

CENTRAL LIMIT THEOREMS IN THE GEOMETRY OF NUMBERS

MICHAEL BJÖRKLUND AND ALEXANDER GORODNIK

ABSTRACT. We investigate in this paper the distribution of the discrepancy of various lattice counting functions. In particular, we prove that the number of lattice points contained in certain domains defined by products of linear forms satisfies a Central Limit Theorem. Furthermore, we show that the Central Limit Theorem holds for the number of rational approximants for weighted Diophantine approximation in \mathbb{R}^d . Our arguments exploit chaotic properties of the Cartan flow on the space of lattices.

1. INTRODUCTION

Let $\Omega_T \subset \mathbb{R}^d$ be an increasing family of compact domains, and let \mathcal{L}_d denote the space of lattices in \mathbb{R}^d with covolume one, endowed with the unique $SL_d(\mathbb{R})$ -invariant probability measure λ_d . We consider the counting function $\Lambda \mapsto |\Lambda \cap \Omega_T|$ on \mathcal{L}_d . Under mild assumptions on the domains Ω_T ,

$$|\Lambda \cap \Omega_T| \sim \frac{\text{vol}(\Omega_T)}{\text{vol}(\mathbb{R}^d/\Lambda)} \quad \text{as } T \rightarrow \infty, \text{ for all } \Lambda \in \mathcal{L}_d.$$

One may ask whether it is possible to derive more precise information about the asymptotic behavior of $|\Lambda \cap \Omega_T|$ for *generic* lattices Λ .

The following naive heuristics might give an idea of what to expect. Let us decompose

$$\Omega_T = \bigsqcup_{i=1}^N \Omega_T^{(i)}$$

into regions $\Omega_T^{(i)}$ with $\text{vol}(\Omega_T^{(i)}) \approx 1$. Then

$$|\Lambda \cap \Omega_T| = \sum_{i=1}^N |\Lambda \cap \Omega_T^{(i)}|,$$

and provided that $\Omega_T^{(i_1)}$ and $\Omega_T^{(i_2)}$ are "far apart", it seems plausible to conjecture that the random variables $\Lambda \mapsto |\Lambda \cap \Omega_T^{(j)}|$, for $j = 1, 2$, on \mathcal{L}_d are "almost independent". Thus, one might wonder whether $|\Lambda \cap \Omega_T|$ behaves like a sum of independent random variables.

Some classical results of W. Schmidt motivated our line of study. In [14, 15], Schmidt showed that for generic lattices $\Lambda \in \mathcal{L}_d$,

$$|\Lambda \cap \Omega_T| = \frac{\text{vol}(\Omega_T)}{\text{vol}(\mathbb{R}^d/\Lambda)} + O_{\Lambda, \varepsilon} \left(\text{vol}(\Omega_T)^{\frac{1}{2} + \varepsilon} \right) \quad \text{for all } \varepsilon > 0.,$$

and thus the counting function $\Lambda \mapsto |\Lambda \cap \Omega_T|$ indeed exhibits cancellations of the same order as a sum of independent random variables. Remarkably, the argument in [14] implicitly follows

Date: Central limit theorems, Diophantine approximation.

2010 Mathematics Subject Classification. Primary: 11H46 ; Secondary: 11K60, 60F05 .

the heuristic approach outlined above and proves some form of pairwise independence using arithmetic considerations.

The aim of this work is to establish a Central Limit Theorem (CLT) in this setting, at least under some additional assumptions on the domains Ω_T . We stress that it is unlikely that a CLT holds for general domains; for instance, the counting of lattice points in the regions

$$\Omega_T = \{(x, y) \in \mathbb{R}^2 : 0 < x < Ty \text{ and } 1 < y < 2\}$$

is closely related to the distribution of averages for the horocyclic flow on \mathcal{L}_2 , which do not admit a CLT (see e.g. [7]).

We shall in this paper consider domains defined by products of linear forms on \mathbb{R}^d . Such domains can be tessellated using images of a small number of regular tiles under a family of diagonal matrices in $SL_d(\mathbb{R})$, and allows us to use dynamical arguments developed in our recent work [4]. The crucial ingredients in our approach are quantitative estimates on higher-order correlations established in our joint work with Einsiedler [3]. Besides generic lattices in the space of lattices \mathcal{L}_d , we also consider the family of lattices

$$\Lambda_{\bar{x}} = \{(p_1 - qx_1, \dots, p_d - qx_d, q) : (\bar{p}, q) \in \mathbb{Z}^d \times \mathbb{Z}\}, \quad (1.1)$$

for $\bar{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, which arises in many problems in the theory of Diophantine approximation.

2. MAIN RESULTS

2.1. Distribution of values for products of linear forms

We fix a collection of linearly independent linear forms $L_1, \dots, L_d : \mathbb{R}^d \rightarrow \mathbb{R}$ with $d \geq 3$ and consider the product form

$$N(x) = L_1(x) \cdots L_d(x).$$

Our aim is to analyze the distribution of the values $N(x)$ when x belongs to a lattice in \mathbb{R}^d . We fix an interval $(a, b) \subset \mathbb{R}^+$, and for $T \geq 1$, we define the domains

$$\Omega_T = \{x \in \mathbb{R}^d : N(x) \in (a, b) \text{ and } |L_1(x)|, \dots, |L_d(x)| < T\}. \quad (2.1)$$

It is not hard to show that

$$\text{vol}(\Omega_T) = c(b - a)(\log T)^{d-1} + O((\log T)^{d-2}),$$

for some $c = c(L_1, \dots, L_d) > 0$, and by [15], almost all unimodular lattices Λ in \mathbb{R}^d satisfy

$$|\Lambda \cap \Omega_T| = \text{vol}(\Omega_T) + O_{\Lambda, \varepsilon} \left((\log T)^{\frac{d-1}{2} + \varepsilon} \right) \text{ for all } \varepsilon > 0.$$

We shall investigate how the error term (also known as the discrepancy) in this formula behaves.

