATED RESULTS

The arguments of the previous section are by no means limited to $SL_{d+r}(\mathbb{R})/SL_{d+r}(\mathbb{Z})$. In particular, we have the following result.

THEOREM 7.1. *Let \mathcal{G} be a C^r diffeomorphism, $r \geq 2$, of a manifold \mathbb{M} and $\mathcal{H} = \{\mathcal{H}_u\}_{u \geq 0}$ be a C^r flow on that space. Suppose that*

(i) *both \mathcal{G} and \mathcal{H} preserve a probability measure μ and there exists $c > 0$ such that*

$$\mathcal{G}^n \mathcal{H}_u = \mathcal{H}_{e^{cn}u} \mathcal{G}^n.$$

(ii) *There are constants $\bar{K} > 0$ and $Q > 1$ such that*

$$(7.1) \quad \|A \circ \mathcal{G}^j\|_{C^r} \leq \bar{K} Q^j \|A\|_{C^r}$$

(iii) *\mathcal{H} is polynomially mixing, that is, there exist $\alpha \in (0, r]$, $\kappa > 0$, and $K > 0$ such that if*

$$(7.2) \quad A \in C^\alpha(\mathbb{M}) \text{ and } \mu(A) = 0 \text{ then}$$

$$(7.3) \quad |\mu(AA \circ \mathcal{H}_u)| \leq \frac{K \|A\|_{C^\alpha}^2}{u^\kappa}.$$

Fix $L > 0$ and let U be a random variable uniformly distributed on $[0, L]$. Let A satisfy (7.2). Then for μ almost all $x \in \mathbb{M}$

$$\frac{\sum_{n=0}^{N-1} A(\mathcal{G}^n \mathcal{H}_U x)}{\sqrt{N}}$$

converges as $N \rightarrow \infty$ to a normal random variable \mathcal{Z} variable with zero mean and variance

$$\sigma^2 = \sum_{n=-\infty}^{\infty} \int_{\mathbb{M}} A(y) A(\mathcal{G}^n y) d\mu(y).$$

Moreover for each ε, D there is a constant C such that

$$\mu \left(x : \sup_{z \in \mathbb{R}} \left| \mathbb{P} \left(\frac{\sum_{n=0}^{N-1} A(\mathcal{G}^n \mathcal{H}_U x)}{\sqrt{N}} \leq z \right) - \mathbb{P}(\mathcal{Z} \leq z) \right| > \varepsilon \right) \leq \frac{C}{N^D}.$$

The constant C can be chosen uniformly when L varies over an interval $[\underline{L}, \bar{L}]$ for some $0 < \underline{L} < \bar{L}$.

We note that (ii) is automatic if \mathbb{M} is compact. For non-compact manifolds this assumption means that \mathcal{G} is compatible with the chosen C^r norm on \mathbb{M} in the sense that the composition with \mathcal{G} is a bounded operator.

The proof of Theorem 7.1 is similar to but easier than the proof of Theorems 2.3 and 6.3. Namely since A is bounded we only need to check conditions (H1)–(H3) of Theorem 4.1.

Fix a partition Π of \mathbb{M} into \mathcal{H}_u orbit segments of size L . Given x the element of Π containing x is of the form $\{\mathcal{H}_v x\}_{-u(x) \leq v < w(x)}$ for some positive numbers $u(x), w(x)$. Let Π_x denote the partition of \mathbb{M} of the form $\mathcal{H}_{u(x)} \Pi$. As in Section 6.3 we let \mathcal{P}_x^l be the subpartition of Π_x into segments of size $\delta_n 2^{-l}$.

Consider the following collections.

$$(7.4) \quad (\{k + l\}_{k \geq R_1(\log_2 n + j)}, [A \cdot (A \circ \mathcal{G}^j)])$$

and

$$(7.5) \quad (\{k + l\}_{k \geq R_3 \log_2 n}, \{A\}).$$

We say that x is N -good if for each $l \leq N$ the partition \mathcal{P}_x^l is representative with respect to families (7.4) and (7.5). Lemma 6.4 is easily extended to show that for each D ,

$$\mathbb{P}(x \text{ is not } N\text{-good}) \leq \frac{C}{N^D}$$

provided that R_1, R_3 are large enough. Hence almost every x is N -good for all sufficiently large N . Next let \mathcal{F}_i^x be the filtration corresponding to \mathcal{P}_x^l . If x is N -good then $\{\mathcal{F}_i^x\}_{i \leq N}$ satisfies the conditions (H1)–(H3) of Theorem 4.1 which implies the CLT in view of Theorem 4.2.

We note the following consequence of Theorem 7.1.

