

AN ERROR TERM IN THE CENTRAL LIMIT THEOREM FOR SUMS OF DISCRETE RANDOM VARIABLES.

DMITRY DOLGOPYAT^{1*} AND KASUN FERNANDO²

(*) corresponding author

(1) Department of Mathematics, University of Maryland, College Park MD, USA
email: dmitry@math.umd.edu

(2) Centro De Giorgi, Scuola Normale Superiore, Pisa PI 56126, Italy
email: buddhima.akurugodage@sns.it

ABSTRACT. We consider sums of independent identically distributed random variables whose distributions have $d + 1$ atoms. Such distributions never admit an Edgeworth expansion of order d but we show that for almost all parameters the Edgeworth expansion of order $d - 1$ is valid and the error of the order $d - 1$ Edgeworth expansion is typically of order $n^{-d/2}$.

1. INTRODUCTION.

Let X be a random variable with zero mean and variance σ^2 . Let $S_n = \sum_{j=1}^n X_j$ where X_j are independent identically distributed and have the same distribution as X . The Central Limit Theorem says that for each z

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z \right) = \mathfrak{N}(z)$$

where

$$\mathfrak{N}(z) = \int_{-\infty}^z \mathfrak{n}(y) dy \text{ and } \mathfrak{n}(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

A classical problem in probability theory is computing higher order approximations to $\mathbb{P} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z \right)$. In particular, the order r Edgeworth series of S_n is an expression of the form

$$\mathcal{E}_r(z) = \mathfrak{N}(z) + \mathfrak{n}(z) \sum_{k=1}^r \frac{P_k(z)}{n^{k/2}}$$

2000 *Mathematics Subject Classification.* Primary 60F05; Secondary 11J13, 37A50.

where P_k are polynomials such that the characteristic function $\phi(t) = \mathbb{E}(e^{itX})$ and the Fourier transform $\hat{\mathcal{E}}_r$ of \mathcal{E}_r satisfy

$$\phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n - \hat{\mathcal{E}}_r(t) = o(n^{-r/2}).$$

In particular,

$$\begin{aligned} \mathcal{E}_1(z) &= \mathfrak{N}(z) + \mathfrak{n}(z) \frac{\mathbb{E}(X^3)}{6\sigma^3\sqrt{n}}(1 - z^2), \\ \mathcal{E}_2(z) &= \mathfrak{N}(z) + \mathfrak{n}(z) \left[\frac{\mathbb{E}(X^3)}{6\sqrt{n}\sigma^3}(1 - z^2) + \frac{\mathbb{E}(X^4) - 3\sigma^4}{24n\sigma^4}(3z - z^3) \right. \\ &\quad \left. - \frac{\mathbb{E}(X^3)^2}{72n\sigma^6}(15z - 10z^3 + z^5) \right]. \end{aligned}$$

We say that S_n admits an order r Edgeworth expansion if for all z

$$(1.1) \quad \lim_{n \rightarrow \infty} n^{r/2} \left[\mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq z\right) - \mathcal{E}_r(z) \right] = 0.$$

Recall that a lattice random variable is a discrete random variable taking values on a set of points of the form $a + nh, n \in \mathbb{Z}$, where $h > 0, a \in \mathbb{R}$. It is known that S_n admits the first order Edgeworth expansion if and only if X is non-lattice (see [12]). The problem of higher order expansion is more complicated. For example, a sufficient condition for S_n to admit the order r Edgeworth expansion is that $\mathbb{E}(|X|^{r+2}) < \infty$ and X has a density. But this condition is far from necessary. We refer the reader to [13, Chapter XVI] for discussion of these and related results. We also note that [2, 5] discusses a weak Edgeworth expansion where the LHS of (1.1) is convolved with smooth compactly supported functions.

In this paper, we consider a case which is opposite to X having a density, namely we suppose that X has a discrete distribution with $d + 1$ atoms where $d \geq 2$. $d = 2$ is the simplest non-trivial case since the distributions with two atoms are lattice, and as a result, they do not admit even the first order Edgeworth expansion.

Thus we suppose that X takes values a_1, \dots, a_{d+1} with probabilities p_1, \dots, p_{d+1} , respectively. Since X should have zero mean we suppose that our $2(d + 1)$ -tuple (\mathbf{a}, \mathbf{p}) belongs to the set

$$\Omega = \{p_i > 0, \quad p_1 + \dots + p_{d+1} = 1, \quad p_1 a_1 + \dots + p_{d+1} a_{d+1} = 0\}.$$

It is easy to see that S_n never admits the order d Edgeworth expansion. Indeed,

$$(1.2) \quad \mathbb{P}_{\mathbf{a}, \mathbf{p}}(S_n \leq z) = \sum_{\substack{m_i \geq 0, \sum m_i = n \\ \sum m_i a_i \leq z}} \frac{n!}{m_1! \dots m_{d+1}!} p_1^{m_1} \dots p_{d+1}^{m_{d+1}}.$$

The Local Central Limit Theorem (see [18, Theorem 2.1.1]), applied to the time homogeneous \mathbb{Z}^d -random walk which jumps to \mathbf{e}_i from the origin $\mathbf{0}$ with probability

p_i for $i = 1, \dots, d$ and stays at $\mathbf{0}$ with probability p_{d+1} , gives us that for all ε there is $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ and all $\mathbf{m} \in \mathbb{Z}^d$

$$(1.3) \quad \left| \mathbb{P}(T_n = \mathbf{m}) - \frac{1}{\sqrt{(2\pi n)^d \det \Gamma}} e^{-\frac{(\mathbf{m}-n\mathbf{q}) \cdot \Gamma^{-1}(\mathbf{m}-n\mathbf{q})}{2n}} \right| \leq \frac{\varepsilon}{n^{d/2}}$$

where T_n is the position of the random walk after n steps, Γ is the associated covariance matrix and $\mathbf{q} = (p_1, \dots, p_d)$. Also, if m_1, \dots, m_d, m_{d+1} are integers such that $m_1 + \dots + m_d + m_{d+1} = n$ and $m_i \geq 0$, then, taking $\mathbf{m} = (m_1, \dots, m_d)$, we have

$$\mathbb{P}(T_n = \mathbf{m}) = \frac{n!}{m_1! \dots m_{d+1}!} p_1^{m_1} \dots p_{d+1}^{m_{d+1}}.$$

As a result, if

$$\sum m_i a_i = n \sum a_i p_i + \mathcal{O}(\sqrt{n}),$$

then in (1.3), the exponent of e is $\mathcal{O}(1)$, and hence, for sufficiently large n ,

$$n^{d/2} \mathbb{P}(T_n = \mathbf{m}) = n^{d/2} \frac{n!}{m_1! \dots m_{d+1}!} p_1^{m_1} \dots p_{d+1}^{m_{d+1}}$$

is uniformly bounded from below. Accordingly, from (1.2), it follows that $\mathbb{P}_{\mathbf{a}, \mathbf{p}}(S_n \leq z)$ has jumps of order $n^{-d/2}$. On the other hand, $\mathcal{E}_d(z)$ is a smooth function of z . So, it can not approximate both $\mathbb{P}_{\mathbf{a}, \mathbf{p}}(S_n \leq z - 0)$ and $\mathbb{P}_{\mathbf{a}, \mathbf{p}}(S_n \leq z + 0)$ at the points of jumps without making an error of $\mathcal{O}(n^{-d/2})$. This means that it is *not* true that $\mathbb{P}_{\mathbf{a}, \mathbf{p}}(S_n \leq z) = \mathcal{E}_d(z) + o(n^{-d/2})$ for all z , showing that the order d Edgeworth expansion fails.

However, in this paper, we show that for *typical* (\mathbf{a}, \mathbf{p}) this failure of the order d Edgeworth expansion happens just barely. We present two results in this direction. For the first result, let

$$b_j = a_j - a_1, \text{ for } j = 2, \dots, d+1.$$

Then, the characteristic function of X , ϕ , satisfies

$$\phi(s) = e^{i s a_1} \psi(s) \quad \text{where} \quad \psi(s) = p_1 + p_2 e^{i s b_2} + \dots + p_{d+1} e^{i s b_{d+1}}.$$

Set

$$d(s) = \max_{j \in \{2, \dots, d+1\}} \text{dist}(b_j s, 2\pi \mathbb{Z}).$$

We say that \mathbf{a} is β -Diophantine if there is a constant K such that for $|s| > 1$,

$$d(s) \geq \frac{K}{|s|^\beta}.$$

It follows from the classical Khinchine-Groshev Theorem (see e.g. [17, Theorem 1.1] or [23]) that almost every \mathbf{a} is β -Diophantine provided that $\beta > \frac{1}{d-1}$.

Theorem 1. *If \mathbf{a} is β -Diophantine and*

$$(1.4) \quad 2 \left(R - \frac{1}{2} \right) \beta < 1$$

then

$$\lim_{n \rightarrow \infty} n^R \left[\mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma \sqrt{n}} \leq z \right) - \mathcal{E}_{d-1}(z) \right] = 0$$

uniformly in $z \in \mathbb{R}$.

Thus, for almost every \mathbf{a} the order $(d-1)$ Edgeworth expansion approximates the distribution of $\frac{S_n}{\sigma \sqrt{n}}$ with error $\mathcal{O}(n^{\varepsilon-d/2})$ for any ε .

Note that Theorem 1 applies for all β s, and in particular, for β s which are much larger than $\frac{1}{d-1}$. However, if β is large, then the statement of the theorem can be simplified. Namely, let r be the integer such that $r < 2R \leq r+1$. (Note that (1.4) can be rewritten as $2R < \frac{1}{\beta} + 1$. So, provided that $2R$ is sufficiently close to $\frac{1}{\beta} + 1$ we can take $r = \left\langle \frac{1}{\beta} \right\rangle + 1$ where $\langle s \rangle$ denotes the largest integer which is strictly smaller than s .) Then,

$$\begin{aligned} \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma \sqrt{n}} \leq z \right) &= \mathcal{E}_{d-1}(z) + o \left(\frac{1}{n^R} \right) \\ &= \mathcal{E}_r(z) + o \left(\frac{1}{n^R} \right) + \mathcal{O}(\mathcal{E}_{d-1}(z) - \mathcal{E}_r(z)). \end{aligned}$$

Since $\frac{r+1}{2} > R$ the first error term dominates the second and we obtain the following result.

Corollary 1.1. *Suppose that \mathbf{a} is β -Diophantine, $r = 1 + \left\langle \frac{1}{\beta} \right\rangle$, and $r < 2R < \frac{1}{\beta} + 1$.*

Then

$$\lim_{n \rightarrow \infty} n^R \left[\mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma \sqrt{n}} \leq z \right) - \mathcal{E}_r(z) \right] = 0$$

uniformly in $z \in \mathbb{R}$.

Corollary 1.1 shows that for almost every \mathbf{a} and for $r \in \{1, \dots, d-1\}$, the order r Edgeworth expansion is valid. Our next results show that

$$(1.5) \quad \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma \sqrt{n}} \leq z \right) - \mathcal{E}_d(z)$$

is typically of order $\mathcal{O}(n^{-d/2})$ but the $\mathcal{O}(n^{-d/2})$ term has wild oscillations. To formulate this result precisely, we suppose that our $2(d+1)$ -tuple is chosen at random

according to an absolutely continuous distribution \mathbf{P} on Ω . Thus, (1.5) becomes a random variable.

Theorem 2. *There exists a smooth function $\Lambda(\mathbf{a}, \mathbf{p})$ such that for each z the random variable*

$$e^{z^2/2} \frac{n^{d/2}}{\Lambda(\mathbf{a}, \mathbf{p})} \left[\mathcal{E}_d(z) - \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z \right) \right]$$

converges in law to a non-trivial random variable \mathcal{X} (defined below in Lemma 1.2) whose distribution is independent of z and \mathbf{P} .

The formulas for the normalizing factor $\Lambda(\mathbf{a}, \mathbf{p})$ and the limiting random variable \mathcal{X} are quite complicated and the next few pages are devoted to their definitions.

The normalization is defined as follows:

$$(1.6) \quad \Lambda(\mathbf{a}, \mathbf{p}) = \frac{|a_{d+1} - a_1|}{(2\pi)^{d+\frac{1}{2}} \sqrt{\det(D_{\mathbf{a}, \mathbf{p}})} \sigma(\mathbf{a}, \mathbf{p})}$$

where $\sigma(\mathbf{a}, \mathbf{p})$ denotes the standard deviation of the distribution of the random variable taking value a_j with probability p_j and $D_{\mathbf{a}, \mathbf{p}}$ is a $(d-1) \times (d-1)$ matrix defined as follows.

The matrix $D_{\mathbf{a}, \mathbf{p}}$. Fix p_1, \dots, p_{d+1} and consider a map

$$\zeta(\mathbf{y}) = \left| \sum_{j=1}^{d+1} p_j e^{iy_j} \right|^2$$

where $\mathbf{y} = (y_1, \dots, y_{d+1}) \in \mathbb{R}^{d+1}$. Let Y be a random variable taking values y_j with probability p_j . Then, for small \mathbf{y} we have

$$\mathbf{E}(e^{iY}) = 1 - \frac{\mathbf{E}(Y^2)}{2} + i\mathbf{E}(Y) + \mathcal{O}(\|\mathbf{y}\|^3).$$

Hence,

$$(1.7) \quad \zeta(\mathbf{y}) = 1 - \mathbf{E}(Y^2) + \mathbf{E}(Y)^2 + \mathcal{O}(\|\mathbf{y}\|^3) = 1 - V(Y) + \mathcal{O}(\|\mathbf{y}\|^3)$$

where $V(\cdot)$ is the variance.

Next, consider the quadratic form given by $Q(\mathbf{y}, \mathbf{y}) = V(Y(\mathbf{y}))$. Let $\mathbf{x}, \mathbf{a} \in \mathbb{R}^{d+1}$ be fixed. In order to maximize $s \mapsto \zeta(\mathbf{x} + s\mathbf{a})$, we want to minimize $s \mapsto Q(\mathbf{x} + s\mathbf{a}, \mathbf{x} + s\mathbf{a})$. We have

$$Q(\mathbf{x} + s\mathbf{a}, \mathbf{x} + s\mathbf{a}) = Q(\mathbf{x}, \mathbf{x}) + 2sQ(\mathbf{x}, \mathbf{a}) + s^2Q(\mathbf{a}, \mathbf{a}).$$

It follows that the minimum is achieved at $s^* = -\frac{Q(\mathbf{a}, \mathbf{x})}{Q(\mathbf{a}, \mathbf{a})}$ and its value is

$$D(\mathbf{x}, \mathbf{x}) = Q(\mathbf{x}, \mathbf{x}) - \frac{Q(\mathbf{x}, \mathbf{a})^2}{Q(\mathbf{a}, \mathbf{a})} = \frac{Q(\mathbf{x}, \mathbf{x})Q(\mathbf{a}, \mathbf{a}) - Q(\mathbf{a}, \mathbf{x})^2}{Q(\mathbf{a}, \mathbf{a})}$$

$$= \frac{V(Y(\mathbf{x}))V(X) - \text{Cov}^2(X, Y(\mathbf{x}))}{V(X)}$$

where X is the random variable taking values a_j with probability p_j . Note that $D(\mathbf{x}, \mathbf{x}) > 0$ on the subspace $x_1 = x_{d+1} = 0$ since $\text{Cov}^2(X, Y(\mathbf{x})) = V(X)V(Y(\mathbf{x}))$ iff $Y(\mathbf{x}) = c_1X + c_2$. Note that the RHS takes $(d+1)$ different values if $c_1 \neq 0$ and it takes a single value if $c_1 = 0$. On the other hand, the LHS takes at most d different values on $\{x_1 = x_{d+1} = 0\}$ and it takes a single value only at $\mathbf{0}$. This implies that $Q(\mathbf{x}) \neq 0$ unless $\mathbf{x} = \mathbf{0}$, and hence, Q is non degenerate. Then $D_{\mathbf{a}, \mathbf{p}}$ is the $(d-1) \times (d-1)$ positive definite matrix such that $-4D_{\mathbf{a}, \mathbf{p}}$ is the Hessian of $\mathbb{R}^{d-1} \ni \tilde{\mathbf{x}} \mapsto \zeta(0, \tilde{\mathbf{x}}, 0)$. The formula for $D_{\mathbf{a}, \mathbf{p}}$ will be proven in Section 6 (see (6.8)).

We note that the infinitesimal computation described above is relevant because we will show, in the course of proving Theorem 2, that the main contribution to the error term come from the resonant points where the Taylor expansion could be used. See Section 6 for more details.

To define \mathcal{X} , we need some notation. Let \mathbb{M} be the space of pairs (\mathcal{L}, χ) where \mathcal{L} is a unimodular lattice in \mathbb{R}^d and χ is a character, that is, a homomorphism $\chi : \mathcal{L} \rightarrow \mathbb{T}$.

The Haar measure on \mathbb{M} . The Haar measure μ on \mathbb{M} can be defined in two equivalent ways. First, note that χ is of the form $\chi(\mathbf{w}) = e^{2\pi i \tilde{\chi}(\mathbf{w})}$ for some linear functional $\tilde{\chi} \in (\mathbb{R}^d)^*$. $SL_d(\mathbb{R})$ acts on $\mathbb{R}^d \oplus (\mathbb{R}^d)^*$ by the formula

$$A(\mathbf{w}, \tilde{\chi}) = (A\mathbf{w}, \tilde{\chi}A^{-1}).$$

Observe that if $A(\mathbf{w}, \tilde{\chi}) = (\hat{\mathbf{w}}, \hat{\chi})$ then

$$(1.8) \quad \tilde{\chi}(\mathbf{w}) = \hat{\chi}(\hat{\mathbf{w}}).$$

The above action of $SL_d(\mathbb{R})$ induces the following action of $SL_d(\mathbb{R}) \times (\mathbb{R}^d)^*$ on \mathbb{M}

$$(A, \tilde{\chi})(\mathcal{L}, \chi) = (A\mathcal{L}, e^{2\pi i \tilde{\chi}} \cdot (\chi \circ A^{-1})).$$

This action is transitive because $SL_d(\mathbb{R})$ acts transitively on unimodular lattices and $(\mathbb{R}^d)^*$ acts transitively on characters. This allows us to identify \mathbb{M} with

$$(SL_d(\mathbb{R}) \times \mathbb{R}^d) / (SL_d(\mathbb{Z}) \times \mathbb{Z}^d)$$

and so \mathbb{M} inherits the Haar measure from $SL_d(\mathbb{R}) \times \mathbb{R}^d$.

The second way to define the Haar measure is to note that the space \mathcal{M} of unimodular lattices is naturally identified with $SL_d(\mathbb{R})/SL_d(\mathbb{Z})$, and so, it inherits the Haar measure from $SL_d(\mathbb{R})$. Next, for a fixed \mathcal{L} the set of homomorphisms $\chi : \mathcal{L} \rightarrow \mathbb{T}$ is a d dimensional torus. So, it comes with its own Haar measure. Now, if we want to compute the average of a function $\Phi(\mathcal{L}, \chi)$ with respect to the Haar measure then we can first compute its average $\bar{\Phi}(\mathcal{L})$ in each fiber and then integrate the result with respect to the Haar measure on the space of lattices. In the proof of Lemma 1.2 given in Section 10, the averaging inside a fiber will be denoted by \mathbf{E}_χ and the averaging with respect to the Haar measure on the space of lattices will be denoted by $\mathbf{E}_\mathcal{L}$.