Our first result shows that the error term admits a Central Limit Theorem. We have currently verified our argument for $d \geq 4$, but it might be possible to optimize the estimates to deal with the case $d = 3$ as well.

Theorem 2.1 (CLT for lattice counting). *For $d \geq 4$, there exists explicit $\sigma > 0$ such that for every $u \in \mathbb{R}$,*

$$\lambda_d \left(\left\{ \Lambda \in \mathcal{L}_d : \frac{|\Lambda \cap \Omega_T| - \text{vol}(\Omega_T)}{\text{vol}(\Omega_T)^{1/2}} < u \right\} \right) \rightarrow \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^u e^{-t^2/(2\sigma)} dt$$

as $T \rightarrow \infty$.

A more general version of this theorem can be established along similar lines. Instead of considering linear forms, let $L_i : \mathbb{R}^d \rightarrow \mathbb{R}^{d_i}$, $1 \leq i \leq k$, be a family of linear maps, and set

$$N(x) = \prod_{i=1}^k \|L_i(x)\|^{d_i},$$

where $\|\cdot\|$ denotes the Euclidean norm. We shall assume that $d_1 + \cdots + d_k = d$ and the map (L_1, \dots, L_k) defines a bijection $\mathbb{R}^d \rightarrow \prod_{i=1}^k \mathbb{R}^{d_i}$. For a fixed interval $(a, b) \subset \mathbb{R}^+$, we define the domains

$$\Omega_T = \{x \in \mathbb{R}^d : N(x) \in (a, b) \text{ and } \|L_1(x)\|, \dots, \|L_k(x)\| < T\}. \quad (2.2)$$

We show that a version of Theorem 2.2 still holds for these domains. In the case $k = 2$, such a result was also established in [5], using a different method, which does not seem to generalize to $k \geq 3$.

2.2. Spiralling

Motivated by the paper [1], we shall also study “spiralling” of the lattice points contained in the regions (2.2), that is, the distribution of their angular components. We denote by $\omega_i : \mathbb{R}^{d_i} \setminus \{0\} \rightarrow S^{d_i-1}$ the radial projections, and for a lattice Λ in \mathbb{R}^d and a Borel set $D \subset S$, we define

$$S_T(\Lambda, D) = \{x \in \Omega_T \cap \Lambda : (\omega_1(x), \dots, \omega_k(x)) \in D\}.$$

It is not hard to show (see [1]) that for almost every unimodular lattice Λ in \mathbb{R}^d , and for any Borel subset $D \subset S$,

$$\frac{|S_T(\Lambda, D)|}{\text{vol}(\Omega_T)} \rightarrow \text{vol}(D) \quad \text{as } T \rightarrow \infty.$$

We prove that if some regularity is imposed on D , then a suitable Central Limit Theorem also holds. Currently, the argument has been verified for $d \geq 4$, but it might be possible to optimize the estimates further to deal with the case $d = 3$.

Theorem 2.2 (CLT for spiralling). *For $d \geq 4$ and for every domain $D \subset S$ with piecewise smooth boundary, there exists explicit $\sigma = \sigma(D) > 0$ such that for every $u \in \mathbb{R}$,*

$$\lambda_d \left(\left\{ \Lambda \in \mathcal{L}_d : \frac{|S_T(\Lambda, D)| - \text{vol}(D)\text{vol}(\Omega_T)}{\text{vol}(\Omega_T)^{1/2}} < u \right\} \right) \rightarrow \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^u e^{-t^2/(2\sigma)} dt$$

as $T \rightarrow \infty$.

2.3. Diophantine approximation

Let us now discuss the distribution of integral solutions of some inequalities which arise in the theory of Diophantine approximation. We start with Diophantine approximation on the real line, which is better understood due to the theory of continued fractions. Fix $c > 0$, and for $x \in \mathbb{R}$, we consider the Diophantine inequality

$$\left| x - \frac{p}{q} \right| < \frac{c}{q^2} \quad (2.3)$$

with $(p, q) \in \mathbb{Z} \times \mathbb{N}$, and the corresponding counting function

$$N_T(x) = |\{(p, q) \in \mathbb{Z} \times \mathbb{N} : 1 \leq q \leq T \text{ and } \frac{p}{q} \text{ is a solution of (2.3)}\}|$$

It is known (see, for instance, [14]) that for almost every $x \in [0, 1]$,

$$N_T(x) = 2c \log T + O_{x,\varepsilon}((\log T)^{1/2+\varepsilon}), \quad \text{for all } \varepsilon > 0.$$

Fuchs showed in [8] that the discrepancy in this formula satisfies the Central Limit Theorem, that is to say, there exists $\sigma > 0$ such that for every $u \in \mathbb{R}$,

$$\left| \left\{ x \in [0, 1] : \frac{N_T(x) - 2c \log T}{(\log T \cdot \log \log T)^{1/2}} < u \right\} \right| \rightarrow \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^u e^{-t^2/(2\sigma)} dt \quad (2.4)$$

as $T \rightarrow \infty$. We stress that the correct normalization in (2.4) has caused some confusion in the previous works [10, 11, 12]; the additional $(\log \log T)$ -factor arises here because a certain counting function on \mathcal{L}_2 is *not* square-integrable. This non-integrability issue does not appear in higher dimensions, whence this additional normalization factor should disappear. An analogue of this result for simultaneous Diophantine approximation has been recently established in [5].

