amagyar2024_220_24

Diophantine equations and discrete harmonic analysis
Diofantikus egyenletek és diszkrét harmonikus analízis

Ákos Magyar

March 2024

A doctoral dissertation
submitted to the Hungarian Academy of Sciences

Contents

Introduction

This dissertation is based on five papers which I consider representative of my research over the past twenty years. Each of these papers is on the interface of analytic number theory and other fields of mathematics, discrete harmonic analysis, ergodic theory and geometric Ramsey theory.

The topic of the first section is the distribution of solutions of general high rank diophantine equations, expressed in terms of pointwise ergodic theorems, the crucial ingredient being certain L^p estimates for maximal averages over integer points on hypersurfaces. Such estimates are considered discrete analogues of certain central results in Euclidean harmonic analysis and has been initiated by the groundbreaking results of Bourgain, as well as systematic study of Stein and Wainger. I have written 8 papers with my co-authors in this area where I think my work has made the most impact. It has brought young researchers to this area, has initiated many further studies on related problems and our papers received a substantial number of references.

The topic of the second and third chapter is geometric Ramsey theory where one studies isometric or similar copies of finite point configurations, the underlying space being the integer lattice or in some instances the Euclidean space the model case of vectorspaces over finite fields. This section is based on 2 papers, which have been mainly motivated by Bourgain's simplex theorem and a conjecture of Graham in Euclidean Ramsey theory. It outlines a new approach, based on some notions and ideas in additive combinatorics and graph theory. My work in this area, consisting of 7 papers, has also raised considerable interest has initiated further research. Although less transparent, methods of analytic number theory play a crucial role here too, in particular the Hardy-Littlewood method of exponential sums and the theory of Siegel theta functions.

The fourth chapter is about finding solutions to diophantine systems in the primes. Here we present one paper which may be viewed as the extension of Hua's work on representing positive integers as sums of a fixed power of primes. This was the first paper which have shown that any sufficiently high rank diophantine system has many solutions in the primes as long as it satisfies certain natural local conditions. This paper has influenced further studies on prime solutions to diophantine equations. In another, somewhat related, in joint work I have extended the celebrated theorem of Green and Tao on the existence of long arithmetic progressions in the primes to the multi-dimensional setting, however the methods of that paper are mainly combinatorial and will not be discussed here.

The topic of the last chapter is our ongoing work toward the so-called Furstenberg-Bergelson-Leibman conjecture in ergodic theory. This conjecture states that general polynomial orbits of actions of nilpotent groups have have almost everywhere as well as norm convergence. This connects back to the first section as the first fundamental results were due to Bourgain, who proved the conjecture for commuting transformations by reducing it to bounds for associated maximal operators over polynomial surfaces and employing methods of analytic number theory. I discuss our most recent paper in this direction, which establishes the conjecture for step-2 nilpotent groups by extending Bourgain's approach to the non-commutative setting. A crucial arithmetic input is to derive Weyl-type estimates for exponential sums arising from diophantine systems which are natural extensions of the so-called Waring-Vinogradov system to free nilpotent groups. As a byproduct, we derive asymptotics for the number of integer solutions to such systems.

Acknowledgements

I'd like to express my gratitude to my Ph.D. thesis advisor Elias M. Stein who has introduced me to this fascinating area of research on the interface of analytic number theory and discrete harmonic analysis, as well as for his continued support long after I have received degree. I'd also like to thank Endre Szemerédi for his friendship and support, as well as for the many stimulation discussions about mathematics over the years. I'm grateful to Stephen Wainger for his insights and ongoing encouragement during my postdoctoral years at the University of Wisconsin-Madison.

I thank my doctoral students and postdoctoral associates the opportunity to work together, in particular to Brian Cook, Hans Parshall and Neil Lyall whom I continued to collaborate over the years. The same goes to my other coauthors, among them Alexandru Ionescu and Mariusz Mirek for their hospitality during my visits for collaboration.

My home institution the University of Georgia has provided a supportive environment for many years, while my visits to other institutions in particular to the Institute of Advanced Study and the Mathematical Sciences Research Institute has been inspiring and conducive to research.

Finally, let me thank to my wife Edit and to my family for their patience and continued support.

1 Diophantine equations and ergodic theorems

In this chapter we discuss ergodic theorems showing the uniformity of distribution of the solutions to certain diophantine equations, proved in [81]. In the first section we recall Bourgain's polynomial ergodic theorem [15, 13, 14], as well as its relation to the l^p boundedness of certain discrete maximal operators. Next we present our main results and in the remaining sections their proof.

1.1 Polynomial ergodic theorems and discrete harmonic analysis

Let (X, μ) be a finite measure space, without loss of generality we will assume $\mu(X) = 1$, i.e. it is a probability space. If $T : X \to X$ is a measure preserving ergodic transformation, then a fundamental theorem in ergodic theory, due Birkhoff [12], states that the orbits $O_x := \{T^n x : \ n \in \mathbb{N}\}$ are uniformly distributed on X. The precise statement is as follows. Let $p \geq 1$, $f \in L^p(X, \mu)$. Then there exists a function $f_* \in L^p(X, \mu)$ such that

$$A_N f(x) := \frac{1}{N} \sum_{n=1}^{N} f(T^n x) \ \to \ f_*(x), \tag{1.1.1}$$

for μ-a.e. $x \in X$. Moreover if T is ergodic, i.e. $E = T^{-1}(E)$ implies that $\mu(E) = 0$ or $\mu(X \backslash E) = 0$, then f_* is constant and equal to $\int_X f \, d\mu$. An approach to Birkhoff's theorem, due to M. Riesz, is first to reduce proving (1.1.1) to the maximal inequality

$$\|A^* f\|_p \leq C_p \, \|f\|_p, \quad \textit{where} \ \ A^* f(x) = \sup_{N \in \mathbb{N}} |A_N f(x)|, \quad (p > 1) \tag{1.1.2}$$

then, using a general transfer principle, due to Calderon, reduce inequality (1.1.2) to the shift $X = \mathbb{Z}$, $Tx = x - 1$. Then one arrives to the so-called Hardy-Littlewood maximal inequality $\|A^* \phi\|_p \leq C_p \, \|\phi\|_p$, where $\phi \in l^p(\mathbb{Z})$ and $A_N \phi(m) = \frac{1}{N} \sum_{n=1}^{N} \phi(m - n)$.

In a series of fundamental papers [15, 13, 14] Bourgain has extended Birkhoff's theorem to polynomial orbits $O_P(x) = \{T^{P(n)}(x); \ n \in \mathbb{N}\}$, $P : \mathbb{Z} \to \mathbb{Z}$ being an integral polynomial. He proved that for every exponent $p > 1$, and for every $f \in L^p(X, \mu)$ there exists a function $f_* \in L^p(X, \mu)$ such that for μ-a.e. $x \in X$,

$$A_{P,N} f(x) := \frac{1}{N} \sum_{n=1}^{N} f(T^{P(n)} x) \ \to \ f_*(x). \tag{1.1.3}$$

Moreover, if T is fully ergodic, that is if T^k ergodic for all $k \in \mathbb{N}$, then f_* constant, thus the polynomial orbits $O_P(x)$ are uniformly distributed for almost every $x \in X$. Bourgain has also proved a multi-dimensional extension, namely that above theorem holds for the averages

$$A_{P_1,\ldots,P_k,N} f(x) := \frac{1}{N} \sum_{n=1}^{N} f(T_1^{P_1 n)} \ldots T_k^{P_k(n)} x), \tag{1.1.4}$$

whenever $T_1, \ldots, T_k$ are commuting, measure preserving transformations on X and $P_1, \ldots, P_k$ are integral polynomials on $\mathbb{Z}$.

As opposed to Birkhoff's theorem there is not a dense set of functions in $L^p(X, \mu)$ for which (1.1.3) holds thus the corresponding maximal inequality

$$\|A_P^* f\|_p \leq C_p \, \|f\|_p, \quad \textit{where} \ \ A_P^* f(x) = \sup_{N \in \mathbb{N}} |A_{P,N} f(x)|, \tag{1.1.5}$$

does not imply point-wise convergence in itself. Nevertheless, proving it is crucial to Bourgain's approach and it is not hard to show that if the family of functions $\{A_N f(x)\}_{n\in\mathbb{N}}$ do not converge almost everywhere on X, then there exists an increasing sequence $\{N_k\}_{k\in N} \subseteq \mathbb{N}$ and a constant $\alpha > 0$, such that for all $k \in \mathbb{N}$,

$$\|A^*_{P,k} f\|_p \geq \alpha, \quad where \quad A^*_{P,k} f(x) = \sup_{N_k \leq N < N_{k+1}} |A_{P,N} f(x) - A_{P,N_k} f(x)|. \tag{1.1.6}$$

Note that the maximal functions $A^*_{P,k} f$ are measuring the oscillation of the sequence of functions $A_{P,N} f$. It is then enough to show that $\|A^*_{P,k} f\|_p \to 0$ as $k \to \infty$, at least when one is taking Cesaro averages that is to prove appropriate maximal inequalities.

The Calderon transference principle holds equally well for polynomial averages, effectively transferring the maximal inequalities to the shift $X = \mathbb{Z}$, $Tx = x - 1$. The key part of Bourgain's proof is to show that for any $\phi \in l^p(\mathbb{Z})$ and for any polynomial $P : \mathbb{Z} \to \mathbb{Z}$ one has that

$$\| \sup_{N\in\mathbb{N}} A^*_{P,N} \phi\|_p \leq C_p \,\|\phi\|_p, \quad where \quad A_{P,N}\phi(m) = \frac{1}{N}\sum_{n=1}^{N} \phi(m - P(n). \tag{1.1.7}$$

Note that $A_{P,N}\phi = f * K_{P,N}$ is a convolution with a kernel $K_{P,N} = \frac{1}{N}\sum_{n=1}^{N} \delta_{P(n)}$, δ_m being the Dirac delta measure concentrated at m. The Fourier transform

$$\widehat{K}_{P,N}(\alpha) = \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i\, P(n)\alpha}, \tag{1.1.8}$$

is an exponential sum extensively studied in analytic number theory, in particular via the so-called Hardy-Littlewood method of exponential sums. A deep study on the structure of $\widehat{K}_{P,N}(\alpha)$ is crucial to Bourgain's proof.

1.2 The distribution of solutions to diophantine equations

A fundamental problem in number theory is to determine asymptotically the number of integer solutions $m = (m_1, \ldots, m_n)$ of a diophantine equation $Q(m_1, \ldots, m_n) = \lambda$ as $\lambda \to \infty$ through the integers, and $Q(m)$ is a positive polynomial with integer coefficients. A general result of this type follows from a variant of the Hardy-Littlewood method of exponential sums developed by Birch [2] and Davenport [4], which is as follows.

Let $Q(m_1, \ldots, m_n)$ be a positive homogeneous polynomial of degree d with integral coefficients, and suppose that it satisfies the non-degeneracy condition

$$n - dim\ V_Q > (d-1)2^d \tag{1.2.1}$$

Here $V_Q = \{z \in \mathbf{C}^n \ : \ \partial_1 Q(z) = \ldots \partial_n Q(z) = 0\}$ is the complex singular variety of the polynomial Q. For simplicity we'll refer to polynomials satisfying all the above conditions as non-degenerate forms.
Then the following asymptotic formula holds for the number of integer solutions $r_Q(\lambda) = |\{m \in \mathbf{Z}^n \ : \ Q(m) = \lambda\}|$,

$$r_Q(\lambda) = c_Q \lambda^{\frac{n}{d}-1} \sum_{q=1}^{\infty} K(q, 0, \lambda) \ + \ O_\delta(\lambda^{\frac{n}{d}-1-\delta}) \tag{1.2.2}$$

for some $\epsilon > 0$. The expression $K(\lambda) = \sum_{q=1}^{\infty} K(q,0,\lambda)$ is called the singular series, the terms are special cases of $(l = 0)$ the exponential sums

$$K(q,l,\lambda) = q^{-n} \sum_{(a,q)=1} \sum_{s\in \mathbf{Z}^n/q\mathbf{Z}} e^{2\pi i \frac{a(Q(s)-\lambda)+s\cdot l}{q}} \tag{1.2.3}$$

that is a goes through the reduced residue classes $(mod\ q)$ and s_j goes through all residue classes $(mod\ q)$ for each j. We remark that $K(q,0,\lambda)$ is a Kloostermann sum if $Q(m)$ is a quadratic form

The asymptotic formula (1.2.2) can be valid just under a condition of type (1.2.1). Indeed consider the polynomial $Q(m) = (m_1^2 + \ldots + m_n^2)^{d/2}$ ($d > 2$ even). Then $r_Q(\lambda) = 0$ unless $\lambda = \mu^{d/2},\ \mu \in \mathbf{N}$, and in that case $r_Q(\lambda) = \mu^{n/2-1} = \lambda^{n/d-2/d}$. Hence formula (0.2) is never valid. The reason is that the complex singular variety: $V_Q = \{z \in \mathbf{C}^n\ :\ z_1^2 + \ldots + z_n^2 = 0\}$ has dimension $n-1$. It is meaningful only if the singular series is nonzero. It can be shown, that if Q is a non-degenerate form, then there exists an arithmetic progression $\Gamma \subseteq \mathbf{N}$ and a constant $0 < A_Q$ such that

$$A_Q \leq K(\lambda)\ ,\quad \text{for every}\quad \lambda \in \Gamma. \tag{1.2.4}$$

We'll refer to such sets Γ as *sets of regular values* of the polynomial Q. Inequality (0.4) is true for all large λ, just under additional assumptions modulo primes. Indeed consider the polynomial $Q(m) = m_1^d + pQ_1(m_2,\ldots m_n)$. For $\lambda = p\lambda_1 + s$ s being a quadratic non-residue, the equation $Q(m) = \lambda$ has no solution, since d is even. Such conditions will be discussed later.

A crucial observation is that a similar approximation formula to (1.2.2) holds for the Fourier transform of the solution set:

$$\hat{\sigma}_{Q,\lambda}(\xi) := \sum_{m\in\mathbf{Z}^n, Q(m)=\lambda} e^{2\pi i m\cdot\xi}\ ,\quad \xi \in \Pi^n,$$

were $\Pi^n = \Re^n/\mathbf{Z}^n$ is the flat torus.

Lemma 1.1. *Let $Q(m)$ be a non-degenerate form, then there exists $\delta > 0$, such that*

$$\hat{\sigma}_{Q,\lambda}(\xi) = c_Q \lambda^{\frac{n}{d}-1} \sum_{q=1}^{\infty} K(q,l,\lambda) \sum_{l\in\mathbf{Z}^n} \psi(q\xi - l) d\tilde{\sigma}_Q(\lambda^{\frac{1}{d}}(\xi - l/q)))\ + \mathcal{E}_\lambda(\xi), \tag{1.2.5}$$

$$\textit{where}\quad \sup_{\xi} |\mathcal{E}_\lambda(\xi)| \leq c_\delta \lambda^{\frac{n}{d}-1-\delta}.$$

Here $\psi(\xi)$ is a smooth cut-off, $\psi(\xi) = 1$ for $\sup_j |\xi_j| \leq 1/8$ and $\psi(\xi) = 0$ for $\sup_j |\xi_j| \geq 1/4$. Moreover

$$\tilde{d\sigma}_Q(\xi) = \int_{\{x\in\Re^n\,:\,Q(x)=1\}} e^{2\pi i x\cdot\xi}\, d\sigma_Q(x). \tag{1.2.6}$$

where $d\sigma_Q(x) = \frac{dS_Q(x)}{|Q'(x)|}$, $dS_Q(x)$ being the Euclidean surface area measure of the level surface $Q(x) = 1$, and $|Q'(x)|$ the magnitude of the gradient of the form Q.

The approximation formula (1.2.5) means, that the Fourier transform of the indicator function of the solution set $Q(m) = \lambda$ is asymptotically a sum over all rational points, of pieces of the Fourier transform of a surface measure of $Q(x) = \lambda$, multiplied by arithmetic factors and shifted by rationals. This formula in the special case $Q(m) = m_1^2 + \ldots + m_n^2$ was proved earlier in [82]. Our main purpose is to study the distribution of the solution sets $\{m \in \mathbf{Z}^n\ :\ Q(m) = \lambda\}$. We have,

Theorem 1.1. *Let $Q(m)$ be a non-degenerate polynomial and Λ is corresponding set of regular values. Then for a test function $\phi(x) \in S(\mathbb{R}^n)$ one has*

$$\lim_{\lambda\in\Lambda,\lambda\to\infty} \frac{1}{r_Q(\lambda)} \sum_{Q(m)=\lambda} \phi(\lambda^{-1/d} m) = \int_{Q(x)=1} \phi(x)\, d\sigma_Q(x). \tag{1.2.7}$$

That is when the solution sets $Q(m) = \lambda$ are projected to the unit surface $Q(x) = 1$ via the dilations $m \to \lambda^{-1/d} m$, they weakly converge to the surface measure $\frac{dS_Q(x)}{|Q'(x)|}$. This is well-known in case $Q(x)$ is a quadratic form.

Our main result concerns the uniform distribution of the images of the solution sets, when mapped to a measure space via an ergodic family of transformations. Let (X, μ) be a probability measure space, and $T = (T_1, \ldots, T_n)$ be a family of commuting, measure preserving and invertible transformations. Suppose for every positive integer q the family $T^q = (T_1^q, \ldots, T_n^q)$ is ergodic. We recall this means, that for every $f \in L^2(X, \mu)$

$$T_1^q f = \ldots T_n^q f = f$$

implies $f =$ *constant*. We'll refer to a family of transformations satisfying all the above conditions as a *strongly ergodic family.*

Theorem 1.2. *Let $Q(m)$ be a non-degenerate form, Γ be a corresponding set of regular values and $T = (T_1, \ldots, T_n)$ a strongly ergodic family of transformations of a measure space (X, μ).*
For $f \in L^2(X, \mu)$ consider the averages

$$A_\lambda f(x) = \frac{1}{r_Q(\lambda)} \sum_{Q(m_1,\ldots m_n)=\lambda} f(T_1^{m_1} T_2^{m_2} \cdots T_n^{m_n} x)$$

Then one has

$$\| \lim_{\lambda\in\Gamma,\lambda\to\infty} (A_\lambda f - \int_X f d\mu)\|_{L^2(X,\mu)} = 0. \tag{1.2.8}$$

This is an L^2 ergodic theorem, it follows from a non-trivial estimate on the exponential sums $\hat{\sigma}_{Q,\lambda}(\xi)$ at irrational points $\xi \notin \mathbb{Q}^n$. More precisely one needs the following

Lemma 1.2. *Let $Q(m)$ be a non-degenerate form, Γ be a corresponding set of regular values. Then for $\xi \notin \mathbb{Q}^n$ one has*

$$\lim_{\lambda\in\Lambda,\lambda\to\infty} \frac{1}{r_Q(\lambda)} |\hat{\sigma}_{Q,\lambda}(\xi)| = 0. \tag{1.2.9}$$

To see the correspondence, suppose that $f \in L^2(x, \mu)$, $f \neq$ *constant* is a joint eigenfunction of the shifts: $T_j f = e^{2\pi i \xi_j} f$ $(T_j f(x) = f(T_j x))$. Then $A_\lambda f = \frac{1}{r_Q(\lambda)} \hat{\sigma}_{Q,\lambda}(\xi) f$, and the strong ergodicity of the family T implies that $\xi \notin \mathbf{Q}^n$.
The main result is the corresponding pointwise ergodic

Theorem 1.3. *Let $Q(m)$ be a non-degenerate form, Γ be a corresponding set of regular values and $T = (T_1, \ldots, T_n)$ a strongly ergodic family of transformations of a measure space (X, μ). Let $f \in L^2(X, \mu)$, Then for μ-almost every $x \in X$ one has*

$$\lim_{\lambda\in\Gamma,\lambda\to\infty} A_\lambda f(x) = \int_X f\, d\mu. \tag{1.2.10}$$

Theorem 1.3 means, that the images of the solution sets

$$U_\lambda = \{m \in \mathbf{Z}^n \ : \ Q(m) = \lambda\}, \tag{1.2.11}$$

under the transformations $T = (T_1, \ldots, T_n)$, that is the sets

$$\Omega_{x,\lambda} = \{(T_1^{m_1} T_2^{m_2} \cdots T_n^{m_n} x) \ : \ m \in U_\lambda\}, \tag{1.2.12}$$

become uniformly distributed on X w.r.t. μ for almost every $x \in X$. Let us mention a special case

Corollary 1.1. *Let $\alpha_1, \ldots, \alpha_n$ be a set of irrational numbers ($\alpha_j \notin \mathbf{Q}\ \forall j$). If $Q(m)$ is a non-degenerate form, and Γ is a corresponding set of regular values, then the sets*

$$\Omega_{\lambda,\alpha} = \{(m_1\alpha_1, \ldots, m_n\alpha_n) \in \Pi^n \ : \ Q(m_1, \ldots, m_n) = \lambda\}, \tag{1.2.13}$$

become uniformly distributed on the torus Π^n w.r.t. the Lebesgue measure.

Indeed, if $X = \Pi^n$ and $T_j(x_1, \ldots, x_j, \ldots x_n) \to (x_1, \ldots, x_j + \alpha_j, \ldots x_n)$ and $\alpha_j \notin \mathbb{Q}$, then the family $T = (T_1, \ldots, T_n)$ is strongly ergodic.

The proof of the pointwise ergodic theorem is based on the L^2 boundedness of a corresponding maximal function

Theorem 1.4. *Let $Q(m)$ be a non-degenerate form, Γ be a corresponding set of regular values. For $\phi \in l^2(\mathbf{Z}^n)$ we define the maximal function*

$$N^*\phi(m) = \sup_{\lambda \in \Gamma} \frac{1}{r_Q(\lambda)} \Big| \sum_{Q(l)=\lambda} \phi(m-l) \Big|, \tag{1.2.14}$$

Then one has

$$\|N^*\phi\|_{l^2(\mathbf{Z}^n)} \leq C \|\phi\|_{l^2(\mathbf{Z}^n)}. \tag{1.2.15}$$

Theorem 1.4 is a discrete analogue of a maximal theorem on $\mathbb{R}^n$, corresponding to the level surfaces of the form $Q(x)$.

Theorem 1.5. *Let $Q(x)$ be a non-degenerate form and $f \in L^2(\mathbb{R}^n)$. Then for the maximal function*

$$M^*f(x) = \sup_{\lambda > 0} \lambda^{-\frac{n}{d}+1} \Big| \int_{Q(y)=\lambda} f(x-y) \ \frac{dS_{Q,\lambda}(y)}{|Q'(y)|} \Big|, \tag{1.2.16}$$

one has

$$\|M^*f\|_{L^2(\mathbb{R}^n)} \leq C \|f\|_{L^2(\mathbb{R}^n)}. \tag{1.2.17}$$

For the polynomial $Q(x) = \sum_{j=1}^n x_j^2$ this is the spherical maximal theorem of E.M.Stein [101]. For general forms Q Theorem 1.5 does not seem to be in the literature [102]. This is due the lack of decay estimates for the Fourier transform of surface measure $dS_{Q,\lambda}$, which we will derive from appropriate estimates from closely related exponential sums.