The random variable \mathcal{X} . Given a vector $\mathbf{w} \in \mathbb{R}^d$, we denote by $y(\mathbf{w})$ its first coordinate and by $\mathbf{x}(\mathbf{w})$ its last $d - 1$ coordinates. We also denote by $\|\cdot\|$ the standard Euclidean norm.

Lemma 1.2. *For almost every pair $(\mathcal{L}, \chi) \in \mathbb{M}$ with respect to the Haar measure the following limit exists*

$$(1.9) \quad \mathcal{X}(\mathcal{L}, \chi) = \lim_{R \rightarrow \infty} \sum_{\mathbf{w} \in \mathcal{L} \setminus \{0\}, \|\mathbf{w}\| \leq R} \frac{\sin(2\pi\chi(\mathbf{w}))}{y(\mathbf{w})} e^{-\|\mathbf{x}(\mathbf{w})\|^2}.$$

In this formula and below, we identify \mathbb{T} with segment $[0, 1)$ equipped with addition modulo one, and thus, the characters $\chi(\mathbf{w})$ are (after this identification) real valued.

In particular, the proof of Lemma 1.2 shows that for almost every \mathcal{L} , whenever $\mathbf{w} \neq 0$, $y(\mathbf{w}) \neq 0$ (see Section 10) and that each individual summand in (1.9) is finite almost everywhere on \mathbb{M} . In order to simplify the notation, we will abbreviate expressions such as (1.9) by

$$(1.10) \quad \mathcal{X}(\mathcal{L}, \chi) = \sum_{\mathbf{w} \in \mathcal{L} \setminus \{0\}} \frac{\sin(2\pi\chi(\mathbf{w}))}{y(\mathbf{w})} e^{-\|\mathbf{x}(\mathbf{w})\|^2}$$

even though (1.10) does not converge absolutely.

If we assume that the pair (\mathcal{L}, χ) is distributed according to the Haar measure on \mathbb{M} then \mathcal{X} , defined by (1.9), becomes a random variable. This is the variable mentioned in Theorem 2. Note that the distribution of \mathcal{X} depends neither on \mathbf{P} nor on z .

Next, we describe how we can use the second representation of Haar measure to describe \mathcal{X} . Let $\mathbf{w}_1, \dots, \mathbf{w}_d$ be the shortest spanning set of \mathcal{L} , i.e., \mathbf{w}_1 is the shortest non zero vector in \mathcal{L} and, for $j > 1$, \mathbf{w}_j is the shortest vector in \mathcal{L} that is linearly independent of $\mathbf{w}_1, \dots, \mathbf{w}_{j-1}$. Given $\mathbf{m} = (m_1, \dots, m_d) \in \mathbb{Z}^d$, let

$$(1.11) \quad (y, \mathbf{x})(\mathbf{m}) := m_1 \mathbf{w}_1 + \dots + m_d \mathbf{w}_d \in \mathcal{L}$$

where $y \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^{d-1}$. Let $\theta_j = \chi(\mathbf{w}_j)$. Then θ_j are uniformly distributed on \mathbb{T} and independent of each other. Set

$$(1.12) \quad \theta(\mathbf{m}) := m_1 \theta_1 + \dots + m_d \theta_d.$$

Now, \mathcal{X} (see definition in Lemma 1.2) can be rewritten as

$$(1.13) \quad \mathcal{X} = \sum_{\mathbf{m} \in \mathbb{Z}^d \setminus \{0\}} \frac{\sin(2\pi\theta(\mathbf{m}))}{y(\mathbf{m})} e^{-\|\mathbf{x}(\mathbf{m})\|^2}$$

where \mathcal{L} is uniformly distributed on the space of lattices, $(y, \mathbf{x})(\mathbf{m})$ is defined by (1.11), and $(\theta_1, \dots, \theta_d)$ is uniformly distributed on \mathbb{T}^d and independent of \mathcal{L} . We will use the representation (1.13) in Sections 8 and 9 in our proofs and in Section 10 when establishing the convergence of \mathcal{X} .

Theorems 1 and 2 have analogues in case we are interested in probability that S_n belongs to a finite interval. In particular, our results have applications to Local Limit Theorems.

Theorem 3. *Let $z_1(n)$ and $z_2(n)$ be two uniformly bounded sequences such that $|z_1(n) - z_2(n)|n^{d/2} \rightarrow \infty$. Then the random vector*

$$(1.14) \quad \frac{n^{d/2}}{\Lambda(\mathbf{a}, \mathbf{p})} \left(e^{z_1^2/2} \left[\mathcal{E}_d(z_1) - \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z_1 \right) \right], e^{z_2^2/2} \left[\mathcal{E}_d(z_2) - \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z_2 \right) \right] \right)$$

converges in law to a random vector $(\mathcal{X}(\mathcal{L}, \chi_1), \mathcal{X}(\mathcal{L}, \chi_2))$ where $\mathcal{X}(\mathcal{L}, \chi)$ is defined by (1.10) and the triple $(\mathcal{L}, \chi_1, \chi_2)$ is uniformly distributed on $(SL_d(\mathbb{R})/SL_d(\mathbb{Z})) \times \mathbb{T}^d \times \mathbb{T}^d$.

Here and below the uniform distribution of $(\mathcal{L}, \chi_1, \chi_2)$ means that \mathcal{L} is uniformly distributed on the space of lattices, and for a given lattice, χ_1 and χ_2 are chosen independently and uniformly from the space of characters.

Theorem 4. *Let $z_1(n) < z_2(n)$ be two uniformly bounded sequences such that $l_n = z_2(n) - z_1(n) \rightarrow 0$.*

(a) *If $l_n \geq Cn^{\varepsilon-d/2}$ for some $\varepsilon > 0$ then*

$$\frac{\mathbb{P}_{\mathbf{a}, \mathbf{p}}(z_1 < \frac{S_n}{\sigma\sqrt{n}} < z_2)}{l_n \mathbf{n}(z_1)} \rightarrow 1$$

almost surely.

(b) *If $l_n n^{d/2} \rightarrow \infty$ then*

$$\frac{\mathbb{P}_{\mathbf{a}, \mathbf{p}}(z_1 < \frac{S_n}{\sigma\sqrt{n}} < z_2)}{l_n \mathbf{n}(z_1)} \Rightarrow 1$$

(here and below “ \Rightarrow ” denotes the convergence in law).

(c) *If $l_n = \frac{c|a_{d+1} - a_1|}{\sigma(\mathbf{a}, \mathbf{p})n^{d/2}}$ then $H(\mathbf{a}, \mathbf{p}) \left[\frac{\mathbb{P}_{\mathbf{a}, \mathbf{p}}(z_1 < \frac{S_n}{\sigma\sqrt{n}} < z_2)}{l_n \mathbf{n}(z_1)} - 1 \right] \Rightarrow \mathcal{Y}$ where*

$$H(\mathbf{a}, \mathbf{p}) = (2\pi)^d \sqrt{\det(D_{\mathbf{a}, \mathbf{p}})}$$

and

$$\mathcal{Y}(\mathcal{L}, \chi, c) = \frac{1}{c} \sum_{\mathbf{w} \in \mathcal{L} \setminus \{0\}} \frac{\sin(2\pi\chi(\mathbf{w})) - \sin(2\pi[\chi(\mathbf{w}) - cy(\mathbf{w})])}{y(\mathbf{w})} e^{-\|\mathbf{x}(\mathbf{w})\|^2},$$

\mathcal{L}, χ are as in Theorem 2 and $D_{\mathbf{a}, \mathbf{p}}$ is from (1.6).

Remark 1.3. The normalization in Theorem 4(c) comes from the following computation. Denote $\Delta_n(z) = \mathcal{E}_d(z) - \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq z\right)$. Then, Theorem 2 can be informally restated as

$$\Delta_n(z) \approx \frac{\Lambda(\mathbf{a}, \mathbf{p})\sqrt{2\pi} \mathbf{n}(z)}{n^{d/2}} \mathcal{X}.$$

Then under the assumption of part (c) of Theorem 4 we have

$$\frac{\Delta_n(z_2) - \Delta_n(z_1)}{l_n} \approx \frac{\Lambda(\mathbf{a}, \mathbf{p})\sqrt{2\pi}}{l_n n^{d/2}} [\mathbf{n}(z_2)\mathcal{X}_2 - \mathbf{n}(z_1)\mathcal{X}_1].$$

Since $\frac{\Lambda(\mathbf{a}, \mathbf{p})\sqrt{2\pi}}{l_n n^{d/2}} = \frac{1}{cH(\mathbf{a}, \mathbf{p})}$ we can rewrite the above equation as

$$cH(\mathbf{a}, \mathbf{p}) \frac{\Delta_n(z_2) - \Delta_n(z_1)}{l_n \mathbf{n}(z_1)} \approx \frac{\mathbf{n}(z_2)}{\mathbf{n}(z_1)} \mathcal{X}_2 - \mathcal{X}_1.$$

Thus, the proof of Theorem 4 proceeds by describing the asymptotics of the joint distributions of $n^{d/2}\Delta_n(z_1)$ and $n^{d/2}\Delta_n(z_2)$ while Theorem 2 gives the marginal distributions.

The intuition behind the results of Theorem 4 is the following. Call y_n δ -plausible if $\mathbb{P}(S_n = y_n) \geq \delta n^{-d/2}$. The discussion following (1.2) shows that for each δ there are about $C(\delta)n^{d/2}$ δ -plausible values. Therefore, if $l_n \ll n^{-d/2}$ then the interval $[z_1(n), z_2(n)]$ would typically contain no plausible values. Hence, we should not expect a Local Limit Theorem (LLT) to hold on that scale. Theorem 4 shows that as soon as interval $[z_1(n), z_2(n)]$ contains many plausible values then an LLT typically holds for this interval.

Recall that

$$\mathbb{P}_{\mathbf{a}, \mathbf{p}}(S_n \in [z_1, z_2]) = \sum_{\substack{m_i \geq 0, \sum m_i = n \\ z_1 \leq \sum m_i a_i \leq z_2}} \frac{n!}{m_1! \dots m_{d+1}!} p_1^{m_1} \dots p_{d+1}^{m_{d+1}}.$$

So, in Theorem 4, we just count the number of visits of a random linear form $\sum m_i a_i$ to a finite interval with weights given by multinomial coefficients. It is also interesting to consider counting with equal weight. In this case the analogue of Theorem 4(c) is obtained in [19] while for longer intervals only partial results are available, see [10, 15].

The layout of the paper is the following. Theorem 1 is proven in Section 2. The proof is a minor modification of the arguments of [13, Chapter XVI]. The bulk of the paper is devoted to the proof of Theorem 2. In Section 3, we provide an equivalent formula for \mathcal{X} . This formula looks more complicated than (1.10) but it is easier to identify with the limit of the error term. Section 4 contains preliminary reductions. Namely, we show that the integration in the Fourier inversion formula could be restricted to a finite domain. In Section 6, we show that the main contribution to the error term

comes from resonances where the characteristic function of S_n is close to 1 in absolute value. The proof relies on the asymptotic analysis of the resonant term performed in Section 5. Several technical estimates used in our analysis are established in Section 7. In Section 8, we use dynamics on homogenous spaces in order to show that the contribution of resonances converges to (1.10) completing the proof of Theorem 2. The proofs of Theorems 3 and 4 are similar to the proof of Theorem 2. The necessary modifications are explained in Section 9. Finally, Section 10 contains the proof of Lemma 1.2.

As a notational remark, in the paper the constants denoted by c , C , or other implied constants may change from line to line or even within the same line.

2. EDGEWORTH EXPANSION UNDER DIOPHANTINE CONDITIONS.

Theorem 1 is a consequence of Lemma 2.1 and Theorem 5 below.

Note that the characteristic function of X is given by

$$(2.1) \quad \phi(s) = p_1 e^{i s a_1} + \dots + p_{d+1} e^{i s a_{d+1}}$$

and recall that $d(s) = \max_{j \in \{2, \dots, d+1\}} \text{dist}(b_j s, 2\pi\mathbb{Z})$ where $b_j = a_j - a_1$.

Lemma 2.1. *There exists a positive constant c such that*

$$(2.2) \quad |\phi(s)| \leq 1 - c d(s)^2.$$

Proof. Since

$$1 - |\phi(s)| = \frac{1 - |\phi(s)|^2}{1 + |\phi(s)|} \geq \frac{1 - |\phi(s)|^2}{2},$$

it suffices to show that

$$(2.3) \quad |\phi(s)|^2 \leq 1 - 2c d(s)^2.$$

Note that

$$|\phi(s)|^2 = \sum_j p_j^2 + 2 \sum_{j < k} p_j p_k \cos((a_j - a_k)s).$$

Taking a constant \bar{c} such that $\cos(t) \leq 1 - \bar{c}t^2$ for $|t| \leq \pi$ and letting $c = \bar{c}(\min_j p_j)^2$ we obtain

$$|\phi(s)|^2 \leq 1 - 2c \sum_{j < k} \text{dist}^2((a_j - a_k)s, 2\pi\mathbb{Z})$$

proving (2.3). □

Theorem 5. *If the distribution of X has $d+2$ moments and its characteristic function ϕ satisfies*

$$(2.4) \quad |\phi(s)| \leq 1 - \frac{K}{|s|^\gamma}$$

and $R < \frac{d}{2}$ is such that

$$(2.5) \quad \left(R - \frac{1}{2}\right) \gamma < 1$$

then

$$(2.6) \quad \lim_{n \rightarrow \infty} n^R \left[\mathbb{P} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z \right) - \mathcal{E}_{d-1}(z) \right] = 0.$$

In particular, if X is discrete with $d+1$ atoms $\mathbf{a} = (a_1, \dots, a_{d+1})$, \mathbf{a} is β -Diophantine and $\left(R - \frac{1}{2}\right) \beta < \frac{1}{2}$, then (2.6) holds.

Theorem 5 follows easily from the estimates in [13, ChapterXVI] but we provide the proof here for completeness.

Proof. Denoting

$$\bar{\Delta}_n = \mathbb{P} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z \right) - \mathcal{E}_{d-1}(z)$$

we get, from the estimate (3.13) in [13, Chapter XVI], that for each T

$$(2.7) \quad |\bar{\Delta}_n| \leq \frac{1}{\pi} \int_{-\frac{T}{\sigma\sqrt{n}}}^{\frac{T}{\sigma\sqrt{n}}} \left| \frac{\phi^n(s) - \hat{\mathcal{E}}_{d-1}(s\sigma\sqrt{n})}{s} \right| ds + \frac{C}{T}.$$

Choose $T = Bn^R$ with $B = \frac{C}{\varepsilon}$. Then, $\frac{C}{T} = \frac{\varepsilon}{n^R}$. Take a small δ and split the integral in the RHS of (2.7) into two parts.

$$(2.8) \quad \frac{1}{\pi} \int_{-\delta}^{\delta} \left| \frac{\phi^n(s) - \hat{\mathcal{E}}_{d-1}(s\sigma\sqrt{n})}{s} \right| ds + \frac{1}{\pi} \int_{\delta < |s| < Bn^{R-1/2}/\sigma} \left| \frac{\phi^n(s) - \hat{\mathcal{E}}_{d-1}(s\sigma\sqrt{n})}{s} \right| ds.$$

From the proof of Theorem 2 in Section 2 and Theorem 3 in Section 4 of [13, Chapter XVI], we have that the first integral of (2.8) is $\mathcal{O}(n^{-d/2})$.

Also, $\int_{|s| > \delta} \left| \frac{\hat{\mathcal{E}}_{d-1}(s\sigma\sqrt{n})}{s} \right| ds$ has exponential decay as $n \rightarrow \infty$. Put

$$J = \{s : \delta < |s| < B\sigma^{-1}n^{R-1/2}\}.$$

Thus, we only need to estimate

$$(2.9) \quad \int_J \left| \frac{\phi^n(s)}{s} \right| ds \leq \frac{1}{\delta} \int_J |\phi^n(s)| ds \leq \frac{C}{\delta} \int_J \exp\left(-bn^{1-(R-\frac{1}{2})\gamma}\right) ds$$

where the last inequality is due to (2.4). By (2.5) the integral decay faster than any power of n . The result follows. \square

Remark 2.2. The fact that the Edgeworth expansion of order $(d - 1)$ holds for almost every \mathbf{a} is obtained in [1, Section 4] (with a weaker error bound). [5] shows, among other things, that a Diophantine condition with any exponent is sufficient for obtaining a weak Edgeworth expansion for sufficiently smooth functions. [6, 14] obtain similar results for dependent random variable including finite state Markov chains. The relation between the Edgeworth expansions and Diophantine approximations are utilized in [1, 3, 4] to show that Edgeworth expansions hold for almost every member of several multi-parameter families.

3. CHANGE OF VARIABLES.

Here, we deduce Theorem 2 from:

Theorem 2* *For each z the random variable*

$$n^{d/2} \left[\mathcal{E}_d(z) - \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma \sqrt{n}} \leq z \right) \right]$$

converges in law to $\hat{\mathcal{X}}$ where

$$(3.1) \quad \hat{\mathcal{X}}(\mathbf{a}, \mathbf{p}, \mathcal{L}, \chi) = e^{-z^2/2} \frac{|\mathbf{a}_{d+1} - \mathbf{a}_1|}{2\sigma(\mathbf{a}, \mathbf{p})\sqrt{2\pi^3}} \sum_{\mathbf{w} \in \mathcal{L} \setminus \{0\}} \frac{\sin 2\pi\chi(\mathbf{w})}{y(\mathbf{w})} e^{-4\pi^2 \mathbf{x}(\mathbf{w}) D_{\mathbf{a}, \mathbf{p}} \cdot \mathbf{x}(\mathbf{w})},$$

(\mathcal{L}, χ) is distributed according to μ , the Haar measure on \mathbb{M} , $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_{d+1})$, $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_{d+1})$ and $(\mathbf{a}, \mathbf{p}) \in \Omega$ are distributed according to \mathbf{P} , (\mathbf{a}, \mathbf{p}) and (\mathcal{L}, χ) are independent, and $D_{\mathbf{a}, \mathbf{p}}$ and $\sigma(\mathbf{a}, \mathbf{p})$ are defined immediately after (1.6).

We note that the convergence of (3.1) for almost every (\mathcal{L}, χ) follows¹ from Lemma 1.2, see Step 1 in the proof of Theorem 2 below.

Proof of Theorem 2 assuming Theorem 2.* We divide the proof into three steps.