In this paper we consider the following more general problem in *weighted* Diophantine approximation. Let us fix a collection of weights

$$0 < w_1, \dots, w_d < 1 \quad \text{and} \quad w_1 + \dots + w_d = 1,$$

and constants $c_1, \dots, c_d > 0$. Given a vector $\bar{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, we are interested in understanding the asymptotics of solutions for the system of Diophantine inequalities defined by

$$\left| x_1 - \frac{p_1}{q} \right| < \frac{c_1}{q^{1+w_1}}, \quad \dots, \quad \left| x_d - \frac{p_d}{q} \right| < \frac{c_d}{q^{1+w_d}} \quad (2.5)$$

with $(\bar{p}, q) \in \mathbb{Z}^d \times \mathbb{N}$. The number of solutions is given by

$$N_T(\bar{x}) = |\{(\bar{p}, q) \in \mathbb{Z}^d \times \mathbb{N} : 1 \leq q < T, \text{ and (2.5) holds}\}|.$$

One can show, using Schmidt's arguments in [14], that for almost every $\bar{x} \in [0, 1]^d$,

$$N_T(\bar{x}) = 2^d c_1 \cdots c_d \log T + O_{x, \varepsilon}((\log T)^{1/2+\varepsilon}) \quad \text{for all } \varepsilon > 0. \quad (2.6)$$

We prove here that the error term in (2.6) satisfies the Central Limit Theorem. In the special case when all weights (w_i) are equal, this result was established in [5].

Theorem 2.3 (CLT for Diophantine approximation). *For $d \geq 2$, there exists explicit $\sigma > 0$ such that for every $u \in \mathbb{R}$,*

$$\text{Leb} \left(\left\{ \bar{x} \in [0, 1]^d : \frac{N_T(\bar{x}) - 2^d c_1 \cdots c_d \log T}{(\log T)^{1/2}} < u \right\} \right) \rightarrow \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^u e^{-t^2/(2\sigma)} dt$$

as $T \rightarrow \infty$.

In this note we outline the proofs of Theorem 2.1–2.3. Details will be published elsewhere.

3. INGREDIENTS IN THE PROOFS

3.1. The space of lattices and Siegel transforms

We denote by \mathcal{L}_d the space of lattices in \mathbb{R}^d with covolume one. We recall that \mathcal{L}_d can be realised as a homogeneous space $\mathcal{L}_d \simeq \text{SL}_d(\mathbb{R})/\text{SL}_d(\mathbb{Z})$, so that it is equipped with the unique invariant probability measure λ_d . Given a bounded Borel measurable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with compact support, we define its *Siegel transform* $\hat{f} : \mathcal{L}_d \rightarrow \mathbb{R}$ by

$$\hat{f}(\Lambda) = \sum_{v \in \Lambda \setminus \{0\}} f(v) \quad \text{for } \Lambda \in \mathcal{L}_d.$$

The starting point of our approach is the observation that the counting functions in Theorems 2.1–2.3 can be realized as certain averages of suitable Siegel transforms. This idea is simpler to explain in the setting of Theorem 2.3, so let us begin by focusing on this case. Let

$$\mathfrak{a} = \text{diag}(2^{w_1}, \dots, 2^{w_d}, 2^{-1}) \in \text{SL}_{d+1}(\mathbb{R})$$

and let χ denote the characteristic function of the domain

$$\{(\bar{x}, y) \in \mathbb{R}^d \times \mathbb{R} : 1 \leq y < 2 \text{ and (2.5) holds}\}.$$

Then one can readily check that for every $T = 2^N$ with $N \geq 1$,

$$N_T(\bar{x}) = \sum_{n=0}^{N-1} \hat{\chi}(\mathfrak{a}^n \Lambda_{\bar{x}}), \quad (3.1)$$

where the lattice $\Lambda_{\bar{x}}$ is defined in (1.1). This basic formula allows to study the distribution of $N_T(\bar{x})$ using dynamics on the space of lattices. More precisely, we shall use an approximation of the form (3.1) with χ replaced by a smooth function f_ε that approximates χ well in the L^1 - and L^2 -senses.

The counting functions in Theorem 2.1 and 2.2 can also be approximated along similar lines, but the formulas are more complicated; in particular, one-parameter subgroups of diagonal matrices are no longer enough to achieve an approximation of N_T as in (3.1). Let A_d denote the subgroup of diagonal matrices in $\text{SL}_d(\mathbb{R})$, and set $\theta_r = \text{diag}(1, \dots, 1, e^r)$. We shall show that for suitably chosen smooth compactly supported functions $f_{\varepsilon, T}$ on \mathbb{R}^d and finite subsets $B(r, T)$ of A_d ,

$$|\Lambda \cap \Omega_T| \approx \int_{\log \mathfrak{a}}^{\log b} \left(\sum_{\mathfrak{a} \in B(r, T)} \hat{f}_{\varepsilon, T}(\theta_r \mathfrak{a} \Lambda) \right) dr, \quad (3.2)$$

in the L^1 - and L^2 -norms for $(\mathcal{L}_d, \lambda_d)$. Our arguments from now on depend crucially on the fact (which will be explained in more detail below) that for a smooth compactly supported function ϕ on \mathcal{L}_d , the collections of functions

$$\{\phi(\mathfrak{a} \Lambda) : \mathfrak{a} \in A_d\}$$

are "weakly independent".