For the special case $Q(m) = \sum_{j=1}^n m_j^2$, Theorem 1.4, was proved by Magyar, stein and Wainger in [82], moreover there the $l^p \to l^p$ boundedness of the discrete maximal operator was shown for the sharp range of exponents $p > \frac{n}{n-2}$. The non-degeneracy condition (0.1) is also, sharp in the sense, that for the form $Q(m) = m_1^2 + m_2^2 + m_3^2 + m_4^2$ (where *codim* $V_Q = 4 = (d-1)2^d$), Theorem 4. is not true, taking averages on any arithmetic progression Γ, see section 1.3 below. Hence the work presented in this chapter is the continuation of [82] to some extent.

As mentioned above, our work is motivated by Bourgain's polynomial ergodic theorem [15, 13, 14] where the Hardy-Littelwood method of exponential sums was used in a crucial way to obtain l^p-bounds for discrete maximal operators from known l^p bounds for analogues maximal operators in Euclidean spaces. However in the present case, the averages are over disjoint sets, the *strong ergodicity* condition is also necessary, and is actually a condition on the joint spectrum of the transformations $(T_1, \ldots, T_n)$. Thus we will need the Spectral Theorem even in case of the pointwise convergence, i.e. in the proof of Theorem 1.3.

1.3 Exponential sums and oscillatory integrals

We recall some estimates going back to Birch [11] on exponential sums, and prove the estimates and properties of oscillatory integrals needed later. In particular we give a proof of Theorem 1.5.

Let $Q(m)$ be a non-degenerate form of degree d, that is a positive homogeneous polynomial with integer coefficients, satisfying the non-degeneracy condition (1.2.1). Let $P > 1$, $0 < \theta \leq 1$ be fixed.

Definition 1.1. *For $1 \leq q \leq P^{(d-1)\theta}$, $1 \leq a < q$, $(a,q) = 1$ we define the major arcs*

$$L_{a,q}(\theta) = \{\alpha : 2|\alpha - a/q| < q^{-1} P^{-d+(d-1)\theta}\} \tag{1.3.1}$$

$$L(\theta) = \bigcup_{q \leq P^{(d-1)\theta},\, (a,q)=1} L_{a,q}(\theta)$$

If $\alpha \notin L(\theta)$ then α belongs to the minor arcs.

The following properties of the major arcs are immediate from the definition, see [2, Sec.4] for the proof.

Proposition 1.1. *If*

(i) $\theta_1 < \theta_2$ then $L(\theta_1) \subseteq L(\theta_2)$.

(ii) $\theta < \frac{d}{3(d-1)}$ then the intervals $L_{a,q}(\theta)$ are disjoint for different values of a and q.

(iii) $\theta < \frac{d}{3(d-1)}$ then $|L(\theta)| \leq P^{-d+3(d-1)\theta}$.

We'll use the notation $\kappa = \frac{codim\ V_Q}{2^{d-1}}$ throughout this chapter, and it is understood that $\frac{\kappa}{d-1} > 2$ which follows from condition (1.2.1).
Let $Q_1(m)$ be a polynomial of degree d, such that its d-degree homogeneous part $Q(m)$ is a non-degenerate form. For a real α and a smooth cut-off function $\phi(x)$, consider the exponential sum

$$S(\alpha) = \sum_{m \in \mathbf{Z}^n} e^{2\pi i \alpha\, Q_1(m)}\, \phi(m/P). \tag{1.3.2}$$

This is a Weyl type sum, the trivial estimate is $S(\alpha) \leq P^n$. The following estimates due to Birch [11] are of basic importance

Lemma 1.3. *Suppose $\alpha \notin L(\theta)$, then for any $\epsilon > 0$, one has*

$$|S(\alpha)| \leq C_\epsilon P^{n-\kappa\theta+\epsilon}. \tag{1.3.3}$$

If $\delta < \frac{\kappa-2(d-1)}{12d(d-1)}$ *and* $\frac{2\delta\kappa}{d-1} - 2 < \theta < \frac{1}{6d}$ *then one has for the average over the minor arcs,*

$$\int_{\alpha\notin L(\theta)} |S(\alpha)|\, d\alpha \leq C_\delta P^{n-d-\delta}. \tag{1.3.4}$$

The constants C_ϵ, and C_δ depend just on the homogeneous part $Q(m)$, on the cut-off ϕ, on ϵ and δ.

Remark. Estimate (1.3.3) is proved in [11], see Lemma 4.3, when the cut-off ϕ is replaced by the characteristic function χ of a cube of side length ≈ 1. Choosing χ s.t. $\chi\phi = \phi$ and applying Plancherel's identity, one has

$$\sum_{m\in\mathbf{Z}^n} e^{2\pi i\alpha Q_1(m)}\, \phi(m/P)\chi(m/P) = \int_{\Pi^n} (\sum_{m\in\mathbf{Z}^n} e^{2\pi i\alpha Q_1(m)-m\cdot\xi}\chi(m/P)\, (P^n\hat{\phi}(P\xi))\, d\xi$$

Here Π^n is the flat torus and can be identified with $[-1/2, 1/2]^n$. Estimate (1.3.3) holds for the first term of the integral uniformly in ξ and it is easy to see that $\|P^n\hat{\phi}(P\xi)\|_1 \leq c_\phi$.
To see (1.3.4) one uses (1.3.3) for $\alpha \notin L(\theta)$, with $\theta < \theta'$, together with the fact that the measure of $L(\theta')$ is small by (1.3.1), which gives $\int_{\alpha\in L(\theta')\setminus L(\theta)} |S(\alpha)|\, d\alpha \leq C_\delta P^{n-d-\delta}$, see [11, Lemma 4.4].

Corollary 1.2. *Let $Q(m)$ be a non-degenerate form, and $1 \leq a < q$ be natural numbers s.t. $(a, q) = 1$. Consider the Weyl sum,*

$$S(a, q) = \sum_{m\in\mathbb{Z}/q\mathbb{Z}^n} e^{2\pi i\, \frac{a}{q} Q(m)}. \tag{1.3.5}$$

One has

$$|S(a, q)| \leq C_{Q,\epsilon}\, q^{n-\frac{\kappa}{d-1}+\epsilon}. \tag{1.3.6}$$

Proof. Choose $\alpha = a/q$, $P = q$ and notice $\alpha \notin L(\theta)$ for $\theta < \frac{1}{d-1}$. Indeed for $q_1 \leq q^{(d-1)\theta} < q$: $|a/q - a_1/q_1| \geq (qq_1)^{-1} \geq q_1^{-1}q^{-d+(d-1)\theta}$. The estimate follows from (1.3.3). $\square$

Corollary 1.3. *If $|\alpha| < P^{-2d/3}$ then $|S(\alpha)| \leq C_{Q,\epsilon}\, P^{n+\epsilon}(P^d|\alpha|)^{-\frac{\kappa}{d-1}}$*

Proof. Choose θ s.t. $|\alpha| = P^{-d+(d-1)\theta}$, that is $(P^d|\alpha|)^{\frac{1}{d-1}} = P^\theta$. The major arcs $L_{a,q}(\theta)$ are disjoint since $(d-1)\theta < d/3$, moreover α is an endpoint of the interval $L_{0,1}(\theta)$ hence $\alpha \notin L_{a,q}(\theta - \epsilon)$ for every $\epsilon > 0$. By (1.3.3)

$$|S(\alpha)| \leq C_{Q,\epsilon}\, P^{n-\kappa\theta+\epsilon} = C_{Q,\epsilon}\, P^{n+\epsilon}(P^d|\alpha|)^{-\frac{\kappa}{d-1}} \quad \square$$

Let $Q(x)$ be a non-degenerate form of degree d, $\kappa = \frac{codim\, V_Q}{2^{d-1}}$, $L > 0$, and $\eta \in \Re^n$.

Lemma 1.4. *Consider the oscillatory integral*

$$I_Q(L, \eta) = \int e^{2\pi i(LQ(x)+x\cdot\eta)}\phi(x)\, dx. \tag{1.3.7}$$

One has, for every $\epsilon > 0$,

$$I_Q(L, \eta) \leq C_{Q,\epsilon}(1 + L)^{-\frac{\kappa}{d-1}+\epsilon}, \tag{1.3.8}$$

where the constant C_ϵ is independent of L and η.

Proof. The estimate is clear for $L < 1$. Let $L \geq 1$, the gradient of the phase: $|LQ'(x) + \eta| \geq L$ if $|\eta| \geq CL$ on the support of $\phi(x)$ for large enough constant $C > 0$, and (1.3.8) follows by partial integration.
Suppose $|\eta| \leq CL$ and introduce the parameters P, θ, α s.t. $\alpha = P^{-d}L$, $L = P^{(d-1)\theta}$ and $P > L^{\frac{3\kappa}{d-1}}$. Changing variables $y = Px$ one has,

$$I_Q(L,\eta) = P^{-n} \int e^{2\pi i\alpha\,(Q(y)+P^{d-1}y\cdot\eta)}\phi(y/P)\,dy$$

We compare the integral to a corresponding exponential sum

$$P^{-n}S(\alpha) = P^{-n} \sum_{m\in\mathbf{Z}^n} e^{2\pi i\alpha\,(Q(m)+P^{d-1}m\cdot\eta)}\phi(m/P)$$

If $y = m + z$ where $m \in \mathbf{Z}^n$ and $z \in [0,1]^n$, then

$$|e^{2\pi i\alpha\,(Q(y)+P^{d-1}y\cdot\eta)} - e^{2\pi i\alpha\,(Q(m)+P^{d-1}m\cdot\eta)}| \leq C|\alpha|(|Q(m+z) - Q(m)| + P^{d-1}|\eta|),$$

which is bounded by $CP^{-1+(d-1)\theta}$, since $|\alpha| = P^{-d+(d-1)\theta}$ and $|\eta| \leq P^{(d-1)\theta}$. Thus $|I_Q(L,\eta) - P^{-n}S(\alpha)| \leq C_Q P^{-1+2(d-1)\theta} \leq C_Q P^{-\frac{1}{3}}$. Corollary 1.3 implies that

$$|P^{-n}S(\alpha)| \leq C_\epsilon (P^d\alpha)^{-\frac{\kappa}{d-1}+\epsilon} C_\epsilon L^{-\frac{\kappa}{d-1}+\epsilon}$$

and (1.3.8) follows using $P^{-\frac{1}{3}} \leq L^{-\frac{\kappa}{d-1}}$. $\square$

The level surfaces of a non-degenerate form $S_{Q,\lambda} = \{x \in \Re^n : Q(x) = \lambda\}$ are compact smooth hypersurfaces (for $\lambda > 0$). Indeed $Q(x) = \lambda$ implies that $|x| \approx \lambda^{1/d}$, moreover $Q'(x) \neq 0$ for every $x \neq 0$. There is a unique $n-1$-form $d\sigma_Q(x)$ on $\Re^n - 0$ for which

$$dQ \wedge d\sigma_Q = dx_1 \wedge \ldots \wedge dx_n, \tag{1.3.9}$$

called the Gelfand-Leray form, see [1, Sec.7.1]. To see this, suppose that $\partial_1 Q(x) \neq 0$ on some open set U. By a change of coordinates: $y_1 = \partial_1 Q(x), y_j = x_j$ for $j \geq 2$, equation (1.3.9) takes the form

$$dy_1 \wedge d\sigma_Q(y) = \partial_1 H(y)\ dy_1 \wedge \ldots \wedge dy_n, \tag{1.3.10}$$

where $x_1 = H(y), x_j = y_j$ is the inverse map. Thus the form: $d\sigma_Q(y) = \partial_1 H(y)\,dy_2 \bigwedge \ldots \bigwedge dy_n$ satisfies equation (1.8).

We define the measure $d\sigma_{Q,\lambda}$ as the restriction of the $n-1$ form $d\sigma_Q$ to the level surface $S_{Q,\lambda}$. This measure is absolutely continuous w.r.t. the Euclidean surface are measure $dS_{Q,\lambda}$, more precisely one has

Proposition 1.2.

$$d\sigma_{Q,\lambda}(x) = \frac{dS_{Q,\lambda}(x)}{|Q'(x)|}. \tag{1.3.11}$$

Proof. Choose local coordinates y as before, in coordinates y level surface and surface area measure takes the form:

$$S_{Q,\lambda} = \{x_1 = H(\lambda, y_2, \ldots, y_n) : x_j = y_j\}$$

and

$$dS_{Q,\lambda}(y) = (1+\sum_{j=2}^{n}\partial_j^2 H(\lambda,y))^{1/2}\ dy_2\wedge\cdots\wedge dy_n$$

Using the identity $F(H(y),y_2,\dots,y_n)=y_1$ one has

$$\partial_1 F(x)\partial_1 H(y) = 1\ ,\quad \partial_1 F(x)\partial_j H(y)+\partial_j F(x)=0$$

This implies that $\partial_1 H(y) = (1+\sum_{j=2}^{n}\partial_j^2 H(y))^{1/2}\cdot|F'(x)|^{-1}$. Then (1.3.11) follows by taking $y_1=\lambda$. $\square$

A key observation is that the measure $d\sigma_{Q,\lambda}$, considered as a distribution on $\mathbb{R}^n$, has a simple oscillatory integral representation

Lemma 1.5. *Let $Q(x)$ be a non-degenerate form and $\lambda>0$. Then in the sense of distributions*

$$d\sigma_{Q,\lambda}(x) = \int_{\mathbb{R}} e^{2\pi i(Q(x)-\lambda)t}\,dt. \tag{1.3.12}$$

This means that for any smooth cut-off function $\chi(t)$ and test function $\phi(x)$ one has

$$\lim_{\epsilon\to 0}\int\int e^{2\pi i(Q(x)-\lambda)t}\chi(\epsilon t)\phi(x)\,dxdt\ =\ \int\phi(x)d\sigma_{Q,\lambda}(x). \tag{1.3.13}$$

Proof. Let U be an open set on which $\partial_1 Q\neq 0$, and by a partition of unity we can suppose, that $supp\,\phi\subseteq U$. Changing variables $y_1=Q(x), y_j=x_j$ the left side of (1.3.13) becomes

$$\lim_{\epsilon\to 0}\int\int e^{2\pi i(y_1-\lambda)t}\chi(\epsilon t)\tilde{\phi}(y)|\partial_1 H(y)|\,dydt = \int\tilde{\phi}(\lambda,y')|\partial_1 H(\lambda,y')|dy'$$

where $y'=(y_2,\dots y_n)$. The last equality can be seen by integrating in t and in y_1 first, and using the Fourier inversion formula:

$$\lim_{\epsilon\to 0}\int\int e^{2\pi i(y_1-\lambda)t}\chi(\epsilon t)g(y_1)\,dy_1dt = g(\lambda)$$

On the other hand $S_{Q,\lambda}\cap U=\{x_1=H(\lambda,y_2,\dots y_n): x_j=y_j\}$ and $d\sigma_{Q,\lambda}(y)=|\partial_1 H(\lambda,y')|\,dy'$ in parameters y'. $\square$

Lemma 1.6. *Let $Q(x)$ be a non-degenerate form of degree d, $\kappa=\frac{codim\,V_Q}{2^{d-1}}$. Then one has for the Fourier transform of the measure $d\sigma_{Q,1}=d\sigma_Q$*

$$|d\tilde{\sigma}_Q(\xi)|\leq C_{Q,\epsilon}(1+|\xi|)^{-\frac{\kappa}{d-1}+1+\epsilon}. \tag{1.3.14}$$

Proof. Suppose $|\xi|>1$. Using the fact that $\phi\,d\sigma_Q=d\sigma_Q$ if $\phi=1$ on a neighborhood of 0 and formula (1.3.13), we have

$$d\tilde{\sigma}_Q(\xi)=\int e^{-2\pi i\,x\cdot\xi}\phi(x)\,dx=\lim_{\delta\to 0}\int\int e^{-2\pi i\,x\cdot\xi}e^{2\pi i(Q(x)-1)t}\phi(x)\chi(\delta t)\,dxdt.$$

We decompose the range of integration into two parts

$$\tilde{d\sigma}_Q(\xi)=\int_{|t|\geq C|\xi|}\int_{\mathbb{R}^n}\qquad+\int_{|t|\leq C|\xi|}\int_{\mathbb{R}^n}\qquad = I_1+I_2$$

Since for fixed $|t| \leq C|\xi|$ the gradient of the phase: $|tQ'(x) - \xi| \geq |\xi|/2$ if $C > 0$ is small enough, integration by parts gives $|I_2| \leq C_N\,(1+|\xi|)^{-N+1}$ for every $N > 0$. For $|t| \geq C|\xi|$ Lemma 1.3 implies,

$$|\int e^{2\pi i(tQ(x)-x\cdot\xi)}\phi(x)\,dx| \leq C_\epsilon|\,|t|^{-\frac{\kappa}{d-1}+\epsilon},$$

hence

$$I_1 \leq C_\epsilon \int_{|t|\geq C|\xi|} |t|^{-\frac{K}{d-1}+\epsilon}\,dt \leq C_\epsilon\,|\xi|^{-\frac{\kappa}{d-1}+1+\epsilon}. \quad \square$$

First we prove a dyadic version of Theorem 1.5., together with a refinement which will be needed in the proof of Theorem 1.3.

Lemma 1.7. *Let $\Lambda > 0$ be fixed, $\omega(\xi)$ be a smooth function with supported on the set $\{\Lambda^{-\frac{1}{2d}} \leq \|\xi\| \leq \frac{1}{4}\}$, where $\|\xi\| = max_j|\xi_j|$.*

Let M_λ and $M_{\omega,\lambda}$ be the multipliers acting on $L^2(\mathbb{R}^n)$ defined by

$$\widetilde{M_\lambda f}(\xi) = d\tilde{\sigma}(\lambda^{1/d}\xi) \quad \textit{and} \quad \widetilde{M_{\omega,\lambda} f}(\xi) = \omega(\xi)d\tilde{\sigma}(\lambda^{1/d}\xi).$$

Then one has for the maximal operators,

$$\|\sup_{\Lambda\leq\lambda<2\Lambda} |M_\lambda f|\,\|_{L^2} \leq C\|f\|_{L^2}. \tag{1.3.15}$$

$$\|\sup_{\Lambda\leq\lambda<2\Lambda} |M_{\omega,\lambda} f|\,\|_{L^2} \leq C\,\Lambda^{-\frac{1}{2d}}\|f\|_{L^2}. \tag{1.3.16}$$

Note that $M_\lambda f = \lambda^{-\frac{n}{d}+1}\,(f * d\sigma_\lambda)$.

Proof. Using the integral representation (1.3.12) one has

$$d\tilde{\sigma}(\lambda^{1/d}\xi) = \lambda^{-\frac{n}{d}+1}[d\tilde{\sigma}_\lambda(\xi) = \lambda^{-\frac{n}{d}+1}\int_{\Re}\int_{\Re^n} e^{2\pi i(Q(x)-\lambda)t+m\cdot\xi}\phi(x/\Lambda^{\frac{1}{d}})\,dx\,dt.$$

This means

$$M_\lambda f = \lambda^{-\frac{n}{d}+1}\int e^{-2\pi i\lambda t}H_{\Lambda,t}f\,dt,$$

where $H_{\Lambda,t}$ is the multiplier corresponding to

$$h_{\Lambda,t}(\xi) = \int e^{2\pi i(Q(x)t+m\cdot\xi)}\phi(x/\Lambda^{\frac{1}{d}})\,dx.$$

Then taking the absolute values, and using Minkowski's integral inequality

$$\|\sup_{\Lambda\leq\lambda<2\Lambda} |M_\lambda f|\,\|_{L^2} \leq C\Lambda^{-\frac{n}{d}+1}\int \|H_{\Lambda,t}f\|_{L^2}\,dt.$$

Using estimate (1.3.14) and the fact that $-\frac{\kappa}{d-1}+\epsilon < -2$, one has

$$|h_{\Lambda,t}(\xi)| \leq C\Lambda^{\frac{n}{d}}\min\{(1+\Lambda|\xi|)^{-N},(1+\Lambda|t|)^{-2}\},$$

and (1.3.15) follows as $\Lambda\int(1+\Lambda t)^{-2}\,dt \leq C$.

To prove (1.3.16) we have to replace $h_{\Lambda,t}(\xi)$ by $\omega(\xi)h_{\Lambda,t}(\xi)$. Then one has better uniform estimates in ξ; indeed for $\Lambda t \leq \Lambda^{\frac{1}{2d}}$ it follows

$$|\omega(\xi)h_{\Lambda,t}(\xi)| \leq C(1+\Lambda|\xi|)^{-N} \leq (1+\Lambda^{\frac{1}{2d}})^{-N}, \quad \text{hence}$$

$$\Lambda \int \sup_{\xi} |\omega(\xi)h_{\Lambda,t}(\xi)|\, dt \leq C\Lambda \int_{\Lambda t \leq \Lambda^{\frac{1}{2d}}} \Lambda^{-\frac{N}{2d}} dt + C\Lambda \int_{\Lambda t \geq \Lambda^{\frac{1}{2d}}} (\Lambda t)^{-2} dt \leq \ C\Lambda^{-\frac{1}{2d}}$$

This proves (1.3.16). $\square$

Proof of Theorem 1.5. If $Q(x)$ is a positive, non-degenerate form of degree d, then $Q(x) \approx |x|^d$. Then the maximal function: $\bar{M}f(x) = \sup_{\lambda>0} \lambda^{-n/d}|\bar{A}_\lambda f(x)|$, where

$$\bar{A}_\lambda f(x) = \int_{Q(y)\leq\lambda} f(x-y)\, dy$$

is majorized by the standard Hardy-Littlewood maximal function, hence is bounded from $L^2(\mathbb{R}^n)$ to itself. Note that,

$$\int_{Q(y)\leq\lambda} g(y)\, dy = \int_0^\lambda \int_{Q(y)=s} g(y)\, d\sigma_{Q,s}(y) ds$$

hence

$$\bar{A}_\lambda f(x) = \lambda^{-1} \int_0^\lambda Af(x)\, ds$$

Then, using estimate (1.3.14), Theorem 1.5 follows by the standard proof of Stein's spherical maximal theorem, see [101]. $\square$

1.4 Fourier transform of integer points on hypersurfaces.

First we rewrite formula (1.2.5) in the form

$$\hat{\sigma}_{Q,\lambda}(\xi) = c_Q \sum_{q=1}^{\infty} \sum_{(a,q)=1} m_\lambda^{a/q}(\xi) + \mathcal{E}_\lambda(\xi), \tag{1.4.1}$$

where

$$m_\lambda^{a/q}(\xi) = \sum_{l\in\mathbf{Z}^n} e^{-2\pi i\lambda a/q} G(a/q,l)\, \psi(q\xi-l) d\tilde{\sigma}_{Q,\lambda}(\xi-l/q) \tag{1.4.2}$$

$$\text{and} \qquad G(a/q,l) = q^{-n} \sum_{s\in\mathbf{Z}^n/q\mathbf{Z}^n} e^{2\pi i \frac{a(Q(s)-\lambda)+s\cdot l}{q}}.$$

Here we used the fact, that $d\tilde{\sigma}_{Q,\lambda}(\eta) = \lambda^{n/d-1} d\tilde{\sigma}_Q(\lambda^{1/d}\eta)$, which follows by scaling, since $|Q'(x)|$ is homogeneous of degree $d-1$.