Step 1. We will show that $e^{z^2/2} \frac{\hat{\mathcal{X}}}{\Lambda(\mathbf{a}, \mathbf{p})}$ has the same distribution as \mathcal{X} (see (1.9)). To this end, we rewrite the sum in (3.1) as

$$(3.2) \quad \frac{1}{(2\pi)^{d-1} \det(\sqrt{D_{\mathbf{a}, \mathbf{p}}})} \sum_{\mathbf{w} \in \mathcal{L} \setminus \{0\}} \frac{\sin(2\pi\chi(\mathbf{w}))}{y(\mathbf{w}) / ((2\pi)^{d-1} \det(\sqrt{D_{\mathbf{a}, \mathbf{p}}}))} e^{-4\pi^2 \|\sqrt{D_{\mathbf{a}, \mathbf{p}}} \mathbf{x}(\mathbf{w})\|^2}.$$

Let A be the linear map such that

$$(3.3) \quad A(y, \mathbf{x}) = \left(\frac{y}{(2\pi)^{d-1} \sqrt{\det(D_{\mathbf{a}, \mathbf{p}})}}, 2\pi \sqrt{D_{\mathbf{a}, \mathbf{p}}} \mathbf{x} \right).$$

¹Lemma 1.2 shows that the convergence holds if the sum in (3.1) is understood as a limit as $R \rightarrow \infty$ of the sums restricted to the domain $\|A\mathbf{w}\| \leq R$ where A is the matrix given by (3.3). However, the proof of Lemma 1.2 shows that this sum could also be understood as the limit of sums over domains $\|\mathbf{w}\| \leq R$.

Put $(\bar{\mathcal{L}}, \bar{\chi}) = A(\mathcal{L}, \chi)$. Then, using (1.8), (3.2) can be rewritten as:

$$\frac{1}{(2\pi)^{d-1} \det(\sqrt{D_{\mathbf{a}, \mathbf{p}}})} \sum_{\bar{\mathbf{w}} \in \bar{\mathcal{L}} \setminus \{0\}} \frac{\sin(2\pi \bar{\chi}(\bar{\mathbf{w}}))}{y(\bar{\mathbf{w}})} e^{-\|\mathbf{x}(\bar{\mathbf{w}})\|^2}.$$

Since $\det(A) = 1$, the pair $(\bar{\mathcal{L}}, \bar{\chi})$ is distributed according to the Haar measure on \mathbb{M} . Thus, using (1.6),

$$\hat{\mathcal{X}}(\mathbf{a}, \mathbf{p}, \mathcal{L}, \chi) = e^{-z^2/2} \Lambda(\mathbf{a}, \mathbf{p}) \sum_{\bar{\mathbf{w}} \in \bar{\mathcal{L}} \setminus \{0\}} \frac{\sin(2\pi \bar{\chi}(\bar{\mathbf{w}}))}{y(\bar{\mathbf{w}})} e^{-\|\mathbf{x}(\bar{\mathbf{w}})\|^2}$$

where $(\bar{\mathcal{L}}, \bar{\chi})$ is distributed according to the Haar measure on \mathbb{M} . So, from (1.9) $e^{z^2/2} \frac{\hat{\mathcal{X}}}{\Lambda(\mathbf{a}, \mathbf{p})}$ and \mathcal{X} have the same distribution.

Step 2. Denote

$$(3.4) \quad \Omega_\kappa^M = \{(\mathbf{a}, \mathbf{p}) \in \Omega : \forall i \ \kappa \leq p_i, \quad |a_i| \leq M \quad \text{and} \quad \forall i \neq j \ |a_i - a_j| \geq \kappa\},$$

$$\Delta_n = \mathcal{E}_d(z) - \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq z\right), \quad \tilde{\Delta}_n = e^{z^2/2} \frac{\Delta_n}{\Lambda(\mathbf{a}, \mathbf{p})}.$$

We claim that it is enough to prove Theorem 2 under the assumption that \mathbf{P} has smooth density supported on Ω_κ^M for some κ and M . Indeed, let p be the original density of \mathbf{P} . Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a smooth compactly supported function. Given ε there exists a smooth density \tilde{p} supported on some Ω_κ^M such that $\|\tilde{p} - p\|_{L^1} \leq \frac{\varepsilon}{2\|f\|_\infty}$.

If Theorem 2 holds for smooth compactly supported densities then we can find $n_0 \in \mathbb{N}$ such that for $n \geq n_0$

$$\left| \iint f\left(n^{d/2} \tilde{\Delta}_n\right) \tilde{p} \, d\mathbf{a} \, d\mathbf{p} - \mathbb{E}(f(\mathcal{X})) \right| \leq \frac{\varepsilon}{2}.$$

Since

$$\left| \iint f\left(n^{d/2} \tilde{\Delta}_n\right) \tilde{p} \, d\mathbf{a} \, d\mathbf{p} - \iint f\left(n^{d/2} \tilde{\Delta}_n\right) p \, d\mathbf{a} \, d\mathbf{p} \right| \leq \|p - \tilde{p}\|_{L^1} \|f\|_{L^\infty} \leq \frac{\varepsilon}{2}$$

it follows that

$$\left| \iint f\left(n^{d/2} \tilde{\Delta}_n\right) p \, d\mathbf{a} \, d\mathbf{p} - \mathbb{E}(f(\mathcal{X})) \right| \leq \varepsilon.$$

Since ε is arbitrary, Theorem 2 follows for an arbitrary L^1 density.

Step 3. By Step 2, we can and will assume that (\mathbf{a}, \mathbf{p}) is distributed according to a smooth density supported on Ω_κ^M for some κ and M . Let f be a smooth compactly supported test function. Divide Ω_κ^M into small cubes $\{Q_j\}$ such that if $(\mathbf{a}_j, \mathbf{p}_j)$ denotes the center of Q_j , then for each j , each $(\mathbf{a}, \mathbf{p}) \in Q_j$ and each $\Delta \in \mathbb{R}$ we have

$$\left| f\left(\frac{\Delta}{\Lambda(\mathbf{a}, \mathbf{p})}\right) - f\left(\frac{\Delta}{\Lambda(\mathbf{a}_j, \mathbf{p}_j)}\right) \right| \leq \varepsilon.$$

Such a partition exists since Λ is continuous and bounded away from 0 on Ω_κ^M . Then

$$\begin{aligned} \iint f\left(n^{d/2}\tilde{\Delta}_n\right) p \, d\mathbf{a} \, d\mathbf{p} &= \iint_{\Omega_\kappa^M} f\left(e^{z^2/2}\frac{n^{d/2}\Delta_n}{\Lambda(\mathbf{a}, \mathbf{p})}\right) p \, d\mathbf{a} \, d\mathbf{p} \\ &= \sum_j \iint_{Q_j} f\left(e^{z^2/2}\frac{n^{d/2}\Delta_n}{\Lambda(\mathbf{a}_j, \mathbf{p}_j)}\right) p \, d\mathbf{a} \, d\mathbf{p} + \delta(n) \end{aligned}$$

where $|\delta(n)| \leq \varepsilon$ for large n .

Applying Theorem 2* in the case where (\mathbf{a}, \mathbf{p}) is distributed according to \mathbf{P} conditioned on Q_j , we get

$$\begin{aligned} \lim_{n \rightarrow \infty} \iint_{Q_j} f\left(e^{z^2/2}\frac{n^{d/2}\Delta_n}{\Lambda(\mathbf{a}_j, \mathbf{p}_j)}\right) p \, d\mathbf{a} \, d\mathbf{p} &= \mathbf{P}(Q_j)\mathbb{E}\left(f\left(e^{z^2/2}\frac{\hat{\mathcal{X}}}{\Lambda(\mathbf{a}_j, \mathbf{p}_j)}\right)\right) \\ &= \iint_{Q_j} \mathbb{E}\left(f\left(e^{z^2/2}\frac{\hat{\mathcal{X}}}{\Lambda(\mathbf{a}, \mathbf{p})}\right)\right) p \, d\mathbf{a} \, d\mathbf{p} + \delta_j \end{aligned}$$

where $|\delta_j| \leq \varepsilon\mathbf{P}(Q_j)$.

By Step 1,

$$\iint_{Q_j} \mathbb{E}\left(f\left(e^{z^2/2}\frac{\hat{\mathcal{X}}}{\Lambda(\mathbf{a}, \mathbf{p})}\right)\right) p \, d\mathbf{a} \, d\mathbf{p} = \mathbf{P}(Q_j)\mathbb{E}(f(\mathcal{X})).$$

Summing over j we conclude that for large n

$$\left|\iint f\left(n^{d/2}\tilde{\Delta}_n\right) p \, d\mathbf{a} \, d\mathbf{p} - \mathbb{E}(f(\mathcal{X}))\right| \leq 3\varepsilon.$$

Since ε is arbitrary, Theorem 2 follows. \square

Remark 3.1. The argument of Step 3 provides the following extension of Theorem 2:

The triple $(n^{d/2}\tilde{\Delta}_n(\mathbf{a}, \mathbf{p}), \mathbf{a}, \mathbf{p})$ converges in law as $n \rightarrow \infty$ to the triple $(\mathcal{X}, \mathbf{a}, \mathbf{p})$ where \mathcal{X} has the distribution described in Theorem 2, and (\mathbf{a}, \mathbf{p}) is distributed according to \mathbf{P} and is independent from \mathcal{X} .

Remark 3.2. The argument of Step 2 shows that it suffices to prove Theorem 2* in the case \mathbf{P} has smooth density supported on Ω_κ^M for some κ and M .

Sections 4–8 are devoted to the proof of Theorem 2*. Note that similarly to (1.13) we have

$$\hat{\mathcal{X}} = e^{-z^2/2} \frac{|\mathbf{a}_{d+1} - \mathbf{a}_1|}{2\sigma(\mathbf{a}, \mathbf{p})\sqrt{\pi^3}} \sum_{\mathbf{m} \in \mathbb{Z}^d \setminus \{0\}} \frac{\sin 2\pi\theta(\mathbf{m})}{y(\mathbf{m})} e^{-4\pi^2 \mathbf{x}(\mathbf{m})D_{\mathbf{a}, \mathbf{p}} \mathbf{x}(\mathbf{m})}.$$

Remark 3.3. The statements of Theorems 2 and 2* look similar, however, there is an important distinction. Namely, the proof of Theorem 2* is constructive. In the course of the proof, given n , \mathbf{a} and z , we construct a lattice $\mathcal{L}(\mathbf{a}, n)$ and a character $\chi(\mathbf{a}, \mathbf{p}, n, z)$ such that the expression $n^{-d/2} \mathcal{X}(\mathbf{a}, \mathbf{p}, \mathcal{L}(\mathbf{a}, n), \chi(\mathbf{a}, \mathbf{p}, n, z))$ well-approximates the error in the Edgeworth expansion. We believe that such a construction could be made for more general distributions where the Edgeworth expansion fails, and this will be a subject of a future investigation. So, the difference between Theorems 2 and 2* is that in the first case, we have only an approximation in law, while in the second case, we are able to obtain an approximation in probability.

4. CUT OFF.

Here we begin the proof of Theorem 2*. By Remark 3.2, we may and will assume that \mathbf{P} has a smooth density supported on Ω_κ^M for some κ and M . Moreover, all constants, including the implied ones in \mathcal{O} -estimates, may depend on d, κ, M and \mathbf{P} .

As in the previous section, let

$$\Delta_n = \mathcal{E}_d(z) - F_n(z) \quad \text{where} \quad F_n(z) = \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z \right).$$

Denote by $v_T(x) = \frac{1}{\pi} \cdot \frac{1 - \cos Tx}{Tx^2}$ and let $\mathcal{V}(s, T) = \left(1 - \frac{|s|}{T}\right) \mathbb{1}_{|s| \leq T}$ be its Fourier transform.²

We use the approach of [13, Section XVI.3]. Let $T_2 = \sigma n^{2d+6}$. Note that $\sigma = \sigma(\mathbf{a}, \mathbf{p})$ is random. Since we assume that $(\mathbf{a}, \mathbf{p}) \in \Omega_\kappa^M$, σ is uniformly bounded, and bounded away from 0. So, $T_2 = \mathcal{O}(n^{2d+6})$ uniformly in (\mathbf{a}, \mathbf{p}) , i.e., there exist constants $c, C > 0$ such that $\lim_{n \rightarrow \infty} T_2/n^{2d+6} \in (c, C)$.

Decompose

$$(4.1) \quad -\Delta_n = [F_n - \mathcal{E}_d] \star v_{T_2}(z) + [F_n - F_n \star v_{T_2}](z) - [\mathcal{E}_d - \mathcal{E}_d \star v_{T_2}](z).$$

To estimate the last term, we split

$$(4.2) \quad [\mathcal{E}_d - \mathcal{E}_d \star v_{T_2}](z) = \int_{|x| \leq 1} [\mathcal{E}_d(z) - \mathcal{E}_d(z-x)] v_{T_2}(x) dx \\ + \int_{|x| \geq 1} [\mathcal{E}_d(z) - \mathcal{E}_d(z-x)] v_{T_2}(x) dx.$$

The first integral in (4.2) equals to

$$\int_{|x| \leq 1} \mathcal{E}'_d(z) x v_{T_2}(x) dx - \int_{|x| \leq 1} \frac{\mathcal{E}''_d(y(z, x))}{2} x^2 v_{T_2}(x) dx$$

²We use $\int e^{isx} f(s) ds$ as definition of the Fourier transform of $f \in L^1$ as in [13, Chapter XVI].

$$= - \int_{|x| \leq 1} \frac{\mathcal{E}_d''(y(z, x))}{2} \left(\frac{1 - \cos T_2 x}{\pi T_2} \right) dx = \mathcal{O} \left(\frac{1}{T_2} \right) = \mathcal{O}(n^{-(2d+6)})$$

where the first equality uses that v_T is even.

Since both \mathcal{E}_d and cosine are bounded the second integral in (4.2) is bounded by

$$C \int_{|x| \geq 1} \frac{dx}{T_2 x^2} = \frac{C}{T_2} = \mathcal{O}(n^{-(2d+6)}).$$

Thus, the last term in (4.1) is $\mathcal{O}(n^{-(2d+6)})$. Here and below, the constant C do not depend on the choice of (\mathbf{a}, \mathbf{p}) .

To estimate the second term in (4.1), we split the integral in $F_n \star v_{T_2}$ into regions $\{|x| \geq 1/\sqrt{T_2}\}$ and $\{|x| \leq 1/\sqrt{T_2}\}$. The contribution of $\{|x| \geq 1/\sqrt{T_2}\}$ is bounded by

$$C \int_{1/\sqrt{T_2}}^{\infty} \frac{dx}{T_2 x^2} = \frac{C}{\sqrt{T_2}} = \mathcal{O}(n^{-(d+3)}).$$

On the other hand

$$\int_{|x| \leq 1/\sqrt{T_2}} [F_n(z) - F_n(z-x)] v_{T_2}(x) dx = 0$$

unless there is a point of increase of F_n inside the interval

$$J_2 = \left[z - 1/\sqrt{T_2}, z + 1/\sqrt{T_2} \right].$$

The probability that J_2 contains a point of increase of F_n is bounded by

$$(4.3) \quad \sum_{m_1 + \dots + m_{d+1} = n} \mathbf{P}(\mathfrak{B}_{\mathbf{m}})$$

where

$$\mathfrak{B}_{\mathbf{m}} := \left\{ \frac{m_1 a_1 + \dots + m_{d+1} a_{d+1}}{\sigma \sqrt{n}} \in \left[z - 1/\sqrt{T_2}, z + 1/\sqrt{T_2} \right] \right\}.$$

Note that $\mathfrak{B}_{\mathbf{m}} = \left\{ |\mathbf{m} \cdot \mathbf{a} - \sigma z \sqrt{n}| \leq \sigma \sqrt{n/T_2} \right\}$. Since σ is bounded on Ω_{κ}^M there is a constant $L = L(M, \kappa)$ such that $\mathfrak{B}_{\mathbf{m}} \subset \bar{\mathfrak{B}}_{\mathbf{m}} := \left\{ |\mathbf{m} \cdot \mathbf{a} - \sigma z \sqrt{n}| \leq L \sqrt{n/T_2} \right\}$. To estimate $\mathbf{P}(\bar{\mathfrak{B}}_{\mathbf{m}})$ we consider the following variables on Ω_{κ}^M :

$$\zeta = (a_1, \dots, a_{d+1}, p_1, \dots, p_{d-1}).$$

Since ζ is distributed according to the bounded density it suffices to estimate the Lebesgue measure of $\bar{\mathfrak{B}}_{\mathbf{m}}$ in these coordinates. Without loss of generality we may assume that m_1 is the maximal among (m_1, \dots, m_{d+1}) , whence $m_1 > n/(d+1)$. Then for large n we have that

$$\left| \frac{\partial}{\partial a_1} [\mathbf{m} \cdot \mathbf{a} - \sigma z \sqrt{n}] \right| = \left| \left[m_1 - z \sqrt{n} \frac{\partial \sigma}{\partial a_1} \right] \right| \geq \frac{n}{2d}.$$

The last inequality follows because the second term is $\mathcal{O}(\sqrt{n})$. Accordingly, for each fixed value of $(a_2, \dots, a_{d+1}, p_1, \dots, p_{d-1})$ the measure of a_1 such that $\zeta \in \mathfrak{B}_{\mathbf{m}}$ belongs to the segment of length $\mathcal{O}(\sqrt{n}/T_2)$ is $\mathcal{O}(\sqrt{1/nT_2})$. Hence, each term in (4.3) is $\mathcal{O}\left(\frac{1}{\sqrt{nT_2}}\right)$, and so, the sum is $\mathcal{O}\left(\frac{n^d}{\sqrt{nT_2}}\right)$. Thus, with probability $1 - \mathcal{O}\left(\frac{1}{n^{7/2}}\right)$, we have that $-\Delta_n = \Delta_{n,2} + \mathcal{O}\left(T_2^{-1/2}\right)$ where

$$\begin{aligned}\Delta_{n,2} &= \frac{1}{2\pi} \int_{-T_2}^{T_2} \frac{\left[\phi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - \hat{\mathcal{E}}_d(t)\right]}{it} \mathcal{V}(t, T_2) e^{-itz} dt \\ &= \frac{1}{2\pi} \int_{-\frac{T_2}{\sigma\sqrt{n}}}^{\frac{T_2}{\sigma\sqrt{n}}} e^{-isz\sigma\sqrt{n}} \frac{\phi^n(s) - \hat{\mathcal{E}}_d(s\sigma\sqrt{n})}{is} \mathcal{V}(s, n, T_2) ds,\end{aligned}$$

$\mathcal{V}(s, n, T) = 1 - \left|\frac{s\sigma\sqrt{n}}{T}\right|$ and $\phi(s)$ is the characteristic function of X given by (2.1).

Let $T_1 = \sigma K_1 n^{d/2}$ for some constant $K_1 > 0$, and define

$$\Delta_{n,1} = \frac{1}{2\pi} \int_{-\frac{T_1}{\sigma\sqrt{n}}}^{\frac{T_1}{\sigma\sqrt{n}}} e^{-isz\sigma\sqrt{n}} \frac{\phi^n(s) - \hat{\mathcal{E}}_d(s\sigma\sqrt{n})}{is} \mathcal{V}(s, n, T_2) ds.$$

Note that $T_1 = \mathcal{O}(n^{d/2})$ with the implied constant independent of $(\mathbf{a}, \mathbf{p}) \in \Omega_{\kappa}^M$. Let $\Gamma_n = \Delta_{n,2} - \Delta_{n,1}$. Put $J_1 = [T_1/(\sigma\sqrt{n}), T_2/(\sigma\sqrt{n})]$ and

$$\tilde{\Gamma}_n = \frac{1}{2\pi} \int_{|s| \in J_1} e^{-isz\sigma\sqrt{n}} \frac{\phi^n(s)}{is} \mathcal{V}(s, n, T_2) ds.$$

Note that, due to the exponential decay of $\hat{\mathcal{E}}_d$,

$$|\tilde{\Gamma}_n - \Gamma_n| \leq C \int_{|s| \in J_1} \frac{|\hat{\mathcal{E}}_d(s\sigma\sqrt{n})|}{|s|} ds \leq C \int_{|s| \in J_1} \frac{e^{-nc s^2 \sigma^2}}{|s|} ds \leq C e^{-cT_1^2} \log |T_2/T_1|.$$

Hence, there exists $\varepsilon > 0$ such that $\Gamma_n = \tilde{\Gamma}_n + \mathcal{O}\left(e^{-\varepsilon T_1^2}\right)$.