3.2. The method of cumulants

There exists today a plethora of different techniques to establish convergence to the Gaussian distribution. One of the first such techniques - if not the first - is nowadays often referred to as the *Method of Moments*, and was used by Chebyshev to prove the classical Central Limit Theorem. We refer to [2] for a modern exposition of this technique. An essentially equivalent technique, but better tailored for problems pertaining to Gaussian distributions, was later developed by Fréchet and Shohat, and goes under the name "Method of Cumulants". Let us briefly survey this method. Given bounded random variables X_1, \dots, X_r , their *joint cumulant* is defined by the curious expression

$$\text{Cum}^{(r)}(X_1, \dots, X_r) = \sum_{\mathcal{P}} (-1)^{|\mathcal{P}|-1} (|\mathcal{P}|-1)! \prod_{I \in \mathcal{P}} \mathbb{E} \left(\prod_{i \in I} X_i \right),$$

where the sum is taken over all partitions \mathcal{P} of the set $\{1, \dots, r\}$. We also set

$$\text{Cum}^{(r)}(X) = \text{Cum}^{(r)}(X, \dots, X)$$

for a single bounded random variable X . The cumulants have many useful combinatorial properties (see [16]). For instance, if there exists a non-trivial partition $\{1, \dots, r\} = I \sqcup J$ such that the collections $\{X_i : i \in I\}$ and $\{X_j : j \in J\}$ are independent of each other, then

$$\text{Cum}^{(r)}(X_1, \dots, X_r) = 0. \quad (3.3)$$

Furthermore, a bounded random variable X with mean zero is normally distributed if and only if $\text{Cum}^{(r)}(X) = 0$ for $r \geq 3$. In what follows, we shall use the following useful criterion due to Fréchet and Shohat [6] to establish our Central Limit Theorems.

Proposition 3.1 (Method of Cumulants). *Let (Z_T) be a collection of real-valued bounded random variables with mean zero satisfying*

$$\sigma^2 := \lim_{T \rightarrow \infty} \text{Var}(Z_T) < \infty, \quad (3.4)$$

and

$$\lim_{T \rightarrow \infty} \text{Cum}^{(r)}(Z_T) = 0, \quad \text{for all } r \geq 3. \quad (3.5)$$

Then for every $u \in \mathbb{R}$,

$$\text{Prob}(Z_T < u) \longrightarrow \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^u e^{-t^2/(2\sigma)} dt \quad \text{as } T \rightarrow \infty.$$

In our recent work [4], we used the Method of Cumulants to establish a general Central Limit Theorem for group actions which are exponentially mixing of all orders. Here we essentially follow the approach developed in [4], but substantial modifications will have to be made in order to handle more general averaging schemes, as well as unbounded test functions.

3.3. Estimates on the higher-order correlations

Let us now prepare the asymptotic formulas for higher-order correlations that will be used to estimate the cumulants and the variance. These formulas will be formulated in terms of Sobolev norms S_k , $k \geq 1$, defined for smooth compactly supported functions on the space \mathcal{L}_d (see [3]). In the proofs of Theorems 2.1 and 2.2, we will use estimates on correlations for the action on \mathcal{L}_d of the group of diagonal matrices $A_d \subset \text{SL}_d(\mathbb{R})$. If we fix a invariant metric ρ on $A_d \cong \mathbb{R}^{d-1}$, then the following result is a special case of [3, Th. 1.1].

Theorem 3.2 (Exponential multiple mixing of all orders). *For every $r \geq 2$, there exists an integer k_r such that for all $k \geq k_r$, there is $\delta_{r,k} > 0$ with the property that for all $\phi_1, \dots, \phi_r \in C_c^\infty(\mathcal{L}_d)$ and $a_1, \dots, a_r \in A_d$,*

$$\begin{aligned} \int_{\mathcal{L}_d} \phi_1(a_1\Lambda) \cdots \phi_r(a_r\Lambda) d\lambda_d(\Lambda) &= \left(\int_{\mathcal{L}_d} \phi_1 d\lambda_d \right) \cdots \left(\int_{\mathcal{L}_d} \phi_r d\lambda_d \right) \\ &\quad + O_{r,k} \left(e^{-\delta_{r,k} D(a_1, \dots, a_r)} S_k(\phi_1) \cdots S_k(\phi_r) \right), \end{aligned}$$

where $D(a_1, \dots, a_r) = \min\{\rho(a_i, a_j) : i \neq j\}$.

In order to study weighted Diophantine approximation (2.5), we need to analyze the distribution of orbits for the one-parameter semigroup

$$a_w(t) = \text{diag}(e^{w_1 t}, \dots, e^{w_d t}, e^{-t}), \quad t > 0,$$

for lattices contained in the subset

$$\mathcal{Y}_d = \{\Lambda_{\bar{x}} : \bar{x} \in [0, 1]^d\}$$

of \mathcal{L}_{d+1} . We denote by σ_d the measure \mathcal{Y}_d induced by the Lebesgue measure on $[0, 1]^d$. We establish the following asymptotic formula for the higher-order correlations of the measures $\alpha_w(t) * \sigma_d$, generalizing the work of Kleinbock and Margulis [9].

Theorem 3.3. *For every $r \geq 2$ and $k \geq k_r$, there exists $\delta'_{r,k} > 0$ such that for every $\phi_1, \dots, \phi_r \in C_c^\infty(\mathcal{L}_{d+1})$ and $t_1, \dots, t_r > 0$,*

$$\int_{\mathcal{Y}_d} \phi_1(\alpha_w(t_1)y) \cdots \phi_r(\alpha_w(t_r)y) d\sigma_d(y) = \left(\int_{\mathcal{L}_{d+1}} \phi_1 d\lambda_{d+1} \right) \cdots \left(\int_{\mathcal{L}_{d+1}} \phi_r d\lambda_{d+1} \right) + O_{r,k} \left(e^{-\delta'_{r,k} D'(t_1, \dots, t_r)} S_k(\phi_1) \cdots S_k(\phi_r) \right),$$

where $D'(t_1, \dots, t_r) = \min\{t_i, |t_i - t_j| : i \neq j\}$.