Note that in the right side of (1.4.1) there is at most one nonzero term, since the cut-off factor $\psi(q\xi-l)$, and then (3.5.4) implies

$$|m_\lambda^{a/q}(\xi)| \leq C_\epsilon \lambda^{n/d-1} q^{-\frac{\kappa}{d-1}+\epsilon} \leq C_\epsilon \lambda^{n/d-1} q^{-2-\epsilon}, \tag{1.4.3}$$

by (1.2.1) if ϵ is small enough, hence the sum in (1.4.1) is absolutely convergent.

Let N_λ and M_λ denote the convolution operators on $\mathbf{Z}^n$ corresponding to the multipliers $\hat{\sigma}_{Q,\lambda}(\xi)$ and $m_\lambda(\xi) = \sum_q \sum_{(a,q)=1} m_\lambda^{a/q}(\xi)$. The main approximation property we need is the following

Lemma 1.8. *Let $\Lambda > 0$, $\delta > 0$ be amall, fixed and $f \in l^2(\mathbf{Z}^n)$ then*

$$\| \sup_{\Lambda \leq \lambda < 2\Lambda} |(N_\lambda - M_\lambda) f| \, \|_{l^2} \leq C_\delta \Lambda^{\frac{n}{d}-1-\delta} \|f\|_{l^2}. \tag{1.4.4}$$

Lemma 1.8 in the special case $Q(m) = \sum_j m_j^2$ is proved in [82] and note that it immediately implies Lemma 1.1 as for fixed λ ($\Lambda \leq \lambda < 2\Lambda$), one has

$$\|(N_\lambda - M_\lambda) f \, \|_{l^2} \leq C \Lambda^{\frac{n}{d}-1-\delta} \|f\|_{l^2} \quad \forall f \in l^2(\mathbf{Z}^n)$$

is equivalent to

$$\sup_\xi |\hat{\sigma}_{Q,\lambda}(\xi) - m_\lambda(\xi)| \leq C \lambda^{\frac{n}{d}-1-\delta}$$

which is the content of (1.2.5).

Let $P = \Lambda^{1/d}$, and let $\phi(x)$ be smooth cut-off function on $\mathbb{R}^n$ s.t. $\phi(x) = 1$ for $Q(x) \leq 2$. Then

$$\hat{\sigma}_{Q,\lambda}(\xi) = \sum_{m \in \mathbf{Z}^n} e^{2\pi i m \cdot \xi} \phi(m/P) \int_0^1 e^{2\pi\alpha i (Q(m) - \lambda)} \, d\alpha = \int_0^1 S(\alpha, \xi) e^{-2\pi i \lambda \alpha} \, d\alpha,$$

where $S(\alpha, \xi) = \sum_m e^{2\pi i (\alpha Q(m) + m \cdot \xi)} \phi(m/P)$.

Let δ and θ be chosen as in Lemma 3. and integrate separately on the major and minor arcs:

$$\hat{\sigma}_{Q,\lambda}(\xi) = \int_{\alpha \in L(\theta)} S(\alpha, \xi) e^{-2\pi i \lambda \alpha} \, d\alpha + \int_{\alpha \notin L(\theta)} S(\alpha, \xi) e^{-2\pi i \lambda \alpha} \, d\alpha \tag{1.4.5}$$

$$:= a_\lambda(\xi) \; + \; \mathcal{E}_\lambda^1(\xi). \tag{1.4.6}$$

The following proposition is a prototype of the error estimates in this section

Proposition 1.3. *Let $\mathcal{E}_\lambda^1$ be the multiplier corresponding to $\mathcal{E}_\lambda^1(\xi)$ that is: $\widehat{\mathcal{E}_\lambda^1 f} = \mathcal{E}_\lambda^1(\xi) \hat{f}(\xi)$. Then one has*

$$\| \sup_{\Lambda \leq \lambda < 2\Lambda} |\mathcal{E}_\lambda^1 f| \|_{l^2} \leq C_{Q,\delta} \Lambda^{\frac{n}{d}-1-\delta} \|f\|_{l^2} \tag{1.4.7}$$

Proof. Let S_α be defined by $\widehat{S_\alpha f} = S(\alpha, \xi) \hat{f}(\xi)$, then

$$\sup_{\Lambda \leq \lambda < 2\Lambda} |\mathcal{E}_\lambda^1 f| \leq \int_{\alpha \notin L(\theta)} |S_\alpha f| \, d\alpha$$

Taking the l^2 norm one gets (1.4.7) from the minor arc estimate (1.3.4),

$$\int_{\alpha \notin L(\theta)} |S_\alpha(x, \xi)| \leq C_\delta \Lambda^{n/d-1-\delta}. \; \square$$

Suppose $\alpha \in L_{a,q}(\theta)$ for some $(a,q)=1$, $q \leq P^{(d-1)\theta}$, and write $\alpha = a/q+\beta$, $|\beta| \leq P^{-d+(d-1)\theta}$, $m = qm_1+s$. We have

$$S(\alpha,\xi) = \sum_{s\in \mathbf{Z}^n/q\mathbf{Z}^n} e^{2\pi i\frac{aQ(s)}{q}} \sum_{m_1\in\mathbf{Z}^n} e^{2\pi i(\beta Q(qm_1+s)+(qm_1+s)\cdot\xi)}\phi(\frac{qm_1+s}{P})$$

Let $H(x,\beta) = e^{2\pi i\beta Q(m)}\phi(m/P)$, applying Poisson summation for the inner sum

$$\sum_{m_1} H(qm_1+s)e^{2\pi i(qm_1+s)\cdot\xi} = q^{-n}\sum_{l} e^{2\pi i\frac{l\cdot s}{q}}\tilde{H}(\xi-l/q,\beta)$$

Integrating in β and summing in a,q, one has

$$a_\lambda(\xi) = \sum_{q\leq P^{(d-1)\theta}} \sum_{(a,q)=1} a_\lambda^{a/q}(\xi), \tag{1.4.8}$$

where

$$a_\lambda^{a/q}(\xi) = \sum_{l\in\mathbf{Z}^n} G(a,l,q)J_\lambda(\xi-l/q), \tag{1.4.9}$$

and

$$J_\lambda(\xi-l/q) = \int_{|\beta|\leq P^{-d+(d-1)\theta}} \tilde{H}(\xi-l/q,\beta)e^{-2\pi i\lambda\beta}\,d\beta. \tag{1.4.10}$$

We shall approximate the multipliers $a_\lambda^{a/q}(\xi)$ by multipliers $b_\lambda^{a/q}(\xi)$ where the cut-off function $\psi(q\xi-l)$have been inserted in (1.4.9) that is let

$$b_\lambda^{a/q}(\xi) = \sum_{l\in\mathbf{Z}^n} G(a,l,q)\psi(q\xi-l)J_\lambda(\xi-l/q) \tag{1.4.11}$$

Next we extend the integration in β in (1.4.10) and define

$$c_\lambda^{a/q}(\xi) = \sum_{l\in\mathbf{Z}^n} G(a,l,q)\psi(q\xi-l)I_\lambda(\xi-l/q), \tag{1.4.12}$$

with

$$I_\lambda(\xi-l/q) = \int_{\Re} \tilde{H}(\xi-l/q,\beta)e^{-2\pi i\lambda\beta}\,d\beta. \tag{1.4.13}$$

Note that the integral in (1.4.13) is absolute convergent. Indeed by (1.3.8) and (1.2.1)

$$|\hat{H}(\eta,\beta)| \leq C_{Q,\epsilon}P^n(1+P^d|\beta|)^{-\frac{K}{d-1}+\epsilon}. \tag{1.4.14}$$

A crucial point is to identify the the integrals $I_\lambda(\eta)$:

$$\begin{aligned} I_\lambda(\eta) &= \int_{\Re^n}\int_{\Re} e^{-2\pi i(Q(x)-\lambda)\beta}e^{2\pi i x\cdot\eta}\phi(x/P)\,d\beta\,d\eta \\ &= \int_{\Re^n} d\sigma_{Q,\lambda}(x)e^{2\pi i x\cdot\eta}\phi(x/P)\,d\eta = d\tilde{\sigma}_{Q,\lambda}(\eta), \end{aligned} \tag{1.4.15}$$

by (1.3.12). This means that $c_\lambda^{a/q}(\xi) = m_\lambda^{a/q}(\xi)$.

Let $A_\lambda^{a/q}$, $B_\lambda^{a/q}$, $M_\lambda^{a/q}$ denote the multipliers, corresponding to $a_\lambda^{a/q}(\xi)$, $b_\lambda^{a/q}(\xi)$, and $m_\lambda^{a/q}(\xi)$.

Proposition 1.4. *We have the following estimates,*

$$\sum_{q\leq P^{(d-1)\theta}} \sum_{(a,q)=1} \| \sup_{\Lambda\leq\lambda<2\Lambda} |(A_\lambda^{a/q} - B_\lambda^{a/q})f| \,\|_{l^2} \leq C_\delta \Lambda^{\frac{n}{d}-1-\delta} \|f\|_{l^2}. \tag{1.4.16}$$

$$\sum_{q\leq P^{(d-1)\theta}} \sum_{(a,q)=1} \| \sup_{\Lambda\leq\lambda<2\Lambda} |(B_\lambda^{a/q} - M_\lambda^{a/q})f| \,\|_{l^2} \leq C_\delta \Lambda^{\frac{n}{d}-1-\delta} \|f\|_{l^2}. \tag{1.4.17}$$

$$\sum_{q\geq P^{(d-1)\theta}} \sum_{(a,q)=1} \| \sup_{\Lambda\leq\lambda<2\Lambda} |M_\lambda^{a/q})f| \,\|_{l^2} \leq C_\delta \Lambda^{\frac{n}{d}-1-\delta} \|f\|_{l^2}. \tag{1.4.18}$$

Proof. Note that each of the operators $A_\lambda^{a/q}$, $B_\lambda^{a/q}$, $M_\lambda^{a/q}$ are of the form

$$T_\lambda f = \int_I e^{-2\pi i\lambda\beta}\, U_\beta f \, d\beta$$

where U_β is some convolution operator acting on functions on $\mathbf{Z}^n$: $\widehat{U_\beta f} = \mu_\beta(\xi)\hat{f}(\xi)$, and I is some interval. Then one has the point-wise estimate

$$\sup_{\Lambda\leq\lambda<2\Lambda} |T_\lambda f| \leq \int_I |U_\beta f| \, d\beta,$$

and taking the l^2 norm

$$\| \sup_{\Lambda\leq\lambda<2\Lambda} |T_\lambda f| \,\|_{l^2} \leq \int_I |\sup_\xi| \,(\mu_\beta(\xi)|)\, d\beta \cdot \|f\|_{l^2}.$$

For the operator $A_\lambda^{a/q} - B_\lambda^{a/q}$, we have

$$\mu_\beta(\xi) = \sum_{l\in\mathbf{Z}^n} G(a,l,q)(1-\psi(q\xi - l))\hat{H}(\xi - l/q, \beta),$$

and $I = \{|\beta| \leq P^{-d+(d-1)\theta}\}$. Let $\eta = \xi - l/q$ and estimate $\hat{H}(\eta,\beta)$ by partial integration:

$$|\hat{H}(\eta,\beta)| = P^n | \int (e^{2\pi i P^d \beta Q(x)} \phi(x)) e^{2\pi i P x\cdot\eta} \, dx| \leq$$

$$C_N P^n |P\eta|^{-N} \int |(d/d\eta)^N (e^{2\pi i P^d\beta Q(x)}\phi(x))| \, dx \leq$$

$$\leq C_N P^n |P\eta|^{-N} (1 + P^d|\beta|)^N.$$

Using the facts that $|P\eta| = P/q|(q\xi - l)| \geq cP^{1-(d-1)\theta}(1 + |(q\xi - l)|)$ on the support of $1 - \psi(q\xi - l)$ (for small $c > 0$), $(d-1)\theta \leq 1/3$ and $|G(a,l,q)| \leq 1$, one has for $|\beta| \leq P^{-d+(d-1)\theta}$

$$|\sup_\xi \mu_\beta(\xi)| \leq C_N P^n P^{-N(1-2(d-1)\theta)} \sum_{l\in\mathbf{Z}^n} (1 + |q\xi - l|)^{-N} \leq C_N P^{n-N/3}$$

Then choosing N large enough, (1.4.16) follows since the total length of integration for different values of a a and q is at most 1.

For the operator $B_\lambda^{a/q} - M_\lambda^{a/q}$, we have

$$\mu_\beta(\xi) = \sum_{l\in\mathbf{Z}^n} G(a,l,q)\psi(q\xi - l)\hat{H}(\xi - l/q, \beta),$$

but we are integrating now on $|\beta| \geq P^{-d+(d-1)\theta}$. Note that $\psi(q\xi - l) \neq 0$ for at most one values of l By estimates (1.4.16) and (1.3.12): $|G(a,l,q)| \leq Cq^{-2-\epsilon}$, we have

$$|\sup_\xi \mu_\beta(\xi)| \leq C_N q^{-2-\varepsilon} P^n (1 + P^d|\beta|)^{-\frac{K}{d-1}+\epsilon}$$

hence by changing variables $\beta_1 = P^d\beta$ one has

$$\begin{aligned}\| \sup_{\Lambda\leq\lambda<2\Lambda} |(B_\lambda^{a/q} - M_\lambda^{a/q})f|\,\|_{l^2} &\leq C_\epsilon q^{-2-\varepsilon} P^{n-d} \int_{|\beta_1|\geq P^{(d-1)\theta}} |\beta_1|^{-2} d\beta \cdot \|f\|_{l^2} \\ &\leq C_\epsilon q^{-2-\varepsilon} P^{n-d-\delta}.\end{aligned}$$

Summing in $a \leq q$ and in $q = 1$ to ∞ proves (1.4.17).

For $M_\lambda^{a/q}$ the multiplier $\mu_\beta(\xi)$ is the same, but now the range of integration is the whole real line. Thus

$$\| \sup_{\Lambda\leq\lambda<2\Lambda} |M_\lambda^{a/q} f|\,\|_{l^2} \leq C_\epsilon q^{-2} \int_{\beta\in\Re} (1 + P^d|\beta|)^{-2} d\beta \cdot \|f\|_{l^2} \leq$$

$$\leq C_\epsilon q^{-2} P^{n-d} \cdot \|f\|_{l^2}$$

Summing for $a \leq q$ and $q \geq P^{(d-1)\theta}$ one gets the estimate $P^{n-d-(d-1)\theta} \leq P^{n-d-\delta}$. $\square$

Lemma 1.8 immediately follows from the above said, indeed for fixed λ

$$|(N_\lambda - M_\lambda)f| \leq \sum_{q\leq P^{(d-1)\theta}} \sum_{(a,q)=1} |(A_\lambda^{a/q} - M_\lambda^{a/q})f| + \sum_{q\geq P^{(d-1)\theta}} |M_\lambda^{a/q} f| + |\mathcal{E}_\lambda^1 f|.$$

We will need the following "dyadic" discrete maximal theorem,

Proposition 1.5. *Let $\Lambda > 0$ be fixed, then for the operator:*

$$N_\lambda f(m) = \sum_{Q(l)=\lambda} f(m - l),$$

one has

$$\| \sup_{\Lambda\leq\lambda<2\Lambda} |N_\lambda f|\,\|_{l^2} \leq C\Lambda^{\frac{n}{d}-1} \|f\|_{l^2}, \tag{1.4.19}$$

where the constant C is independent of Λ.

Proof. Note that $\widehat{N_\lambda f}(\xi) = \hat{\sigma}_{Q,\lambda}(\xi)\hat{f}(\xi)$ hence

$$N_\lambda f = \sum_{q,a} M_\lambda^{a/q} f + \sum_{q,a} (A_\lambda^{a/q} - M_\lambda^{a/q}) f + \mathcal{E}_\lambda^1 f$$

By Proposition 1.4 it is enough to show that,

$$\sum_{q,a} \| \sup_{\Lambda \leq \lambda < 2\Lambda} |M_\lambda^{a/q} f| \, \|_{l^2} \leq C \|f\|_{l^2}.$$

In proving (1.4.18) we showed

$$\| \sup_{\Lambda \leq \lambda < 2\Lambda} |M_\lambda^{a/q} f| \, \|_{l^2} \leq C q^{-\frac{\kappa}{d-1}} P^{n-d} \|f\|_{l^2} = C q^{-\frac{\kappa}{d-1}} \Lambda^{\frac{n}{d}-1} \|f\|_{l^2}$$

The sum in a, q is convergent and the Proposition is proved. $\Box$.

1.5 The singular series

In this section we analyse the normalising factor $r_Q(\lambda)$ which is the number of solutions $m \in \mathbb{Z}^n$ for which $Q(m) = \lambda$. We start by showing the existence of a regular set of values Γ, defined in Section 1, corresponding to a non-degenerate form Q. Taking $\xi = 0$ formula (1.2.5) one has that

$$r_Q(\lambda) = c_Q \lambda^{\frac{n}{d}-1} \sum_{q=1}^{\infty} K(q, 0, \lambda) + O(\lambda^{\frac{n}{d}-1-\delta}).$$

By the well-known multiplicative property $K(q_1, 0, \lambda) K(q_2, 0, \lambda) = K(q_1 q_2, 0, \lambda)$ for q_1 and q_2 being relative primes, we have

$$K(\lambda) = \sum_{q=1}^{\infty} K(q, 0, \lambda) = \prod_{p \ prime} \left(\sum_{r=0}^{\infty} K(p^r, 0, \lambda)\right) = \prod_{p \ prime} K_p(\lambda)$$

.

Note that $K(1, 0, \lambda) = 1$, then by estimate (1.5) it follows that $K_p(\lambda) = 1 + O(p^{-\frac{\kappa}{d-1}+1+\epsilon}) = 1 + O(p^{-1-\varepsilon})$, by our assumption $\kappa = \frac{codim\, V_Q}{2^{-(d-1)}} > 2$. Thus there exists $R = R_Q$ s.t.

$$1/2 \leq \prod_{p > R \ prime} |K_p(\lambda)| \leq 2. \tag{1.5.1}$$

We recall that $K_p(\lambda)$ is the density of solutions of the equation $Q(m) = \lambda$ among the p-adic integers, see [11]. More precisely,

Proposition 1.6. *Let $r_Q(p^N, \lambda) = |\{m \in \mathbf{Z}^n / p^N \mathbf{Z}^n : \ Q(m) = \lambda \ (mod\ p^N)\}|$, that is the number of solutions of the equation $Q(m) = \lambda \ (mod\ p^N)$. Then one has*

$$\sum_{r=0}^{N} K(p^r, 0, \lambda) = p^{-n(N-1)} r_Q(p^N, \lambda). \tag{1.5.2}$$

Proof. First

$$r_Q(p^N, \lambda) = \sum_{m \ (mod\ p^N)} p^{-N} \sum_{b=1}^{p^N} e^{2\pi i (Q(m)-\lambda)\frac{b}{p^N}},$$

since the inner sum is equal to p^N or 0 according to $Q(m) = \lambda\ (mod\ p^N)$ or not. Next, one writes $b = ap^{N-r}$, where $(a, p) = 1$, $a < p^r$ and $r = 0, \dots, N$, and collects the terms corresponding to a fixed r which turns out to be $K(p^r, 0, \lambda)$. $\square$

We remark that this implies: $lim_{n\to\infty} p^{-n(N-1)} r_Q(p^N, \lambda) = K_p(\lambda)$. To count the number of solutions $(mod\ p^N)$, one uses the p-adic version of Newton's method, referred to as Hensel's Lemma, see [51, Chapter 4].

Lemma 1.9. *Let p be a prime, λ and k, l be natural numbers s.t. $l > 2k$. Suppose there is an $m_0 \in \mathbf{Z}^n$ for which*

$$Q(m_0) \equiv \lambda\ (mod\ p^l), \tag{1.5.3}$$

moreover suppose that p^k is the highest power of p which divides all the partial derivatives $\partial_j Q(m_0)$. Then for $N \geq l$, one has:

$$p^{-N(n-1)} r_Q(p^N, \lambda) \geq p^{-l(n-1)}. \tag{1.5.4}$$

Proof. For $N = l$ this is obvious. Suppose it is true for N, and consider all the solutions $m_1\ (mod\ p^{N+1})$ of the form $m_1 = m + p^{N-k}s$ where $s\ (mod\ p)$. Then

$$Q(m + p^{N-k}s) - \lambda = Q(m) - \lambda + p^{n-k} Q'(m) \cdot s = 0 \ \ (mod\ p^{N+1}),$$

that is $a + b \cdot s = 0\ (mod\ p)$ where $ap^N = Q(m) - \lambda$ and $bp^k = Q'(m)$. Then $b_j \neq 0\ (mod\ p)$ for some j hence there are p^{n-1} solutions of this form. All obtained solutions are different $mod\ (p^{N+1})$, and m_1 satisfy the hypothesis of the lemma. $\square$

We remark that in case of $m = 1$, $k = 0$ the above argument shows that there are exactly $p^{(N-1)(n-1)}$ solutions m for which $m = m_0\ (mod\ p)$ and $q(m) = \lambda (mod\ p^N)$.

Lemma 1.10. *Let $Q(m)$ be a non-degenerate form, then there exists a set of regular values in the sense of* (1.2.4).

Proof. Let $\lambda_0 = Q(m_0) \neq 0$ for some fixed $m_0 \neq 0$. Let $p_1, \dots, p_J$ be the set of primes less then R (R is defined in (3.1)). Let k be an integer s.t. p_j^k does not divide $d\lambda_0$, for all $j \leq J$, where d is degree of $Q(m)$. By the homogeneity relation $Q'(m_0) \cdot m = d\lambda_0$ it follows that p_j^k does not divide some partial derivative $\partial_i Q(m_0)$. Fix l s.t. $l > 2k$ and define the arithmetic progression: $\Gamma = \{\lambda_0 + k \prod_{j=1}^J p_j^l :\ k \geq k_Q\}$. We claim that Γ is a set of regular values. Indeed, by Lemma 1.9 one has for $\lambda \in \Gamma$

$$K_{p_j}(\lambda) = \lim_{N\to\infty} p_j^{-n(N-1)} r_Q(p_j^N, \lambda) \geq p_j^{-l(N-1)}.$$

This together with (4.6.1) ensures that the singular series $K(\lambda)$ remains bounded from below, and the error term becomes negligible by choosing k_Q large enough. $\square$

Let us remark that along the same lines it can be shown, that all large numbers are regular values of $Q(m)$, if for each prime $p < R$ and each residue class $s\ (mod\ p)$, there is a solution of the equations $Q(m) = s\ (mod\ p)$ s.t. $Q'(m) \neq 0\ (mod\ p)$. This is the case for example for $Q(m) = \sum_j m_j^d$.