COROLLARY 7.2. *Let G be a semisimple Lie group without compact factors, and $\Gamma \subset G$ be an irreducible lattice. Let h_u be a unipotent subgroup which is expanded by an element $g \in G$ in the sense that*

$$g^n h_u = h_{e^{cn}u} g^n$$

for some $c > 0$. Fix $L > 0$ and let U be a random variable uniformly distributed on $[0, L]$. Let A be a C^α function for some $\alpha > 0$ with zero mean. Then for Haar almost all $g_0 \in G/\Gamma$

$$\frac{\sum_{n=0}^{N-1} A(g^n h_U g_0)}{\sqrt{N}}$$

converges as $N \rightarrow \infty$ to a normal random variable \mathcal{Z} with zero mean and variance

$$\sigma^2 = \sum_{n=-\infty}^{\infty} \int_{G/\Gamma} A(g_0) A(g^n g_0) d\mu(g_0).$$

Moreover for each ε, D there is a constant C such that

$$\mu \left(g_0 : \sup_{z \in \mathbb{R}} \left| \mathbb{P} \left(\frac{\sum_{n=0}^{N-1} A(g^n h_U g_0)}{\sqrt{N}} \leq z \right) - \mathbb{P}(\mathcal{Z} \leq z) \right| > \varepsilon \right) \leq \frac{C}{N^D}.$$

The constant C can be chosen uniformly when L varies over an interval $[\underline{L}, \bar{L}]$ for some $0 < \underline{L} < \bar{L}$.

Corollary 7.2 follows from Theorem 7.1 with $\mathbb{M} = G/\Gamma$ and \mathcal{G} and \mathcal{H} actions of g and h respectively. To apply the Theorem we need to check the polynomial mixing for C^α functions. If r_0 is large enough then (namely $r_0 = 2l$ where l is the constant from [22, Theorem 3.4]) then (7.3) with C^{r_0} functions follows from [22]. For $\alpha < r_0$ we use a standard approximation argument. Let \mathfrak{g} be the Lie algebra of G , $k = \dim(G)$, and $\phi: \mathfrak{g} \rightarrow \mathbb{R}$ be a nonnegative C^∞ function with integral 1. Set

$$A_\varepsilon(x) = \int_{\mathfrak{g}} A(\exp(z)x) \phi\left(\frac{z}{\varepsilon}\right) \frac{dz}{\varepsilon^k}.$$

Then,

$$\|A_\varepsilon - A\|_{C^0} \leq C_1 \varepsilon^{\min\{1, \alpha\}} \|A\|_{C^{\min\{1, \alpha\}}}, \quad \|A_\varepsilon\|_{C^{r_0}} \leq C_2 \|A\|_{C^0} \varepsilon^{-r_0}.$$

Therefore we have

$$|\mu(AA \circ \mathcal{H}_u)| \leq |\mu(A_\varepsilon A_\varepsilon \circ \mathcal{H}_u)| + C_3 \varepsilon^{\min\{1, \alpha\}} \|A\|_{C^\alpha}^2 \leq C_4 (\varepsilon^{-2r_0} u^{-\kappa} + \varepsilon^{\min\{1, \alpha\}}) \|A\|_{C^\alpha}^2.$$

Choosing ε appropriately as a function of u^κ we obtain (7.3) with smaller κ .

8. VARIANCES

8.1. Variance of U_N . Here we establish (1.8). We prove the formula for σ_1^2 , the computation for σ_2^2 is the same. By (4.4) it suffices to compute

$$\sigma_1^2 = \lim_{N \rightarrow \infty} \frac{1}{\ln N} \text{Var} \left(\sum_{n=0}^{\lfloor \log_2 N \rfloor - 1} \mathcal{S}(\mathbb{1}_{E_c})(g^n \mathcal{L}) \right)$$

where the variance is taken with respect to the Haar measure on the space of lattices. Note that $\sum_{n=0}^{\lfloor \log_2 N \rfloor - 1} \mathcal{S}(\mathbb{1}_{E_c})(g^n \mathcal{L})$ can be replaced by $\mathcal{S}(\mathbb{1}_{E_c(N)})(\mathcal{L})$ where

$$E_c(N) = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^r : |x| \in [1, N], |x|^{d/r} y_j \in [0, c]\}.$$