We note that \mathcal{Y}_d is an unstable manifold for the one-parameter semigroup

$$g(t) = \text{diag}(e^{t/d}, \dots, e^{t/d}, e^{-t}), \quad t > 0,$$

but not for semigroup $\alpha_w(t)$, unless the weights w_i are all equal. The quantitative equidistribution for the translates $\alpha_w(t)\mathcal{Y}_d$ has been established by Kleinbock and Margulis in [9]. We refine their argument to deal with higher-order correlations. The proof of Theorem 3.3 was inspired by [9]. It goes by induction on r and uses quantitative equidistribution of the measures $\alpha_w(t) * \sigma_d$ combined with non-divergence estimates for the unipotent flows.

From Theorem 3.3, we deduce the following non-divergence estimate:

Corollary 3.4 (Non-divergence). *Let f be a continuous compactly supported function on \mathbb{R}^d . Then there exists $c > 0$ such that for every $L \geq 1$ and $n \geq c \log L$,*

$$\sigma_d \left(\{y \in \mathcal{Y}_d : \widehat{f}(\alpha^n y) > L\} \right) \ll_f L^{-d-1}.$$

This corollary will be used to construct bounded approximations for Siegel transforms.

3.4. Bounded approximations

It might now be tempting to try to apply Proposition 3.1, combined with Theorems 3.2 and 3.3, to the approximations (3.1) and (3.2) directly. However, we stress that the Siegel transform \widehat{f} of a smooth compactly supported function f on \mathbb{R}^d gives an *unbounded* function on the space of lattices \mathcal{L}_d . Moreover, the Sobolev norms $S_k(f)$ in §3.3 are infinite. In order to deal with these issues, we shall use that $\widehat{f} \in L^p(\lambda_d)$ for $p < d$ and show that one can approximate \widehat{f} by a family of functions $\phi_L \in C_c^\infty(\mathcal{L}_d)$ satisfying

$$\|\phi_L\|_\infty = O(L) \quad \text{and} \quad S_k(\phi_L) = O_k(L^{d+1}), \quad (3.6)$$

and

$$\|\widehat{f} - \phi_L\|_1 = O_q(L^{-q}) \quad \text{for all } q < d-1, \quad (3.7)$$

and

$$\|\widehat{f} - \phi_L\|_2 = O_q(L^{-q}) \quad \text{for all } q < (d-2)/2. \quad (3.8)$$

This observation will allow us to exploit the estimates from Subsection 3.3 to analyze the variance and the cumulants of higher orders. In the setting of Theorems 2.1 and 2.2, we shall use

the approximation (3.2) and consider

$$Z_T(\Lambda) = \int_{\log a}^{\log b} \left(\sum_{\alpha \in B(r, T)} \widehat{f}_{\varepsilon, T}(\theta_r \alpha \Lambda) \right) dr,$$

and

$$Z_T^*(\Lambda) = \int_{\log a}^{\log b} \left(\sum_{\alpha \in B(r, T)} \phi_{\varepsilon, T}^L(\theta_r \alpha \Lambda) \right) dr,$$

where $\phi_{\varepsilon, T}^L$ is the bounded approximation for $\widehat{f}_{\varepsilon, T}$. Because of (3.7), the parameter $L = L(T)$ can be chosen so that

$$\|Z_T - Z_T^*\|_{L^1(\lambda_d)} \rightarrow 0 \quad \text{as } T \rightarrow \infty.$$

After this choice has been made, it suffices to analyze convergence in distribution of Z_T^* .

In the proof of Theorem 2.3, we consider

$$Z_T(y) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \left(\widehat{\chi}(a^n y) - \int_{y_d} \widehat{\chi}(a^n y) d\sigma_d(y) \right)$$

and its approximation

$$Z_T^*(y) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \left(\phi_{\varepsilon}^L(a^n y) - \int_{y_d} \phi_{\varepsilon}^L(a^n y) d\sigma_d(y) \right),$$

where ϕ_{ε}^L denotes the bounded approximation to $\widehat{f}_{\varepsilon}$ as above, and f_{ε} is a smooth approximation to the characteristic function χ (recall the notation from Subsection 3.1). Here the parameters $\varepsilon = \varepsilon(T)$ and $L = L(T)$ can be chosen so that

$$\|Z_T - Z_T^*\|_{L^1(\sigma_d)} \rightarrow 0 \quad \text{as } T \rightarrow \infty. \quad (3.9)$$

To arrange (3.9), we use the non-divergence estimate established in Corollary 3.4 and the following uniform bound

$$\sup_{n \geq 1} \|\widehat{f}_{\varepsilon} \circ a^n\|_{L^2(\sigma_d)} < \infty. \quad (3.10)$$

In order to prove (3.10), we interpret the L^2 -norm arithmetically and reduce this estimate to a problem of counting solutions of certain Diophantine equations.

Ultimately, we shall show that Z_T^* converges to the Normal Law using Proposition 3.1. Our main tool is the estimates on higher-order correlations from §3.3. We note the bounds in our computations will depend on the parameters ε, T, L , and thus the explicit forms of error terms in Theorems 3.2 and 3.3 are essential for this purpose.