Let us fix a set of regular values Γ, and a rational point $k/p \neq 0$ in Π^n , where $k = (k_1, \dots, k_n) \in \mathbf{Z}^n$. Define the measure space X to be the set of residue classes $(mod\ p)$, with each element having measure $1/p$. Let $T_j(x) = x + k_j\ (mod\ p)$, then the family of transformations $T = (T_1 \dots T_n)$ is commuting, measure preserving and ergodic. Indeed for some j, $k_j \neq 0\ (mod\ p)$ and then T_j is ergodic. The function $f(x) = e^{2\pi i x/p}$ is a joint eigenfunction : $T_j f = e^{2\pi i k_j/p} f$, hence

$$A_\lambda f = \frac{1}{r_Q(\lambda)} \hat{\sigma}_{Q,\lambda}(k/p) f, \tag{1.5.5}$$

where $A_\lambda f$ are the averages defined in (1.2.7). We'll show below that the mean ergodic theorem (1.2.7) is not valid in this setting, and hence the condition *strong ergodicity* is necessary. Note that $T = (T_1, \ldots, T_n)$ is not strongly ergodic as $T_1^p = \ldots = T_n^p = I$.

Lemma 1.11. *Let Γ be a set of regular values. Let p be a large enough prime: $p > d$, $p > R$, $p > \lambda_0$ (where λ_0 is the smallest element of Γ), and $k \in \mathbb{Z}^n$. Then for $\lambda \in \Gamma$, $\lambda = \lambda_0$ (mod p) one has*

$$\frac{1}{r_Q(\lambda)}\hat{\sigma}_{Q,\lambda}(k/p) = \frac{1}{r_Q(p,\lambda)} \sum_{\substack{Q(m)=\lambda \pmod p,\\ m\in\mathbb{Z}^n/p\mathbb{Z}^n}} e^{2\pi i\frac{m\cdot k}{p}} + O(\lambda^{-\delta}). \tag{1.5.6}$$

Taking Lemma 1.11 granted for a moment, note that the expression:

$$S_k = \sum_{\substack{Q(m)=\lambda \pmod p,\\ m\in\mathbb{Z}^n/p\mathbb{Z}^n}} e^{2\pi i\frac{m\cdot k}{p}},$$

is the Fourier transform on the group $\mathbb{Z}^n/p\mathbb{Z}^n$ of the indicator function of the set of solutions to $Q(m) = \lambda \pmod p$, thus $\sum_k |S_k|^2 = p^n |r_Q(p,\lambda_0)|$ by Plancherel's formula. This implies that $S_k \neq 0$ for at least one $k \neq 0$, since otherwise the equation $Q(m) = \lambda = \lambda_0$ $(mod\ p)$ would have p^n or no solution, both cases are impossible for p being large enough.
Thus (1.2.7) is not true, assuming only that the family of transformations is ergodic. We prove now Lemma (1.11)

Proof. For a regular value $r_Q(\lambda) = c_Q K(\lambda)\lambda^{n/d-1} + O(\lambda^{n/d-1-\delta})$ where $|K(\lambda)| \gg 1$, hence by (1.2.5), it is enough to show

$$c_Q^{-1}\frac{1}{K(\lambda)}\sum_{q=1}^{\infty}\sum_{l\in\mathbf{Z}^n} K(q,l,\lambda)\psi(qk/p - l)d\tilde{\sigma}_Q(\lambda^{1/d}(k/p - l/q)) = \tag{1.5.7}$$

$$= \frac{1}{r_Q(p,\lambda)} \sum_{\substack{Q(m)=\lambda \pmod p,\\ m\in\mathbb{Z}^n/p\mathbb{Z}^n}} e^{2\pi i\frac{m\cdot k}{p}} + O(\lambda^{-\delta}).$$

For q not divisible by p, $|\frac{k}{p} - \frac{l}{q}| \geq \frac{1}{pq}$, hence each term in the sum is bounded by $q^{-\frac{\kappa}{d-1}+\epsilon}\lambda^{-\kappa/(d-1)+1+\epsilon}$ by (1.5) and (1.13). There is at most one nonzero term in the l sum for fixed q, and thus the total sum for q not divisible by p is of $O(\lambda^{-\delta})$.

For $q = bp$, only those terms for which $k/p = l/q$ are nonzero in (1.5.7), hence the sum becomes

$$\frac{1}{K(\lambda)}\sum_{b=1}^{\infty} K(bp, bk, \lambda).$$

We write $q = cp^r$ where $(c,p) = 1$ and use the multiplicative property

$$K(cp^{r+1}, ckp^r, \lambda) = K(c,0,\lambda)K(p^{r+1}, kp^r, \lambda),$$

which follows by the Chinese remainder theorem. At this point it is enough to show

$$\frac{1}{K(\lambda)}(\sum_{(c,p)=1} K(c,0,\lambda))(\sum_{r=1}^{\infty} K(p^{r+1}, kp^r, \lambda)) = \frac{1}{r_Q(p,\lambda)} \sum_{\substack{Q(m)=\lambda \pmod p,\\ m\in\mathbb{Z}^n/p\mathbb{Z}^n}} e^{2\pi i\frac{m\cdot k}{p}}. \tag{1.5.8}$$

Again, by multiplicativity,

$$\sum_{(c,p)=1} K(c,0,\lambda) \cdot \sum_{r=1}^{\infty} K(p^r,0,\lambda) = \sum_{q=1}^{\infty} K(q,0,\lambda). \tag{1.5.9}$$

For the other factor in (1.5.8) one has

$$\sum_{r=1}^{\infty} K(p^{r+1},kp^r,\lambda) = p^{-(n-1)} \sum_{\substack{Q(m)=\lambda \pmod p, \\ m\in\mathbb{Z}^n/p\mathbb{Z}^n}} e^{2\pi i \frac{m\cdot k}{p}}. \tag{1.5.10}$$

Similarly as in (3.2)

$$\sum_{m \ (mod\ p^N)} p^{-N} \sum_{b=1}^{p^N} e^{2\pi i (Q(m)-\lambda)\frac{b}{p^N}} e^{2\pi i \frac{m\cdot k}{p^N}}$$

and writes $b = ap^{N-r}$, where $(a,p)=1$, $a<p^r$ and $r=0,\dots,N$. Each term corresponding to a fixed r is $K(p^r,kp^{r-1},\lambda)$ for $r\geq 1$, while the term corresponding to $r=0$ is zero.

Next, let m_0 be a solution of $Q(m)=\lambda\ (mod\ p)$. Then by homogeneity $Q'(m_0)\cdot m_0 = d\lambda = d\lambda_0 \neq 0$ it follows by the remark after Lemma 9. that the number of solutions: $m\ (mod\ p^N)$ for which $m=m_0\ (mod\ p)$ and $Q(m)=\lambda\ (mod\ p^N)$ is exactly $p^{(n-1)(N-1)}$. Thus

$$\sum_{m \ (mod\ p^N)} p^{-N} \sum_{b=1}^{p^N} e^{2\pi i (Q(m)-\lambda)\frac{b}{p^N}} e^{2\pi i \frac{m\cdot k}{p^N}} = p^{-(n-1)} \sum_{\substack{Q(m)=\lambda \pmod p, \\ m\in\mathbb{Z}^n/p\mathbb{Z}^n}} e^{2\pi i \frac{m\cdot k}{p}},$$

this proves (3.10). By the same argument,

$$K_p(\lambda) = p^{-(n-1)}\, r_Q(p,\lambda), \tag{1.5.11}$$

and (1.5.8) follows immediately from (1.5.9), (1.5.10) and (1.5.11). $\square$

1.6 Proof of the main results.

In this section, first we prove Theorems 1.1- 1.2 and Lemma 1.2. the

Proof of Theorem 1.1 Let $\phi_\lambda(x) = \phi(x/\lambda^{1/d})$, the one has

$$\sum_{Q(m)=\lambda} \phi_\lambda(m) = \int_{\Pi^n} \hat{\sigma}_{Q,\lambda}(\xi)\hat{\phi}_\lambda(\xi)\, d\xi$$

where

$$\hat{\phi}_\lambda(\xi)\, d\xi = \sum_{m\in\mathbf{Z}^n} \phi_\lambda(m) e^{-2\pi i m\cdot\xi} = \sum_{m\in\mathbf{Z}^n} \tilde{\phi}_\lambda(\xi+m)$$

by Poisson summation (here $\tilde{\phi}_\lambda(\xi)$ denotes the Fourier transform on $\Re^n$. Since the exponential sum $\hat{\sigma}_{Q,\lambda}(\xi)$ is a smooth periodic function on $\mathbb{R}^n$ it follows

$$\sum_{Q(m)=\lambda} \phi_\lambda(m) = \int_{\Re^n} \hat{\sigma}_{Q,\lambda}(\xi)\tilde{\phi}_\lambda(\xi)\, d\xi. \tag{1.6.1}$$

Write $\hat{\sigma}_{Q,\lambda}(\xi) = m_\lambda(\xi) + \mathcal{E}_\lambda(\xi)$ and estimate the contribution of the error term

$$\int_{\Re^n} |\mathcal{E}_\lambda(\xi)\tilde{\phi}_\lambda(\xi)|\, d\xi \leq C_\delta \lambda^{n/d-1-\delta} \|\tilde{\phi}_\lambda\|_1 \leq C_\delta \lambda^{n/d-1-\delta}. \tag{1.6.2}$$

We used the error estimate in (1.2.5) and the fact that $\|\tilde{\phi}_\lambda\|_1 = \|\tilde{\phi}\|_1 \leq C$. Recall that

$$m_\lambda(\xi) = \sum_{q=1}^{\infty} \sum_{l} K(q,l,\lambda)\psi(q\xi - l) d\tilde{\sigma}_{Q,\lambda}(\xi - l/q)$$

Next we estimate the contribution of the terms corresponding to $l \neq 0$. For $q \geq \lambda^{\frac{1}{2d}}$ we use

$$\begin{aligned} \sum_{q \geq \lambda^{\frac{1}{2d}}} \sum_{l \neq 0} |K(q,l,\lambda)\psi(q\xi - l) d\tilde{\sigma}_{Q,\lambda}(\xi - l/q)| \leq \\ \leq C\lambda^{n/d-1} \sum_{q \geq \lambda^{\frac{1}{2d}}} q^{-2} \leq C_\delta \lambda^{n/d-1-\delta}, \end{aligned} \tag{1.6.3}$$

and after integrating we get the same estimate as in (1.6.2) ($\frac{\kappa}{d-1} > 2$). For $q \leq \lambda^{\frac{1}{2d}}$ we give the estimate

$$\sum_{q \leq \lambda^{1/2d}} \sum_{l \neq 0} \int_{\Re^n} |K(q,l,\lambda)\psi(q\xi - l) d\tilde{\sigma}_{Q,\lambda}(\xi - l/q)\, \tilde{\phi}_\lambda(\xi)|\, d\xi \leq c_N \lambda^{-N}, \tag{1.6.4}$$

for any $N > 0$ integer. For fixed $l \neq 0$, on the support of the cut-off factor $\psi(q\xi - l)$, one has $\|\xi - l/q\| \leq 1/(4q)$, which implies $\|\xi\| \geq 1/(2q)$, and also $\|\xi\| \geq \|l\|/(2q)$. Thus

$$\begin{aligned} |\tilde{\phi}_\lambda(\xi)| \leq\ & C_N \lambda^{n/d}(1 + \lambda^{1/d}|\xi|)^{-2N} \leq \\ \leq C_N &\lambda^{n/d}(1 + \lambda^{1/d}/2q)^{-N}(1 + c|l|/2q)^{-N}. \end{aligned} \tag{1.6.5}$$

Integrating in ξ over the region $\|\xi - l/q\| \leq 1/(4q)$, and then summing in l and in $q \leq \lambda^{\frac{1}{2d}}$ one obtains (1.6.4).
Estimates (1.6.3) and (1.6.4) imply together that the total contribution of the terms corresponding to $l \neq 0$ in (1.6.1), is $O(\lambda^{n/d-1-\delta}$. Finally, we note that

$$\sum_{q=1}^{\infty} \int |K(q,0,\lambda)(1 - \psi(q\xi)) d\tilde{\sigma}_\lambda(\xi)\tilde{\phi}_\lambda(\xi)|\, d\xi \leq C_\delta \lambda^{\frac{n}{d}-1-\delta}, \tag{1.6.6}$$

by the same argument as used in proving (1.6.3) and (1.6.4). Indeed, the range of integration is $|\xi| \geq c/q$ where both for $q \geq \lambda^{1/2d}$ and for $q \leq \lambda^{1/2d}$, one has a gain, using the decay of the the factor $K(q,0,\lambda)$ for small, and the decay of $\tilde{\phi}_\lambda$ for large values of q.
Using (1.6.3), (1.6.4) and (1.6.6), one has

$$\begin{aligned} \int_{\Re^n} \hat{\sigma}_{Q,\lambda}(\xi)\tilde{\phi}_\lambda(\xi)\, d\xi &= c_Q\, K(\lambda) \int_{\Re^n} \tilde{\sigma}_{Q,\lambda}(\xi)\tilde{\phi}_\lambda(\xi)\, d\xi + O(\lambda^{\frac{n}{d}-1-\delta}) \\ &= r_Q(\lambda) \int_{Q(y)=1} \phi(y)\, d\sigma_Q(y) + O(\lambda^{\frac{n}{d}-1-\delta}). \end{aligned} \tag{1.6.7}$$

Indeed, one replaces the singular series $c_Q K(\lambda)$ by $\lambda^{-n/d+1} r_Q(\lambda)$, use Plancherel's formula, and a change of variables $x = \lambda^{1/d} y$. This proves the Theorem, since $r_Q(\lambda) \geq C_Q \lambda^{n/d-1}$ for regular values λ. $\square$

Proof of Lemma 2. One writes

$$frac1r_Q(\lambda)|\hat{\sigma}_{Q,\lambda}(\xi)| \leq C_\delta \lambda^{-n/d+1}|m_\lambda(\xi)| + O(\lambda^{-\delta}). \tag{1.6.8}$$

For q fixed and $\xi \notin \mathbf{Q}^n$ (i.e. when ξ_j is irrational for some j)

$$\begin{aligned} \lambda^{-n/d+1}|m_{q,\lambda}(\xi)| &= c_Q \sum_l |K(q,l,\lambda)\psi(q\xi - l)d\tilde{\sigma}_{Q,\lambda}(\xi - l/q)| \\ &\leq C_Q q^{-\frac{\kappa}{d-1}+\epsilon}|d\tilde{\sigma}_{Q,\lambda}(\lambda^{1/d}\{q\xi\}/q)|, \end{aligned} \tag{1.6.9}$$

where $\{\xi\} = min\,|\xi - l|$. Indeed in the l sum only term corresponding to the closest lattice point to $q\xi$ is nonzero.

Note that $\{q\xi\} \neq 0$ for every q, since otherwise $\xi \in \mathbf{Q}^n$. Then by (1.13) and (4.10) for $q \leq \lambda^{1/2d}$ we have the estimate $\lambda^{-n/d+1}|m_{q,\lambda}(\xi)| \leq Cq^{-1-\epsilon}\lambda^{-\delta}$, while for $q \geq \lambda^{1/2d}$ one uses the bound $q^{-1-\epsilon}$. The lemma follows by summing in q. $\square$

In both the mean and pointwise ergodic theorem the Spectral Theorem will play an essential role. Also, strong (or full) ergodidity is a condition on joint spectrum of the shifts T_j ($T_j f(x) = f(T_j x)$). To see that let (X,μ) be a probability measure space, $T = (T_1 \dots T_n)$ be a family of commuting, measure preserving and invertible transformations. By the Spectral theorem there exists a positive Borel measure ν_f on the torus Π^n, s.t.

$$\langle P(T_1,\dots,T_n)f, f\rangle = \int_{\Pi^n} p(\xi)d\nu_f(\xi), \tag{1.6.10}$$

for every polynomial $P(z_1,\dots,z_n)$, where

$$(\xi) = p(\xi_1,\dots,\xi_n) = P(e^{2\pi i\xi_1},\dots,e^{2\pi i\xi_n}),$$

and $\langle,\rangle$ denotes the inner product on $L^2(X,\mu)$. We recall two basic facts

i) For $r \in \Pi^n$, $\nu_f(r) > 0$ if and only if r is a joint eigenvalue of the shifts T_j, (i.e. there exists $g \in L^2(X)$ s.t. $T_j g = e^{2\pi i r_j} g$ for each j.

ii) If the family $T = (T_1,\dots,T_n)$ is ergodic, then $\nu_f(0) = |\langle f, \mathbf{1}\rangle|^2 = |\int_X f d\mu|^2$.

Proposition 1.7. *Suppose the family $T = (T_1,\dots,T_n)$ is ergodic. Then it is strongly ergodic if and only if $\nu_f(r) = 0$ for every $r \in \mathbf{Q}^n,\ r \neq 0$.*

Proof. Suppose $\nu_f(l/q) > 0$ for some $l \neq 0$, then there exists $g \in L^2(X,\mu)$ s.t. $T_j g = e^{2\pi i l_j/q} g\ \forall j$. But then $T_j^q g = g\ \forall j$ but $g \neq constant$ since $l \neq 0$.
On the other hand suppose that $T_j^q g = g,\ \forall j$ for some $g \neq constant$. Then the functions $g_{s_1 \dots s_n}$ for $s \in \mathbf{Z}^n/q\mathbf{Z}^n$ defined by

$$g_{s_1\dots s_n} = \sum_{m \in \mathbf{Z}^n/q\mathbf{Z}^n} e^{-2\pi i \frac{m\cdot s}{q}} T_1^{m_1} \dots T_n^{m_n} g$$

are joint eigenfunctions of with eigenvalues s_j/q. They cannot vanish for all $s \neq 0$ ($mod\ q$), because then one would have $T_j g = g\ \forall j$, as can be seen easily by expressing $T_j g$ in terms of the functions $g_{s_1\dots s_n}$. $\square$

Proof of Theorem 2. We start by

$$\|A_\lambda f - \langle f, \mathbf{1}\rangle \mathbf{1}\|_2^2 = \|A_\lambda f\|_2^2 - |\langle f, \mathbf{1}\rangle|^2 = \int_{\Pi^n/\{0\}} \frac{|\hat{\sigma}_{Q,\lambda}(\xi)|^2}{r_Q(\lambda)^2}\, d\nu_f(\xi).$$

The point is that $\nu_f(\mathbf{Q}^n/\{0\}) = 0$ by the strong ergodicity condition, moreover the integrand pointwise tends to zero on the irrationals by Lemma 2, and is majorized by $\mathbf{1}$. It follows from the Lebesgue dominant convergence theorem, that the integral also tends to 0 as $\lambda \to \infty$. This proves the theorem. □

We prove Theorem 1.4 i.e. the L^2 boundedness of the discrete maximal function associated to the form $Q(m)$ now, which plays a crucial role in the proof of the pointwise ergodic theorem.

Let $\phi \in l^2\mathbf{Z}^n$, the averages we are interested in: $\frac{1}{r_Q(\lambda)} \sum_{Q(l)=\lambda} \phi(m-l)$ will be replaced by

$$N_\lambda \phi(m) = \frac{1}{\lambda^{n/d-1}} \sum_{Q(l)=\lambda} \phi(m-l). \tag{1.6.11}$$

Indeed it is enough to prove the maximal theorem for the averages N_λ, since for regular values: $r_Q(\lambda) \geq c_Q \lambda^{n/d-1}$. We write

$$N_\lambda \phi = M_\lambda \phi + \mathcal{E}_\lambda \phi = \sum_{q=1}^{\infty} \sum_{(a,q)=1} M_\lambda^{a/q} \phi + \mathcal{E}_\lambda \phi, \tag{1.6.12}$$

where M_λ, $M_\lambda^{a/q}$, $\mathcal{E}_\lambda$ denote the mulitpliers corresponding to the functions $\lambda^{-n/d+1} m_\lambda(\xi)$, $m_\lambda^{a/q}(\xi)$, $\mathcal{E}_\lambda(\xi)$. We denote by M_*, $M_*^{a/q}$, $\mathcal{E}_*$ the corresponding maximal operators.
By Lemma 8. we have,

$$\|\mathcal{E}_* \phi\|_{l^2} \leq \sum_{k=0}^{\infty} \| \sup_{2^k \leq \lambda < 2^{k+1}} |\mathcal{E}_\lambda \phi|\, \|_{l^2} \leq C_\delta \sum_{k=0}^{\infty} 2^{-k\delta} \|\phi\|_{l^2} \leq C_\delta \|\phi\|_{l^2}. \tag{1.6.13}$$

The same shows, that

$$\| \sup_{\Lambda \leq \lambda} |\mathcal{E}_\lambda \phi|\, \|_{l^2} \leq C_\delta \Lambda^{-\delta} \|\phi\|_{l^2}. \tag{1.6.14}$$

Thus to prove Theorem 1.4 it is enough to show

Lemma 1.12. *Let $q \geq 1$, and a s.t. $(a,q) = 1$ be given. The one has*

$$\|M_*^{a/q}\|_{l^2} \leq C_\epsilon q^{-\frac{\kappa}{d-1}+\epsilon} \|\phi\|_{l^2}. \tag{1.6.15}$$

It is understood that $Q(m)$ is a non-degenerate form, hence $\kappa = \frac{1}{2^(d-1)} codim\ V_Q > 2$ and $\epsilon > 0$ can be taken arbitrary small. Hence in the right side of (1.6.15) we can take the bound $Cq^{-2-\epsilon}$, but we'd like to emphasize the explicit dependence on κ.
Assuming Lemma1.12 for a moment, by sub-additivity it follows:

$$\|M_* \phi\|_{l^2} \leq C \sum_{q=1}^{\infty} q \cdot q^{-2-\epsilon} \|\phi\|_{l^2} \leq C \|\phi\|_{l^2}$$

Together with estimate (5.2) this proves Theorem 1.4 □

The proof of the Lemma 1.3.13 is based on a general transfer principle proved in [82].

Lemma 1.13. *Let $q \geq 1$ be a fixed integer and B be a finite dimensional Banach space. Let $m(\xi)$ be a bounded measurable function on $\Re^n$, taking values in B, and supported in the cube $[-\frac{1}{2q}, -\frac{1}{2q}]^n$. Define the periodic extension by*

$$m^q_{per}(\xi) = \sum_{l \in \mathbf{Z}^n} m(\xi - l/q)$$

Let $T : L^2(\Re^n) \to L^2_B(\Re^n)$ (where $L^2_B(\Re^n)$ is the space of square integrable functions taking values in the space B), be the multiplier operator corresponding to the function $m_\lambda(\xi)$.
Similarly let $T^q_{dis} : L^2(\mathbf{Z}^n) \to L^2_B(\mathbf{Z}^n)$ be the multiplier operator corresponding to the periodic function $m^q_{per}(\xi)$. Then one has

$$\|T^q_{dis}\|_{L^2(\mathbf{Z}^n) \to L^2_B(\mathbf{Z}^n)} \leq C \|T\|_{L^2(\Re^n) \to L^2_B(\Re^n)}, \tag{1.6.16}$$

where the constant C does not depend on the Banach space B, and is also independent of q.