Now Proposition 5.1(b) gives

$$\begin{aligned} \sigma_1^2 &= \lim_{N \rightarrow \infty} \frac{1}{\ln N} \sum_{\gcd(p,q)=1} \int_{\mathbb{R}^{d+r}} \mathbb{1}_{E_c(N)}(px, py) \mathbb{1}_{E_c(N)}(qx, qy) dx dy \\ &= 2 \lim_{N \rightarrow \infty} \frac{1}{\ln N} \sum_{\substack{\gcd(p,q)=1 \\ p < q}} \int_{\mathbb{R}^{d+r}} \mathbb{1}_{E_c(N)}(px, py) \mathbb{1}_{E_c(N)}(qx, qy) dx dy \end{aligned}$$

Since $p < q$, the last integral equals to

$$\begin{aligned} &\int_{\mathbb{R}^{d+r}} \mathbb{1}_{[1,N]}(|px|) \mathbb{1}_{[1,N]}(|qx|) \prod_j (\mathbb{1}_{[0,c]}(p^{1+d/r} |x|^{d/r} y_j) \mathbb{1}_{[0,c]}(q^{1+d/r} |x|^{d/r} y_j)) dx dy \\ &= \int_{\mathbb{R}^{d+r}} \mathbb{1}_{[1/p, N/q]}(|x|) \prod_j (\mathbb{1}_{[0,c]}(q^{1+d/r} |x|^{d/r} y_j)) dx dy = \frac{c^r}{q^{d+r}} \int_{\mathbb{R}^d} \mathbb{1}_{[1/p, N/q]}(|x|) \frac{dx}{|x|^d} \end{aligned}$$

To evaluate the last integral we pass to the polar coordinates $x = \rho s$ where s is a unit vector in the Euclidean norm. Then,

$$\int_{\mathbb{R}^d} \mathbb{1}_{[1/p, N/q]}(|x|) \frac{dx}{|x|^d} = \int_{\mathbb{S}^{d-1}} ds \int_{1/p|s|}^{N/q|s|} \frac{\rho^{d-1} d\rho}{|s|^d \rho^d} = \ln \left(\frac{Np}{q} \right) \int_{\mathbb{S}^{d-1}} \frac{ds}{|s|^d}.$$

The second factor here equals to

$$\int_{\mathbb{S}^{d-1}} \frac{ds}{|s|^d} = d \int_{\mathbb{S}^{d-1}} ds \int_0^{1/|s|} \rho^{d-1} d\rho = d \int_{|x| < 1} dx = d \text{Vol}(\mathcal{B}).$$

Therefore,

$$\sigma_1^2 = 2c^r d \sum_{q=1}^{\infty} \frac{\varphi(q)}{q^{d+r}} \text{Vol}(\mathcal{B}).$$

By [16, Theorem 288]

$$\sum_{q=1}^{\infty} \frac{\varphi(q)}{q^{d+r}} = \frac{\zeta(d+r-1)}{\zeta(d+r)}$$

so

$$\sigma_1^2 = 2c^r d \frac{\zeta(d+r-1)}{\zeta(d+r)} \text{Vol}(\mathcal{B})$$

as claimed in Theorem 1.1.

8.2. **Variance of V_N .** Here we compute the limiting variance for V_N . As in Section 8.1 we consider the case of boxes, the computations for balls being similar.

The same computation as in Section 8.1 shows that we need to compute

$$\lim_{N \rightarrow \infty} \frac{1}{\widehat{V}_N} \text{Var} \left(\mathcal{S}(\mathbb{1}_{E_c(N)})(\tilde{\mathcal{L}}) \right)$$

where the variance is taken with respect to the Haar measure on the space of affine lattices. By Proposition 5.1(d) this variance equals to

$$\int_{\mathbb{R}^{d+r}} [\mathbb{1}_{E_c(N)}]^2(x, y) dx dy = \int_{\mathbb{R}^{d+r}} \mathbb{1}_{E_c(N)}(x, y) dx dy = \widehat{V}_N.$$

REMARK 8.1. The fact that the variance of V_N has a simpler form than the variance of U_N has the following explanation. Let

$$\eta_k = \mathbb{1}_{B(k, d, r, c)}(k\mathbf{a}), \quad \tilde{\eta}_k = \mathbb{1}_{B(k, d, r, c)}(\mathbf{x} + k\mathbf{a})$$

so that

$$U_N = \sum_{|k| < N} \eta_k, \quad V_N = \sum_{|k| < N} \tilde{\eta}_k.$$

Then $\tilde{\eta}_k$'s are pairwise independent (even though triples $\tilde{\eta}_{k'}, \tilde{\eta}_{k''}, \tilde{\eta}_{k'''}$ are strongly dependent) and hence uncorrelated (see e.g. [31]) while η_k 's are not pairwise independent.