3.5. Well-separated tuples and estimating the cumulants

By linearity, the estimates on cumulants arising in the proofs of Theorems 2.1 and 2.2 reduce to the following basic problem, which is discussed in more detail in our paper [4]. Given $\phi_1, \dots, \phi_r \in C_c^\infty(\mathcal{L}_d)$ and $(a_1, \dots, a_r) \in A_d^r$, we wish to estimate averages of cumulants of the form

$$\text{Cum}_{\lambda_d}^{(r)}(\phi_1 \circ a_1, \dots, \phi_r \circ a_r) = \sum_{\mathcal{P}} (-1)^{|\mathcal{P}|-1} (|\mathcal{P}|-1)! \prod_{I \in \mathcal{P}} \left(\int_{\mathcal{L}_d} \left(\prod_{i \in I} \phi_i \circ a_i \right) d\lambda_d \right), \quad (3.11)$$

as (a_1, \dots, a_r) varies over certain subsets of A_d^r .

The idea is to decompose A_d^r into finitely many regions where the cumulants can be estimated separately. These regions are defined as follows. Recall that ρ is a fixed invariant metric on A_d . For $I, J \subset [r]$ and $\bar{a} = (a_1, \dots, a_r) \in A_d^r$, we set

$$\rho^I(\bar{a}) = \max \{ \rho(a_i, a_j) : i, j \in I \} \quad \text{and} \quad \rho_{I,J}(\bar{a}) = \min \{ \rho(a_i, a_j) : i \in I, j \in J \}.$$

If \mathcal{Q} is a partition of $\{1, \dots, r\}$, we define

$$\rho^{\mathcal{Q}}(\bar{a}) = \max \{ \rho^I(\bar{a}) : I \in \mathcal{Q} \} \quad \text{and} \quad \rho_{\mathcal{Q}}(\bar{a}) = \min \{ \rho_{I,J}(\bar{a}) : I \neq J, I, J \in \mathcal{Q} \}.$$

For $0 \leq \alpha < \beta$, we define

$$\Delta_{\mathcal{Q}}(\alpha, \beta) = \{ \bar{a} \in A_d^r : \rho^{\mathcal{Q}}(\bar{a}) \leq \alpha, \text{ and } \rho_{\mathcal{Q}}(\bar{a}) > \beta \}$$

and

$$\Delta(\beta) = \{ \bar{a} \in A_d^r : \rho(a_i, a_j) \leq \beta \text{ for all } i, j \}.$$

We shall think of the tuples in $\Delta_{\mathcal{Q}}(\alpha, \beta)$, for some partition \mathcal{Q} with $|\mathcal{Q}| \geq 2$, as being “well-separated”, while we think of the tuples in $\Delta(\beta)$ as being “clustered”.

We estimate the cumulants on $\Delta_{\mathcal{Q}}(\alpha, \beta)$ using the exponential multiple mixing property established in Theorem 3.2.

Lemma 3.5 (Proposition 6.1, [4]). *For all $r \geq 3$ and $k > k_r + r$, there exist $c_{r,k}, \delta_{r,k} > 0$ such that for any partition \mathcal{Q} of $\{1, \dots, r\}$ with $|\mathcal{Q}| \geq 2$ and $s > 0$, we have*

$$|\text{Cum}_{\lambda_d}^{(r)}(\phi_1 \circ a_1, \dots, \phi_r \circ a_r)| \ll_{r,k} e^{-\delta_{r,k}s} S_k(\phi_1) \cdots S_k(\phi_r)$$

when $(a_1, \dots, a_r) \in \Delta_{\mathcal{Q}}(s, c_{r,k}s)$.

To prove this lemma, we introduce a cumulants “conditioned” on a given partition \mathcal{Q} . For partitions \mathcal{P} and \mathcal{Q} , we set $\mathcal{P} \wedge \mathcal{Q} = \{P \cap Q : P \in \mathcal{P}, Q \in \mathcal{Q}\}$. We define

$$\text{Cum}_{\lambda_d, \mathcal{Q}}^{(r)}(\phi_1 \circ a_1, \dots, \phi_r \circ a_r) = \sum_{\mathcal{P}} (-1)^{|\mathcal{P}|-1} (|\mathcal{P}|-1)! \prod_{J \in \mathcal{P} \wedge \mathcal{Q}} \left(\int_{\mathcal{L}_d} \left(\prod_{i \in J} \phi_i \circ a_i \right) d\lambda_d \right). \quad (3.12)$$

Comparing (3.11) and (3.12), we realize that they are approximately equal for tuples $(a_1, \dots, a_r) \in \Delta_{\mathcal{Q}}(\alpha, \beta)$ with suitably chosen α and β because

$$\int_{\mathcal{L}_d} \left(\prod_{i \in I} \phi_i \circ a_i \right) d\lambda_d \approx \prod_{K \in \mathcal{Q}} \int_{\mathcal{L}_d} \left(\prod_{i \in I \cap K} \phi_i \circ a_i \right) d\lambda_d, \quad \text{for all } I, K \subset [r],$$

according to Theorem 3.2. The second step in the proof of Lemma 3.5 utilizes the fact that when \mathcal{Q} is a non-trivial partition of $[r]$, then

$$\text{Cum}_{\lambda_d, \mathcal{Q}}^{(r)}(\phi_1 \circ a_1, \dots, \phi_r \circ a_r) = 0,$$

which is a combinatorial version of (3.3) (see Proposition 8.1 in [4]). This leads to the estimate in Lemma 3.5.