Proof of Lemma 12. Choose a smooth function ψ' supported in $[-1/2, 1/2]^n$ for which $\psi = \psi' \psi$. Then $m^{a/q}_\lambda(\xi)$ can be written as the product of the functions

$$m^{a/q}(\xi) = \sum_{l \in \mathbf{Z}^n} G(a, l, q) \psi'(\xi - l/q), \tag{1.6.17}$$

and

$$m^q_\lambda(\xi) = \sum_{l \in \mathbf{Z}^n} \psi(\xi - l/q) d\tilde{\sigma}_\lambda(\xi - l/q). \tag{1.6.18}$$

For the first multiplier operator $M^{a/q}$ it is bounded from l^2 to itself with norm:
$\sup_\xi |m^{a/q}(\xi)| \leq C_\epsilon q^{-\frac{\kappa}{d-1}+\epsilon}$. The sequence of functions $m^q_\lambda(\xi)$ defined by (5.6) can be considered as a function mapping from $\Re^n$ to the banach space B_Λ which is the l^∞ space of functions of $1 \leq \lambda \leq \Lambda$ for some fixed Λ.

The multiplier corresponding to $\psi(q\xi) d\tilde{\sigma}_\lambda(\xi)$ is a bounded operator from $L^2(\Re^n)$ to $L^2_B(\Re^n)$ (B being the l^∞ space of functions of $\lambda > 0$), which is the content of Theorem 5. Then one applies Lemma 13. to see that the multiplier $m^q_\lambda(\xi)$ is bounded from $l^2 \mathbf{Z}^n$ to $l^2_{B_\Lambda} \mathbf{Z}^n$ with norm independent of Λ. This implies (5.2). $\square$

The proof of our main result, the Theorem 1.3, consists of a number of reductions. The argument is motivated by that of Bourgain's polynomial ergodic theorem corresponding to arithmetic subsets of integers [15, 14]. However in our case the averages are taken over disjoint sets, a condition on the joint spectrum must be imposed, and the Spectral Theorem will play an essential even in the proof of the pointwise ergodic theorem.

Let $f \in L^2(X, \mu)$, we can suppose $\int_X f \, d\mu = 0$, and then we have to show that $|A_\lambda f(x)| \to 0$ for μ almost every x, as $\lambda \to \infty$ and $\lambda \in \Gamma$. Then again we can replace the factor $r_Q(\lambda)$ by $\lambda^{n/d-1}$ in the averages.

i) We start with a standard reduction to shifts on $\mathbf{Z}^n$. Let (X, μ) be a probability measure space, $T = (T_1, \ldots, T_n)$. For $x \in X$ and $L > 0$ and define: $\phi_{L,x}(m) = f(T^m x)$ if $\|m\| \leq L$ and to be 0 otherwise. Here $m = (m_1, \ldots m_n) \in \mathbf{Z}^n$, $\|m\| = \sup_j |m_j|$ and $T^m x = T_1^{m_1} \cdot \ldots \cdot T_n^{m_n} x$. Notice that for fixed $\Lambda < L$

$$A_\lambda^* f(T^l x) = \sup_{\lambda\le\Lambda} |A_\lambda f(T^l x)| = \sup_{\lambda\le\Lambda} |N_\lambda \phi_{L,x}(l) = |N_\lambda^* \phi_{L,x}(l), \tag{1.6.19}$$

for $\|l\| \le c(L-\Lambda)$ Thus taking the square, summing in l (for $\|l\| \le c(L-\Lambda)$), and integrating over the space X one obtains

$$c(L-\Lambda)^n \|A_\lambda^* f\|_{L^2(X)} \le \int_X \|N_\lambda^* \phi_{L,x}\|_{l^2}\, d\mu \tag{1.6.20}$$

using the fact that the transformations T^l are measure preserving. Also

$$\int_X \|\phi_{L,x}\|_{l^2}^2\, d\mu = c_n L^n \|f\|_{L^2(x)}^2. \tag{1.6.21}$$

Then letting $\Lambda \to \infty$, it follows that the $L^2(X) \to L^2(X)$ norm of the maximal operator A_* is majorized by the $l^2 \to l^2$ norm of the discrete maximal operator N_*. Then it is enough to prove the pointwise ergodic theorem for a dense subset of $L^2(X)$, p.e. for $L^\infty(X)$.

ii) Following [15], one reduces pointwise convergence to L^2 bounds for "truncated" maximal operators. Suppose indirect, that

$$\mu\{x:\ \limsup |A_\lambda f(x)| > 0\} > 0$$

then the same is true with a small constant $\alpha > 0$ inserted:

$$\mu\{x:\ \limsup |A_\lambda f(x)| > 2\alpha\} > 2\alpha$$

and using the definition of the upper limit it is easy to see, that to each λ_k if λ_{k+1} is chosen large enough then

$$\mu\{x:\ A_k^* f(x) = sup_{\lambda_k\le\lambda\le\lambda_{k+1}} |A_\lambda f(x)| > \alpha\} > \alpha$$

which implies $\|A_*^k f\|_2^2 > \alpha^3,\ \forall\ k$. Lets fix such a sequence λ_k which is quickly increasing: $\lambda_{k+1} > 4{\lambda_k}^{4d}$. Then it is enough to prove

$$\frac{1}{K} \sum_{k\le K} \|A_*^k f\|_2^2 < \alpha^3, \tag{1.6.22}$$

for $K > K_(\alpha)$. This means that the Cesaro averages converges in (6.4) tends to 0 (the terms themselves may not converge to 0).
Now fix K and choose $L > \lambda_{K+1}$. The reasoning in i) leads to

$$c(L-\Lambda)^n \frac{1}{K} \sum_{k\le K} \|A_k^* f\|_2 \le \int_X \frac{1}{K} \sum_{k\le K} \|N_k^* \phi_{L,x}\|_{l^2}\, d\mu, \tag{1.6.23}$$

where N_*^k is defined analogously to A_*^k. Thus it is enough to prove

$$\int_X \Big(\frac{1}{K} \sum_{k\le K} \|N_k^* \phi_{L,x}\|_{l^2}^2\Big)\, d\mu \le c_n \alpha^3 L^n \|f\|_2^2 \tag{1.6.24}$$

for $K > K(\alpha)$ and $L > L(K,\alpha)$. By (1.6.21), inequality (6.6) would follow, if the same would be true pointwise, that is $1/K \sum_{k\le K} \|N_k^* \phi_{L,x}\|_{l^2}^2 \to 0$ for every x, however this seems to be true just in average, and has to do with the fact that nearby averages cannot be compared.

i3) We use the approximations to N_λ introduced in Section 2., and the transfer principle (1.6.16) to reduce the estimates to that of $L^2 \to\ L^2$ norms of the corresponding maximal operators acting on $\mathbb{R}^n$.

We often use the following notations; if $\gamma_\lambda(\xi)$ are continuous functions on Π^n, then denote by Γ_λ the corresponding multipliers and by Γ_k^* the maximal operator: $\Gamma_k^*\phi = \sup_{\lambda_k\leq\lambda<\lambda_{k+1}} |\Gamma_\lambda\phi|$.
Since

$$\lambda^{-n/d+1}\hat{\sigma}_\lambda(\xi) = \sum_{q=1}^{\infty} \lambda^{-n/d+1} m_{q,\lambda}(\xi) + \lambda^{-n/d+1}\mathcal{E}_\lambda(\xi)$$

then by estimates (1.4.7) and (1.6.13)

$$\|\mathcal{E}_k^*\|_{l^2\to l^2} \leq C_\delta \lambda_k^{-\delta} \tag{1.6.25}$$

and

$$\| \sum_{q\geq q_\alpha} M_{q,k}^* \|_{l^2\to l^2} \leq C q_\alpha^{-\epsilon}. \tag{1.6.26}$$

If we apply (1.6.25) and (1.6.26) to the function $\phi_{L,x}$ integrate the square over X and average for $k \leq K$, the total contribution to the L^2 norm is less then:

$$(q_\alpha^{-\epsilon} + c_\delta K^{-1}) \int_X \|\phi_{L,x}\|_{l^2}^2 \, d\mu(x) \leq \alpha^3 L^n \|f\|_{L^2(X)}$$

by choosing K and q_α large enough w.r.t. α and ϵ.

Thus enough to deal with the finitely many maximal operators attached to the functions $m_\lambda^{a/q}(\xi)$, for $q \leq q_\alpha$ and $a \leq q,\ (a,q) = 1$. Then we can fix a and q, and write

$$\begin{aligned} \lambda^{-n/d+1} m_\lambda^{a/q}(\xi) &= \sum_{l\in\mathbf{Z}^n} G(a,l,q)\psi(q\xi - l) d\tilde{\sigma}(\lambda^{1/d}(\xi - l/q)) \\ &= \sum_{s\in\mathbf{Z}^n/q\mathbf{Z}^n} G(a,s,q)\, \psi(q\xi - s) d\tilde{\sigma}_Q(\lambda^{1/d}(\xi - s/q))_{per}, \end{aligned} \tag{1.6.27}$$

where $\gamma_{per}(\xi) = \sum_{l_1\in\mathbf{Z}^n} \gamma(\xi - l_1)$ denotes the periodization of γ. Indeed write $l = ql_1 + s$ and use the fact that $G(a,l,q) = G(a,s,q)$. Again we can fix s (there are at most $q^n \leq q_\alpha^n$ choice for each q). We remark that for $\phi \in l^2$ and $\phi_{s/q}(m) = e^{-2\pi i m s/q}\phi(m)$ i.e. $\hat{\phi}_{s/q}(\xi) = \hat{\phi}(\xi + s/q)$, one has

$$M_{s/q,k}^*\phi = M_k^*\phi_{s/q}$$

where $M_{s/q,k}^*$ is the maximal operator corresponding to the function $\psi(q\xi - s)d\tilde{\sigma}(\lambda^{1/d}(\xi - s/q))_{per}$, while M_k^* corresponds to $\psi(q\xi)d\tilde{\sigma}(\lambda^{1/d}(\xi))_{per}$. Indeed one changes variables $(\xi - s/q) \to \xi$ in evaluating the multipliers (the factors $e^{2\pi i m s/q}$ vanish when taking absolute values).

We are in a position to apply the continuous spherical maximal theorem, and further decompose the functions $\psi(q\xi)d\tilde{\sigma}(\lambda^{1/d}(\xi))$ to get decay estimates. Let $\mathbf{1} = \omega_{k,0} + \omega_{k,1} + \omega_{k,2}$ be smooth partition of unity on $\|\xi\| = \sup_j |\xi|_j \leq 1/2$ such that
$\omega_{k,0}(\xi) = 0$ unless $\|\xi\| \geq \frac{1}{2}\lambda_{k+1}^{-2}$,
$\omega_{k,1}(\xi) = 0$ unless $\frac{1}{2}\lambda_{k+1}^{-2}\|\xi\| \leq \lambda_k^{-\frac{1}{2d}}$ and
$\omega_{k,2}(\xi) = 0$ unless $\lambda_k^{-\frac{1}{2d}} \leq \|\xi\|$

Accordingly we have the decomposition: $M^*_k \leq M^*_{k,0}+M^*_{k,1}+M^*_{k,2}$ and estimate each term separately. For fixed λ, using the fact that $|d\tilde{\sigma}(\lambda^{1/d}\xi)-c_Q| \leq \lambda^{1/d}|\xi|$ ($c_Q = d\tilde{\sigma}(0)$), one has

$$|\omega_{k,0}(\xi)\psi(q\xi)d\tilde{\sigma}(\lambda^{1/d}\xi) - c_Q\omega_{k,0}(\xi)\psi(q\xi)| \leq C\lambda^{1/d}\lambda_{k+1}^{-2}. \tag{1.6.28}$$

Thus by the standard square function estimate the $l^2 \to l^2$ norm of the maximal operator (taking the sup over $\lambda_k \leq \lambda < \lambda_{k+1}$) corresponding to the functions in (1.6.27) is bounded by:

$$\left(\sum_{\lambda<\lambda_{k+1}} \lambda^{2/d}\lambda_{k+1}^{-4}\right)^{1/2} \leq \lambda_{k+1}^{-1}.$$

To estimate the maximal operator $M^*_{k,1}$ corresponding to the functions $\omega_{k,1}(\xi)\psi(q\xi)d\tilde{\sigma}(\lambda^{1/d}(\xi))_{per}$, we first use the transfer principle to see that it is bounded by the $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ norm of the maximal operator corresponding to the functions $\omega_{k,1}(\xi)\psi(q\xi)d\tilde{\sigma}(\lambda^{1/d}(\xi))$. Notice that the maximal operator (the sup taken over all $\lambda > 0$) corresponding to the functions $d\tilde{\sigma}(\lambda^{1/d}(\xi))$ is bounded from $L^2 \to L^2$ by Theorem 1.5.

Thus for $\phi_{s/q} = \phi_{L,x,s/q}$ one has

$$\|M^*_{k,1}\phi_{s/q}\|_{l^2} \leq C_Q \int_{\Pi^n} |\omega_{k,1}(\xi)|^2|\hat{\phi}(\xi+s/q)|^2\,d\xi. \tag{1.6.29}$$

The point is that since the sequence λ_k is quickly increasing $\lambda_{k+1} > 4\lambda^{4d}$ each point can belong to at most 3 intervals I_k on which $\omega_{k,1}$ supported. Hence averaging over $k \leq K$ the right side of (1.6.28), gives a contribution of $3/K\|\phi\|^2_{l^2}$.
Finally, the family of functions $\omega_{k,2}(\xi)\psi(q\xi)d\tilde{\sigma}(\lambda^{1/d}(\xi))$ satisfy the conditions of Lemma 7. Then (1.16) and (5.4) imply the bound

$$\|M^*_{k,2}\phi_{s/q}\|_{l^2} \leq C_Q\lambda_k^{-\frac{1}{2d}}\|\phi\|_{l^2}. \tag{1.6.30}$$

Note that (1.6.27)-(1.6.29) mean, that the maximal function

$$\frac{1}{K}\sum_{k\leq K}\|M^*_k\phi_{s/q}\|^2_{l^2} \leq C\int_{\Pi^n}|\psi(q\xi)\omega_{k,1}(\xi)|^2|\hat{\phi}(\xi+s/q)|^2\,d\xi + O(K^{-1})\|\phi_{s/q}\|_{l^2}.$$

i4) It is enough to prove now for fixed $r = s/q$, that

$$L^{-n}\int_X\int_{\Pi^n}\omega_{k,1}(\xi)|\hat{\phi}(\xi+s/q)|^2\,d\xi\,d\mu(x)) < |\alpha|^3\|f\|_2^2, \tag{1.6.31}$$

if $k > k(\alpha)$ and $L > L(k,\alpha)$, where we wrote $\omega_k(\xi) = |\omega_{k,1}(\xi)|^2$ for simplicity of notation.
By applying Plancherel for the inner integral in (1.6.33), one obtains

$$\begin{aligned} &L^{-n}\int_X\sum_{m,m'}\phi_{L,x}(m)\bar{\phi_{L,x}}(m')\hat{\omega}_k(m-m')e^{2\pi i(m-m')s/q}\,d\xi\,d\mu(x) \qquad (1.6.32)\\ &= L^{-n}\sum_{\|m\|\leq L,\,\|m'\|\leq L}\langle T^{m-m'}f,f\rangle\hat{\omega}_k(m-m')e^{2\pi i(m-m')s/q}\\ &= L^{-n}\int_{\Pi^n}\sum_{\|m\|\leq L,\,\|m'\|\leq L}\hat{\omega}_k(m-m')e^{2\pi i(m-m')(\theta+s/q)}\,d\nu_f(\theta)\\ &= L^{-n}\int_{\Pi^n}\sum_{l\in\mathbf{Z}^n}a_L(l)\hat{\omega}_k(l)e^{2\pi i(\theta+s/q)}\,d\nu_f(\theta), \end{aligned}$$

by the spectral theorem, where $a_L(l) = |\{(m,m');\ \|m\| \leq L,\ \|m'\| \leq L,\ m-m'=l\}|$. Finally, one has

$$\int_{\Pi^n} (L^{-n}\hat{a}_L * \omega_k)\,(\theta + s/q)\,d\nu_f, \tag{1.6.33}$$

where $*$ denotes the convolution on Π^n with respect to the Lebesgue measure. Note that

$$L^{-n}\hat{a}_L(\theta) = L^{-n}|\sum_{m=-L}^{L} e^{2\pi i m\theta}|^{2n} \leq L^n \min(1, \frac{1}{L\{\theta\}})^{2n}.$$

This means that $L^{-n}\hat{a}_L$ is a δ-sequence (i.e. weakly converges to a Dirac delta) as $L \to \infty$. Indeed it is easy to see that: $L^{-n}\hat{a}_L * \omega_k \leq c\omega_k + \epsilon$ for every $\epsilon > 0$ if L is large enough w.r.t. to λ_k and ϵ.

Finally, if we substitute this estimate into (1.6.33), then using the fact that $\omega_k(\theta) = 0$ unless $\|\theta\| \leq \lambda_k^{-1/2d}$, one has

$$\begin{aligned}\int_{\Pi^n} (L^{-n}\hat{a}_L * \omega_k)\,(\theta + s/q)\,d\nu_f \ &\leq c d\nu_f\{\theta :\ \|\theta + s/q\| < \lambda_k^{-1/2d}\} \\ &+ \epsilon\, d\nu_f(\Pi^n) \leq \alpha^3 \|f\|^2_{L^2(X)},\end{aligned}$$

if k is large enough w.r.t. α and L is large enough w.r.t. k and α.

Indeed $d\nu_f(\Pi^n) = \|f\|^2_{L^2(X)}$, and only here we use the condition strong ergodicity, that is the condition that $d\nu_f\{s/q\} = 0$ for every rational point $s/q \neq 0$, note that by our assumption $d\nu_f\{0\} = \int_X f d\mu = 0$ as well. This implies that $d\nu_f\{\theta :\ \|\theta + s/q\| < \lambda_k^{-1/2d}\} \to 0$ as $k \to \infty$.

This finishes the proof of our main result, Theorem 1.4. $\Box$.

2 Quadratic systems and simplices in sets positive density of $\mathbb{Z}^d$

In this chapter we present some results on the existence of geometric point configurations in sets of positive density of the integer lattice $\mathbb{Z}^d$.
Such patterns are determined by quadratic systems of equations hence number theoretic methods developed to count integer solutions of such systems will play an important role. In particular, using the theory if Siegel theta functions referred also to as "Siegel's generalised circle method", we prove a discrete analogue of Bourgain's simplex theorem, with the underlying Euclidean space replaced by the integer lattice.

We then extend our results to more complex patterns such as k-dimensional boxes or direct products of simplices, by utilizing some constructs and techniques from additive combinatorics and hypergraph theory. In particular we will develop and make use if the so-called Gowers box-norms[48] and a weak hypergraph regularity lemma [48, 108] in the context both Euclidean spaces and the integer lattice.

2.1 A conjecture of Graham in geometric Ramsey theory

Geometric (or Euclidean) Ramsey theory was pioneered in a series of papers [34, 35, 50] by Erdős at al. They define a finite point configuration $X \subseteq \mathbb{R}^k$ to be *Ramsey* if for every number of colors r, there is a large enough dimension $d = d(r, X)$ such that every r-coloring of R^d contains a monochromatic, congruent (isometric) copy of X. The fundamental problem in the area is to classify which point configurations are Ramsey.

While this problem has been studied via combinatorial means [41, 69, 70], a density of analogue was proved by Bourgain [16] using Fourier analysis. He has shown that if $X = \{x_1, \ldots, x_k\}$ is non-degenerate simplex, then any set $A \subseteq \mathbb{R}^k$ of positive upper density contains a congruent copy of *all* of its sufficiently large dilates. Again a basic problem is to classify which finite point configurations X, besides simplices, have this property to which we will refer to as being *density Ramsey.*

For $k = 2$ Bourgain's theorem says that the distance set $D(A) := \{|x_1 - x_2| : \; x_1, x_2 \in A\}$ of a set $A \subseteq \mathbb{R}^k$ $(k \geq 2)$ of positive upper density contains all large numbers $\lambda \geq \lambda(A)$, which is already a highly non-trivial result first shown by Katznelson and Weiss [45] via ergodic means. The threshold $\lambda(A)$ must depend on the set A not just on its density as can be seen by taking the set $A = 2\lambda \cdot \mathbb{Z}^k + B_{\lambda/2}$ i.e. the set of point of distance less than $\lambda/2$ from the points of the grid $2\lambda \cdot \mathbb{Z}^d$. Indeed, such a set has positive density depending only on k but will not contain any two points at distance λ.

It was observed by by Erdős at al. [34] and by independently by Bourgain [16] that Ramsey sets must be spherical in both context. This has led Graham to conjecture that a finite point configuration X is Ramsey or density Ramsey if and only if it is spherical, i.e. if it can be inscribed into a sphere [50].

2.2 Simplices in sets of positive density of the integer lattice

We prove a variant of Bourgain's result for subsets of the lattice $A \subseteq \mathbb{Z}^n$ of positive density. Let us recall that a subset A of $\mathbb{Z}^n$ has upper density at least ε, and write $\delta(A) \geq \varepsilon$, if there exists a sequence of cubes B_{R_j} of sizes $R_j \to \infty$, not necessarily centered at the origin, such that for all $j \in \mathbb{N}$

$$|A \cap B_{R_j}| \geq \varepsilon\, R_j^n$$

moreover the upper density $\delta(A)$ is defined to be the supremum of all $\varepsilon > 0$ satisfying the above condition.

In so doing, we must avoid certain natural obstructions, which we describe below. Consider a simplex $\triangle = \{v_0, \ldots, v_k\} \subseteq \mathbb{R}^n$, where $k < n$. Associated to the simplex is a positive definite matrix $k \times k$ matrix $T_\triangle = (t_{ij})$, with entries

$$t_{ij} = (v_i - v_0) \cdot (v_j - v_0), \tag{2.2.1}$$

where " $\cdot$ " denotes the dot product. Nota that $T_\triangle$ is independent of the rigid motions of the simplex.

It is clear that the simplex $\triangle$ can be embedded in a set $A \subseteq \mathbb{Z}^n$ only if $T_\triangle$ has all integral entries. In this case we will call the simplex *integral.* It follows that we can consider dilates of the simplex of the form $\sqrt{\lambda}\triangle$, for positive integers λ. Let $A = (q\mathbb{Z})^n$ for some positive integer q then $\delta(A) = q^{-n}$ and $\sqrt{\lambda}\triangle \subseteq A$ only if q^2 divides λ. We can state now our main result.

Theorem 2.1. *Let $k \geq 2$, and let the dimension $n > 2k + 4$. For each $A \subseteq \mathbb{Z}^n$ with $\delta(A) = \delta > 0$, the following holds for all integral k-dimensional simplices $\triangle$.*
There is a positive integer $Q = Q(\delta)$, and a number $\Lambda = \Lambda(A, \triangle)$ so that for all integers $\lambda > \Lambda$, there is a simplex $\triangle' \subseteq A$, which is, up to a rigid motion, $\sqrt{\lambda}Q\,\triangle$.