Further, note that $T_1/(\sigma\sqrt{n})$ and $T_2/(\sigma\sqrt{n})$ do not depend on (\mathbf{a}, \mathbf{p}) . Thus, $\tilde{\Gamma}_n$ is an integral over the union of the two intervals J_1 and $-J_1$ whose lengths are independent of (\mathbf{a}, \mathbf{p}) .

The main result of Section 4 is the following.

Proposition 4.1.

$$(4.4) \quad \left\| \tilde{\Gamma}_n \right\|_{L^2} \leq \frac{C}{\sqrt{T_1 n^{d/2}}}.$$

The proof of Proposition 4.1 relies on the following estimates.

Lemma 4.2. *For each integer l there is a constant $C = C(l)$ such that*

$$\mathbf{E}(|\phi^{n-l}(s)|) \leq \frac{C}{n^{d/2}}$$

for all $|s| \geq 1$.

Lemma 4.3. *If I is a finite interval with length of order 1 and l be an integer then*

$$\int_I |\phi^{n-l}(s)| ds = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$$

(where the implicit constant depends on l and on the length of I but not on its location).

Lemmas 4.2 and 4.3 will be proven in Section 7 and Section 5, respectively.

Proof of Proposition 4.1. Note that \mathcal{V} is an even function in s and $\overline{\phi(s)} = \phi(-s)$. Therefore, the complex conjugate of $\tilde{\Gamma}_n$ is

$$\begin{aligned} \overline{\tilde{\Gamma}_n} &= \frac{1}{2\pi} \int_{|s| \in J_1} e^{isz\sigma\sqrt{n}} \frac{\overline{\phi^n(s)}}{-is} \mathcal{V}(s, n, T_2) ds \\ &= \frac{1}{2\pi} \int_{|-s| \in J_1} e^{i(-s)z\sigma\sqrt{n}} \frac{\overline{\phi^n(-s)}}{is} \mathcal{V}(-s, n, T_2) ds \\ &= \frac{1}{2\pi} \int_{|s| \in J_1} e^{-isz\sigma\sqrt{n}} \frac{\phi^n(s)}{is} \mathcal{V}(s, n, T_2) ds = \tilde{\Gamma}_n. \end{aligned}$$

To estimate the L^2 -norm of $\tilde{\Gamma}_n$, we write

$$\begin{aligned} \mathbf{E}(\tilde{\Gamma}_n^2) &= \frac{1}{4\pi^2} \mathbf{E} \left(\int_{|s| \in J_1} e^{-isz\sigma\sqrt{n}} \frac{\phi^n(s)}{is} \mathcal{V}(s, n, T_2) ds \right)^2 \\ &= -\frac{1}{4\pi^2} \iint_{|s_1|, |s_2| \in J_1} \mathbf{E} \left(e^{-i(s_1+s_2)z\sigma\sqrt{n}} \phi^n(s_1) \phi^n(s_2) \right) \frac{\mathcal{V}_n(s_1) ds_1}{s_1} \frac{\mathcal{V}_n(s_2) ds_2}{s_2} \end{aligned}$$

where

$$(4.5) \quad \mathcal{V}_n(s) = \mathcal{V}(s, n, T_2) = 1 - \left| \frac{s\sigma\sqrt{n}}{T_2} \right| = \left(1 - \frac{|s|}{n^{2d+\frac{11}{2}}} \right)$$

is independent of σ , and $0 \leq \mathcal{V}_n \leq 1$ on J_1 .

We split this integral into two parts.

(1) In the region where $|s_1 + s_2| \leq 1$, we use Lemma 4.3 to estimate the integral by

$$\begin{aligned}
& \mathbf{E} \left(\int_{|s_1| \in J_1} |\phi^n(s_1)| \int_{-1-s_1}^{1-s_1} |\phi^n(s_2)| \frac{ds_2}{|s_2|} \frac{ds_1}{|s_1|} \right) \\
(4.6) \quad & = \mathcal{O} \left(\int_{|s_1| \in J_1} \frac{1}{\sqrt{n}s_1^2} \mathbf{E}(|\phi^n(s_1)|) ds_1 \right).
\end{aligned}$$

Plugging the estimate of Lemma 4.2 into (4.6) and integrating we see that the contribution of the first region to $\mathbf{E}(\tilde{\Gamma}_n^2)$ is $\mathcal{O}\left(\frac{1}{T_1 n^{d/2}}\right)$.

(2) Consider now the region where $|s_1 + s_2| \geq 1$.

Recall that on Ω ,

$$(4.7) \quad p_1 + \cdots + p_{d+1} = 1, \quad \text{and} \quad p_1 a_1 + \cdots + p_{d+1} a_{d+1} = 0.$$

We use the $2d$ -dimensional coordinates $(a_1, \boldsymbol{\nu})$ where $\boldsymbol{\nu} := (p_1, p_3, \dots, p_d, b_2, \dots, b_{d+1})$.

Then there exists a compactly supported density $\rho = \rho(a_1, \boldsymbol{\nu})$ such that the contribution of the second region is

$$\iint_{\substack{|s_1|, |s_2| \in J_1 \\ |s_1 + s_2| \geq 1}} \left(\iint g(s_1, s_2, a_1, \boldsymbol{\nu}) \rho da_1 d\boldsymbol{\nu} \right) \frac{\mathcal{V}_n(s_1) ds_1}{s_1} \frac{\mathcal{V}_n(s_2) ds_2}{s_2}$$

where

$$g(s_1, s_2, a_1, \boldsymbol{\nu}) = e^{-i(s_1+s_2)z\sigma\sqrt{n}} e^{in(s_1+s_2)a_1} \psi^n(s_1) \psi^n(s_2).$$

To estimate this integral, we integrate by parts with respect to a_1 . Note that for each k we have

$$e^{isna_1} = \left[\frac{1}{isn} \frac{d}{da_1} \right]^k e^{isna_1}.$$

Fix a large k (for example, we can take $k = 8d + 25$). The integration by parts amounts to applying $\left(\frac{d}{da_1}\right)^k$ to $e^{-i(s_1+s_2)z\sigma\sqrt{n}} \rho[\psi(s_1)\psi(s_2)]^n$ which leads to terms formed by products of

$$\left\{ \left(\frac{d}{da_1}\right)^{k_1} \left[e^{-i(s_1+s_2)z\sigma\sqrt{n}} \right] \right\}, \quad \left\{ \left(\frac{d}{da_1}\right)^{k_2} [\rho] \right\}, \quad \text{and} \quad \left\{ \left(\frac{d}{da_1}\right)^{k_3} [\psi(s_1)\psi(s_2)]^n \right\}$$

where $k_1 + k_2 + k_3 = k$. Note that all of the above expressions depend implicitly on a_1 because p_2 and p_{d+1} depend on a_1 due to the second equation in (4.7). Rewriting that equation in the form

$$a_1 + \sum_{j=2}^{d+1} p_j b_j = 0,$$

we obtain $\frac{\partial p_j}{\partial a_1} = -1/b_j$, $j = 2$ or $d+1$. We also observe that when we integrate by parts, the boundary terms vanish because ρ is smooth and has compact support.

Thus, the contribution of the above term to the integral is bounded by the expectation of

$$C \iint_{\substack{|s_1|, |s_2| \in J_1 \\ |s_1 + s_2| \geq 1}} \frac{n^{(k_1/2)+k_3}}{|s_1 + s_2|^{k-k_1} n^k} |\phi^{n-k_3}(s_1)| |\phi^{n-k_3}(s_2)| \frac{ds_1 ds_2}{|s_1| |s_2|}.$$

To estimate the above integral we consider two cases, $k_1 \geq k - 3$ and $k_1 < k - 3$.

In the first case, we use trivial bounds $|s_1| \geq 1$, $|s_2| \geq 1$, $|s_1 + s_2|^{k-k_1} \geq 1$ and $|\phi^{n-k_3}(s_2)| \leq 1$, and Lemma 4.2 to estimate³ $\mathbb{E}(|\phi^{n-k_3}(s_1)|)$ to obtain the upper bound:

$$\frac{C}{n^{d/2+k-k_1/2-k_3}} \iint_{|s_1|, |s_2| \in J_1} ds_1 ds_2 \leq \frac{C|J_1|^2}{n^{(d+k-3)/2}} \leq \frac{CT_2^2}{n^{(d+k-1)/2}} = \frac{CT_2^2}{n^{9d/2+12}} \leq \frac{C}{\sqrt{T_1} n^{d/2}}.$$

Since $T_1 = \mathcal{O}(n^{d/2})$, $T_2 = \mathcal{O}(n^{2d+6})$ and $k = 8d + 25$, we have the last inequality. In the second case, we observe that $|s_1 + s_2|^{k-k_1} \geq |s_1 + s_2|^3$. We divide the integration region into two parts.

(a) $|s_1 + s_2| \geq 0.1|s_2|$. In this case the integrand is bounded by

$$\frac{C}{|s_1||s_2|^4} |\phi^{n-k_3}(s_1)|.$$

Using Lemma 4.2 to estimate the expectation of the last term and then performing the integration, we obtain the bound

$$\frac{n^{3/2} \ln n}{n^{d/2} T_1^3} = \frac{1}{n^{d/2} T_1} \times \frac{n^{3/2} \ln n}{T_1^2}.$$

The second factor is smaller than 1 since $T_1^2 = K_1^2 \sigma^2 n^d$ and $d \geq 2$.

(b) $|s_1 + s_2| \leq 0.1|s_2|$. In this case s_i 's are of the same order:

$$\frac{1}{2} \leq \left| \frac{s_1}{s_2} \right| \leq 2.$$

Accordingly, the integrand is bounded by

$$\frac{1}{s_1^2} |s_2 + s_1|^{-3} |\phi^{n-k_3}(s_1)| |\phi^{n-k_3}(s_2)|.$$

To perform the integration over s_2 , we divide the domain of integration into segments $I_l(s_1)$ of length of order 1, so that there exists $c, C > 0$ such that on I_l ,

$$c|l| \leq |s_2 + s_1| \leq C|l|.$$

Using Lemma 4.3 on each segment, we obtain

$$(4.8) \quad \int_{s_2 \in J_2, |s_1 + s_2| < 0.1s_2} \frac{|\phi^{n-k_3}(s_2)|}{|s_1 + s_2|^3} ds_2 \leq \sum_l \frac{C}{l^3 \sqrt{n}} \leq \frac{C}{\sqrt{n}}$$

³Here we use the fact that Lemma 4.2 applies to any absolutely continuous distribution of (\mathbf{a}, \mathbf{p}) . In particular, it applies to the integration with respect to the (normalized) Lebesgue measure.

where the constant C does not depend on s_1 . We now perform the integration over s_1 . Using Lemma 4.2 we bound the expectation of the integral by

$$(4.9) \quad \frac{C}{n^{d/2}} \int_{|s_1| \geq T_1/(\sigma\sqrt{n})} \frac{ds_1}{s_1^2} = \frac{C\sqrt{n}}{n^{d/2}T_1}.$$

Multiplying the bounds of (4.8) and (4.9), we obtain that the integral over region (b) is also within the bounds of Proposition 4.1. \square

Proposition 4.1 shows that (by taking K_1 sufficiently large) the contribution from $\tilde{\Gamma}_n$ to the L^2 -limit of $n^{d/2}\Delta_n$ can be made arbitrarily small. On $|s| \leq T_1/\sigma\sqrt{n}$, due to (4.5), we have

$$\mathcal{V}(s, n, T_2) = \left(1 - \frac{|s|}{n^{2d+\frac{11}{2}}}\right).$$

Hence, $\Delta_{n,1} = \hat{\Delta}_n + o(n^{-3d/2})$ where

$$(4.10) \quad \hat{\Delta}_n := \frac{1}{2\pi} \int_{|s| \leq T_1/\sigma\sqrt{n}} \frac{\phi^n(s) - \hat{\mathcal{E}}_d(s\sigma\sqrt{n})}{is} e^{-isz\sigma\sqrt{n}} ds.$$

In summary, the analysis of Section 4 shows that $n^{d/2}\|\hat{\Delta}_n - \Delta_n\|_{L^2} \rightarrow 0$ as $n \rightarrow \infty$. Hence, we only need to analyze $n^{d/2}\hat{\Delta}_n$ for large n .

5. CONTRIBUTION OF RESONANT INTERVALS.

5.1. Definition of resonant intervals. Denote

$$s_k = \frac{2\pi k}{|b_{d+1}|}$$

and let I_k be the segment of length $\frac{2\pi}{|b_{d+1}|}$ centered at s_k . Let K_2 be a constant such that $K_2 \gg K_1$. Due to the results of the previous section, it is sufficient to study

$$\hat{\Delta}_n = \sum_{|k| \leq K_2 n^{(d-1)/2}} \tilde{\mathcal{I}}_k$$

where

$$\tilde{\mathcal{I}}_k = \frac{1}{2\pi i} \int_{I_k} e^{-isz\sigma\sqrt{n}} \frac{\phi^n(s) - \hat{\mathcal{E}}_d(s\sigma\sqrt{n})}{s} \mathbb{1}_{|s| \leq T_1/\sigma\sqrt{n}} ds.$$

$\tilde{\mathcal{I}}_0 = o(n^{-d/2})$ due to [13, Section XVI.2]. Next, $\hat{\mathcal{E}}_d(s\sigma\sqrt{n})$ decays exponentially with respect to n outside of I_0 . So, its contribution to $\tilde{\mathcal{I}}_k$ is negligible for $k \neq 0$. Accordingly,

$$\hat{\Delta}_n = \sum_{0 < |k| \leq K_2 n^{(d-1)/2}} \mathcal{I}_k + o\left(\frac{1}{n^{d/2}}\right)$$

where

$$\mathcal{I}_k = \frac{1}{2\pi i} \int_{I_k} e^{-isz\sigma\sqrt{n}} \frac{\phi^n(s)}{s} \mathbb{1}_{|s| \leq T_1/\sigma\sqrt{n}} ds.$$

Write

$$\bar{s}_k = \arg \max_{s \in I_k} |\phi(s)|, \quad \phi(\bar{s}_k) = r_k e^{i\phi_k}.$$

Call the interval I_k *resonant* if $r_k^n \geq n^{-100d}$ and call it *non-resonant* otherwise. By definition, if the interval I_k is non-resonant, then $\mathcal{I}_k = \mathcal{O}(n^{-100d})$. Since there are $\mathcal{O}(n^{(d-1)/2})$ number of intervals (both resonant and non-resonant), the total contribution of the non-resonant intervals is at most $\mathcal{O}(n^{-(199d+1)/2})$ which is negligible. So, from now on, we focus only on the contribution of the resonant intervals.

5.2. Asymptotics of the resonant terms. The following lemma is similar to the results of [8, Section 5.2].

Lemma 5.1. *Suppose that*

$$(5.1) \quad r_k^n \geq n^{-100d}$$

and

$$(5.2) \quad \pm \frac{T_1}{\sigma\sqrt{n}} \notin I_k.$$

Then

$$\mathcal{I}_k = \frac{1}{i\sqrt{2\pi n\sigma}} \frac{r_k^n}{\bar{s}_k} e^{-z^2/2} e^{in\phi_k - i\bar{s}_k z \sigma\sqrt{n}} \left(1 + \mathcal{O}\left(\frac{\ln^3 n}{\sqrt{n}}\right) \right).$$

Proof. Let $e^{i\bar{s}_k a_j} = e^{i(\phi_k + \beta_j(k))}$ with $|\beta_j(k)| \leq \pi$. Then

$$(5.3) \quad r_k = \sum_{j=1}^{d+1} p_j \cos \beta_j(k),$$

and

$$(5.4) \quad \sum_{j=1}^{d+1} p_j \sin \beta_j(k) = 0.$$

From (5.1), we have

$$(5.5) \quad r_k \geq 1 - \frac{C \ln n}{n}.$$

(5.5) along with (5.3) give

$$(5.6) \quad \sum_{j=1}^{d+1} p_j \beta_j(k)^2 \leq \frac{C \ln n}{n},$$

and hence, $|\beta_j(k)| \leq C\sqrt{\frac{\ln n}{n}}$. Combining this with (5.4) we obtain

$$(5.7) \quad \sum_{j=1}^{d+1} p_j \beta_j(k) = \mathcal{O}\left(\frac{\ln^{3/2} n}{n^{3/2}}\right).$$

Next, by the definition of \bar{s}_k , $\frac{\partial}{\partial \delta} \Big|_{\delta=0} \phi(\bar{s}_k + \delta)$ is perpendicular to $\phi(\bar{s}_k)$ and whence

$$(5.8) \quad \sum_j p_j a_j \sin \beta_j(k) = 0.$$

Let $s \in I_k$, then $s = \bar{s}_k + \delta$ for some δ . Using Taylor expansion,

$$\begin{aligned} e^{i(\bar{s}_k + \delta)a_j} &= e^{i\phi_k} e^{i\beta_j(k)} \left(1 + ia_j \delta - \frac{a_j^2 \delta^2}{2}\right) + \mathcal{O}(\delta^3) \\ &= e^{i\phi_k} \left(\cos \beta_j(k) + i \sin \beta_j(k) + i\delta a_j \cos \beta_j(k) - \delta a_j \sin \beta_j(k)\right) \\ &\quad - e^{i\phi_k} (\cos \beta_j(k) + i \sin \beta_j(k)) \frac{a_j^2 \delta^2}{2} + \mathcal{O}(\delta^3). \end{aligned}$$

Thus,

$$\begin{aligned} \phi(\bar{s}_k + \delta) &= \sum_{j=1}^{d+1} p_j e^{i(\bar{s}_k + \delta)a_j} = \\ &= e^{i\phi_k} r_k + e^{i\phi_k} \sum_{j=1}^{d+1} p_j \cos \beta_j(k) \left(ia_j \delta - \frac{a_j^2 \delta^2}{2}\right) + \mathcal{O}\left(\frac{\ln^{3/2} n}{n^{3/2}} + \delta^3\right) \\ (5.9) \quad &= r_k e^{i\phi_k} \left(1 - \frac{\sigma^2 \delta^2}{2}\right) + i\delta e^{i\phi_k} \sum_{j=1}^{d+1} p_j a_j (\cos \beta_j(k) - 1) \\ &\quad - \frac{\delta^2}{2} e^{i\phi_k} \sum_{j=1}^{d+1} p_j a_j^2 (\cos \beta_j(k) - r_k) + \mathcal{O}\left(\frac{\ln^{3/2} n}{n^{3/2}} + \delta^3\right) \end{aligned}$$

where we have used (5.4), (5.7), (5.8) as well as

$$p_1 a_1 + \cdots + p_{d+1} a_{d+1} = 0 \quad \text{and} \quad p_1 a_1^2 + \cdots + p_{d+1} a_{d+1}^2 = \sigma^2.$$

The main term in (5.9) is the first one since

$$|\cos \beta_j(k) - r_k| \leq |\cos \beta_j(k) - 1| + |1 - r_k| = \mathcal{O}\left(\beta_j(k)^2 + \frac{\ln n}{n}\right).$$

Hence, using (5.6), we obtain

$$\begin{aligned}\phi(\bar{s}_k + \delta) &= r_k e^{i\phi_k} \left(1 - \frac{\sigma^2 \delta^2}{2}\right) + \mathcal{O}\left((\delta + \delta^2) \frac{\ln n}{n}\right) + \mathcal{O}\left(\frac{\ln^{3/2} n}{n^{3/2}} + \delta^3\right) \\ &= r_k e^{i\phi_k} \left(1 - \frac{\sigma^2 \delta^2}{2}\right) + \mathcal{O}\left(\frac{\ln^{3/2} n}{n^{3/2}} + \delta^3\right).\end{aligned}$$

In summary,

$$(5.10) \quad \phi(\bar{s}_k + \delta) = r_k e^{i\phi_k} \left(1 - \frac{\sigma^2 \delta^2}{2}\right) + \mathcal{O}\left(\frac{\ln^{3/2} n}{n^{3/2}} + \delta^3\right).$$

Next, split $I_k = I'_k \cup I''_k$ where I'_k is the part of I_k where $\left\{|\delta| \leq \frac{C \ln n}{\sqrt{n}}\right\}$ and $I''_k = I_k \setminus I'_k$. Note that, if (5.5) holds, then Lemma 2.1 shows that \bar{s}_k is close to s_k . So, the set $\left\{|\delta| \leq \frac{C \ln n}{\sqrt{n}}\right\}$ is completely contained in I_k . Lemma 2.1 also shows that for $s \in I''_k$, $|\phi(s)|^n \leq n^{-c \ln n}$. So, the contribution of I''_k to \mathcal{I}_k is negligible.