In order to apply Lemma 3.5, we decompose A_d^r into regions where the tuples (a_1, \dots, a_r) are “well-separated” or “clustered” on certain scales. We show (cf. [4, Prop. 6.2]) that for suitably chosen parameters

$$0 = \alpha_0 < \beta_0 < \alpha_1 < \cdots < \beta_{r-1} < \alpha_r,$$

we have a decomposition

$$A_d^r = \Delta(\alpha_r) \cup \left(\bigcup_{j=0}^{r-1} \bigcup_{|Q| \geq 2} \Delta_Q(\alpha_j, \beta_j) \right), \quad (3.13)$$

where the union is taken over the partitions Q of $\{1, \dots, r\}$ with $|Q| \geq 2$. It turns out possible to choose the parameters α_j, β_j in such a way that Lemma 3.5 can be applied to the averages of the cumulants over subsets of $\Delta_Q(\alpha_j, \beta_j)$ to conclude that they are negligible. Now it remains to estimate the average over a subset of $\Delta(\alpha_r)$. Since we can choose α_r quite small, the latter average can be estimated by bounding the number of terms.

The above argument requires some modifications for the proof of Theorem 2.3 because we need to take into account the estimator D' in Theorem 3.3. It will be convenient to embed A_d^r in A_d^{r+1} by $\bar{a} \mapsto (e, \bar{a})$ and define subsets $\Delta_Q(\alpha, \beta)$ of A_d^r with respect to this embedding for partitions Q of $\{0, 1, \dots, r\}$. As before, we use the decomposition (3.13). When the partition Q is non-trivial and different from $\{\{0\}, \{1, \dots, r\}\}$, we are able to modify the proof of Lemma 3.5 using Theorem 3.3 and estimate the cumulants $\text{Cum}_{\sigma_d}^{(r)}(\phi_1 \circ a^{n_1}, \dots, \phi_r \circ a^{n_r})$ when $(a^{n_1}, \dots, a^{n_r}) \in \Delta_Q(s, c_{r,k} s)$. When $Q = \{\{0\}, \{1, \dots, r\}\}$, we observe that Theorem 3.3 implies that

$$\text{Cum}_{\sigma_d}^{(r)}(\phi_1 \circ a^{n_1}, \dots, \phi_r \circ a^{n_r}) \approx \text{Cum}_{\lambda_d}^{(r)}(\phi_1 \circ a^{n_1}, \dots, \phi_r \circ a^{n_r})$$

when $(a^{n_1}, \dots, a^{n_r}) \in \Delta_Q(s, c_{r,k} s)$, and the latter cumulant has already been estimated.

Finally, we have to deal with the average over $(a^{n_1}, \dots, a^{n_r}) \in \Delta(\alpha_r)$. For this purpose, we modify the function Z_T^* in such a way that its convergence in distribution is not affected. We set

$$Z_T^{**}(y) = \frac{1}{\sqrt{N}} \sum_{n=M}^{N-1} \left(\phi_\varepsilon^L(a^n y) - \int_{y_d} \phi_\varepsilon^L(a^n y) d\sigma_d(y) \right),$$

where the parameter $M = M(N) \rightarrow \infty$ is chosen so that

$$\|Z_T^* - Z_T^{**}\|_{L^1(\sigma_d)} \rightarrow 0 \quad \text{as } T \rightarrow \infty.$$

In particular Z_T^* and Z_T^{**} have the same distributional limits, and thus it suffices to establish convergence in distribution of Z_T^{**} . Choosing $M = M(N)$ appropriately, we can further arrange so that averages over subsets of $\Delta(\alpha_r)$ in the cumulant calculations $\text{Cum}_{\sigma_d}^{(r)}(Z_T^{**})$ tend to zero.

3.6. Estimating the variance

In order to estimate the variance in the setting of Theorems 2.1 and 2.2, we need to consider sums of the form

$$\frac{1}{|B_T|} \sum_{a, b \in B_T} \int_{\mathcal{L}_d} \psi(a\Lambda) \psi(b\Lambda) d\lambda_d(\Lambda), \quad (3.14)$$

where B_T 's are finite subsets of a lattice $\Delta_d \subset A_d$ and $\psi = \phi - \int_{\mathcal{L}_d} \phi d\lambda_d$ with a smooth compactly supported function ϕ on \mathcal{L}_d (compare with (3.1)). Using invariance of the measure λ_d , we rewrite this expression as

$$\frac{1}{|B_T|} \sum_{a \in B_T, c \in a^{-1}B_T} \int_{\mathcal{L}_d} \psi(a\Lambda) \psi(ac\Lambda) d\lambda_d(\Lambda) = \sum_{c \in B_T^{-1}B_T} \frac{|B_T \cap B_T c^{-1}|}{|B_T|} \int_{\mathcal{L}_d} \psi(\psi \circ c) d\lambda_d.$$

Then using Theorem 3.2 with $r = 2$, we deduce that (3.14) converges to

$$\sum_{c \in \Delta_d} \int_{\mathcal{L}_d} \psi(\psi \circ c) \, d\lambda_d. \quad (3.15)$$

It should be noted that this argument have to be applied to the family of functions $\phi_{\varepsilon, T}^L \circ \theta_r$, introduced in §3.4, with suitably chosen parameters. The explicit form of the error term in Theorem 3.2 still allows to justify convergence of (3.14).