Again, taking $A = (q\mathbb{Z})^n$ for all $1 \le q \le \delta^{-1/n}$ (so $\delta(A) \ge \delta$), we see that the factor $Q(\delta)$ must be divisible by the least common multiple of positive integers at most $\delta^{-1/n}$, thus it follows from elementary estimates on primes that $Q(\delta) \ge \exp(c\,\delta^{-1/n})$. The number $Q(\delta)$ will be constructed explicitly and will satisfy the upper bound $Q(\delta) \le \exp(C\,\delta^{-4(k+1)/n-2k-4})$.
For $k = 1$, Theorem 2.1 translates to the fact that the distance set of A, $d(A) = \{|m-l| :\ m \in A,\ l \in A\}$ contains all large distances of the form $\sqrt{\lambda}Q(\delta)$. This was proved earlier in [84] in dimensions $n > 4$.

To introduce our terminology, let us call two simplices $\triangle,\ \triangle' \subseteq \mathbb{R}^n$ *isometric*, and write $\triangle' \simeq \triangle$, if one is obtained from the other via a rigid motion, that is when $\triangle' = x + U(\triangle)$ for some $x \in \mathbb{R}^n$ and $U \in SO(n)$. It is clear that " $\simeq$ " is an equivalence relation, we call the equivalence classes $k+1$-*point configurations*. Thus a k-*point configuration* preserves only the geometry, or the "shape", of k points. Theorem 2.1 says, roughly, that a set $A \subseteq \mathbb{Z}^n$ of positive upper density contains all large dilates of any given configuration of k points in general position, satisfying the above mentioned natural conditions.

We emphasize that the above result is proved only under the assumption that the simplex $\triangle = \{0, v_1, \ldots, v_k\}$ is non-degenerate. A counter-example is shown in [16] in the continuous case, when $n = k = 2$, $\triangle = \{0, e_1, 2e_1\}$. In our settings when $\triangle = \{0, e_1, 2e_1, \ldots, ke_1\}$, the existence of an embedding of $\triangle$ in A follows from Szemerédi's theorem on arithmetic progressions [104], however it is not true that all large dilates of $\triangle$ can be embedded in A in the sense of Theorem 2.1

We will turn now to some quantitative results. These will depend on the *eccentricity* $e(T)$ (with $T = T_\triangle$) of the simplex $\triangle$, defined by

$$e(T) = \frac{|T|}{\mu(T)}, \quad \text{where} \quad \mu(T) = \inf_{|x|=1} Tx \cdot x, \quad |T| = (\sum_{i,j=1}^{k} |t_{ij}|^2\ )^{\frac{1}{2}}. \tag{2.2.2}$$

Note that $|T|^{1/2}$ is comparable to the diameter of $\triangle$, and the quantity $e(T)$ may be viewed as a measure of how close the simplex $\triangle$ is to being degenerate.

Theorem 2.2. *Let $k \ge 2,\ n > 2k+4,\ \varepsilon > 0$. Let $A \subseteq \mathbb{Z}^n \cap B_R$ such that $|A| \ge \varepsilon R^n$, and let $\triangle \subset \mathbb{R}^n$ be a k-dimensional integral simplex and let $T = T_\triangle$. If*

$$R \ \ge\ C_1\,|T|^{\frac{1}{2}}\ \exp\Big(C_2\,\varepsilon^{-\frac{11}{2}\,(k+1)}\ \log\left(e(T)\right)\Big), \tag{2.2.3}$$

for some positive constants C_1 and C_2 depending only on the dimensions n and k, then there exists a simplex $\triangle' \subseteq A$ and a $\lambda \in \mathbb{N}$ such that $\triangle' \simeq \sqrt{\lambda} \cdot \triangle$.

In other words, if $A \subseteq B_R \cap \mathbb{Z}^n$ contains an ε-portion of the points in the cube B_R and if R is large enough, then the set A contains a "copy" of the simplex $\triangle$, obtained by a translation, a rotation and a dilation. Both of the above theorems are consequences of the following.

Theorem 2.3. *Let $k \ge 2,\ n > 2k+4,\ \varepsilon > 0$. Let $A \subseteq \mathbb{Z}^n \cap B_R$ such that $|A| \ge \varepsilon R^n$, and let $\triangle \subset \mathbb{R}^n$ be a k-dimensional integral simplex and let $T = T_\triangle$.*
Then there exists a pair of integers $Q = Q(\varepsilon)$, $J = J(\varepsilon)$ such that for any sequence of integers $\mathcal{C}_0 \le \lambda_1 < \lambda_2 < \ldots < \lambda_{J(\varepsilon)}$, satisfying

$$\lambda_{j+1} > 2\,e(T)\,\lambda_j, \qquad \text{and} \qquad \lambda_{J(\varepsilon)}^{\frac{1}{2}}\,|T|^{\frac{1}{2}} \le R$$

there exists a simplex $\triangle' \subseteq A$ *such that* $\triangle' \simeq \sqrt{\lambda_j} Q \triangle$ *for some* $1 \leq j \leq J(\varepsilon)$.
Moreover the numbers $Q(\varepsilon)$, $J(\varepsilon)$ *satisfy the inequalities*

$$Q(\varepsilon) \leq \exp\,(C\,\varepsilon^{-\frac{4(k+1)}{n-2k-4}}), \qquad J(\varepsilon) \leq C\,\varepsilon^{-\frac{11}{2}\,(k+1)}, \tag{2.2.4}$$

for some positive constant C *depending only on the dimensions* n *and* k.

It is not hard to see that, for $k < n$, $\triangle' \simeq \triangle$, if and only if $T_\triangle = T_{\triangle'}$. Indeed, if $\triangle' = \{0, v_1', \dots, v_k'\}$ and $\triangle = \{0, v_1, \dots, v_k\}$ then there is a rotation U_0 which takes v_1 to v_1', hence assume that $v_1 = v_1'$. If P stands for the projection to the orthogonal complement of v_1, then it is easy to see that $T_{\bar{\triangle}} = T_{\bar{\triangle}'}$ where $\bar{\triangle} = P(\{v_2, \dots, v_k\})$ and $\bar{\triangle}' = P(\{v_2', \dots, v_k'\})$. Thus, by induction, there is a rotation $U \in SO(n)$ such that $U(P(v_i)) = U(P(v_i'))$ for $i \geq 2$, and $U(v_1) = U(v_1') = v_1 = v_1'$, hence $U(\triangle) = \triangle'$. Thus $k+1-$point configurations are in one to one correspondence with positive definite $k \times k$ matrices.

We emphasize that the above results are proved only under the assumption that the simplex $\triangle = \{0, v_1, \dots, v_k\}$ is non-degenerate, that is the vectors $v_1, \dots, v_k$ are linearly independent in $\mathbb{R}^n$. A counter-example is shown in [16] in the continuous case when $n = k = 2$, $\triangle = \{0, e_1, 2e_1\}$, which can adapted to our settings as follows. If $\triangle' \simeq \sqrt{\lambda}\triangle$, then $\triangle' = \{x - y, x, x + y\}$ with $|y|^2 = \lambda$. Let α be an irrational number and let

$$A := \{x \in \mathbb{Z}^n :\ \|\alpha|x|^2\| \leq \frac{1}{100}\}, \quad \textit{where} \quad \|\beta\| := \min_{m \in \mathbb{Z}} |\beta - m|.$$

It is easy to see that A has positive upper density. However, if $\triangle' = \{x - y, x, x + y\} \subseteq A$ then by the parallelogram identity $2\lambda = 2|y|^2 = |x - y|^2 + |x + y|^2 - 2|x|^2$ we have that $\|\lambda^2 2\alpha\| \leq \frac{1}{10}$. For any $Q \in \mathbb{N}$ $Q\lambda^2 2\alpha$ is uniformly distributed $\pmod 1$ as λ^2 runs through the positive integers thus $\|Q\lambda^2 2\alpha\| \leq \frac{1}{10}$ cannot hold for all sufficiently large $\lambda > \Lambda$. The above construction can extended to any non-degenerate simplex in fact to any non-spherical configuration, see section 2.4.

The problem of embedding $\sqrt{\lambda}\triangle$ into $\mathbb{Z}^n$ is equivalent of finding integer solutions $x_1, \dots, x_k \in \mathbb{Z}^n$ of the quadratic system of equations $x_i \cdot x_j = \lambda t_{ij}$ for $1 \leq i \leq j \leq k$, which is further equivalent of representing the quadratic form $\lambda T y \cdot y = l_1(y)^2 + \ldots + l_n(y)^2$ i.e. as a sum of squares of n integral linear forms. This problem has been extensively studied [98, 95, 67] with the strongest results due to Kitaoka [67] whi obtained an asymptotic formula for the number of representations in dimensions $n > 2K + 2$.

We remark that the existence of a dilate $\lambda\triangle$ which can be embedded in A follows from Kitaoka's theorem [67] together with the so-called multi-dimensional Szemerédi theorem [39, 48, 108], which implies that for every finite set $S \subseteq \mathbb{Z}^n$ there is an $m \in \mathbb{Z}^n$ and $\lambda \in \mathbb{N}$ such that $S' = m + \lambda\, S \subseteq A$. However at present, the multi-dimensional Szemerédi theorem has no Fourier analytic proof, quantitative versions with reasonable bounds, while Theorem 2.2 provides a single-exponential quantitative bound. Also, the emphasis in Theorem 2.1 is in the fact that, in a sense, all large dilates of $\triangle$ can be embedded in A which is not possible to obtain via this route.

2.2.1 Outline of the proofs of the main results.

Let us start by observing that Theorem 2.3 implies both Theorem 2.1 and Theorem 2.2. Indeed assuming that the conclusion of Theorem 2.1 is not true, it follows that there is a set $A \subseteq \mathbb{Z}^n$

with upper density $\delta(A) \geq \varepsilon$, and an infinite lacunary sequence λ_j such that $\sqrt{\lambda_j} Q(\varepsilon)\triangle$ cannot be embedded in A for all $j \in \mathbb{N}$. Choosing a cube B_R of size $R \geq C\,(\lambda_{J(\varepsilon)}|T|)^{1/2}$ such that $|A \cap B_R| \geq \varepsilon R^n$ contradicts Theorem 2.3. Also, choosing $Q(\varepsilon)$ and $J(\varepsilon)$ is in Theorem 2.3, and a lacunary sequence $\lambda_1 < \ldots < \lambda_{J(\varepsilon)}$ such that $\lambda_{J(\varepsilon)} \leq \exp\left(J(\varepsilon)\log\left(e(T)\right)\right)$, it follows from (2.2.4) that $\sqrt{\lambda}\triangle$ can be embedded in A for some $\lambda = \lambda_j Q(\varepsilon)^2$ as long as $A \subseteq \mathbb{Z}^n \cap B_R$ with $|A| \geq \varepsilon R^n$ and R satisfies (2.2.6) thus Theorem 2.2 follows.

Let us outline now, the proof of Theorem 2.3. We'll use a variant of the density increment approach of Roth. In our settings this amounts to showing that the set A contains an isometric copy of $\sqrt{\lambda}\triangle$ for some $\lambda \in \mathbb{N}$, or the density of A increases on a large cubic grid by a fixed amount $c(\varepsilon) > 0$, depending only on ε. We'll prove a somewhat stronger statement; namely if for a *fixed* λ the simplex $\sqrt{\lambda}\triangle$ cannot be embedded in A, then either the density of A increases to $(1+c)\varepsilon$ on a large grid of common difference $q = q(\varepsilon)$, or the Fourier transform $\hat{\mathbf{1}}_A$, $\mathbf{1}_A$ being the indicator function of the set A, is concentrated on a small set $\mathbb{T}_{\lambda,q}$. Moreover if $\lambda' \gg \lambda$, then the sets $\mathbb{T}_{\lambda',q}$ and $\mathbb{T}_{\lambda,q}$ are disjoint, thus if $\lambda_1 < \lambda_2 < \ldots < \lambda_J$ is a lacunary sequence with $J \geq J(\varepsilon)$ is large enough and if A does not contain an isometric copy of any simplex $\sqrt{\lambda_j}\triangle$, then A must have increased density on a large grid of difference $q = q(\varepsilon)$. Iterating this, will prove Theorem 2.3.

To formulate the above statements precisely, let us introduce some notations. We'll denote by $c > 0$ resp. $C > 0$, small resp. large constants depending only on the dimensions n and k, whose value can change from place to place. If they depend on other parameters like ε, δ and so on, we indicate those in parenthesis $c(\varepsilon), c(\varepsilon,\delta)$. The least common multiple of a set of integers $q_1,\ldots,q_l$ will be denoted by $lcm\,\{q_1,\ldots,q_l\}$. To a given $0 < \varepsilon \leq 1$ we attach the integer

$$q(\varepsilon) = lcm\,\{1 \leq q \leq C\varepsilon^{-\frac{4(k+1}{n-2k-4}}\}. \tag{2.2.5}$$

The importance of this number is in the fact that the grid $(\frac{1}{q(\varepsilon)}\mathbb{Z})^n = \{\frac{m}{q(\varepsilon)};\ m \in \mathbb{Z}^n\}$ contains all rational points $a/q \in \mathbb{R}^n$ with denominator $q \leq C\varepsilon^{-\frac{4(k+1}{n-2k-4}}$. For given $s \in \mathbb{Z}^n$, $q \in \mathbb{N}$ and $L > q$ we define the cubic grid of size L and common difference q

$$B_L(q,s) = (s + (q\mathbb{Z})^n) \cap B_L, \tag{2.2.6}$$

where B_L is a cube of size L. In the Fourier space $\mathbb{T}^n = (\mathbb{R}/\mathbb{Z})^n$, a key role will be played by the sets

$$\mathbb{T}_{(L_1,L_2,q)} = \left(\frac{1}{q}\mathbb{Z}\right)^n + D_{L_1,L_2} \quad \text{where} \quad D_{L_1,L_2} = \left[-\frac{1}{2L_1},\frac{1}{2L_1}\right]^n \setminus \left[-\frac{1}{2L_2},\frac{1}{2L_2}\right]^n, \tag{2.2.7}$$

where $q \in \mathbb{N}$ and $q < L_1 < L_2$. Here by $S + T$ we denote the sumset of the sets S and T. The key is to obtain the following

Lemma 2.1. *Let $n > 2k+4$, $0 < \varepsilon \leq 1$, let $A \subseteq B_R \cap \mathbb{Z}^n$ such that $|A| \geq \varepsilon R^n$, and let $\triangle$ be an integral k-dimensional simplex.*

If for a given $\lambda \in \mathbb{N}$ the simplex $\sqrt{\lambda}\triangle$ cannot be embedded in A, then either there exists a cubic grid $B_L(q,s)$ with $q = q(\varepsilon)$ defined in (2.1), and $L \geq C\sqrt{\lambda}|\triangle|$, such that

$$(i) \qquad |A \cap B_L(q,s)| \geq (1+\alpha)\varepsilon\,|B_L(q,s)| \quad \textit{with} \quad \alpha = \frac{1}{10(k+1)}, \quad \textit{or} \tag{2.2.8}$$

$$(ii) \qquad \int_{\mathbb{T}_{\lambda,\varepsilon}} |\hat{1}_A(\xi)|^2\,d\xi \geq c\,\varepsilon^{2k+2}R^n, \tag{2.2.9}$$

where $\mathbb{T}_{\lambda,\varepsilon} = \mathbb{T}_{(L_1(\lambda,\varepsilon),L_2(\lambda,\varepsilon),q(\varepsilon))}$ *is the set defined in* (2.2.7) *and*

$$L_1(\lambda,\varepsilon) = C^{-1}\, e(T)^{-4}\varepsilon^{9(k+1)}(\lambda|T|)^{1/2}, \qquad L_2(\lambda,\varepsilon) = C\,\varepsilon^{-(k+1)}\,(\lambda|T|)^{1/2}, \tag{2.2.10}$$

as long as the parameters λ *and* R *satisfy* $q(\varepsilon) < L_1(\lambda,\varepsilon) < L_2(\lambda,\varepsilon) < R.$

We now describe how repeated applications of Lemma 2.1 implies Theorem 1.3. main result.

Proof of Theorem 2.3 For $r = 0, 1, 2, \ldots$ define,

$$\varepsilon_r = (1+\alpha)^{-r} \quad \text{with} \quad \alpha = \frac{1}{10(k+1)}, \tag{2.2.11}$$

moreover let $q_r = q(\varepsilon_r)$ given in (2.2.5) and $Q_r = q_1 q_2 \ldots q_r$, we set $Q_0 = 1$. We define the numbers J_r inductively with $J_0 = 1$ and J_r being the smallest positive integer satisfying

$$J_r \geq \gamma J_{r-1} + \bar{C}\,{\varepsilon_r}^{-4(k+1)}\,\log\left({\varepsilon_r}^{-1}\right) \quad \text{with} \quad \gamma = e^{1/2}. \tag{2.2.12}$$

We will show by induction on r, that Theorem 1.3 holds for $\varepsilon_{r-1} > \varepsilon \geq \varepsilon_r$. This amounts to showing that if $A \subseteq B_R \cap \mathbb{Z}^n$ with $|A| \geq \varepsilon_r R^n$, and if $C < \lambda_1 < \ldots < \lambda_{J_r}$ is a given lacunary sequence with $\lambda_{i+1} > 2e(T)\lambda_i$, then A contains an isometric copy of a simplex $\sqrt{\lambda_i} Q_r \triangle$ for some $1 \leq i \leq J_r$. For $r = 0$, $\varepsilon = \varepsilon_0 = 1$ thus $A = B_R \cap \mathbb{Z}^n$ and Theorem 2.3 follows from Kiatoke's theorem (with $Q_0 = J_0 = 1$), as explained in the introduction.

Now, assume indirectly that there exists an $r \in \mathbb{N}$, such that the conclusion of Theorem 1.3 holds for the triple $\varepsilon_{r-1}, Q_{r-1}, J_{r-1}$, but not for ε_r, Q_r, J_r. Then none of the simplices $\sqrt{\lambda_i} Q_r \triangle$ can be embedded in A. Since $J_r \geq \bar{C}\varepsilon_r^{-4(k+1)} \log(\varepsilon_r^{-1})$, one may choose a subsequence $\{\mu_1, \ldots, \mu_t\}$ of the sequence $\{\lambda_j;\ J_r/\gamma \leq j \leq J_r\}$ such that $t > (c\,\varepsilon^{2k+2})^{-1}$ and for all $1 \leq i \leq t$ one has $L_1(\mu_{i+1}, \varepsilon_r) > L_2(\mu_i, \varepsilon_r)$, as long as the constant $\bar{C}$ is chosen large enough with respect to c and C given in (2.2.9) and in (2.2.10). It follows that the sets $\mathbb{T}_{\lambda,\varepsilon}$ for $\lambda = \mu_i Q_r^2$ are disjoint, and thus inequality (2.2.9) cannot hold simultaneously for all $1 \leq i \leq t$ as it would imply that: $|A| = \int_{\mathbb{T}^n} |\hat{\mathbf{1}}_A(\xi)|^2\, d\xi > R^n$. By Lemma 2.1 there must exist a positive integer $\lambda = \mu_i Q_r^2 = \lambda_j Q_r^2$ with $J_r/\gamma \leq j \leq J_r$, such that

$$|A \cap B_L(q,s)| \geq (1+\alpha)\varepsilon_r\,|B_L(q,s)| = \varepsilon_{r-1}\,|B_L(q,s)|, \tag{2.2.13}$$

for a grid $B_L(q_r, s)$ of size $L > C(\lambda|T|)^{1/2}$. The affine map $\Phi(m) = q_r m + s$ identifies the set $B_L(q_r, s)$ with $B_{R'} \cap \mathbb{Z}^n$ ($R' = L/q_r$) and also $A \cap B_L(q_r, s)$ with a set $A' \subseteq B_{R'} \cap \mathbb{Z}^n$.

By (2.2.13) one has that $|A'| \geq \varepsilon_{r-1}(R')^n$ and one may apply the induction hypothesis for the set A' and the sequence $\lambda_1 < \lambda_2 < \ldots < \lambda_{J_{r-1}}$. Indeed, it is easy to check that the size of the box $B_{R'}$ satisfies

$$R' = L/q_r \geq C\,(\lambda_j|T|)^{1/2}\, Q_r/q_r \geq C\,(\lambda_{J_{r-1}}|T|)^{1/2}\, Q_{r-1},$$

as $j \geq J_r/\gamma \gg J_{r-1}$ It follows that A' contains a simplex $\triangle'$ isometric to $\sqrt{\lambda_i} Q_{r-1} \triangle$ for some $1 \leq i \leq J_{r-1}$, hence A contains the simplex $\Phi(\triangle') = s + q_r \triangle'$ which is isometric to $\sqrt{\lambda_i} Q_{r-1} q_r \triangle = \sqrt{\lambda_i} Q_r \triangle$.

To finish the proof one only needs to check that $J(\varepsilon)$ and $Q(\varepsilon)$ satisfy the quantitative bounds (2.2.4). If $\varepsilon_r \leq \varepsilon < \varepsilon_{r-1}$, then $Q(\varepsilon) = Q_r = \prod_{l=1}^{r} q_l$ where $q_l \leq \exp\left(C\,\varepsilon_l^{-4(k+1)/\,(n-2k-4)}\right)$ by well-known estimates on the primes. Thus, also $Q(\varepsilon) \le \exp(\bar{C}\,\varepsilon^{-4(k+1)/(n-2k-4)})$ for a slightly larger constant $\bar{C}$. To estimate $J(\varepsilon) = J_r$ where $\varepsilon_r \le \varepsilon < (1+\alpha)\varepsilon_r$, note that dividing (2.8) by γ^r one obtains

$$\frac{J_r}{\gamma^r} - \frac{J_{r-1}}{\gamma^{r-1}} \le C\,\frac{\varepsilon_r^{-4(k+1)}\ \log(\varepsilon_r^{-1})+1}{\gamma^{r-1}}. \tag{2.2.14}$$

Since

$$\varepsilon_r^{-4(k+1)} = \left(1+\frac{1}{10(k+1)}\right)^{4(k+1)r} \le e^{4r/10}$$

it follows that the sum in (2.10) converges in r and hence $J_r \le C\,\gamma^r = C\,e^{r/2}$. Also $\log(1+\alpha) = \log(1+\frac{1}{10(k+1)}) \ge \frac{1}{11(k+1)}$, thus

$$J(\varepsilon) = J_r \le C\,\gamma^r = \varepsilon_r^{-\frac{1}{2\log(1+\alpha)}} \le C\varepsilon_r^{-\frac{11}{2}(k+1)} \le C'\,\varepsilon^{-\frac{11}{2}(k+1)}$$

This proves estimate (2.2.4). □

It remains to prove Lemma 2.1. To do that, similarly as in case of arithmetic progressions, one introduces a multilinear form to count the number of embeddings of a given simplex $\sqrt{\lambda}\triangle$ into the set A. For a given $k\times k$ integral positive matrix $T=(t_{ij})$, let $S_T : \mathbb{Z}^{nk} \to \{0,1\}$ denote the function

$$S_T(m_1,\dots,m_k) = \begin{cases} 1 & \text{if } \ m_i\cdot m_j = t_{ij} \quad \forall\ 1\le i\le j\le k \\ 0 & \text{otherwise} \end{cases} \tag{2.2.15}$$

where $m_i \in \mathbb{Z}^n$ for $1\le i\le k$. For functions $f_i : \mathbb{Z}^n \to \mathbb{C}$, $(0\le i\le k)$ of finite support and for a given $\lambda\in\mathbb{N}$ define the corresponding form

$$N_{\lambda T}(f_0,f_1,\dots,f_k) = \sum_{m,m_1,\dots,m_k\in\mathbb{Z}^n} f_0(m)f_1(m+m_1)\dots f_k(m+m_k)\,S_{\lambda T}(m_1,\dots,m_k). \tag{2.2.16}$$

The point is that if $T = T_\triangle$, that is the inner product matrix of the simplex $\triangle$, and if $f_0 = f_1 = \dots = f_k = \mathbf{1}_A$ the indicator function of the set A, then $N_{\lambda T}(\mathbf{1}_A,\dots,\mathbf{1}_A)$ is the number of simplices $\triangle' \subseteq A$ such that $\triangle' \simeq \sqrt{\lambda}\triangle$.