Next, on I'_k the error term in (5.10) is $\mathcal{O}\left(\frac{\ln^3 n}{n^{3/2}}\right)$. Hence, the contribution to \mathcal{I}_k from I'_k is

$$\begin{aligned}& \frac{r_k^n}{2\pi i \bar{s}_k} e^{i(n\phi_k - \sqrt{n}\sigma z \bar{s}_k)} \int_{|\delta| < C \ln n / \sqrt{n}} \left(1 - \frac{\sigma^2 \delta^2}{2} + \mathcal{O}\left(\frac{\ln^3 n}{n^{3/2}}\right)\right)^n (1 + \mathcal{O}(\delta)) e^{-i\sigma z \delta \sqrt{n}} d\delta \\ &= \frac{r_k^n}{2\pi i \bar{s}_k} e^{i(n\phi_k - \sqrt{n}\sigma z \bar{s}_k)} \int_{|\delta| < C \ln n / \sqrt{n}} e^{-\sigma^2 \delta^2 n / 2 - i\sigma \delta \sqrt{n} z} e^{\mathcal{O}(\ln^3 n / \sqrt{n})} (1 + \mathcal{O}(n\delta^4 + \delta)) d\delta \\ &= \frac{r_k^n}{2\pi i \bar{s}_k} e^{i(n\phi_k - \sqrt{n}\sigma z \bar{s}_k)} \left(1 + \mathcal{O}\left(\frac{\ln^3 n}{\sqrt{n}}\right)\right) \int_{|\delta| < C \ln n / \sqrt{n}} e^{-\sigma^2 \delta^2 n / 2 - i\sigma \delta \sqrt{n} z} d\delta.\end{aligned}$$

Making the change of variables $\sigma \delta \sqrt{n} = t$, we can rewrite the last expression as

$$\begin{aligned}& \frac{r_k^n e^{-z^2/2}}{2\pi i \bar{s}_k \sigma \sqrt{n}} e^{i(n\phi_k - \sqrt{n}\sigma z \bar{s}_k)} \left(1 + \mathcal{O}\left(\frac{\ln^3 n}{\sqrt{n}}\right)\right) \int_{|\delta| < C \sigma \ln n} e^{-(t+iz)^2/2} dt \\ &= \frac{r_k^n e^{-z^2/2}}{2\pi i \bar{s}_k \sigma \sqrt{n}} e^{i(n\phi_k - \sqrt{n}\sigma z \bar{s}_k)} \left(1 + \mathcal{O}\left(\frac{\ln^3 n}{\sqrt{n}}\right)\right) \int_{\mathbb{R}} e^{-(t+iz)^2/2} dt \\ &= \frac{r_k^n e^{-z^2/2}}{\sqrt{2\pi i \bar{s}_k \sigma \sqrt{n}}} e^{i(n\phi_k - \sqrt{n}\sigma z \bar{s}_k)} \left(1 + \mathcal{O}\left(\frac{\ln^3 n}{\sqrt{n}}\right)\right).\end{aligned}$$

This completes the proof of the lemma. \square

5.3. Proof of Lemma 4.3.

Proof. Note that if $|\phi^n(s)| \leq n^{-100d}$ then $|\phi^{n-l}(s)| \leq n^{-50d}$ and if $|\phi^n(s)| \geq n^{-100d}$ then $|\phi(s)| \geq 1 - \frac{C \ln n}{n}$, and hence, $|\phi^{-l}(s)| \leq 2$. Therefore

$$(5.11) \quad |\phi^{n-l}(s)| \leq 2|\phi^n(s)| + \frac{1}{n^{50d}}.$$

Thus it suffices to prove the result for $l = 0$. We can cover I by a finite number of intervals I_k . For the intervals where $r_k^n < n^{-100d}$, we have

$$\sum_{r_k^n < n^{-100d}} \int_{I_k} |\phi^n(s)| ds \leq C \frac{|I|}{n^{100d}}.$$

For resonant intervals where $r_k^n \geq n^{-100d}$ and $k \neq 0$, the proof of Lemma 5.1 shows that

$$\sum_{r_k^n \geq n^{-100d}} \int_{I_k} |\phi^n(s)| ds \leq C \int_{|\delta| < C \ln n / \sqrt{n}} (1 - c\delta^2)^n d\delta + \mathcal{O}(n^{-c \ln n}) = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Finally, the case $k = 0$ is analyzed in [13, Section XVI.2]. \square

6. SIMPLIFYING THE ERROR TERM.

As noted above, in the resonant case, Lemma 2.1 gives $d(\bar{s}_k) \leq C\sqrt{\frac{\ln n}{n}}$. In particular, $\text{dist}(b_{d+1}\bar{s}_k, b_{d+1}s_k) \leq C\sqrt{\frac{\ln n}{n}}$ because $b_{d+1}s_k \in 2\pi\mathbb{Z}$. So, $\xi_k := \bar{s}_k - s_k$ satisfies

$$|\xi_k| \leq C\sqrt{\frac{\ln n}{n}}.$$

Since $d(s_k) = d(\bar{s}_k) + \mathcal{O}(s_k - \bar{s}_k)$, we also have $d(s_k) \leq C\sqrt{\frac{\ln n}{n}}$.

Noting that $b_j s_k = \frac{2\pi k b_j}{|b_{d+1}|}$ we define $\eta_{j,k} = \frac{2\pi k b_j}{|b_{d+1}|} + 2\pi l_{j,k}$, for $j = 2, \dots, d+1$, where $l_{j,k}$ is the unique integer such that

$$(6.1) \quad -\pi < \frac{2\pi k b_j}{|b_{d+1}|} + 2\pi l_{j,k} \leq \pi.$$

Then, $\eta_{d+1,k} = 0$ and the foregoing discussion gives

$$(6.2) \quad |\eta_{j,k}| \leq C\sqrt{\frac{\ln n}{n}}.$$

Define the random vector

$$X_k = \sqrt{n}\boldsymbol{\eta}_k$$

where $\boldsymbol{\eta}_k$ is the vector with components $(\eta_{2,k}, \dots, \eta_{d,k})$, and let

$$Y_k = \frac{k}{n^{(d-1)/2}}.$$

Also, for the remainder of the paper, we fix a constant α :

$$(6.3) \quad \alpha = \frac{1}{2(d-1)}.$$

The main result of Section 6 is the following.

Proposition 6.1. *Let $\tilde{\Delta}_n(\delta, K) :=$*

$$\frac{|b_{d+1}|e^{-z^2/2}}{n^{d/2}\sigma\sqrt{2\pi^3}} \sum_{k \in S(n, \delta, K)} \frac{\sin\left(\frac{2\pi n^{d/2}}{|b_{d+1}|}(\sqrt{n}a_1 - z\sigma)Y_k + (\sqrt{n}\mathbf{q} + z\sigma\boldsymbol{\omega}) \cdot X_k\right)}{Y_k} e^{-X_k D_{\mathbf{a}, \mathbf{p}} \cdot X_k}$$

where

$$(6.4) \quad S(n, \delta, K) = \{k > 0 \mid \delta < Y_k < K, |Y_k|^\alpha \|X_k\| < 2^{K+1}\}$$

and the vectors $\boldsymbol{\omega} = (\boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_d)$ and $\mathbf{q} = (\mathbf{q}_2, \dots, \mathbf{q}_d)$ satisfy

$$(6.5) \quad \boldsymbol{\omega}_m = \frac{2 \sum_{l=1}^{d+1} p_l p_m (b_l - b_m)}{\sum_{j=1}^{d+1} \sum_{l=1}^{d+1} p_l p_j (b_l - b_j)^2}, \quad \mathbf{q}_m = p_m, \quad m = 2, \dots, d.$$

Then, given ε we can find δ, K such that

$$\mathbb{P}\left(|\hat{\Delta}_n - \tilde{\Delta}_n(\delta, K)| > \varepsilon/n^{d/2}\right) < \varepsilon.$$

Before proving this, we obtain an approximation for r_k and use it to obtain an approximation for \mathcal{I}_k .

Sublemma 6.2. *There exists a $(d-1) \times (d-1)$ matrix $D_{\mathbf{a}, \mathbf{p}}$ such that*

$$(6.6) \quad r_k = 1 - \boldsymbol{\eta}_k D_{\mathbf{a}, \mathbf{p}} \cdot \boldsymbol{\eta}_k + \mathcal{O}(\|\boldsymbol{\eta}_k\|^3).$$

Proof. Writing $r_k^2 = \psi(\bar{s}_k)\overline{\psi(\bar{s}_k)}$, $\bar{s}_k = s_k + \xi_k$ and substituting $\eta_{j,k} + b_j \xi_k$ for $b_j \bar{s}_k$, we obtain,

$$\begin{aligned} r_k^2 &= \sum_{j=1}^{d+1} p_j^2 + 2 \sum_{l>j, j \neq 1} p_l p_j \cos[(b_l - b_j)\xi_k + \eta_{l,k} - \eta_{j,k}] + 2p_{d+1}p_1 \cos b_{d+1}\xi_k \\ &\quad + 2 \sum_{j=2}^d p_j p_1 \cos(b_j \xi_k + \eta_{j,k}). \end{aligned}$$

Therefore,

$$\begin{aligned} r_k^2 &= 1 - \sum_{l>j, j \neq 1} p_l p_j [(b_l - b_j)\xi_k + \eta_{l,k} - \eta_{j,k}]^2 - p_{d+1}p_1 b_{d+1}^2 \xi_k^2 \\ &\quad - \sum_{j=2}^d p_j p_1 (b_j \xi_k + \eta_{j,k})^2 + \mathcal{O}\left(\xi_k^3 + \sum_{l=2}^d \eta_{l,k}^3\right). \end{aligned}$$

Note that the implied constants here and below can be chosen to be independent of $(\mathbf{a}, \mathbf{p}) \in \Omega_\kappa^M$.

Taking $\eta_{1,k} = b_1 = 0$, we can write the above as

$$\begin{aligned} r_k^2 &= -\xi_k^2 \sum_{l>j} p_l p_j (b_l - b_j)^2 - 2\xi_k \widehat{\sum} p_l p_j (b_l - b_j) (\eta_{l,k} - \eta_{j,k}) \\ &\quad + 1 - \widehat{\sum} p_l p_j (\eta_{l,k} - \eta_{j,k})^2 + \mathcal{O} \left(\xi_k^3 + \sum_{l=1}^d \eta_{l,k}^3 \right) \end{aligned}$$

where the sum in $\widehat{\sum}$ is taken over the pairs (l, j) such that $l > j$ and $(l, j) \neq (d+1, 1)$.

Since r_k^2 is approximated by a quadratic polynomial in ξ_k (the unknown) we can approximate ξ_k by determining $\operatorname{argmax} r_k^2(\xi)$, obtaining

$$\begin{aligned} (6.7) \quad \xi_k &= -\frac{\widehat{\sum} p_l p_j (b_l - b_j) (\eta_{l,k} - \eta_{j,k})}{\sum_{l>j} p_l p_j (b_l - b_j)^2} + \mathcal{O}(\|\boldsymbol{\eta}_k\|^2) \\ &= -\frac{\sum_{j=1}^{d+1} \sum_{l=1}^{d+1} p_l p_j (b_l - b_j) \eta_{j,k}}{\frac{1}{2} \sum_{j=1}^{d+1} \sum_{l=1}^{d+1} p_l p_j (b_l - b_j)^2} + \mathcal{O}(\|\boldsymbol{\eta}_k\|^2). \end{aligned}$$

We recall that $b_1 = 0$ and $\eta_{1,k} = \eta_{d+1,k} = 0$. Substituting back we find r_k in terms of $\eta_{j,k}$ only. Namely,

$$r_k^2 = 1 - \widehat{\sum} p_l p_j (\eta_{l,k} - \eta_{j,k})^2 + \frac{\left[\widehat{\sum} p_l p_j (b_l - b_j) (\eta_{l,k} - \eta_{j,k}) \right]^2}{\sum_{l>j} p_l p_j (b_l - b_j)^2} + \mathcal{O} \left(\sum_{l=1}^d \eta_{l,k}^3 \right).$$

Put $R = \left[\sum_{l>j} p_l p_j (b_l - b_j)^2 \right]^{-1}$. Then,

$$\begin{aligned} (6.8) \quad r_k^2 &= 1 + \widehat{\sum} p_l p_j [p_l p_j (b_l - b_j)^2 R - 1] (\eta_{l,k} - \eta_{j,k})^2 \\ &\quad + \sum_{\substack{l>j, m>\bar{m} \\ (l,j) \neq (m,\bar{m}) \\ (l,j), (m,\bar{m}) \neq (d+1,1)}} p_l p_j p_m p_{\bar{m}} (b_l - b_j) (b_m - b_{\bar{m}}) R (\eta_{l,k} - \eta_{j,k}) (\eta_{m,k} - \eta_{\bar{m},k}) + \mathcal{O} \left(\sum_{l=1}^d \eta_{l,k}^3 \right) \\ &:= 1 - 2 \sum_{l,j=2}^d D_{l,j}(\mathbf{a}, \mathbf{p}) \eta_{l,k} \eta_{j,k} + \mathcal{O} \left(\sum_{l=1}^d \eta_{l,k}^3 \right). \end{aligned}$$

Thus,

$$r_k = 1 - \sum_{l,j=2}^d D_{l,j}(\mathbf{a}, \mathbf{p}) \eta_{l,k} \eta_{j,k} + \mathcal{O} \left(\sum_{l=1}^d \eta_{l,k}^3 \right) = 1 - \boldsymbol{\eta}_k D_{\mathbf{a}, \mathbf{p}} \cdot \boldsymbol{\eta}_k + \mathcal{O}(\|\boldsymbol{\eta}_k\|^3)$$

where $D_{\mathbf{a}, \mathbf{p}}$ is the $(d-1) \times (d-1)$ matrix with

$$(6.9) \quad [D_{\mathbf{a}, \mathbf{p}}]_{i,j} = D_{i,j}(\mathbf{a}, \mathbf{p}),$$

proving (6.6). \square

Lemma 6.3. *The matrix $D_{\mathbf{a}, \mathbf{p}}$ defined by (6.9) satisfies*

$$\mathcal{I}_k = \frac{e^{-z^2/2}}{i\sqrt{2\pi n\sigma}} \frac{(1 - \boldsymbol{\eta}_k D_{\mathbf{a}, \mathbf{p}} \cdot \boldsymbol{\eta}_k + \mathcal{O}(\|\boldsymbol{\eta}_k\|^3))^n}{\bar{s}_k} e^{in\phi_k - i\bar{s}_k z \sigma \sqrt{n}} (1 + o(1))$$

where $\boldsymbol{\eta}_k = (\eta_{2,k}, \dots, \eta_{d,k})$.

Proof. Follows directly from Lemma 5.1 and (6.6). \square

We next consider the \mathcal{I}_k at the two ends. Let $\mathcal{B}(\mathbf{a}, \mathbf{p})$ be the contribution of these boundary terms, i.e. from k such that $\pm \frac{T_1}{\sigma\sqrt{n}} \in I_k$. By Lemma 4.3,

$$\mathcal{B}(\mathbf{a}, \mathbf{p}) \leq \frac{C}{T_1}.$$

Recalling that $T_1 = K_1 \sigma n^{d/2}$, we see that we can make $n^{d/2} \mathcal{B}(\mathbf{a}, \mathbf{p})$ as small as we wish by taking K_1 large. So, from now on, we ignore these terms.

Lemma 6.4. *Let*

$$\mathcal{I}_{k,l} = \mathcal{I}_k \mathbb{1}_{|k|^{\alpha} n^{1/4} \|\boldsymbol{\eta}_k\| \in [2^l, 2^{l+1})}.$$

For all sufficiently large $K > 0$, there is a constant \tilde{c} such that

$$\mathbf{E} \left(\sum_{0 < |k| < Kn^{(d-1)/2}} \widehat{\sum}_l |\mathcal{I}_{k,l}| \right) = \mathcal{O} \left(\frac{1}{n^{d/2}} 2^{K(d-1)} \exp(-\tilde{c}2^{2K}) \right)$$

where the sum in $\widehat{\sum}_l$ is over l satisfying $l > K$ and $2^l < \frac{Kk^\alpha \sqrt{\ln n}}{n^{1/4}}$.

Remark 6.5. We could restrict to l satisfying $l > K$ and $2^l < \frac{Kk^\alpha \sqrt{\ln n}}{n^{1/4}}$ since, by the discussion at the beginning of Section 6, it is enough to consider the intervals satisfying (6.2) and we can take $K > C$ where C is the constant from (6.2).

The proof of the above lemma will be given in Section 7. Lemma 6.4 shows that we should focus on the contribution of $\mathcal{I}_{k,l}$ with

$$0 < |k| < K_2 n^{(d-1)/2} \quad \text{and} \quad l \leq K_2.$$

Next, we prove a result that allows us to simplify $\hat{\Delta}_n$ even further. Recall that we are dealing with resonant k , that is, we assume that $r_k^n \geq n^{-100d}$.