APPENDIX A. TRUNCATION AND NORMS

For a fixed dimension $p \in \mathbb{N}$, we denote by \mathcal{M} the space of p dimensional lattices. We let $C^s(\mathcal{M})$ denote the space of smooth functions on \mathcal{M} . Namely, let $\mathfrak{U}_1, \mathfrak{U}_2, \dots, \mathfrak{U}_{p^2-1}$ be a basis in the space of left invariant vector fields on \mathcal{M} . We let

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

$C^s(\mathcal{M})$ is the space of functions with finite $\|\cdot\|_{C^s}$ -norm. The space $C^s(\tilde{\mathcal{M}})$ of smooth functions on the space of r -dimensional affine lattices is defined similarly.

We have the following inequality:

$$(A.1) \quad \|\Psi\Phi\|_{C^s} \leq C \|\Psi\|_{C^s} \|\Phi\|_{C^s}.$$

Below we provide an extension to approximately smooth functions.

LEMMA A.1. *There is a constant C such that if Φ_1, Φ_2 are $C^{s,r}$ functions on \mathcal{M} or $\tilde{\mathcal{M}}$ then $\Phi_1\Phi_2$ is a $C^{s,2r}(\mathcal{M})$ function and*

$$(A.2) \quad \|\Phi_1\Phi_2\|_{C^{s,2r}} \leq C \|\Phi_1\|_{C^{s,r}} \|\Phi_2\|_{C^{s,r}}.$$

Proof. Suppose first that $1 \leq \Phi_j \leq 2$. Given ε let Φ_j^\pm be the functions such that

$$\Phi_j^- \leq \Phi_j \leq \Phi_j^+, \quad \|\Phi_j^\pm\|_{C^s} \leq 2\varepsilon^{-r}, \quad \|\Phi_j^+ - \Phi_j^-\|_{L^1} \leq \varepsilon.$$

Without the loss of generality we may assume that

$$0 \leq \Phi_j^-, \quad \Phi_j^+ \leq 3,$$

since otherwise we can replace Φ_j^\pm by $\chi(\Phi_j^\pm)$ where χ is an appropriate cutoff function. Then

$$\Phi_1^- \Phi_2^- \leq \Phi_1 \Phi_2 \leq \Phi_1^+ \Phi_2^+, \quad \|\Phi_1^+ \Phi_2^+ - \Phi_1^- \Phi_2^-\| \leq 6\varepsilon, \quad \|\Phi_1^\pm \Phi_2^\pm\|_{C^s} \leq C\varepsilon^{-2r}.$$

This proves the result in case $1 \leq \Phi_j \leq 2$. To obtain the result without this assumption we may suppose without the loss of generality that

$$\|\Phi_1\|_{C^{s,r}} = \|\Phi_2\|_{C^{s,r}} = \frac{1}{2}.$$

Split $\Phi_j = \tilde{\Phi}_j - \tilde{\tilde{\Phi}}_j$ where $\tilde{\Phi}_j = \frac{3}{2}$, $\tilde{\tilde{\Phi}}_j = \frac{3}{2} - \Phi_j$. Then both $\tilde{\Phi}_j$ and $\tilde{\tilde{\Phi}}_j$ are between 1 and 2. Thus we could apply the foregoing discussion to each term of the product $(\tilde{\Phi}_1 - \tilde{\tilde{\Phi}}_1)(\tilde{\Phi}_2 - \tilde{\tilde{\Phi}}_2)$ and obtain (A.2) in the general case \square

Let

$$\mathbf{a}(\mathcal{L}) = \max\{(\text{covol}(\bar{\mathcal{L}}))^{-1} : \bar{\mathcal{L}} \subset \mathcal{L}\}.$$

The role of this function is explained by the following lemmata.

LEMMA A.2. *For each sufficiently large R there is a constant $C_1 = C_1(R)$ such that if f is supported on the ball of radius R centered at the origin, then*

$$(A.3) \quad \mathcal{S}(f)(\mathcal{L}) \leq C_1 \mathbf{a}(\mathcal{L})$$

and

$$(A.4) \quad \tilde{\mathcal{S}}(f)(\mathcal{L} + \mathbf{x}) \leq C_1 \mathbf{a}(\mathcal{L}).$$

Also

$$(A.5) \quad \mathbf{a}(\mathcal{L}) \leq C_1 \mathcal{S}(\mathbb{1}_{B(0,R)})(\mathcal{L}).$$