Going back to Lemma 2.1, we will assume from now on that that for a given $\lambda\in\mathbb{N}$ the simplex $\sqrt{\lambda}\triangle$ cannot be embedded in A, that is

$$N_{\lambda T}(\mathbf{1}_A,\dots,\mathbf{1}_A) = 0 \tag{2.2.17}$$

and moreover that the set A is uniformly distributed on the grids $B_L(q,s)$ in the sense that

$$|A\cap B_L(q,s)| \le (1+\alpha)\varepsilon\,|B_L(q,s)| \quad \text{with} \quad \alpha = \frac{1}{10(k+1)}, \tag{2.2.18}$$

for all such grids $B_L(q,s)\subseteq B_R$, for some parameters for a given $q\in\mathbb{N}$ and $L > C\,(\lambda|T|)^{1/2}$ (later we will choose $q = q(\varepsilon)$ given in (2.2.5)). we partition $B_R\cap\mathbb{Z}^n$ into grids $B_L(q,s)$ and define the corresponding conditional expectation function $h_{L,q} : B_R\cap\mathbb{Z}^n \to [0,1]$ by

$$h_{L,q}(m) = |A\cap B_L(q,m)|/|B_L(q,m)|, \tag{2.2.19}$$

where $B_L(q,m)$ is the grid in the partition containing the point m. Note that the function $h_{L,q}$ is constant and is equal to the average of $\mathbf{1}_A$ on each grid $B_L(q,s)$ of the partition. Using assumption (2.2.18) on the distribution of A, and Kitaoka's theorem:

$$\|S_{\lambda T}\|_1 = \sum_{m_1,\dots,m_k} S_{\lambda T}(m_1,\dots,m_k) \geq c_0 \det(\lambda T)^{\frac{n-k-1}{2}} \tag{2.2.20}$$

it will be fairly easy to show that

$$N_{\lambda T}(\mathbf{1}_A, h_{L,q},\dots,h_{L,q}) \geq c \det(\lambda T)^{\frac{n-k-1}{2}} \varepsilon^{k+1} R^n. \tag{2.2.21}$$

Indeed, from (2.2.18) it is easy to see that $h_{L,q}(m) \geq c\,\varepsilon$ for all but a small number of $m \in B_R \cap \mathbb{Z}^n$.

It will be more convenient to work with functions of the form $f_{L,q} = \mathbf{1}_A * \psi_{L,q}$ which majorize $h_{L,q}$ and whose Fourier transform is easier to handle. Indeed, if $\psi > 0$ is a strictly positive Schwarz function, and if

$$\psi_{L,q}(m) = \begin{cases} q^n L^{-n}\, \psi(m/L) & \text{if } m \in (q\mathbb{Z})^n \\ 0 & \text{otherwise} \end{cases} \tag{2.2.22}$$

then $f_{L,q} \geq c\, h_{L,q}$, see Proposition 3.2. Thus we get our main estimate from below

$$N_{\lambda T}(\mathbf{1}_A, f_{L,q},\dots,f_{L,q}) \geq c_1 \det(\lambda T)^{\frac{n-k-1}{2}} \varepsilon^{k+1} R^n, \tag{2.2.23}$$

for some constant $c_1 > 0$, see Lemma 2.3 for the precise statement.

The advantage of using the functions $f_{L,q}$ is in that their Fourier transform can be described fairly precisely

$$\hat{f}_{L,q}(\xi) = \hat{\mathbf{1}}_A(\xi)\widehat{\psi}_{L,q}(\xi) = \hat{\mathbf{1}}_A(\xi) \sum_{l\in\mathbb{Z}^n} \widehat{\psi}(L(\xi - l/q)), \tag{2.2.24}$$

moreover if ψ is chosen such that

$$1 = \hat{\psi}(0) \geq \hat{\psi}(\xi) > 0 \ \ \forall\, \xi \quad \text{ and } \quad \mathit{supp}\ \hat{\psi} \subseteq [-1/2, 1/2]^n, \tag{2.2.25}$$

then $\hat{f}_{L,q}(\xi)$ is supported on the set $(\frac{1}{q}\mathbb{Z})^n + [-\frac{1}{2L}, \frac{1}{2L}]^n$ and it essentially equals to $\hat{\mathbf{1}}_A(\xi)$ on a smaller such set.
In Section 2.2.3, we prove our crucial error estimate, namely that if $q = q(\varepsilon)$ and if one chooses $L_1 = L_1(\lambda,\varepsilon)$ given in (2.2.10), with the constant C large enough with respect to c_1 appearing in (2.2.11), then

$$|N_{\lambda T}(\mathbf{1}_A, \mathbf{1}_A,\dots,\mathbf{1}_A) - N_{\lambda T}(\mathbf{1}_A, f_{L_1,q},\dots,f_{L_1,q})| \leq \frac{c_1}{2} \det(\lambda T)^{\frac{n-k-1}{2}} \varepsilon^{k+1} R^n, \tag{2.2.26}$$

see Lemma 2.4. Taking this granted for now, let us sketch the

Proof of Lemma 2.1. Using estimates (2.2.23) for $L = C\,(\lambda|T|)^{1/2}$ and (2.2.26) for $L_1 = L_1(\lambda,\varepsilon)$, it follows from our assumption (2.2.17) that

$$|N_{\lambda T}(\mathbf{1}_A, f_{L_1,q},\dots,f_{L_1,q}) - N_{\lambda T}(\mathbf{1}_A, f_{L,q},\dots,f_{L,q})| \geq \frac{c_1}{2} \det(\lambda T)^{\frac{n-k-1}{2}} \varepsilon^{k+1} R^n. \tag{2.2.27}$$

Now, it is easy to see that the left side of (2.2.27) is bounded by

$$\|S_{\lambda T}\|_1\, \|\mathbf{1}_A\|_2\, \|f_{L_1,q} - f_{L,q}\|_2 \leq C \det(\lambda T)^{\frac{n-k-1}{2}} R^{\frac{n}{2}} \|f_{L_1,q} - f_{L,q}\|_2, \tag{2.2.28}$$

see Proposition 2.3. It follows,

$$\|f_{L_1,q} - f_{L,q}\|^2 = \int_{\mathbb{T}^n} |\hat{\mathbf{1}}_A(\xi)|^2 \, |\widehat{\psi}_{L_1,q} - \widehat{\psi}_{L,q}|^2 \, d\xi \ \geq c\, \varepsilon^{2k+2} \, R^n \tag{2.2.29}$$

This implies inequality (2.2.9) as the function $|\widehat{\psi}_{L_1,q} - \widehat{\psi}_{L,q}|$ is uniformly bounded by $\bar{c}\,\varepsilon^{k+1}$ with a small constant, say $\bar{c} < c/2$, outside the set $\mathbb{T}_{\lambda,\varepsilon} = \mathbb{T}_{(L_1(\lambda,\varepsilon), L_2(\lambda,\varepsilon), q(\varepsilon))}$ given in (2.2.10), and Lemma 2.1 follows. □

The detailed proof of Lemma 2.1 will be given in Section 2.2.3. The proof the crucial estimate (2.2.26) will be based on an estimate of the Fourier transform of the function S_T at points $\mathcal{X} = (\xi_1, \dots, \xi_k)$ which are away from rational points with small denominator. Such estimates can be proved by techniques from analytic number theory, and can be viewed as discrete analogues of stationary phase estimates on the Fourier transforms of surface carried measures in Euclidean spaces [102]. It is summarized in the following lemma.

Using the matrix notation, let $M = (m_1, \dots, m_k) \in \mathbb{Z}^{n\times k}$ and $\mathcal{X} = (\xi_1, \dots, \xi_k) \in \mathbb{T}^{n\times k}$ be $n \times k$ matrices with column vectors $m_i \in \mathbb{Z}^n$ and $\xi_i \in \mathbb{T}^n$ ($\mathbb{T} = \mathbb{R}/\mathbb{Z}$), the Fourier transform of the function S_T given in (2.2.15) is defined by the exponential sum

$$\widehat{S}_T(\mathcal{X}) = \sum_{M \in \mathbb{Z}^{n\times k}} S_T(M) \; e^{-2\pi i \, tr\,(M^t \, \mathcal{X})}, \tag{2.2.30}$$

where $tr\,(M^t\,\mathcal{X}) = m_1 \cdot \xi + \ldots + m_k \cdot \xi_k$ stands for the trace of the product matrix $M^t\,\mathcal{X}$. Let $P/q = (p_{ij}/q)$ denote the dilate of a matrix $P = (p_{ij})$ by the factor of $1/q$.

Lemma 2.2. *Let $n > 2k+2$, $\tau > 0$, and $q_0 > 1$ be a positive integer. Let T be a positive definite integral $k \times k$ matrix. Then one has*

$$\widehat{S}_T(0) \ \leq \ C \, \det(T)^{\frac{n-k-1}{2}} \tag{2.2.31}$$

If $\mathcal{X} = (\xi_1, \dots, \xi_k) \in \mathbb{T}^{n\times k}$ such that for all $P \in \mathbb{Z}^{n\times k}$ and $q \leq q_0$

$$|\mathcal{X} - P/q| \geq \tau.$$

Then one has

$$|\widehat{S}_T(\mathcal{X})| \leq \ C \ \left[\det(T)^{\frac{n-k-1}{2}} \left((\tau^2 \mu(T))^{-\frac{n-2k-2}{4}} + q_0^{-\frac{n-2k-2}{2}} \right) \ + \ |T|^{\frac{(n-k)(k-1)}{2}} \right]. \tag{2.2.32}$$

We remark that if the parameters τ and q_0 is chosen such that $\tau > C(\varepsilon)\lambda^{-1/2}$, $q_0 > C(\varepsilon)$ and if λ is large enough with respect to $|T|$, then estimate (2.2.32) implies that

$$|\widehat{S}_{\lambda T}(\mathcal{X})| \leq c(\varepsilon) \; \det(\lambda T)^{\frac{n-k-1}{2}}, \tag{2.2.33}$$

for a given constant $c(\varepsilon) > 0$, as long as $C(\varepsilon)$ is chosen large enough with respect to $c(\varepsilon)$.

The proof of Lemma 2.2 is purely number theoretic and is independent of the rest of the paper. It will be given in Section 2.2.4 using the theory of theta functions developed by Siegel [98] and later by Kitaoka [67], adapted to our settings.

2.2.2 Lower bounds on the main terms.

From now on we fix $k \in \mathbb{N}$, $n \in \mathbb{N}$, $\varepsilon > 0$ and $R > 1$ and a set $A \subseteq B_R \cap \mathbb{Z}^n$ such that $|A| \geq \varepsilon R^n$. For given parameters $q \in \mathbb{N}$ and $q \leq L < R$ such that $R/L \in \mathbb{N}$, we partition the cube B_R into R^n/L^n cubes B_L of size L, and them further into congruence classes of the modulus q, i.e. into sets of the form

$$B_L(q,s) = B_L \cap (s + (q\mathbb{Z})^n), \tag{2.2.34}$$

where $s \in (\mathbb{Z}/q\mathbb{Z})^n$ is running through the congruence classes of q. With a slight abuse of notation, for given $m \in B_R$ we will denote by $B_L(q,m)$ the unique set $B_L(q,s)$ containing m. For given $\alpha > 0$, we say that the set A is α-uniformly distributed w.r.t. q and L if for each element $B_L(q,s)$ of the partition

$$\delta(A|B_L(q,s)) = \frac{|A \cap B_L(q,s)|}{|B_L(q,s)|} \leq (1+\alpha)\varepsilon. \tag{2.2.35}$$

Here we used the notation $\delta(A|B) = |A \cap B|/|B|$ for the relative density of the set A on the set B. It is immediate from (2.2.35) that $\delta(A|B_L) \leq (1+\alpha)\varepsilon$ for every cube B_L and that $\delta(A) = \delta(A|B_R) \leq (1+\alpha)\varepsilon$. It is also easy to see that $\delta(A|B_L) \geq (1-2\alpha)\varepsilon$ holds for many cubes B_L, such cubes B_L will be called *dense*. Indeed

$$\varepsilon \leq \delta(A) = \frac{L^n}{R^n}\sum_{B_L} \delta(A|B_L) \leq \frac{L^n}{R^n}\sum_{B_L\ dense} (1+\alpha)\varepsilon + (1-2\alpha)\varepsilon \tag{2.2.36}$$

It follows that there are at least $\frac{2\alpha}{(1+\alpha)}\frac{R^n}{L^n}$ dense cubes. We define the function $h_{L,q} : B_R \cap \mathbb{Z}^n \to [0,1]$ by

$$h_{L,q}(m) := \delta(A|B_L(q,m)). \tag{2.2.37}$$

Note that $h_{L,q}$ is constant and is equal to the average of the function $\mathbf{1}_A$ on each set $B_L(q,m)$, thus it is the so-called conditional expectation function of $\mathbf{1}_A$ with respect to the above partition.

Proposition 2.1. *Let $q \in \mathbb{N}$, $L > 0$ be given, and assume that the set A satisfies condition (3.2) with $\alpha = 1/10(k+1)$. If $q \leq \beta L$, with $\beta = \alpha\varepsilon/4n$, then for any $m_1, \ldots, m_k \in \mathbb{Z}^n$ such that $|m_i| \leq \beta L$ for each $1 \leq i \leq k$, then one has*

$$\sum_{m \in \mathbb{Z}^n} \mathbf{1}_A(m)\, h_{L,q}(m+m_1)\, h_{L,q}(m+m_2) \ldots h_{L,q}(m+m_k) \geq c_k\, \varepsilon^{k+1} R^n. \tag{2.2.38}$$

Proof. Let B_L be a dense cube and define the set $G = \{m \in B_L :\ h_{L,q}(m) \geq \alpha\varepsilon\}$. Arguing similarly as in (2.2.36),

$$(1-2\alpha)\varepsilon \leq \delta(A|B_L) \leq L^{-n} \sum_{m \in G} (1+\alpha)\varepsilon + \alpha\varepsilon. \tag{2.2.39}$$

hence $|G| > (1-4\alpha)L^n$. Let $B_{L'}$ denote the cube obtained by dilating B_L from its center with a factor of $1-\beta$. Then $L' = (1-\beta)L$ and $|B_L \backslash B_{L'}| < 2n\beta L^n$. For $m \in G$ one has

$$\delta(A|B_{L'} \cap B_L(q,m)) \geq \frac{q^n}{L^n}\, |A \cap B_L(q,m)| - \frac{1}{L^n}|B_L \backslash B_{L'}| \ \geq\ \alpha\varepsilon - 2n\beta \geq \frac{\alpha\varepsilon}{2}. \tag{2.2.40}$$

For $m \in B_{L'}$, $m + m_i \in B_L$ for each $1 \leq i \leq$ as $|m_i| \leq \beta L$, and the functions $m \to h_{L,q}(m+m_i)$ are constant on the set $B_{L'} \cap B_L(q,m)$. Thus

$$\sum_{m \in B_{L'}} \mathbf{1}_A(m)\, h_{L,q}(m+m_1)\, h_{L,q}(m+m_2) \ldots h_{L,q}(m+m_k)\ = \tag{2.2.41}$$

$$\sum_{m \in B_{L'}} \delta(A|B_{L'} \cap B_L(q,m))\, h_{L,q}(m+m_1)\ h_{L,q}(m+m_2) \ldots h_{L,q}(m+m_k).$$

If $m \in B_{L'} \cap G \cap (G-m_1) \cap \ldots \cap (G-m_k)$ then $m \in G$ and $m+m_i \in G$ for each $i \leq k$ hence by (2.2.40) and the definition of G the expression in (2.2.41) is further estimated from below by: $\frac{(\alpha\varepsilon)^{k+1}}{2} |B_{L'} \cap G \cap (G-m_1) \cap \ldots \cap (G-m_k)|$. Let $G' = B_{L'} \cap G$, then

$$|G'| \geq |G| - |B_L \backslash B_{L'}| \geq (1-4\alpha-2n\beta)L^n > (1-5\alpha)L^n \qquad \text{and}$$

$$\begin{aligned} |B_{L'} \cap G \cap (G-m_1) \cap \ldots \cap (G-m_k)| &\geq |G' \cap (G'-m_1) \cap \ldots \cap (G'-m_k)| \geq \\ &\geq \ (1-5\alpha(k+1))\, L^n \geq L^n/2. \end{aligned}$$

Thus, for a dense cube B_L, the expression in (2.2.31) is bounded below by $c_k\, \varepsilon^{k+1}\, L^n$, and since there are at least $\frac{2\alpha}{(1+\alpha)} \frac{R^n}{L^n}$ dense cubes, (2.2.38) follows. □

Note that the functions $f_{L,q} = \mathbf{1}_A * \psi_{L,q}$ defined in (2.2.19) majorize the functions $h_{L,q}$.

Proposition 2.2. *There exist a constant $c_n > 0$ such that all $m \in \mathbb{Z}^n$*

$$f_{L,q}(m) \ \geq \ c_n\, h_{L,q}(m). \tag{2.2.42}$$

Proof. By definition,

$$\begin{aligned} f_{L,q}(m) = q^n L^{-n} \sum_{l \in \mathbb{Z}^n} \mathbf{1}_A(m-ql)\, \psi(ql/L) \ &\geq \ c_n\, q^n L^{-n} \sum_{l \in \mathbb{Z}^n,\ |ql| \leq \sqrt{n}L} \mathbf{1}_A(m-ql) \ \geq \\ \geq \ c_n\, q^n L^{-n} \sum_{m' \in B_L(q,m)} \mathbf{1}_A(m') \ &\geq \ c_n\, h_{L,q}(m). \end{aligned}$$

The second inequality follows as the diameter of the set $B_L(q,m)$ is at most $\sqrt{n}L$. □

Let $\lambda \in \mathbb{N}$, $\triangle \in \mathbb{R}^n$ be an integral k-dimensional simplex, and let $T = T_\triangle$ be its inner product matrix. We proceed to estimate the expression $N_{\lambda T}(\mathbf{1}_A, f_{L,q}, \ldots, f_{L,q})$ defined in (2.2.13) from below under certain conditions on the parameters q, L, R, λ. To do so one needs a lower bound for the number of integral solutions $m_1, \ldots, m_k \in \mathbb{Z}^n$ of the system of equations $m_i \cdot m_j = \lambda t_{ij}$. This was done in [67], indeed, for $A = I_n$ (the $n \times n$ identity matrix), $B = \lambda T$, Theorems A-C in [67] implies for $n > 2k+2$ and $\lambda > \lambda(n,k)$, that

$$\sum_{m_1,\ldots,m_k \in \mathbb{Z}^n} S_{\lambda T}(m_1, \ldots, m_k) \ \geq \ c_0\, \det(\lambda\, T)^{\frac{n-k-1}{2}}, \tag{2.2.43}$$

for some positive constant c_0 depending only on n and k. Note that the left side of (3.10) is the number of solutions $m_1, \ldots, m_k$ of (1.2). Now it is easy to show

Lemma 2.3. *Let $k \geq 2$, $n > 2k+2$, $\varepsilon > 0$, $R > 0$ and let $A \subseteq B_R \cap \mathbb{Z}^n$ be a set such that $|A| \geq \varepsilon R^n$. Let $\lambda \in \mathbb{N}$ and let $\triangle \subseteq \mathbb{R}^n$ be an integral $k-$simplex with inner product matrix $T = T_\triangle$. Let $q \in \mathbb{N},\ L > 0$ be parameters satisfying*

$$q \leq \beta L, \quad \sqrt{\lambda |T|} \leq \beta L, \quad \text{with} \quad \beta = \frac{\varepsilon}{40n(k+1)}. \tag{2.2.44}$$

If A is α-uniformly distributed w.r.t q and L with $\alpha = \frac{1}{10(k+1)}$ and if $f_{L,q}(m)$ is defined as in (2.2.22), *then*

$$N_{\lambda T}(\mathbf{1}_A, f_{L,q}, \ldots, f_{L,q}) \gtrsim \det(\lambda\, T)^{\frac{n-k-1}{2}}\, \varepsilon^{k+1}\, R^n. \tag{2.2.45}$$

Proof. If $S_{\lambda T}(m_1,\dots,m_k)\neq 0$ then $|m_i|^2=\lambda t_{ii}$ ($\forall\ 1\le i\le k$), hence $|m_i|\le \beta L$. It follows from (2.2.38) and (2.2.42) that

$$\sum_{m\in\mathbb{Z}^n} \mathbf{1}_A(m)\, f_{L,q}(m+m_1)\, f_{L,q}(m+m_2)\dots f_{L,q}(m+m_k) \gtrsim \varepsilon^{k+1}R^n$$

Summing the above bound for all such $m_1,\dots,m_k$, the Lemma follows from inequality (2.2.43). $\square$

Let us point out that the right side of (2.2.45) is the expected value of $N_{\lambda T}(\mathbf{1}_A,\dots,\mathbf{1}_A)$ if $A\subseteq B_R\cap\mathbb{Z}^n$ in a random set of density ε, obtained by choosing each point of $B_R\cap\mathbb{Z}^n$ independently with probability ε. Indeed, for given $m\in B_R\cap\mathbb{Z}^n$ and a solution $m_1,\dots,m_k$, the probability that all points $m,m_1,\dots,m_k$ are in the set A is ε^{k+1}.

2.2.3 Error estimates

In this section we estimate quantities of the following form

$$E_{\lambda T}(f;f_1,f_2)=N_{\lambda T}(f,f_1,\dots,f_1)-N_{\lambda T}(f,f_2,\dots,f_2), \tag{2.2.46}$$

where the functions $f,f_1,f_2:\mathbb{Z}^n\to[-1,1]$ are of finite support or rapidly decreasing. Note that

$$E_{\lambda T}(f;f_1,f_2)=\sum_{i=1}^k E^i_{\lambda T}(f;f_1,f_2),\qquad \text{where}$$

$$E^i_{\lambda T}(f;f_1,f_2)=N_{\lambda T}(f,f_1,\dots,f_1,f_2,\dots,f_2)-N_{\lambda T}(f,f_1,\dots,f_2,f_2,\dots,f_2) \tag{2.2.47}$$

Here, the second term in (2.2.47) is obtained from the first term by replacing the function f_1 with the function f_2 at the $i-$th place.