Lemma 6.6. (a) $\bar{s}_k = s_k - \boldsymbol{\omega} \cdot \boldsymbol{\eta}_k + \mathcal{O}(\|\boldsymbol{\eta}_k\|^2)$ where $\boldsymbol{\omega} = \boldsymbol{\omega}(\mathbf{a}, \mathbf{p})$ is the $1 \times (d-1)$ vector defined in (6.5).

(b) If $\|\boldsymbol{\eta}_k\| = \mathcal{O}\left(\frac{\ln n}{\sqrt{n}}\right)$ then $n\phi_k = ns_ka_1 + np_2\eta_{2,k} + \cdots + np_d\eta_{d,k} + o(1)$.

Proof. Since $\bar{s}_k - s_k = \xi_k$ part (a) follows by (6.7).

Recall that $\phi(\bar{s}_k) = r_k e^{i\phi_k}$, and by (5.10)

$$\phi_k = \arg \phi(s_k) + \mathcal{O}\left(|\bar{s}_k - s_k|^3 + \frac{\ln^{3/2} n}{n^{3/2}}\right).$$

Note that,

$$\phi(s_k) = e^{is_ka_1}(p_1 + p_2 e^{i\eta_{2,k}} + \cdots + p_d e^{i\eta_{d,k}} + p_{d+1}).$$

Thus,

$$\begin{aligned} \arg(\phi(s_k)) &= s_ka_1 + \tan^{-1}\left(\frac{p_2 \sin \eta_{2,k} + \cdots + p_d \sin \eta_{d,k}}{p_1 + p_2 \cos \eta_{2,k} + \cdots + p_d \cos \eta_{d,k} + p_{d+1}}\right) \\ &= s_ka_1 + \sum_{l=2}^d p_l \eta_{l,k} + \mathcal{O}(\|\boldsymbol{\eta}_k\|^3) \end{aligned}$$

since the denominator in the first line is $1 + \mathcal{O}(\|\boldsymbol{\eta}_k\|^2)$. Part (b) now follows easily. \square

Proof of Proposition 6.1. First, we show that it is enough to consider $\mathcal{I}_{k,l}$ when

$$\delta n^{(d-1)/2} \leq |k| < K_2 n^{(d-1)/2} \quad \text{and} \quad l \leq K_2$$

for appropriately chosen δ and K_2 .

Recall from Section 5.1 that

$$\hat{\Delta}_n = \sum_{0 < |k| \leq K_2 n^{(d-1)/2}} \mathcal{I}_k + o\left(\frac{1}{n^{d/2}}\right)$$

for $K_2 \gg K_1$. By Lemma 6.4, the contribution of $\mathcal{I}_{k,l}$ with

$$0 < |k| < K_2 n^{(d-1)/2} \quad \text{and} \quad l > K_2$$

can be made arbitrarily small by choosing K_2 large.

Next, we claim that the distribution of $\boldsymbol{\eta}_k$ has bounded density. Since (\mathbf{a}, \mathbf{p}) has a bounded density on ${}^4\Omega_\kappa^M$, the vector

$$\mathbf{b} = \left(\frac{b_2}{|b_{d+1}|}, \dots, \frac{b_d}{|b_{d+1}|} \right)$$

⁴Recall that Ω_κ^M is defined by (3.4).

has a bounded density on $V_\kappa^M := \{(x_1, \dots, x_{d-1}) \mid \forall j \quad \kappa(2M)^{-1} \leq x_j \leq 2M\kappa^{-1}\}$. Let L denote the supremum of the density of \mathbf{b} . Since $\boldsymbol{\eta}_k$ is obtained by rescaling \mathbf{b} by $2\pi k$ and taking mod 2π , its density is bounded by

$$(6.10) \quad \frac{L}{2\pi k} \times \left\lceil \frac{4\pi M k}{\kappa} \right\rceil \leq \frac{4ML}{\kappa}$$

where the second factor on the LHS accounts for the multiplicity of the fractional part on V_κ^M . Since the RHS of (6.10) is independent of k , the claim follows.

Next, define

$$A_1 = \{(\mathbf{a}, \mathbf{p}) \mid \mathcal{I}_{k,l} = 0 \ \forall k, l \text{ s.t. } 0 < |k| < \delta n^{(d-1)/2} \text{ and } l \leq K_2\}.$$

Then

$$\begin{aligned} A_1^c &= \{(\mathbf{a}, \mathbf{p}) \mid \exists k, l \text{ s.t. } 0 < |k| < \delta n^{(d-1)/2}, l \leq K_2, \mathbb{1}_{|k|^\alpha n^{1/4} \|\boldsymbol{\eta}_k\| \in [2^l, 2^{l+1})} = 1\} \\ &= \{(\mathbf{a}, \mathbf{p}) \mid \exists k \text{ s.t. } 0 < |k| < \delta n^{(d-1)/2}, |k|^\alpha n^{1/4} \|\boldsymbol{\eta}_k\| < 2^{K_2+1}\}. \end{aligned}$$

Thus,

$$(6.11) \quad \begin{aligned} \mathbf{P}(A_1^c) &\leq \sum_{0 < |k| < \delta n^{(d-1)/2}} \mathbf{P}(|k|^\alpha n^{1/4} \|\boldsymbol{\eta}_k\| < 2^{K_2+1}) \\ &\leq \sum_{0 < |k| < \delta n^{(d-1)/2}} \frac{C 2^{(K_2+1)(d-1)}}{|k|^{(d-1)\alpha} n^{(d-1)/4}} = \mathcal{O}\left(\sqrt{\delta} 2^{(K_2+1)(d-1)}\right) \end{aligned}$$

where $\alpha = [2(d-1)]^{-1}$ (see (6.3)) and the probability estimate follows from $\boldsymbol{\eta}_k$ having a bounded density.

Hence, for K_2 and δ such that $\sqrt{\delta} 2^{(K_2+1)(d-1)}$ is small, we can approximate $\hat{\Delta}_n$ by the sum of \mathcal{I}_k 's with $\delta \leq |k| n^{-(d-1)/2} < K_2$ and $|k|^\alpha n^{1/4} \|\boldsymbol{\eta}_k\| < 2^{K_2+1}$.

Combining terms corresponding to k and $-k$, we obtain the following approximation to the distribution of Δ_n for large n

$$\frac{|b_{d+1}| e^{-z^2/2}}{n^{d/2} \sigma \sqrt{2\pi^3}} \sum_{k \in S(n, \delta, K)} \frac{\sin(n\phi_k - \bar{s}_k z \sigma \sqrt{n})}{Y_k} e^{-X_k D_{\mathbf{a}, \mathbf{p}} \cdot X_k}$$

for appropriate choices of K and δ , and where $S(n, \delta, K)$ is defined in (6.4). The restriction $Y > \delta$ in $S(n, \delta, K)$ comes from (6.11), the upper bound $Y < K$ comes from (4.10), and the restriction $|Y_k|^\alpha \|X_k\| < 2^{K+1}$ comes from Lemma 6.4. We have also used Lemma 6.6(a) and the fact that $|s_k| > c\delta n^{(d-1)/2}$ in the region we consider to replace \bar{s}_k by s_k .

Recall (see (6.5)) that $\mathbf{q} := (p_2, \dots, p_d)$. Lemma 6.6(b) shows that

$$\begin{aligned} n\phi_k - \bar{s}_k z \sigma \sqrt{n} &= s_k(na_1 - z\sigma\sqrt{n}) + n\mathbf{q} \cdot \boldsymbol{\eta}_k + z\sigma\sqrt{n}\boldsymbol{\omega} \cdot \boldsymbol{\eta}_k + o(1) \\ &= \frac{2\pi n^{d/2}}{|b_{d+1}|} (\sqrt{n}a_1 - z\sigma) Y_k + (\sqrt{n}\mathbf{q} + z\sigma\boldsymbol{\omega}) \cdot X_k + o(1). \end{aligned}$$

Therefore, for large n and K and δ such that $\sqrt{\delta}2^{(K+1)(d-1)}$ is very small, the distribution of $\hat{\Delta}_n$ is well approximated by $\tilde{\Delta}_n(\delta, K)$ completing the proof of the proposition. \square

7. EXPECTATION OF THE CHARACTERISTIC FUNCTION.

Proof of Lemma 4.2. As in the proof of Lemma 4.3, the inequality (5.11) shows that it suffices to consider the case $l = 0$. Recall that $d(s) = \max_{2 \leq j \leq d+1} d(b_j s, 0)$ where the distance is computed on the torus $\mathbb{R}/(2\pi\mathbb{Z})$. Lemma 2.1 shows that there is a positive constant c such that

$$(7.1) \quad |\phi^n(s)| \leq e^{-cnd(s)^2}.$$

To prove the lemma we decompose $\mathbf{E}(e^{-cnd(s)^2})$ into the pieces where $d(s)\sqrt{n}$ is of order 2^l for some $l \leq (\log_2 n)/2$.

Since \mathbf{P} has a bounded density, the distribution of the $(b_2 s, \dots, b_{d+1} s)$ has bounded density on \mathbb{T}^d where the bound is uniform for $|s| \geq 1$. Hence

$$\mathbf{P}(c_1 \leq d(s) \leq c_2) = \mathcal{O}(c_2^d - c_1^d)$$

for all $0 \leq c_1 < c_2 < 1$ uniformly in $|s| \geq 1$. Therefore,

$$\begin{aligned} \mathbf{E}(|\phi^n(s)|) &\leq C \mathbf{P}\left(d(s) < \frac{1}{\sqrt{n}}\right) + C \sum_{l=0}^{(\log_2 n)/2} \mathbf{P}(d(s)\sqrt{n} \in [2^l, 2^{l+1})) e^{-c4^l} \\ &\leq \frac{C}{n^{d/2}} + C \sum_{l=0}^{(\log_2 n)/2} \frac{2^{dl}}{n^{d/2}} e^{-c4^l} \leq \frac{C}{n^{d/2}} \end{aligned}$$

where the constant C can be chosen uniformly for all $|s| \geq 1$. This completes the proof. \square

Proof of Lemma 6.4. Since, by (6.6), $r_k = 1 - \boldsymbol{\eta}_k D_{\mathbf{a}, \mathbf{p}} \cdot \boldsymbol{\eta}_k + \mathcal{O}(\|\boldsymbol{\eta}_k\|^3)$ where the implied constant is independent of $(\mathbf{a}, \mathbf{p}) \in \Omega_\kappa^M$ and $|k|^\alpha n^{1/4} \|\boldsymbol{\eta}_k\| \in [2^l, 2^{l+1})$, we have

$$r_k \leq 1 - c \frac{4^l}{|k|^{2\alpha} \sqrt{n}},$$

where c is independent of (\mathbf{a}, \mathbf{p}) . Accordingly,

$$r_k^n \leq C e^{-\frac{c2^{2l}\sqrt{n}}{|k|^{2\alpha}}}.$$

Also, similarly to the proof of Lemma 4.2, we get

$$\mathbf{P}(|k|^\alpha n^{1/4} \|\boldsymbol{\eta}\| \in [2^l, 2^{l+1})) \leq \frac{C2^{l(d-1)}}{\sqrt{|k|} n^{(d-1)/4}}.$$

Hence,

$$\mathbf{E}(|\mathcal{I}_{k,l}|) \leq \frac{C e^{-\frac{c2^{2l}\sqrt{n}}{|k|^{2\alpha}}}}{\sqrt{n}|k|} \frac{2^{l(d-1)}}{\sqrt{|k|n^{(d-1)/4}}} = \frac{C 2^{l(d-1)} e^{-\frac{c2^{2l}\sqrt{n}}{|k|^{2\alpha}}}}{|k|^{3/2} n^{(d+1)/4}}.$$

Thus,

$$\widehat{\sum_l} \mathbf{E}(|\mathcal{I}_{k,l}|) \leq \frac{C 2^{K(d-1)} e^{-\frac{c2^{2K}\sqrt{n}}{|k|^{2\alpha}}}}{|k|^{3/2} n^{(d+1)/4}}.$$

Therefore, we need to estimate

$$(7.2) \quad \sum_{0 < |k| < K n^{(d-1)/2}} \frac{C 2^{K(d-1)} e^{-\frac{c2^{2K}\sqrt{n}}{|k|^{2\alpha}}}}{|k|^{3/2} n^{(d+1)/4}} \\ = \frac{C}{n^{d/2}} \sum_{0 < |k| < K n^{(d-1)/2}} \frac{1}{|k|} \sqrt{\frac{2^{2K(d-1)} n^{(d-1)/2}}{|k|}} e^{-\frac{c2^{2K}\sqrt{n}}{|k|^{2\alpha}}}.$$

Split the sum over

$$(7.3) \quad |k| \in \left[\frac{K n^{(d-1)/2}}{2^{q+1}}, \frac{K n^{(d-1)/2}}{2^q} \right)$$

for $q \in \mathbb{N}$. Then, for a fixed q we have

$$|k|^{2\alpha} = \mathcal{O} \left(\frac{K^{\frac{1}{d-1}} \sqrt{n}}{2^{\frac{q}{d-1}}} \right).$$

So, each term in the sum (7.2) is of order

$$\frac{2^{K(d-1)+(3q/2)}}{K^{3/2} n^{(d-1)/2}} \exp \left(-\frac{c2^{2K+\frac{q}{d-1}}}{K^{\frac{1}{d-1}}} \right).$$

The number of terms in (7.3) is $\mathcal{O} \left(\frac{K n^{(d-1)/2}}{2^q} \right)$. Hence, the sum over k in (7.3) is

$$\mathcal{O} \left(\frac{2^{K(d-1)+q/2}}{\sqrt{K}} \exp \left(-\frac{c2^{2K+\frac{q}{d-1}}}{K^{\frac{1}{d-1}}} \right) \right) = \\ \mathcal{O} \left(\frac{2^{K(d-1)}}{\sqrt{K}} \exp \left(-\frac{c2^{2K}}{K^{\frac{1}{d-1}}} \right) \right) \times \mathcal{O} \left(2^{q/2} \exp \left[-\left(\frac{c2^{q/2}}{\sqrt{K}} \right)^{2/(d-1)} \right] \right).$$

Since⁵

$$(7.4) \quad \sum_q 2^{q/2} \exp \left[-\left(\frac{c2^{q/2}}{\sqrt{K}} \right)^{2/(d-1)} \right] \leq C \sqrt{K}$$

⁵To see this, one can, for example compare the sum in (7.4) with the integral $\int_0^\infty \exp \left(-c(x/\sqrt{K})^{2/(d-1)} \right) dx = \mathcal{O}(\sqrt{K})$.

we obtain the lemma upon taking $\tilde{c} < c/K^{\frac{1}{d-1}}$. \square

8. RELATION TO HOMOGENEOUS FLOWS.

Given $\mathbf{u} \in \mathbb{R}^{d-1}$, $v \in \mathbb{R}$ consider the following function on the space \mathcal{M} of unimodular lattices in \mathbb{R}^d :

$$(8.1) \quad \mathcal{Z}(L; \mathbf{u}, v) = \sum_{(y, \mathbf{x}) \in L \setminus \{\mathbf{0}\}} \frac{\sin 2\pi(\mathbf{u} \cdot \mathbf{x} + vy)}{y} e^{-4\pi^2 \mathbf{x} D_{\mathbf{a}, \mathbf{p}} \cdot \mathbf{x}} \mathbb{1}_{\{\delta < y < K, 2\pi y^\alpha \|\mathbf{x}\| < 2^{K+1}\}}.$$

Define $\boldsymbol{\gamma} = \left(\frac{b_2}{|b_{d+1}|}, \dots, \frac{b_d}{|b_{d+1}|} \right)$. Introduce the following matrices

$$H_\boldsymbol{\gamma} = \begin{pmatrix} 1 & \boldsymbol{\gamma} \\ \mathbf{0}^T & I_{d-1} \end{pmatrix}, \quad G_t = \begin{pmatrix} e^{-(d-1)t} & \mathbf{0} \\ \mathbf{0}^T & e^t I_{d-1} \end{pmatrix}.$$

Then, we get

$$(8.2) \quad n^{d/2} \tilde{\boldsymbol{\Delta}}_n(\delta, K) = \frac{|b_{d+1}| e^{-z^2/2}}{\sigma \sqrt{2\pi^3}} \mathcal{Z}(\mathcal{L}(n, \mathbf{a}); \mathbf{u}, v),$$

where

$$(8.3) \quad \mathbf{u} = \sqrt{n} \mathbf{q} + z\sigma \boldsymbol{\omega}, \quad v = \frac{n^{d/2}}{|b_{d+1}|} (\sqrt{n} a_1 - z\sigma),$$

$\boldsymbol{\omega}$ and \mathbf{q} are given by (6.5), and $\mathcal{L}(n, \mathbf{a})$ is the unimodular lattice $\mathbb{Z}^d H_\boldsymbol{\gamma} G_{\frac{\ln(n)}{2}}$.

To see this, note that, for an arbitrary vector $(k, m_{2,k}, \dots, m_{d,k}) \in \mathbb{Z}^d$,

$$\begin{aligned} (k, m_{2,k}, \dots, m_{d,k}) H_\boldsymbol{\gamma} G_{\frac{\ln(n)}{2}} &= (k, m_{2,k}, \dots, m_{d,k}) \begin{pmatrix} n^{-(d-1)/2} & \sqrt{n} \boldsymbol{\gamma} \\ \mathbf{0}^T & \sqrt{n} I_{d-1} \end{pmatrix} \\ &= \left(\frac{k}{n^{(d-1)/2}}, \sqrt{n} k \boldsymbol{\gamma} + \sqrt{n} (m_{2,k}, \dots, m_{d,k}) \right) \\ &= (Y_k, (2\pi)^{-1} X_k + \sqrt{n} (m_{2,k} - \ell_{2,k}, \dots, m_{d,k} - \ell_{d,k})) \end{aligned}$$

where Y_k and X_k are as in Proposition 6.1 and $\ell_{j,k}$ are given by (6.1). Note that the second term has norm at least \sqrt{n} unless $m_{j,k} = \ell_{j,k}$ for $j = 2, \dots, d$. It follows that the only term which contributes to the RHS of (8.1) is the term with $m_{j,k} = \ell_{j,k}$ justifying (8.1) and (8.2).

Let $\mathbf{w}_j(n, \mathbf{a}) = (y_j(n, \mathbf{a}), \mathbf{x}_j(n, \mathbf{a}))$, $j = 1, \dots, d$ with $y_j \in \mathbb{R}$ and $\mathbf{x}_j \in \mathbb{R}^{d-1}$ be the shortest spanning set of $\mathcal{L}(n, \mathbf{a})$. Put

$$\theta_j(n, (\mathbf{a}, \mathbf{p})) = \mathbf{u} \cdot \mathbf{x}_j(n, \mathbf{a}) + v y_j(n, \mathbf{a}), \quad j = 1, \dots, d.$$

Proposition 8.1. *If (\mathbf{a}, \mathbf{p}) is distributed according to \mathbf{P} then the distribution of the random vector*

$$((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a}), \boldsymbol{\theta}(n, (\mathbf{a}, \mathbf{p})))$$

converges to $\mathbf{P} \times \mu$ as $n \rightarrow \infty$, where μ is the Haar measure on

$$[SL_d(\mathbb{R})/SL_d(\mathbb{Z})] \times \mathbb{T}^d.$$

If we restrict our attention only to $((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a}))$ then the result is standard (see [20, Theorem 5.8], as well as [11, 16, 21]). We refer the readers to [7, Theorem 3], [24] and the references therein for extensions to $[SL_d(\mathbb{R})/SL_d(\mathbb{Z})] \times (\mathbb{T}^d)^p$ under various conditions. Our proof of Proposition 8.1 follows the approach of the proof of Proposition 5.1 in [9].