Proof. (A.3) and (A.5) are taken from ([23, Lemma 5.1]). (A.4) follows from (A.3). Indeed suppose that $\tilde{\mathcal{S}}(f)(\mathcal{L} + \mathbf{x}) \neq 0$. Then there exists $\bar{e} \in \mathcal{L} + \mathbf{x}$ such that $f(\bar{e}) \neq 0$. Now we have

$$\tilde{\mathcal{S}}(f)(\mathcal{L} + \mathbf{x}) = \mathcal{S}(\tau_{\bar{e}} f)(\mathcal{L})$$

where $\tau_{\bar{e}}(f)(e) = f(e + \bar{e})$. Applying (A.3) to $\tau_{\bar{e}}(f)$ we get (A.4). \square

LEMMA A.3. *There is a constant C_2 such that*

$$\mu(\mathcal{L} : \mathbf{a}(\mathcal{L}) > t) \leq \frac{C_2}{t^{d+r}}.$$

Proof. The proof follows from (A.5) and the estimate

$$\mu(\mathcal{L} : \mathcal{S}(\mathbb{1}_{B(0,R)}) > t) \leq \frac{C_2}{t^{d+r}}$$

given in [25, Theorem 4.5]. \square

LEMMA A.4. [22, 12] *For each \mathbf{s} there are constants C_3, C_4 such that for each $K \geq 1$ there is a function $\mathfrak{h}_{1,K} : \mathcal{M} \rightarrow \mathbb{R}$ such that*

- (C1) $0 \leq \mathfrak{h}_{1,K} \leq 1$,
- (C2) $\mathfrak{h}_{1,K}(\mathcal{L}) = 1$ if $\mathbf{a}(\mathcal{L}) \geq K$,
- (C3) $\mathfrak{h}_{1,K}(\mathcal{L}) = 0$ if $\mathbf{a}(\mathcal{L}) \leq C_3 K$,

(C4) $\|\mathfrak{h}_{1,K}\|_{C^s(\mathcal{M})} \leq C_4$.

For example, one can take

$$\mathfrak{h}_{1,K} = \int_{SL_p(\mathbb{R})} G(g) \mathbb{1}_{\mathfrak{a}(gL) > C_3 K} (g\mathcal{L}) d\mu(g),$$

where G is a non negative function with integral one supported on the set $C_3^{-1} \leq \|g\| \leq C_3$. We write $\mathfrak{h}_2 = 1 - \mathfrak{h}_1$. We can also regard \mathfrak{h}_j as functions on $\tilde{\mathcal{M}}$ defined by the formula $\mathfrak{h}_j(\mathcal{L}, \mathbf{x}) = \mathfrak{h}_j(\mathcal{L})$. We are ready now to give the proofs of the statements from Section 5.3.

Proof of Lemma 5.4. We prove the estimates for \mathcal{S} , the estimates for $\tilde{\mathcal{S}}$ are similar.

(a) We have

$$|\mathcal{S}(f)(\mathcal{L})| \mathfrak{h}_{2,K}(\mathcal{L}) \leq \mathbb{1}_{\mathfrak{a}(\mathcal{L}) \leq K} |\mathcal{S}(f)|(\mathcal{L}) \leq CK$$

where the last step uses Lemma A.2(a). The derivatives of $\mathcal{S}(f)$ are estimated similarly using the formula

$$\partial_{\mathfrak{u}}(\mathcal{S}(f)) = \mathcal{S}(\partial_{\mathfrak{u}} f) \text{ where } (\partial_{\mathfrak{u}} f)(\mathbf{x}) = \left. \frac{d}{dt} \right|_{t=0} f(e^{t\mathfrak{u}} \mathbf{x}).$$

(b) Given ε consider functions f^\pm such that $f^- \leq f \leq f^+$ and (2.1) holds. Then $\mathcal{S}(f^-) \mathfrak{h}_{2,K} \leq \mathcal{S}(f) \mathfrak{h}_{2,K} \leq \mathcal{S}(f^+) \mathfrak{h}_{2,K}$ so the result follows from already established part (a) and Proposition 5.1(a).

(c) We already know from part (b) that

$$\|\mathcal{S}(f) \mathfrak{h}_{2,K}\|_{C^{s,r}} = O(K).$$

A similar argument shows that

$$\|[\mathcal{S}(f) \mathfrak{h}_{2,K}] \circ g^j\|_{C^{s,r}} = O(2^{js} K).$$

Now the result follows by Lemma A.1. \square

Proof of Lemma 5.5. Property (C2) of Lemma A.4 and Lemma A.3 imply that

$$\begin{aligned} |\mathbb{E}(\Phi - \Phi^K)| &\leq C \int_{\mathfrak{a}(\mathcal{L}) \geq K/C_3} |\Phi(\mathcal{L})| d\mu \\ &\leq C \int_{\mathfrak{a}(\mathcal{L}) \geq K/C_3} \mathfrak{a}(\mathcal{L}) d\mu \\ &\leq \frac{C}{K^{d+r-1}} \end{aligned}$$

which gives (5.1), and

$$\begin{aligned} \mathbb{E}((\Phi - \Phi^K)^2) &\leq C \int_{\mathfrak{a}(\mathcal{L}) \geq K/C_3} \Phi^2(\mathcal{L}) d\mu \\ &\leq C \int_{\mathfrak{a}(\mathcal{L}) \geq K/C_3} \mathfrak{a}^2(\mathcal{L}) d\mu \\ &\leq \frac{C}{K^{d+r-2}} \end{aligned}$$

which gives (5.2).