The computation of variance in Theorem 2.3 reduces to analysing the expressions

$$\begin{aligned} \frac{1}{N} \sum_{n, m=0}^{N-1} \left(\int_{y_d} \psi(n, y) \psi(m, y) \, d\sigma_d(y) \right) &= \frac{1}{N} \sum_{n=0}^{N-1} \left(\int_{y_d} \psi(n, y)^2 \, d\sigma_d(y) \right) \\ &+ \frac{2}{N} \sum_{0 \leq n < m \leq N-1} \left(\int_{y_d} \psi(n, y) \psi(m, y) \, d\sigma_d(y) \right), \end{aligned} \quad (3.16)$$

where $\psi(n, y) = \phi(a^n y) - \int_{y_d} (\phi \circ a^n) \, d\sigma_d$ for smooth compactly supported functions ϕ on \mathcal{L}_{d+1} . We shall show that (3.16) converges to

$$\sum_{k \in \mathbb{Z}} \left(\int_{\mathcal{L}_{d+1}} \phi(\phi \circ a^k) \, d\lambda_{d+1} - \left(\int_{\mathcal{L}_{d+1}} \phi \, d\lambda_{d+1} \right)^2 \right) \quad (3.17)$$

as $N \rightarrow \infty$. First, we observe that by Theorem 3.3 the first term in (3.16) converges to

$$\int_{\mathcal{L}_{d+1}} \phi^2 \, d\lambda_{d+1} - \left(\int_{\mathcal{L}_{d+1}} \phi \, d\lambda_{d+1} \right)^2$$

as $N \rightarrow \infty$. To estimate the second term in (3.16), we rewrite it as

$$\sum_{k=1}^{N-1} \left(\frac{2}{N} \sum_{n=0}^{N-1-k} \left(\int_{y_d} \psi(n, y) \psi(n+k, y) \, d\sigma_d(y) \right) \right). \quad (3.18)$$

It follows from Theorem 3.3 that for fixed k ,

$$\int_{y_d} \psi(n, y) \psi(n+k, y) \, d\sigma_d(y) \longrightarrow \int_{\mathcal{L}_{d+1}} \phi(\phi \circ a^k) \, d\lambda_{d+1} - \left(\int_{\mathcal{L}_{d+1}} \phi \, d\lambda_{d+1} \right)^2$$

as $n \rightarrow \infty$. A more tedious analysis, which utilizes the explicit quantitative bounds from Theorem 3.3 with $r = 1, 2$, allows to conclude that (3.18) converges to

$$2 \sum_{k \geq 1} \left(\int_{\mathcal{L}_{d+1}} \phi(\phi \circ a^k) \, d\lambda_{d+1} - \left(\int_{\mathcal{L}_{d+1}} \phi \, d\lambda_{d+1} \right)^2 \right).$$

This leads to the formula (3.17). More precisely, this argument will be applied to the family of functions ϕ_{ε}^L , introduced in §3.4, but the explicit form of the error term in Theorem 3.3 allows to handle this.

Finally, we note that the expressions (3.15) and (3.17) can be computed explicitly in our setting using Rogers' formula [13]. In particular, we conclude that the obtained variances are positive.

REFERENCES

1. Athreya, J.; Ghosh, A.; Tseng, J. *Spiraling of approximations and spherical averages of Siegel transforms*. J. Lond. Math. Soc. 91 (2015), no. 2, 383–404.
2. Billingsley, P. *Probability and measure*. Third edition. Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995.
3. Björklund, M.; Einsiedler, M.; Gorodnik, A. *Quantitative multiple mixing*. submitted, <https://arxiv.org/abs/1701.00945>
4. Björklund, M.; Gorodnik, A. *Central limit theorem for group actions which are exponentially mixing of all orders*. submitted.
5. Dolgopyat, D.; Fayad, B.; Vinogradov, I. *Central limit theorems for simultaneous Diophantine approximations*. J. Éc. polytech. Math. 4 (2017), 1–36.
6. Fréchet, M.; Shohat, J. *A proof of the generalized second-limit theorem in the theory of probability*. Trans. Amer. Math. Soc. 33 (1931), no. 2, 533–543.
7. Flaminio, L.; Forni, G. *Invariant distributions and time averages for horocycle flows*. Duke Math. J. 119 (2003), no. 3, 465–526.
8. Fuchs, M. *On a problem of W. J. LeVeque concerning metric Diophantine approximation*. Trans. Amer. Math. Soc. 355 (2003), no. 5, 1787–1801.
9. Kleinbock, D. Y.; Margulis, G. A. *On effective equidistribution of expanding translates of certain orbits in the space of lattices*. Number theory, analysis and geometry, 385–396, Springer, New York, 2012.
10. Leveque, W. J. *On the frequency of small fractional parts in certain real sequences I*. Trans. Amer. Math. Soc. 87 (1958) 237–260.
11. Leveque, W. J. *On the frequency of small fractional parts in certain real sequences II*. Trans. Amer. Math. Soc. 94 (1959) 130–149.
12. Philipp, W. *Mixing sequences of random variables and probabilistic number theory*. Memoirs of the American Mathematical Society, No. 114. American Mathematical Society, Providence, R. I., 1971.
13. Rogers, C. A. *Mean values over the space of lattices*. Acta Math. 94 (1955), 249–287.
14. Schmidt, W. M. *A metrical theorem in Diophantine approximation*. Canad. J. Math. 12 (1960), 619–631.
15. Schmidt, W. M. *A metrical theorem in geometry of numbers*. Trans. Amer. Math. Soc. 95 (1960), 516–529.
16. Speed, T. P. *Cumulants and Partition Lattices*. Austral. J. Statist., 25(2) (1983), 378–388.

DEPARTMENT OF MATHEMATICS, CHALMERS, GOTHENBURG, SWEDEN

E-mail address: micbjo@chalmers.se

UNIVERSITY OF BRISTOL, BRISTOL, UK

E-mail address: a.gorodnik@bristol.ac.uk