Proof. We need to show that for each bounded smooth test function f ,

$$(8.4) \quad \int_{\Omega} f((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a}), \boldsymbol{\theta}(n, (\mathbf{a}, \mathbf{p}))) d\mathbf{P} \rightarrow \int_{\Omega \times \mathcal{M} \times \mathbb{T}^d} f((\mathbf{a}, \mathbf{p}), \mathcal{L}, \boldsymbol{\theta}) d\mathbf{P} d\mathcal{L} d\boldsymbol{\theta}$$

as $n \rightarrow \infty$. Write the Fourier series expansion of f with respect to $\boldsymbol{\theta}$

$$(8.5) \quad f((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a}), \boldsymbol{\theta}) = \sum_{\mathbf{k}=(k_1, \dots, k_d) \in \mathbb{Z}^d} f_{\mathbf{k}}((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a})) e^{2\pi i \mathbf{k} \cdot \boldsymbol{\theta}}.$$

Then, it is enough to prove (8.4) for individual terms in (8.5).

If $\mathbf{k} = \mathbf{0}$ then by [20, Theorem 5.8] we can conclude that

$$\int_{\Omega} f_{\mathbf{0}}((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a})) d\mathbf{P} \rightarrow \int_{\Omega \times \mathcal{M} \times \mathbb{T}^d} f_{\mathbf{0}}((\mathbf{a}, \mathbf{p}), \mathcal{L}) d\mathbf{P} d\mathcal{L} d\boldsymbol{\theta}$$

as $n \rightarrow \infty$.

Next, assume that $\mathbf{k} \neq \mathbf{0}$. Since Ω is $2d$ dimensional, we can use

$$(a_1, \boldsymbol{\nu}) := (a_1, (p_1, p_3, \dots, p_d, b_2, \dots, b_{d+1}))$$

as local coordinates. In these coordinates, \mathcal{L} is independent of a_1 . Hence, y_j 's and \mathbf{x}_j 's are independent of a_1 . Note that there exists a compactly supported density $\rho = \rho(a_1, \boldsymbol{\nu})$ such that

$$(8.6) \quad J_{n, \mathbf{k}} = \int f_{\mathbf{k}} e^{2\pi i \mathbf{k} \cdot \boldsymbol{\theta}} d\mathbf{P} = \int \int e^{2\pi i (v \sum y_j k_j + z \sigma \sum k_j \boldsymbol{\omega} \cdot \mathbf{x}_j + \sqrt{n} \sum k_j \mathbf{q} \cdot \mathbf{x}_j)} (f_{\mathbf{k}} \rho) da_1 d\boldsymbol{\nu}$$

where we recall that v is defined in (8.3). Note that

$$\int_{\mathbb{T}^d \times \Omega \times \mathcal{M}} f_{\mathbf{k}} e^{2\pi i \mathbf{k} \cdot \boldsymbol{\theta}} d\theta_1 \dots d\theta_d d\mathbf{P} d\mathcal{L} = 0$$

because

$$\int_{\mathbb{T}^d} e^{2\pi i \mathbf{k} \cdot \boldsymbol{\theta}} d\theta_1 \dots d\theta_d = 0.$$

Therefore, it is enough to prove that $J_{n, \mathbf{k}}$ converges to 0 as $n \rightarrow \infty$.

To this end, we use integration by parts as follows. Define

$$g(a_1, \boldsymbol{\nu}) = e^{2\pi i \frac{n^{(d+1)/2} \sum y_j k_j}{|b_{d+1}|} a_1} = e^{in^{(d+1)/2} \phi(\boldsymbol{\nu}) a_1}$$

where $\phi(\boldsymbol{\nu}) = \frac{2\pi \sum y_j k_j}{|b_{d+1}|}$, and

$$h(a_1, \boldsymbol{\nu}) = (f_{\mathbf{k}} \rho)(a_1, \boldsymbol{\nu}) e^{-2\pi i \left(\left(\frac{n^{d/2} \sum y_j k_j}{|b_{d+1}|} - 4\pi \sum k_j \boldsymbol{\omega} \cdot \mathbf{x}_j \right) z \sigma(a_1, \boldsymbol{\nu}) - \sqrt{n} \sum k_j \mathbf{q} \cdot \mathbf{x}_j \right)}.$$

Then, the inner integral in (8.6) is $\int g(a_1, \boldsymbol{\nu}) h(a_1, \boldsymbol{\nu}) da_1$.

Let $\varepsilon > 0$. On the set $Q_{\mathbf{k}} = \{\phi(\boldsymbol{\nu}) > \varepsilon\}$, we can write

$$g(a_1, \boldsymbol{\nu}) da_1 = \frac{1}{i\phi(\boldsymbol{\nu})n^{(d+1)/2}} de^{ia_1 n^{(d+1)/2} \phi(\boldsymbol{\nu})}.$$

Integrating by parts on $Q_{\mathbf{k}}$ (note that h has compact support) and using trivial bounds on $Q_{\mathbf{k}}^c$, we can conclude that

$$\begin{aligned} |J_{n,\mathbf{k}}| &\leq C \max_{\boldsymbol{\nu}} \left| \int \frac{e^{ia_1 n^{(d+1)/2} \phi(\boldsymbol{\nu})}}{i\phi(\boldsymbol{\nu})n^{(d+1)/2}} \frac{\partial h}{\partial a_1}(a_1, \boldsymbol{\nu}) da_1 \right| + C\mathbf{P}(\{\phi(\boldsymbol{\nu}) \leq \varepsilon\}) \\ &\leq \frac{C}{\varepsilon n^{(d+1)/2}} \int \max_{\boldsymbol{\nu}} \left| \frac{\partial h}{\partial a_1}(a_1, \boldsymbol{\nu}) \right| da_1 + C\mathbf{P}(\{\phi(\boldsymbol{\nu}) \leq \varepsilon\}) \end{aligned}$$

for small enough ε . But

$$\frac{\partial h}{\partial a_1}(a_1, \boldsymbol{\nu}) = \mathcal{O}(n^{d/2}),$$

and hence, the first term is $\mathcal{O}_{\varepsilon}(1/\sqrt{n})$. Therefore, first taking $n \rightarrow \infty$ and then taking $\varepsilon \rightarrow 0$ we have the required result. \square

Recall the definitions of $(y, \mathbf{x})(\mathbf{m})$ and $\theta(\mathbf{m})$ given by (1.11) and (1.12), respectively. With this notation, Proposition 8.1 implies that as $n \rightarrow \infty$ the distribution of $n^{d/2} \tilde{\Delta}_n(\delta, K)$ converges to the distribution of

$$(8.7) \quad \hat{\mathcal{X}}^{(K,\delta)}(\mathcal{L}, \chi) := \frac{|\mathbf{a}_{d+1} - \mathbf{a}_1| e^{-z^2/2}}{2\sigma(\mathbf{a}, \mathbf{p}) \sqrt{2\pi^3}} \sum_{\mathbf{m} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \frac{\sin 2\pi\theta(\mathbf{m})}{y(\mathbf{m})} e^{-4\pi^2 \mathbf{x} D_{\mathbf{a}, \mathbf{p}} \cdot \mathbf{x}} \mathbb{1}_{\mathcal{U}_{K,\delta}}$$

where⁶

$$(8.8) \quad \mathcal{U}_{K,\delta} = \{\delta < |y(\mathbf{m})| < K, 2\pi |y(\mathbf{m})|^\alpha \|\mathbf{x}(\mathbf{m})\| < 2^{K+1}\}$$

and $(\mathcal{L}, \chi) \in \mathbb{M}$ is distributed according to μ . Therefore, Theorem 2* follows from the result below.

Lemma 8.2. $\hat{\mathcal{X}}^{(K,\delta)}$ converges in law as $K \rightarrow \infty$ and $\delta \rightarrow 0$ to the random variable $\hat{\mathcal{X}}$ given by (3.1).

Lemma 8.2, proven in Section 10, completes the proof of Theorem 2*.

⁶Note that (8.7) contains an additional factor of 2 in the denominator comparing with (8.2). This is because in (8.7) the sum is over all lattice vectors (see (8.8)) while in (8.2) we only consider the vectors with positive y coordinate.

9. FINITE INTERVALS.

The proofs of Theorems 3 and 4 are similar to the proofs of Theorems 1 and 2 so we just explain the necessary changes leaving the details to the readers.

Proof of Theorem 3. The random vector (1.14) can be approximated by $(\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)})$ where $\mathcal{Z}^{(i)}$ are defined as in (8.1) but with \mathbf{u} and v replaced by

$$\mathbf{u}^{(i)} = \sqrt{n}\mathbf{q} + z_i\sigma\boldsymbol{\omega} \quad \text{and} \quad v^{(i)} = \frac{n^{d/2}}{|b_{d+1}|}(\sqrt{n}a_1 - z_i\sigma)$$

respectively. Define $\boldsymbol{\theta}^{(i)}$ as in Proposition 8.1 but \mathbf{u} and v replaced by $\mathbf{u}^{(i)}$ and $v^{(i)}$.

To complete the proof, we prove an analogue of Proposition 8.1. Namely, we prove that the distribution of

$$((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a}), \boldsymbol{\theta}^{(1)}(n, (\mathbf{a}, \mathbf{p})), \boldsymbol{\theta}^{(2)}(n, (\mathbf{a}, \mathbf{p})))$$

converges to $\mathbf{P} \times \mu'$ as $n \rightarrow \infty$ where μ' is the Haar measure on $[SL_d(\mathbb{R})/SL_d(\mathbb{Z})] \times \mathbb{T}^d \times \mathbb{T}^d$.

As in the proof of Proposition 8.1, we prove that for individual terms in the Fourier series of a smooth function f on $[SL_d(\mathbb{R})/SL_d(\mathbb{Z})] \times \mathbb{T}^d \times \mathbb{T}^d$

$$\sum_{(\mathbf{k}_1, \mathbf{k}_2) \in \mathbb{Z}^d \times \mathbb{Z}^d} f_{\mathbf{k}_1, \mathbf{k}_2}((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a})) e^{2\pi i[\mathbf{k}_1 \cdot \boldsymbol{\theta}^{(1)} + \mathbf{k}_2 \cdot (\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})]}$$

we have

$$J_{n, \mathbf{k}_1, \mathbf{k}_2} := \int_{\Omega} f_{\mathbf{k}_1, \mathbf{k}_2}((\mathbf{a}, \mathbf{p}), \mathcal{L}(n, \mathbf{a})) e^{2\pi i[\mathbf{k}_1 \cdot \boldsymbol{\theta}^{(1)} + \mathbf{k}_2 \cdot (\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})]} d\mathbf{P}$$

$$\xrightarrow{n \rightarrow \infty} \int_{\Omega \times \mathcal{M} \times \mathbb{T}^d \times \mathbb{T}^d} f_{\mathbf{k}_1, \mathbf{k}_2}((\mathbf{a}, \mathbf{p}), \mathcal{L}) e^{2\pi i[\mathbf{k}_1 \cdot \boldsymbol{\theta}_1 + \mathbf{k}_2 \cdot (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)]} d\mathbf{P} d\mathcal{L} d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2.$$

The case $\mathbf{k}_1 = \mathbf{k}_2 = 0$ follows from [20, Theorem 5.8]. Note that

$$\mathbf{k}_2 \cdot (\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}) = (z_2(n) - z_1(n)) \left(\frac{n^{d/2}}{|b_{d+1}|} \sum y_j k_{2,j} - \sum k_{2,j} \boldsymbol{\omega} \cdot \mathbf{x}_j \right) \sigma.$$

If $\mathbf{k}_1 = 0$ choose appropriate local coordinates in which σ is a coordinate. Integrating by parts with respect to $\sigma = \sigma(\mathbf{a}, \mathbf{p})$ and using $|z_1(n) - z_2(n)|n^{d/2} \rightarrow \infty$, we see that $J_{n, \mathbf{0}, \mathbf{k}_2} \rightarrow 0$ as $n \rightarrow \infty$.

If $\mathbf{k}_1 \neq 0$, then using the same local coordinates $(a_1, \boldsymbol{\nu})$ as in the proof of Proposition 8.1, we can integrate by parts to conclude that $J_{n, \mathbf{k}_1, \mathbf{k}_2} \rightarrow 0$ as $n \rightarrow \infty$. The proof follows through because the leading term of $\mathbf{k}_1 \cdot \boldsymbol{\theta}^{(1)} + \mathbf{k}_2 \cdot (\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})$ is still $n^{(d+1)/2} \phi(\boldsymbol{\nu}) a_1$. \square

Proof of Theorem 4. To prove part (a) pick $\bar{\varepsilon} < \varepsilon$. Applying Theorem 1 we obtain that for almost every (\mathbf{a}, \mathbf{p})

$$\mathbb{P}_{(\mathbf{a}, \mathbf{p})} \left(z_1 \leq \frac{S_n}{\sigma\sqrt{n}} \leq z_2 \right) = \mathcal{E}_{d-1}(z_2) - \mathcal{E}_{d-1}(z_1) + \mathcal{O}(n^{-(d-\bar{\varepsilon})/2})$$

$$= \mathbf{n}(z_1)l_n + \mathcal{O}(l_n^2) + \mathcal{O}(l_n/\sqrt{n}) + \mathcal{O}(n^{-(d-\varepsilon)/2}).$$

According to the assumptions of part (a), the first term is much larger than the remaining terms proving the result.

The proof of part (b) is similar except that we apply Theorem 3 instead of Theorem 1. So, we only get convergence in law.

To prove part (c) we first prove the following analogue of Theorem 3 in the case where $z_2 = z_1 + \frac{c|a_{d+1} - a_1|}{n^{d/2}\sigma}$:

$$\frac{n^{d/2}}{\Lambda(\mathbf{a}, \mathbf{p})} \left(e^{z_1^2/2} \left[\mathcal{E}_d(z_1) - \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z_1 \right) \right], e^{z_2^2/2} \left[\mathcal{E}_d(z_2) - \mathbb{P}_{\mathbf{a}, \mathbf{p}} \left(\frac{S_n}{\sigma\sqrt{n}} \leq z_2 \right) \right] \right)$$

converges in law to a random vector $(\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2)(\mathcal{L}, \theta, c)$ where

$$(\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2)(\mathcal{L}, \theta, c) = \sum_{\mathbf{m} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \frac{e^{-\|\mathbf{x}(\mathbf{m})\|^2}}{y(\mathbf{m})} \left(\sin \theta(\mathbf{m}), \sin(\theta(\mathbf{m}) - cy(\mathbf{m})) \right).$$

Once this convergence is established, the proof of part (c) is the same as the proof of part (b). The proof of convergence is similar to the proof of Theorem 3 except that $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are now not independent. Namely, using the same notation as in the proof of Theorem 3 we have that

$$(9.1) \quad \mathbf{u}^{(2)} = \mathbf{u}^{(1)} + O(n^{-d/2}) \quad \text{and} \quad v^{(2)} = v^{(1)} - c.$$

By Proposition 8.1 $(\mathcal{L}(n, \mathbf{a}), \boldsymbol{\theta}^{(1)}(n, \mathbf{a}))$, converges as $n \rightarrow \infty$ to $(\mathcal{L}^*, \boldsymbol{\theta}^*)$ where $(\mathcal{L}^*, \boldsymbol{\theta}^*)$ is distributed according to the Haar measure on $SL_d(\mathbb{R})/SL_d(\mathbb{Z}) \times \mathbb{T}^d$. Combining this fact with (9.1) we obtain that $(\mathcal{L}(n, \mathbf{a}), \boldsymbol{\theta}^{(1)}(n, \mathbf{a}), \boldsymbol{\theta}^{(2)}(n, \mathbf{a}))$ converges as $n \rightarrow \infty$ to $(\mathcal{L}^*, \boldsymbol{\theta}^*, \hat{\boldsymbol{\theta}}^*)$ where $(\mathcal{L}^*, \boldsymbol{\theta}^*)$ is distributed according to the Haar measure on $SL_d(\mathbb{R})/SL_d(\mathbb{Z}) \times \mathbb{T}^d$ and $\hat{\boldsymbol{\theta}}_j^* = \boldsymbol{\theta}_j^* - cy_j$. This justifies the formula for $(\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2)$. \square

10. CONVERGENCE OF \mathcal{X} .

We need some background information. Given a piecewise smooth compactly supported function $g : \mathbb{R}^d \rightarrow \mathbb{R}$, its Siegel transform is a function on the space of lattices defined by

$$\mathcal{S}(g)(\mathcal{L}) = \sum_{\mathbf{w} \in \mathcal{L} \setminus \{\mathbf{0}\}} g(\mathbf{w}).$$

An identity of Siegel, see ([19, Section 3.7] or [22, Lecture XV]) says that

$$(10.1) \quad \mathbf{E}_{\mathcal{L}}(\mathcal{S}(g)) = \int_{\mathbb{R}^d} g(\mathbf{w}) d\mathbf{w}.$$

In particular, if B is a (bounded) set in \mathbb{R}^d with piecewise smooth boundary not containing $\mathbf{0}$ then

$$(10.2) \quad \mathbf{P}_{\mathcal{L}}(\mathcal{L} \cap B \neq \emptyset) \leq \mathbf{P}_{\mathcal{L}}(\mathcal{S}(\mathbb{1}_B)(\mathcal{L}) \geq 1) \leq \mathbf{E}_{\mathcal{L}}(\mathcal{S}(\mathbb{1}_B)) = \text{Vol}(B).$$

We shall use the following consequence of this result.

Lemma 10.1. *Let $\beta > d$. Then for almost every lattice \mathcal{L} , there exist $C = C(\mathcal{L})$ such that for all $\mathbf{w} \in \mathcal{L} \setminus \{\mathbf{0}\}$ it holds $|y(\mathbf{w})| > C\|\mathbf{w}\|^{-\beta}$.*

Proof. For $k \geq 1$ let $B_k \in \mathbb{R}^d$ denote the following set

$$B_k = \{(\mathbf{x}, y) \in \mathbb{R}^{d-1} \times \mathbb{R} : \|\mathbf{x}\| \in [k, k+1), |y| < k^{-\beta}\},$$

and denote $D_k = \{\mathcal{L} : \mathcal{L} \cap B_k \neq \emptyset\}$. By (10.2), $\mathbf{P}_{\mathcal{L}}(D_k) \leq Kk^{d-1-\beta}$. Since $\beta > d$,

$$\sum_k \mathbf{P}_{\mathcal{L}}(D_k) < \infty.$$

So, by the Borel-Cantelli Lemma, for almost every \mathcal{L} , there exists $k_0(\mathcal{L})$ such that $\mathcal{L} \cap B_k = \emptyset$ for $k \geq k_0$. Since \mathcal{L} contains finitely many vectors satisfying $\|\mathbf{w}\| < k_0$, we can choose $C(\mathcal{L}) \leq 1$ so small that $|y(\mathbf{w})| \geq C\|\mathbf{w}\|^{-\beta}$ for all non-zero \mathbf{w} in the ball of radius k_0 . The result follows. \square

Proof of Lemma 1.2. Let $\mathcal{L}^+ = \{\mathbf{w} \in \mathcal{L} : y(\mathbf{w}) > 0\}$. Since

$$\frac{\sin(2\pi\chi(\mathbf{w}))}{y(\mathbf{w})}$$

is even, and almost every lattice contains no vectors \mathbf{w} with $y(\mathbf{w}) = 0$ and $\mathbf{w} \neq 0$ (this follows immediately from Lemma 10.1), it is enough to restrict the attention to $\mathbf{w} \in \mathcal{L}^+$. Throughout the proof we fix two numbers $\varepsilon > 0$ and $\tau < 1$ such that $\varepsilon \ll (1 - \tau) \ll 1$.