Now we deal with the case of affine lattices and $(r, d) = (1, 1)$. Let \mathcal{L} be such that $\mathfrak{a}(\mathcal{L}) = t \gg 1$. We claim that

$$(A.6) \quad \int_{\mathbb{R}^2/\mathcal{L}} \Phi(\mathcal{L}, \mathbf{x}) d\mathbf{x} \leq C, \quad \int_{\mathbb{R}^2/\mathcal{L}} \Phi^2(\mathcal{L}, \mathbf{x}) d\mathbf{x} \leq Ct.$$

This gives the required improvement of an extra power of t that is sufficient to verify (5.3) and (5.4) using Lemma A.3.

To show (A.6) let e_1 be the shortest vector in \mathcal{L} . Note that $|e_1|$ is of order $1/t$. Thus \mathcal{L} is contained in a union of lines going in the direction of e_1 so that the distance between the lines is almost t . If we shift \mathbf{x} in the direction perpendicular to e_1 then the probability that one of the shifted lines intersects the ball of a fixed radius around the origin is $O(1/t)$. Since $\Phi(\mathcal{L}, \mathbf{x}) = O(t)$ due to (A.4), the estimate (A.6) follows.

We now show that (5.1) and (5.2) hold if Haar measure is replaced by a measure having a C -centrally smoothable density with respect to Haar measure. We just prove (5.1) in the lattice case, since the proofs of (5.2) as well as the proofs for affine lattices are exactly the same.

$$\begin{aligned} |\mathbb{E}_\rho(\Phi - \Phi^K)| &= \int_{\mathcal{M}} 1_{\Phi(g^l \mathcal{L}) > K} |\Phi(g^l \mathcal{L})| \rho(\mathcal{L}) d\mu \\ &\leq C \int_{\mathcal{M}} 1_{\mathfrak{a}(g^l \mathcal{L}) > K/C} \mathfrak{a}(g^l \mathcal{L}) \rho(\mathcal{L}) d\mu \end{aligned}$$

where the inequality follows from Lemma A.2 since $\Phi = \mathcal{S}(f)$ with f having a compact support. Next, the K -central smoothability of ρ and Lemma A.3 imply that

$$\begin{aligned} \int_{\mathcal{M}} 1_{\mathfrak{a}(g_l \mathcal{L}) > K/C} \mathfrak{a}(g_l \mathcal{L}) \rho(\mathcal{L}) d\mu &= \int_{\mathcal{A}} \phi(a) \int_{\mathcal{M}} 1_{\mathfrak{a}(g_l a \mathcal{L}) > K/C} \rho(a \mathcal{L}) \mathfrak{a}(g_l(a \mathcal{L})) d\mu da \\ &\leq C \int_{\mathcal{M}} 1_{\mathfrak{a}(g_l \mathcal{L}) > K/C} \mathfrak{a}(g_l \mathcal{L}) \left(\int_{\mathcal{A}} \phi(a) \rho(a \mathcal{L}) da \right) d\mu \\ &\leq C \int_{\mathcal{M}} 1_{\mathfrak{a}(g_l \mathcal{L}) > K/C} \mathfrak{a}(g_l \mathcal{L}) d\mu \\ &\leq C \int_{\mathfrak{a}(\mathcal{L}) > K/C} \mathfrak{a}(\mathcal{L}) d\mu \\ &\leq CK^{1-(d+r)}. \end{aligned}$$

Inequality (5.1) is thus proved.

As for the case of affine lattices and $(r, d) = (1, 1)$, (5.3) and (5.4) can be proved as in the case of Haar measure, if one makes the following two observations:

- (1) Equation (A.6) still holds for a measure with density ρ ;
- (2) The tail estimate of Lemma A.3 can be proved for measures with centrally smoothable densities following the same lines as the proof of (5.1).

□

Proof of Lemma 5.6. Let U be the set of points obtained by issuing local center-stable manifolds through all points of Π . That is

$$U = \bigcup_{\tilde{\mathcal{L}}' \in \Pi, |\sigma| \leq 1, |\mathbf{b}| \leq 1} D_t \bar{\Lambda}_{\mathbf{b}} \tilde{\mathcal{L}}'.$$

Let $\mu_{\mathbb{L}}$ and $\mu_{\mathbb{G}}$ denote the Haar measures on \mathbb{L} and \mathbb{G} respectively. Then

$$\begin{aligned} \mathbb{P}_{\Pi}(|\Phi(g^l \tilde{\mathcal{L}}')| > K) &\leq C_1 \mu_{\mathbb{G}}(\hat{\mathcal{L}} \in U : |\mathbf{a}(g^l \hat{\mathcal{L}})| > K) \\ &\leq C_2 \mu_{\mathbb{G}}(\hat{\mathcal{L}} \in \tilde{\mathcal{M}} : |\mathbf{a}(g^l \hat{\mathcal{L}})| > K) \\ (A.7) \quad &\leq \frac{C}{K^{d+r}}. \end{aligned}$$

Similarly for $q \in \{1, 2\}$,

$$\begin{aligned} \int_{\Pi} |\Phi_l(\tilde{\mathcal{L}}') - \Phi_l^K(\tilde{\mathcal{L}}')|^q d\mu_{\mathbb{L}}(\tilde{\mathcal{L}}') &\leq C \int_{\Pi} |\Phi(g^l \tilde{\mathcal{L}}')|^q \mathbb{1}_{\mathbf{a}(g^l \tilde{\mathcal{L}}') > K} d\mu_{\mathbb{L}}(\tilde{\mathcal{L}}') \\ &\leq C \int_U |\mathbf{a}(g^l \tilde{\mathcal{L}}')|^q \mathbb{1}_{\mathbf{a}(g^l \tilde{\mathcal{L}}') > K} d\mu_{\mathbb{G}}(\tilde{\mathcal{L}}') \end{aligned}$$

where the last step follows from (A.4). Since the integrand depends only on projection of $\tilde{\mathcal{L}}'$ to \mathcal{M} the integral can be estimated by

$$C \int_{\mathcal{M}} |\mathbf{a}(g^l \mathcal{L})|^q \mathbb{1}_{\mathbf{a}(g^l \mathcal{L}) > K} d\mu_{\mathbb{G}}(\mathcal{L}) = C \int_{\mathcal{M}} |\mathbf{a}(\mathcal{L})|^q \mathbb{1}_{\mathbf{a}(\mathcal{L}) > K} d\mu_{\mathbb{G}}(\mathcal{L}).$$

Thus (5.6) and (5.7) follow Lemma A.3. \square

Proof of Lemma 6.2. We use the notation from the proof of Lemma 5.6 In view of (A.4)

$$\begin{aligned} \mathbb{P}_{\hat{\Pi}(\tilde{\mathcal{L}}^*)} \left(\tilde{\mathcal{S}}(f)(g^n \tilde{\mathcal{L}}) \neq \Phi^K(g^n \tilde{\mathcal{L}}) \quad \text{or} \quad \tilde{\mathcal{S}}(f)(g^n \hat{\mathcal{L}}) \neq \Phi^K(g^n \hat{\mathcal{L}}) \right) \\ \leq C_1 \mu_{\mathbb{G}}(\bar{\mathcal{L}} \in U : |\mathbf{a}(g^n \bar{\mathcal{L}})| > K) \end{aligned}$$

so the result follows from (A.7). \square

REFERENCES

- [1] J. S. Athreya, A. Ghosh, J. Tseng *Spiraling of approximations and spherical averages of Siegel transforms*, Journal LMS **91** (2015) 383–404.
- [2] J. S. Athreya, A. Ghosh, J. Tseng *Spherical averages of Siegel transforms for higher rank diagonal actions and applications*, arXiv:1407.3573
- [3] J. S. Athreya, A. Parrish, J. Tseng *Ergodic Theory and Diophantine approximation for linear forms and translation surfaces*, arXiv:1401.4148
- [4] Badziahin D., Beresnevich V., Velani S. *Inhomogeneous theory of dual Diophantine approximation on manifolds*, Adv. Math. **232** (2013) 1–35.
- [5] V. Beresnevich, V. Bernik, D. Kleinbock, G. A. Margulis, *Metric Diophantine approximation: The Khintchine–Groshev theorem for non-degenerate manifolds*, Moscow Math. J. **2** (2002), 203–225.
- [6] Beresnevich V., Velani S. *Classical metric Diophantine approximation revisited: the Khintchine–Groshev theorem*, IMRN (2010) no. 1, 69–86.
- [7] J. W. S. Cassels, *An introduction to Diophantine approximation*, Cambridge Tracts in Math., vol. 45, Cambridge Univ. Press, Cambridge, 1957.