By applying Lemma 10.1 with $\beta = d + 1$, for almost all \mathcal{L} we have

$$\left| \sum_{\mathbf{w} \in \mathcal{L}^+ : \|x(\mathbf{w})\| \geq \|\mathbf{w}\|^\varepsilon} \frac{\sin 2\pi\chi(\mathbf{w})}{y(\mathbf{w})} e^{-\|x(\mathbf{w})\|^2} \right| \leq \sum_{\mathbf{w} \in \mathcal{L}^+} C\|\mathbf{w}\|^{d+1} e^{-\|\mathbf{w}\|^{2\varepsilon}}$$

converges absolutely. Hence, it suffices to establish the convergence of

$$\bar{\mathcal{X}}_R := \sum_{\mathbf{w} \in \mathcal{L}^+ : \|x(\mathbf{w})\| \leq \|\mathbf{w}\|^\varepsilon < R^\varepsilon} \frac{\sin 2\pi\chi(\mathbf{w})}{y(\mathbf{w})} e^{-\|x(\mathbf{w})\|^2}.$$

Let $R_{j,k} = 2^k + j2^{\tau k}$, $j = 0, \dots, \lfloor 2^{(1-\tau)k} \rfloor$. To prove the convergence of \mathcal{X} , we will show that almost all \mathcal{L} and almost all χ satisfy the two estimates below:

$$(10.3) \quad \forall \text{ sequences } \{j_k\}, \bar{\mathcal{X}}_{R_{j_k, k}} \text{ converges as } k \rightarrow \infty,$$

$$(10.4) \quad \max_j \sup_{R_{j,k} \leq R < R_{j+1,k}} |\bar{\mathcal{X}}_R - \bar{\mathcal{X}}_{R_{j,k}}| \rightarrow 0 \text{ as } k \rightarrow \infty.$$

To prove (10.3), let

$$S_{j,k} = \sum_{\mathbf{w} \in \mathcal{L}^+ : \|x(\mathbf{w})\| \leq \|\mathbf{w}\|^\varepsilon, R_{j,k} \leq \|\mathbf{w}\| < R_{j+1,k}} \frac{\sin 2\pi\chi(\mathbf{w})}{y(\mathbf{w})} e^{-\|x(\mathbf{w})\|^2}.$$

Using that $\mathbf{E}_\chi(\sin(2\pi(\chi(\mathbf{w})))) = 0$ and for $\mathbf{w}_1 \neq \pm\mathbf{w}_2$,

$$\mathbf{E}_\chi(\sin(2\pi(\chi(\mathbf{w}_1))) \sin(2\pi(\chi(\mathbf{w}_2)))) = 0,$$

we see that $\mathbf{E}_\chi(S_{j,k}) = 0$ and

$$\begin{aligned} \text{Var}_\chi(S_{j,k}) &= \sum_{\mathbf{w} \in \mathcal{L}^+ : \|x(\mathbf{w})\| \leq \|\mathbf{w}\|^\varepsilon, R_{j,k} \leq \|\mathbf{w}\| < R_{j+1,k}} \frac{e^{-2\|x(\mathbf{w})\|^2}}{2y^2(\mathbf{w})} \\ &\leq \frac{1}{2^{2k+1}} \text{Card}(\mathbf{w} : \|x(\mathbf{w})\| \leq \|\mathbf{w}\|^\varepsilon, R_{j,k} \leq \|\mathbf{w}\| < R_{j+1,k}) \\ &\leq \frac{C(\mathcal{L})}{2^{2k}} \text{Vol}(\mathbf{w} : \|x(\mathbf{w})\| \leq \|\mathbf{w}\|^\varepsilon, R_{j,k} \leq \|\mathbf{w}\| < R_{j+1,k}) \\ &\leq \bar{C}(\mathcal{L}) 2^{(\tau+\varepsilon(d-1)-2)k}. \end{aligned}$$

Hence, by Chebyshev's inequality for each j

$$\mathbf{P}_\chi(|S_{j,k}| \geq 2^{-(1-\tau+\varepsilon)k}) \leq \bar{C}(\mathcal{L}) 2^{(\varepsilon(d+1)-\tau)k}.$$

Therefore

$$\mathbf{P}_\chi(\exists j : |S_{j,k}| \geq 2^{-(1-\tau+\varepsilon)k}) \leq \bar{C}(\mathcal{L}) 2^{(1+\varepsilon(d+1)-2\tau)k}.$$

Thus, if ε is sufficiently small and τ is sufficiently close to 1 then Borel-Cantelli Lemma shows that for almost every χ , if k is large enough, then for all j , $|S_{j,k}| \leq 2^{-(1-\tau+\varepsilon)k}$, and thus, $\sum_j |S_{j,k}| \leq 2^{-\varepsilon k}$ proving (10.3). Likewise,

$$\begin{aligned} \sup_{R_{j,k} \leq R \leq R_{j+1,k}} |\bar{\mathcal{X}}_R - \bar{\mathcal{X}}_{R_{j,k}}| &\leq \sum_{\mathbf{w} \in \mathcal{L}^+ : \|x(\mathbf{w})\| \leq \|\mathbf{w}\|^\varepsilon, \|\mathbf{w}\| \in [R_{j,k}, R_{j+1,k})} \frac{1}{|y(\mathbf{w})|} e^{-\|x(\mathbf{w})\|^2} \\ &\leq C(\mathcal{L}) 2^{-k} \text{Vol}(\mathbf{w} : \|x(\mathbf{w})\| \leq \|\mathbf{w}\|^\varepsilon, R_{j,k} \leq \|\mathbf{w}\| < R_{j+1,k}) \leq \bar{C}(\mathcal{L}) 2^{(\tau+\varepsilon(d-1)-1)k} \end{aligned}$$

proving (10.4). Lemma 1.2 is established. \square

Proof of Lemma 8.2. Given a domain $\mathcal{U} \in \mathbb{R}^d$ let

$$\bar{\mathcal{X}}_{\mathcal{U}}(\mathcal{L}, \chi) = \sum_{\mathbf{m} \in \mathbb{Z}^d \setminus \{0\}} \frac{\sin 2\pi\theta(\mathbf{m})}{y(\mathbf{m})} e^{-4\pi^2 \mathbf{x} D_{\mathbf{a}, \mathbf{p}} \mathbf{x}} \mathbb{1}_{\mathcal{U}}.$$

Then

$$\hat{\mathcal{X}}^{(K, \delta)} = \frac{|\mathbf{a}_{d+1} - \mathbf{a}_1| e^{-z^2/2}}{2\sigma(\mathbf{a}, \mathbf{p}) \sqrt{2\pi^3}} \bar{\mathcal{X}}_{\mathcal{U}_{K, \delta}}$$

where $\mathcal{U}_{K, \delta}$ is given by (8.8). Let $\Gamma_R = AB_R$ where B_R is the ball of radius R centered at the origin and A is the linear map given by (3.3). Lemma 1.2 (after the change of variables $\mathcal{L} \mapsto A\mathcal{L}$) tells us that $\bar{\mathcal{X}}_{\Gamma_R}(\mathcal{L}, \chi) \rightarrow \bar{\mathcal{X}}(\mathcal{L}, \chi)$ as $R \rightarrow \infty$ almost surely where $\bar{\mathcal{X}} = \bar{\mathcal{X}}_{\mathbb{R}^d}$. Therefore it suffices to show that for each η there exist δ_0 and K_0 such if $\delta < \delta_0$ and $K \geq K_0$ then

$$\mathbf{P}\left(\left|\bar{\mathcal{X}}_{\Gamma_{2K/\delta^2}} - \bar{\mathcal{X}}_{\mathcal{U}_{K, \delta}}\right| > \eta\right) < \eta.$$

Note that $\mathcal{U}_{K,\delta} \subset \Gamma_{2K/\delta^2}$ (for sufficiently small δ) and so

$$(10.5) \quad \bar{\chi}_{\Gamma_{2K/\delta^2}} - \bar{\chi}_{\mathcal{U}_{K,\delta}} = \sum_{\mathbf{m} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \frac{\sin 2\pi\theta(\mathbf{m})}{y(\mathbf{m})} e^{-4\pi^2 \mathbf{x} D_{a,p} \mathbf{x}} \mathbb{1}_{\Gamma_{2K/\delta^2} \setminus \mathcal{U}_{K,\delta}}.$$

Below, we choose K so large and δ so small that

$$(10.6) \quad \mathbf{P}_{\mathcal{L}}(\mathfrak{A}_{K,\delta}^c) \leq \eta/100$$

where $\mathfrak{A}_{K,\delta}$ is the set of lattices \mathcal{L} satisfying the following conditions:

- (i) the shortest non-zero vector in \mathcal{L} is longer than $3\delta^\varepsilon$;
- (ii) $\mathbf{w} = (\mathbf{x}, y) \in \mathcal{L}$ then $|y| \geq \max(K, 1/\delta) \|\mathbf{w}\|^{-(d+1)}$;
- (iii) \mathcal{L} contains no vectors (\mathbf{x}, y) with $|y| \leq \delta$ and $2\pi \|\mathbf{x}\| \leq \delta^{-1/2d}$.

It is easy to see using (10.2) that the measure of lattices *not* satisfying at least one of the above conditions is small (cf. the proof of Lemma 10.1). We now estimate the contribution to (10.5) coming from six different regions in $\Gamma_{2K/\delta^2} \setminus \mathcal{U}_{K,\delta}$.

(1) Consider first the terms with $\|\mathbf{x}\| \geq \|\mathbf{w}\|^\varepsilon$. Then for $\mathcal{L} \in \mathfrak{A}_{K,\delta}$ each term in the sum is bounded by $C \|\mathbf{w}\|^{d+1} e^{-c\|\mathbf{w}\|^{2\varepsilon}}$. We now consider several cases depending on the restrictions on y .

(a) If $|y| \geq K$ then $\|\mathbf{w}\| \geq K$ and so the sum over this region is bounded (in absolute value) by

$$W_{1a} := \sum_{\mathbf{w} \in \mathcal{L}: \|\mathbf{w}\| \geq K} C \|\mathbf{w}\|^{d+1} e^{-c\|\mathbf{w}\|^{2\varepsilon}}.$$

By the Siegel identity, (10.1), $\mathbf{E}_{\mathcal{L}}(W_{1a} \mathbf{1}_{\mathfrak{A}_{K,\delta}}) \leq \bar{C} e^{-\bar{c}K^{2\varepsilon}}$ and so the contribution coming from domain (1a) is negligible in view of the Markov inequality.

(b) If $\delta < |y| < K$, then $2\pi |y|^\alpha \|\mathbf{x}\| \geq 2^{K+1}$ whence $\|\mathbf{x}\| \geq 2^K / (\pi K^\alpha)$. Denoting by W_{1b} the contribution from this region we obtain using property (ii) of the definition of $\mathfrak{A}_{K,\delta}$ that

$$W_{1b} \leq \sum_{\mathbf{w} \in \mathcal{L}: \|\mathbf{w}\| \geq 2^K / (\pi K^\alpha)} C \|\mathbf{w}\|^{d+1} e^{-c\|\mathbf{w}\|^{2\varepsilon}}.$$

Hence, by (10.1), $\mathbf{E}_{\mathcal{L}}(W_{1b} \mathbf{1}_{\mathfrak{A}_{K,\delta}}) \leq \bar{C} e^{-\bar{c}2^{2\varepsilon K} / K^{\varepsilon\alpha}}$ which shows that the contribution from the region (1b) is negligible in view of the Markov inequality.

(c) If $|y| \leq \delta$, then $\|\mathbf{w}\| \geq \|\mathbf{x}\| \geq \delta^{-1/2d} / (2\pi)$ because $\mathcal{L} \in \mathfrak{A}_{K,\delta}$. Hence, if W_{1c} denotes the contribution of the terms from (1c), then, similarly to the case (1b), $\mathbf{E}_{\mathcal{L}}(W_{1c} \mathbf{1}_{\mathfrak{A}_{K,\delta}}) \leq \bar{C} e^{-\bar{c}\delta^{-\varepsilon/d}}$ and the contribution of region (1c) is negligible as well.

(2) Now we discuss the terms with $\|\mathbf{x}\| \leq \|\mathbf{w}\|^\varepsilon$.

Again, we shall consider three cases:

(a) $|y| \geq K$. Note that in case (2) we have $\|\mathbf{x}\|^{1/\varepsilon} \leq \|\mathbf{w}\| \leq \|\mathbf{x}\| + |y|$. Therefore $|y| \geq \|\mathbf{x}\|^{1/\varepsilon} - \|\mathbf{x}\| \geq (\|\mathbf{x}\|/2)^{1/\varepsilon}$ for sufficiently large $\|\mathbf{x}\|$. So, $\|\mathbf{x}\| \leq 2|y|^\varepsilon$. Thus taking the L^2 -norm and integrating first with respect to χ we get the L^2 -norm of the terms in (2a) is bounded by

$$\mathbf{E}_{\mathcal{L}} \left(\sum_{(\mathbf{x}, y) \in \mathcal{L} \cap (\Gamma_{2K/\delta^2} \setminus \mathcal{U}_{K, \delta}) \cap \{|y| \geq K\}} \frac{C}{y^2} \right) \leq C \int_{\|\mathbf{x}\| < 2|y|^\varepsilon, |y| \geq K} \frac{1}{y^2} d\mathbf{w} \leq \frac{C}{K^{1-\varepsilon d}}$$

where the first inequality relies on (10.1).

(b) $\delta < |y| < K$. In this case, $\|\mathbf{x}\|^{1/\varepsilon} < \|\mathbf{x}\| + K$. Thus, $K > \|\mathbf{x}\|^{1/\varepsilon} - \|\mathbf{x}\| \geq (\|\mathbf{x}\|/2)^{1/\varepsilon}$ for sufficiently large $\|\mathbf{x}\|$. So, $\|\mathbf{x}\| \leq 2K^\varepsilon \leq K$, and hence, $(\mathbf{x}, y) \in \mathcal{U}_{K, \delta}$. So, the region (2b) does not contribute to our sum.

(c) $|y| \leq \delta$. In this case, $\|\mathbf{w}\| \leq \|\mathbf{x}\| + |y| \leq 3\delta^\varepsilon$. However, (ii) implies that $\|\mathbf{w}\| \geq \delta^{-2/(d+1)}$. Hence, this case is impossible.

Combining the six cases considered above, we obtain the result. \square

REFERENCES

- [1] Angst J., Poly G. *A weak Cramér condition and application to Edgeworth expansions*, Electron. J. Probab. **22** (2017) paper 59, 24 pp.
- [2] Barbour A. D. *Asymptotic expansions based on smooth functions in the central limit theorem*, Probab. Theory Relat. Fields **72** (1986) 289–303.
- [3] Bobkov, S. G. *Central limit theorem and Diophantine approximations*, J. Theoret. Probab. **31** (2018) 2390–2411.
- [4] Bobkov S. G. *Khinchine’s theorem and Edgeworth approximations for weighted sums*, Ann. Statist. **47** (2019) 1616–1633.
- [5] Breuillard E. *Distributions diophantiennes et theoreme limite local sur \mathbb{R}^d* , Prob. Th. Related Fields **132** (2005) 39–73.
- [6] Coelho Z., Parry W. *Central limit asymptotics for shifts of finite type*, Israel J. Math. **69** (1990) 235–249.
- [7] Dettmann C. P., Marklof J., Strombergsson A. *Universal hitting time statistics for integrable flows*, J. Stat. Phys. **166** (2017) 714–749.
- [8] Dolgopyat D. *A Local Limit Theorem for sum of independent random vectors*, Electronic J. Prob. **21** (2016) paper 39.
- [9] Dolgopyat D., Fayad B. *Deviations of ergodic sums for toral translations: Convex bodies*, GAFA **24** (2014) 85–115.
- [10] Dolgopyat D., Fayad B. *Limit theorems for toral translations*, Proc. Sympos. Pure Math **89** (2015) 227–277.
- [11] Eskin A., McMullen C. *Mixing, counting, and equidistribution in Lie groups*, Duke Math. J. **71** (1993) 181–209.
- [12] Esseen C.-G. *Fourier analysis of distribution functions: A mathematical study of the Laplace-Gaussian law*, Acta Math. **77** (1945) 1–125.
- [13] Feller W. *An introduction to probability theory and its applications*, Vol. II. 2nd ed. John Wiley & Sons, New York-London-Sydney (1971) xxiv+669 pp.
- [14] Fernando K., Liverani C. *Edgeworth expansions for weakly dependent random variables*, Ann. Inst. Henri Poincaré Probab. Stat. **57** (2021) 469–505.

- [15] Kesten H. *Uniform distribution mod 1*, part I: Ann. of Math. **71** (1960) 445–471, part II: Acta Arith. **7** (1961/1962) 355–380.
- [16] Kleinbock D. Y., Margulis G. A. *Bounded orbits of nonquasiunipotent flows on homogeneous spaces*, AMS Transl. **171** (1996) 141–172.
- [17] Kleinbock D. Y., Margulis G. A. *Logarithm laws for flows on homogeneous spaces*, Invent. Math. **138** (1999) 451–494.
- [18] Lawler, G., Limic, V. *Random Walk: A Modern Introduction*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (2010)
- [19] Marklof J. *The n -point correlations between values of a linear form*, Erg. Th. Dynam. Sys. **20** (2000) 1127–1172.
- [20] Marklof J., Strombergsson A. *The distribution of free path lengths in the periodic Lorentz gas and related lattice point problems*, Ann. of Math. **172** (2010) 1949–2033.
- [21] Shah N. A. *Limit distributions of expanding translates of certain orbits on homogeneous spaces*, Proc. Indian Acad. Sci. Math. Sci. **106** (1996) 105–125.
- [22] Siegel C. L. *Lectures on the geometry of numbers*, Springer, Berlin (1989) x+160 pp.
- [23] Sprindzuk V. G. *Metric theory of Diophantine approximations*, Scripta Series in Math. V. H. Winston & Sons, Washington D.C. (1979) xiii+156 pp.
- [24] Strombergsson A. *Effective Ratner equidistribution for $SL(2, \mathbb{R}) \ltimes \mathbb{R}^2$* , Duke Math. J. **164** (2015) 843–902.