- [8] Cornfeld I. P., Fomin S. V., Sina, Ya. G. *Ergodic theory*, Grundlehren der Mathematischen Wissenschaften **245** (1982) Springer, New York, x+486 pp.
- [9] Dani S. G. *Divergent trajectories of flows on homogeneous spaces and Diophantine approximation*, J. Reine Angew. Math. **359** (1985) 55–89.
- [10] D. Dolgopyat *Limit Theorems for partially hyperbolic systems*, Trans. AMS **356** (2004) 1637–1689.
- [11] D. Dolgopyat, B. Fayad, *Limit Theorems for toral translations*, Proc. Symp. Pure Math. **89** (2015) 227–277.
- [12] D. Dolgopyat, B. Fayad *Deviations of Ergodic sums for Toral Translations: Boxes*, preprint.
- [13] Duffin R. J., Schaeffer A.C. *Khintchine’s problem in metric Diophantine approximation*, Duke Math. J. **8** (1941) 243–255.
- [14] S. Edwards *The rate of mixing for diagonal flows on spaces of affine lattices*, preprint.
- [15] D. El-Baz, J. Marklof, I. Vinogradov *The distribution of directions in an affine lattice: two-point correlations and mixed moments*, IMRN **2015** (2015) 1371–1400.
- [16] G. H. Hardy, E. M. Wright *An Introduction to the Theory of Numbers*, 5th ed., (1979), Oxford University Press, Oxford, UK.
- [17] N. Haydn, M. Nicol, S. Vaienti, L. Zhang *Central limit theorems for the shrinking target problem*, J. Stat. Phys. **153** (2013) 864–887.
- [18] Groshev A. *A theorem on a system of linear forms*, Doklady Akademii Nauk SSSR **19** (1938) 151–152.
- [19] A. Khintchine *Ein Satz uber Kettenbruche, mit arithmetischen Anwendungen*, Mathematische Zeitschrift **18** (1923) 289–306.
- [20] A. Khintchine *Einige Satze uber Kettenbruche, mit Anwendungen auf die Theorie der Diophantischen Approximationen*, Math. Ann. **92** (1924) 115–125.
- [21] 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.
- [22] D. Y. Kleinbock, G. A. Margulis *Logarithm laws for flows on homogeneous spaces*, Invent. Math. **138** (1999) 451–494.
- [23] D. Kleinbock, R. Shi, B. Weiss *Pointwise equidistribution with an error rate and with respect to unbounded functions*, to appear in Math. Ann.; arXiv:1505.06717.
- [24] Le Borgne S. *Principes d’invariance pour les flots diagonaux sur $SL(d, \mathbb{R})/SL(d, \mathbb{Z})$* , Ann. Inst. H. Poincaré Prob. Stat. **38** (2002) 581–612.
- [25] J. Marklof *The n -point correlations between values of a linear form*, Erg. Th., Dynam. Sys. **20** (2000) 1127–1172.
- [26] J. Marklof *Distribution modulo one and Ratner’s theorem*, in *Equidistribution in number theory, an introduction*, NATO Sci. Ser. II Math. Phys. Chem., **237** (2007) 217–244, Springer, Dordrecht.
- [27] W. Philipp, *Mixing Sequences of Random Variables and Probabilistic Number Theory*, AMS Memoirs **114**, 1971.
- [28] Ratner M. *The central limit theorem for geodesic flows on n -dimensional manifolds of negative curvature*, Israel J. Math. **16** (1973) 181–197.
- [29] Samur J., *A functional central limit theorem in Diophantine approximation*, Proc. Amer. Math. Soc. **111** (1991) 901–911.
- [30] Schmidt W. M. *A metrical theorem in diophantine approximation*, Canadian J. Math. **12** (1960) 619–631.
- [31] Schmidt W. *Metrical theorems on fractional parts of sequences*, Trans. Amer. Math. Soc. **110** (1964) 493–518.
- [32] W. Schmidt, *Badly approximable systems of linear forms*, J. Number Theory **1** (1969), 139–154.
- [33] Sinai Ya. G. *The central limit theorem for geodesic flows on manifolds of constant negative curvature*, Soviet Math. Dokl. **1** (1960) 983–987.
- [34] A. Strombergsson *An Effective Ratner Equidistribution Result for $SL(2, \mathbb{R}) \times \mathbb{R}^2$* , Duke Math. J. **164** (2015), 843–902.