For fixed $1\le i\le k$, let T_i denote the $(k-1)\times(k-1)$ matrix obtained from the matrix T by deleting the $i-$th row and column. Note that $T_i=T_{\triangle_i}$, where $\triangle_i$ is the $k-1$-dimensional face of the simplex $\triangle=\{0,v_1,\dots,v_k\}$ which does not contain the $i-$th vertex v_i. For given $\mathbf{m}=(m_1,\dots,m_k)\in\mathbb{Z}^{nk}$ let us introduce the notation $\mathbf{m}^i=(m_1,\dots,m_{i-1},m_{i+1},\dots,m_k)\in\mathbb{Z}^{n(k-1)}$ and define the function $S_{\lambda T,\mathbf{m}^i}:\mathbb{Z}^n\to\{0,1\}$ by

$$S_{\lambda T,\mathbf{m}^i}(m_i)=\begin{cases}1 & \text{if}\ \ m_i\cdot m_j=\lambda t_{ij}\ \ \forall\ 1\le j\le k\\ 0 & \text{otherwise}\end{cases}$$

Then, clearly

$$S_{\lambda T}(\mathbf{m})=S_{\lambda T_i}(\mathbf{m}^i)\,S_{\lambda T,\mathbf{m}^i}(m_i), \tag{2.2.48}$$

where the function $S_{\lambda T_i}$ is defined in (2.2.15).

Proposition 2.3. *Let $k\ge 2$, $n>2k+2$ and let $f,f_1,f_2:\mathbb{Z}^n\to[-1,1]$ be given functions. Then one has*

$$|E_{\lambda T}(f;f_1,f_2)|\lesssim \det(\lambda T)^{\frac{n-k-1}{2}}\,\|f\|_2\,\|f_1-f_2\|_2 \tag{2.2.49}$$

Proof. For fixed $1\le i\le k$, using $S_{\lambda T}(\mathbf{m})=S_{\lambda T}(-\mathbf{m})$, one may write

$$N_{\lambda T}(f_0,f_1,\dots,f_k)=\sum_{\mathbf{m}^i}\sum_m\sum_{m_i} S_{\lambda T_i}(\mathbf{m}^i)\,f(m)\prod_{j\neq i} f_j(m-m_j)\,f_i(m-m_i)S_{\lambda T,\mathbf{m}^i}(m_i)$$

$$= \sum_{\mathbf{m}^i} \sum_m S_{\lambda T_i}(\mathbf{m}^i)\, f(m)\, g_i(m, \mathbf{m}^i)\, (f_i * S_{\lambda T, \mathbf{m}^i})(m).$$

Thus,

$$\begin{aligned} |E^i_{\lambda T}(f; f_1, f_2)| &\le \sum_{\mathbf{m}^i, m} S_{\lambda T_i}(\mathbf{m}^i)\, |f(m)|\, |(f_1 - f_2) * S_{\lambda T, \mathbf{m}^i})(m)| \\ &\le \|f\|_2 \sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\, \|(f_1 - f_2) * S_{\lambda T, \mathbf{m}^i}\|_2 \\ &\le \|f\|_2\, \|f_1 - f_2\|_2 \sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\, \|S_{\lambda T, \mathbf{m}^i}\|_1 \end{aligned} \tag{2.2.50}$$

where the second line follows from Cauchy-Schwarz and the third line from Minkowski's integral inequality. Finally, by inequality (2.2.31)

$$\sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\, \|S_{\lambda T, \mathbf{m}^i}\|_1 = \sum_{\mathbf{m}} S_{\lambda T}(\mathbf{m}) \lesssim \det(\lambda T)^{\frac{n-k-1}{2}}.$$

and (2.2.49) follows. □

Next, we give a different estimate on the quantity $E_{\lambda T}(f; f_1, f_2)$.

Proposition 2.4. *Let $k \ge 2$, $n > 2k + 2$ and let $f, f_1, f_2 : \mathbb{Z}^n \to [-1, 1]$ be given functions. Then for fixed $1 \le i \le k$, one has*

$$|E^i_{\lambda T}(f; f_1, f_2)| \lesssim \det(\lambda T_i)^{\frac{n-k}{4}}\, \|f\|_2 \left(\int_{\mathbb{T}^n} |(\hat{f}_1 - \hat{f}_2)(\xi)|^2 \sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\, |\widehat{S}_{\lambda T, \mathbf{m}^i}(\xi)|^2\, d\xi \right)^{\frac{1}{2}}. \tag{2.2.51}$$

Proof. Using the matrix formulation, the support of the function $S_{\lambda T_i}$ consists of those integral matrices $M \in \mathbb{Z}^{n \times (k-1)}$ which satisfy the equation $M^t \cdot M = \lambda T_i$, hence by (2.2.31) the size of its support is bounded by $C \det(\lambda T_i)^{\frac{n-k}{2}}$.
Starting with the second line of (2.2.50) and using the Cauchy-Schwarz inequality, one obtains

$$|E^i_{\lambda T}(f; f_1, f_2)| \lesssim \|f\|_2 \det(\lambda T_i)^{\frac{n-k}{4}} \left(\sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\, \|(f_1 - f_2) * S_{\lambda T, \mathbf{m}^i}\|_2^2 \right)^{\frac{1}{2}}$$

Inequality (2.2.51) follows by applying Plancherel's formula to the above expression in parenthesis, and by interchanging the summation and integration. □

Expanding the sum in formula (2.2.51), one obtains

$$\sum_{\mathbf{m}^i \in \mathbb{Z}^{n(k-1)}} \sum_{m_i, m_{k+1} \in \mathbb{Z}^n} S_{\lambda T_i}(\mathbf{m}^i) S_{\lambda, \mathbf{m}^i}(m_i) S_{\lambda, \mathbf{m}^i}(m_{k+1})\, e^{2\pi i (m_i \cdot \xi - m_{k+1} \cdot \xi)}. \tag{2.2.52}$$

If one defines $G^i_{\lambda,T}(\mathbf{m}, m_{k+1}) = S_{\lambda T_i}(\mathbf{m}^i) S_{\lambda, \mathbf{m}^i}(m_i) S_{\lambda, \mathbf{m}^i}(m_{k+1}) : \mathbb{Z}^{n(k+1)} \to \{0, 1\}$, where $\mathbf{m} = (m_1, \dots, m_k) \in \mathbb{Z}^{nk}$, then the expression in (2.2.52) is equal to $\widehat{G^i}_{\lambda,T}(0, \dots, 0, \xi, 0, \dots, 0, -\xi) = \widehat{G^i}_{\lambda,T}(\mathcal{X})$ with $\mathcal{X} = (0, \dots, 0, \xi, 0, \dots, 0, -\xi) \in \mathbb{R}^{n \times (k+1)}$, where the entries ξ and $-\xi$ appear the the

i-th and $k+1$-th place. Note that $G^i_{\lambda,T}(m_1,\ldots,m_{k+1})=1$ if and only the vectors $m_1,\ldots,m_{k+1}$ satisfy the system of equations:

$$m_j\cdot m_l=\lambda t_{jl},\quad m_{k+1}\cdot m_l=m_i\cdot m_l=\lambda t_{il}\ (l\neq i),\quad m_{k+1}\cdot m_{k+1}=m_i\cdot m_i=\lambda t_{ii},\tag{2.2.53}$$

for all $1\leq l\leq j\leq k$. If one writes $\lambda t=m_{k+1}\cdot m_i$, and defines the symmetric $(k+1)\times(k+1)$ matrix $T^i(t)=(\tau_{j,l})$ with entries $(1\leq l\leq j\leq k)$

$$\tau_{j,l}=t_{jl},\ \tau_{k+1,l}=t_{il}\ (l\neq i),\ \tau_{k+1,k+1}=t_{ii},\ \tau_{k+1,i}=t,\tag{2.2.54}$$

then it is clear that

$$G^i_{\lambda,T}(\mathbf{m},m_{k+1})=\sum_t S_{\lambda T^i(t)}(\mathbf{m},m_{k+1}).\tag{2.2.55}$$

Note that the summation in (4.10) is finite as the function $S_{\lambda T^i(t)}$ is constant 0 unless there exists an $\tilde{M}\in\mathbb{Z}^{n\times(k+1)}$ such that $\tilde{M}^t\tilde{M}=\lambda T^i(t)$, in which case we will call the number t *admissible*. Thus if t is admissible then, in particular, $t^2\leq t_{ii}^2$ and $\lambda t\in\mathbb{Z}$. To summarize, we have for $1\leq i\leq k$ and $\xi\in\mathbb{R}^n$

$$\sum_{\mathbf{m}^i\in\mathbb{Z}^{n(k-1)}} S_{\lambda T_i}(\mathbf{m}^i)\,|\widehat{S}_{\lambda T,\mathbf{m}^i}(\xi)|^2\ =\sum_{t\ admissible}\widehat{S}_{\lambda T^i(t)}(\mathcal{X}),\tag{2.2.56}$$

with $\mathcal{X}=(0,\ldots,0,\xi,0,\ldots,0,-\xi)$. We need to collect some geometric facts about the matrices $T^i(t)$, to estimate the right side of (2.2.56)

Proposition 2.5. *Let $T>0$ be a fixed integral $k\times k$ matrix and let $1\leq i\leq k$. Then,*

(i) The number of admissible values of t is bounded by: $2\det(\lambda T)/\det(\lambda T_i)+1$.

(ii) For each $M=(m_1,\ldots,m_k)$ such that $M^tM=\lambda T$, there are at most 2 vectors $m_{k+1}\in\mathbb{Z}^n$ such that $\det(T^i(t))=0$, *where the vectors $m_1,\ldots,m_{k+1}$ and the matrix $T^i(t)$ satisfy* (2.2.53) *and* (2.2.54).

(i3) Let t be admissible, and let $\tilde{M}=(m_1,\ldots,m_k,m_{k+1})$ be such that $\tilde{M}^t\tilde{M}=\lambda T^i(t)$. Let d denote the distance of the vector m_{k+1} to the subspace $Span\{m_1,\ldots,m_k\}$, that is to the subspace spanned by the vectors $m_1,\ldots,m_k$. Then

$$\mu(\lambda T^i(t))\geq\frac{d^2\mu(T)}{8|T|}.\tag{2.2.57}$$

Here $\mu(T)$ is defined in (2.2.2) *and* $|T|=(\sum_{i,j}t_{ij}^2)^{1/2}$.

(i4) Let $0<\delta<e(T)^{-4}/64$, where $e(T)$ is defined in (1.5). Then

$$|\{t\ admissible:\ \mu(T^i(t))\leq\ |T|\,\delta\}|\ \lesssim\ \delta^{\frac12}\ \det(\lambda T)/\det(\lambda T^i).\tag{2.2.58}$$

(i5) Let t be admissible, then one has

$$\det(\lambda T^i(t))\leq\det(\lambda T)^2/\det(\lambda T_i)\tag{2.2.59}$$

Proof. Let t be admissible, and let $\tilde{M}=(m_1,\ldots,m_k,m_{k+1})$ be such that $\tilde{M}^t\tilde{M}=\lambda T^i(t)$. If P denotes the orthogonal projection to the subspace spanned by the vectors $m_1,\ldots,m_{i-1},m_{i+1},\ldots,m_k$ then by (4.8) $Pm_i=Pm_{k+1}$. Denote this vector by u, and write $m_i=u+w$, $m_{k+1}=u+w'$. If one considers the vectors $m_1,\ldots,m_k$ as elements of the k-dimensional subspace $Span\{m_1,\ldots,m_k\}$ then the quantity $|\det(m_1,\ldots,m_k)|$ is well-defined and is equal to the volume of the parallelepiped spanned by these vectors. Moreover it is easy to see that $\det(\lambda T)=|\det(m_1,\ldots,m_k)|^2$, and also that $|w'| = |w| = |\det(m_1,\dots,m_k)|/|\det(m_1,\dots,m_{i-1},m_{i+1},\dots,m_k)|$. Since $\lambda t = m_{k+1}\cdot m_k = |u|^2 + w\cdot w'$ it follows that

$$|\lambda t - |u|^2| \;\leq\; |w|^2 \;=\; \det(\lambda T)/\det(\lambda T^i) \tag{2.2.60}$$

and (i) is proved.
If $\det(T^i(t)) = 0$ then m_{k+1} is linearly dependent of the vectors $m_1,\dots,m_k$, thus w' is also linearly dependent of the vectors $m_1,\dots,m_{i-1},m_{i+1},\dots,m_k$ and w, hence $w' = \pm w$. It follows $m_{k+1} = u \pm w$ and (ii) is proved.
Let $x = (x_1,\dots,x_k,x_{k+1}) \in \mathbb{R}^{k+1}$, $|x| = 1$ such that

$$\mu(\lambda T^i(t)) = \lambda T^i(t)x\cdot x = |m_1x_1 + \ldots + m_{k+1}x_{k+1}|^2$$

It is clear that $\mu(\lambda T^i(t)) \geq d^2\,|x_{k+1}|^2$, thus if $|x_{k+1}|^2 \geq \mu(T)/4|T|$ then inequality(4.12) holds. Otherwise $|x_{k+1}|^2 \leq \mu(T)/4|T|$ and one estimates

$$\mu(\lambda T^i(t)) \geq (|m_1x_1 + \ldots + m_kx_k| - |m_{k+1}||x_{k+1}|)^2 \geq \frac{1}{8}\mu(\lambda T),$$

as $|m_{k+1}|^2 = |m_i|^2 = \lambda t_{ii} \leq |\lambda T|$ and $x_1^2 + \ldots + x_k^2 \geq 3/4$. Also $d^2 \leq |m_{k+1}|^2 \leq |\lambda T|$ thus $d^2\mu(T)/8|T| \leq \mu(\lambda T)/8$ and (2.2.57) follows.
Writing $u = m_1y_1 + \ldots + m_{i-1}y_{i-1} + m_{i+1}y_{i+1} + \ldots m_ky_k$, it follows

$$|w|^2 = |u - m_i|^2 \;\geq\; (1 + y_1^2 + \ldots + y_k^2)\mu(\lambda T) \;\geq\; |\lambda T|\, e(T)^{-1}$$

If v denotes the orthogonal projection of the vector m_{k+1} to the subspace spanned by the vectors $m_1,\dots,m_k$, the it is easy to see that $v = u + w\,\frac{w\cdot w'}{|w|}$. Thus

$$\frac{(w\cdot w')^2}{|w|^2} + d^2 = |w|^2 \qquad \text{substituting} \qquad \lambda t - |u|^2 = w\cdot w'$$

$$|w|^2 \;\geq\; |\lambda t - |u|^2| \;\geq\; |w|^2\,(1 - d^2/|w|^2)^{\frac{1}{2}}$$

If $\mu(T^i(t)) \leq |T|\,\delta$ then by (2.2.57) and the assumption on δ

$$\frac{d^2}{|w|^2} \;\leq\; d^2 e(T)\,|\lambda T|^{-1} \;\leq\; 8\delta\, e(T)^2 \;\leq\; \delta^{\frac{1}{2}}.$$

Since $\delta < 1$, it follows that $|w|^2 \geq |\lambda t - |u|^2| \geq |w|^2(1 - \delta^{1/2})$ and this implies (2.2.58).
Finally, arguing as in (4.15) one has

$$\det(\lambda T^i(t))/\det(\lambda T) = d^2 \;\leq\; |w|^2 = \det(\lambda T)/\det(\lambda T_i),$$

and (2.2.59) follows. □

Using Lemma 2.1, in dimensions n and $k+1$ it is now not hard to estimate the right side of (2.2.56). We remark that it is here where the stronger condition $n > 2k+4$ is needed.

Proposition 2.6. *Let $k \geq 2$, $n > 2k+4$, and let $T \in \mathbb{Z}^{k\times k}$ be a positive matrix. Let $q_0 \in \mathbb{N}$ and $0 < \delta < e(T)^{-4}/64$ be given parameters. Then for $1 \leq i \leq k$*

$$\sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\,|\widehat{S}_{\lambda T,\mathbf{m}^i}(\xi)|^2 \lesssim \frac{\det(\lambda T)^{n-k-1}}{\det(\lambda T_i)^{\frac{n-k}{2}}}\,\left(1 + \lambda^{-\frac{n-k}{2}}\, e(T)^{\frac{(n-k)(k-1)}{2}}\right) \tag{2.2.61}$$

holds uniformly for $\xi \in \mathbb{R}^n$.

If $|\xi - l/q| \geq \delta^{-1}\lambda^{-1/2}\,|T|^{-1/2}$ for all $l \in \mathbb{Z}^n$ and $q \leq q_0$, then one has

$$\sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\,|\widehat{S}_{\lambda T,\mathbf{m}^i}(\xi)|^2 \lesssim \frac{\det(\lambda T)^{n-k-1}}{\det(\lambda T_i)^{\frac{n-k}{2}}} \left(q_0^{-\frac{n-2k-4}{2}} + \delta^{\frac{1}{4}} + \lambda^{-\frac{n-2k-2}{2}} e(T)^{(n-k-1)k} \right). \quad (2.2.62)$$

Proof. Let us first estimate the sum in (2.2.52) over those $k+1$ tuples $(m_1, \dots, m_k, m_{k+1})$ for which m_{k+1} is linearly dependent on the vectors $m_1, \dots, m_k$. By Proposition 2.5 (i), there are at most 2 possible choices for the vector m_{k+1}. Thus one estimates the contribution of such $k+1$ tuples to the sum in (2.2.52) by

$$\widehat{S}_{\lambda T}(0) \lesssim \det(\lambda T)^{\frac{n-k-1}{2}} \lesssim \frac{\det(\lambda T)^{n-k-1}}{\det(\lambda T_i)^{\frac{n-k}{2}}}\; \lambda^{-\frac{n-2k}{2}} e(T)^{\frac{(n-k)(k-1)}{2}}. \quad (2.2.63)$$

The first inequality in (2.2.52) follows from (2.2.31), while the second follows from the facts that $\det(\lambda T_i) \leq |\lambda T_i|^{k-1} \leq |\lambda T|^{k-1}$ and $|\lambda T|^k = \mu(\lambda T)^k e(T)^k \leq \det(\lambda T) e(T)^k$.

Summing over the $k+1$-tuples $(m_1, \dots, m_k, m_{k+1})$ in formula (2.2.52) which are linearly independent, is equal to the sum on the right side of (2.2.56) over those admissible values of t for which $\det(T^i(t)) > 0$, and one may apply Lemma 1 to the matrix $\lambda T^i(t)$, for each such value of t. Thus by (2.2.31) and (2.2.59), one has uniformly in $\xi \in \mathbb{R}^n$

$$|\widehat{S}_{\lambda T^i(t)}(\xi)| \lesssim \det(\lambda T^i(t))^{\frac{n-k-2}{2}} \leq \det(\lambda T)^{n-k-2} \det(\lambda T_i)^{-\frac{n-k-2}{2}}. \quad (2.2.64)$$

By Proposition 2.5 (i), the number of admissible values t (for which $\det(T^i(t)) \neq 0$) is bounded by $2\det(\lambda T)/\det(\lambda T_i)$ and (2.2.61) follows from (2.2.56) and (2.2.63).
Let us assume now that $|\xi - l/q| \geq \delta^{-1}\lambda^{-1/2}\,|T|^{-1/2}$, for all $l \in \mathbb{Z}^n$ and $1 \leq q \leq q_0$, and hence $|\mathcal{X} - P/q| \geq \delta^{-1}\lambda^{-1/2}$ for all $P \in \mathbb{Z}^{n\times(k+1)}$ and $q \leq q_0$ (where $\mathcal{X} = (0, \dots, 0, \xi, 0, \dots, 0, -\xi)$ as before). Then one may use inequality (3.28) in Lemma 1 with $\tau = \delta^{-1}\lambda^{-1/2}\,|T|^{-1/2} > 0$ to estimate the left side of (2.2.52):

$$\begin{aligned} |\widehat{S}_{\lambda T^i(t)}(\mathcal{X})| &\lesssim \det(\lambda T^i(t))^{\frac{n-k-2}{2}} q_0^{-\frac{n-2k-4}{2}} + \det(\lambda T^i(t))^{\frac{n-k-2}{2}} (\delta^{-2}|T|^{-1}\mu(T^i(t)))^{-\frac{n-2k-4}{4}} \\ &\quad + (\lambda |T^i(t)|)^{\frac{(n-k-1)k}{2}} = S_1(t) + S_2(t) + S_3(t) \end{aligned} \quad (2.2.65)$$

Summing the fist terms over admissible values of t is estimated exactly as in (2.2.61) and one gets

$$\sum_t S_1(t) \lesssim \det(\lambda T)^{n-k-1} \det(\lambda T_i)^{-\frac{n-k}{2}}\, q_0^{-\frac{n-2k-4}{2}}.$$

If t is such that $\mu(T^i(t)) \geq \delta\,|T|$ then $(\delta^{-2}|T|^{-1}\mu(T^i(t)))^{-(n-2k-4)/4} \leq \delta^{1/4}$ as $n - 2k - 4 \geq 1$ and summing over such t's gives the second term of the right side of (2.2.62). By Proposition 4.3, the number of admissible t's such that $\mu(T^i(t)) \leq \delta\,|T|$ is bounded by $2\delta^{1/2}\det(T)/\det(T_i)$ and one get a gain by a factor of $\delta^{1/2}$ over the estimate in (2.2.61), thus

$$\sum_t S_2(t) = \sum_{t:\,\mu(T^i(t))\geq\delta\,|T|} S_2(t) + \sum_{t:\,\mu(T^i(t))<\delta\,|T|} S_2(t) \lesssim \delta^{\frac{1}{4}} \det(\lambda T)^{n-k-1} \det(\lambda T_i)^{-\frac{n-k}{2}}.$$

Finally, using the facts $|\lambda T^i(t)| \lesssim |\lambda T| \leq \det(\lambda T)^{1/k}\, e(T)$ and $\det(\lambda T_i) \leq \det(\lambda T)^{(k-1)/k}\, e(T)^{k-1}$ a straightforward calculation shows, that summing the third terms on the right side of inequality (2.2.65), one gets

$$\sum_t S_3(t) \lesssim \det(\lambda T)\, \det(\lambda T_i)^{-1}\, |\lambda T|^{\frac{(n-k-1)k}{2}} \lesssim \frac{\det(\lambda T)^{n-k-1}}{\det(\lambda T_i)^{\frac{n-k}{2}}}\, \lambda^{-\frac{n-2k-2}{2}} e(T)^{(n-k-1)k}.$$

This proves the proposition. $\square$

We will apply inequalities (2.2.51), (2.2.61) and (2.2.62) to functions of the form $f_i = f_{L_i,q}$ $(i = 1, 2)$ defined in (2.2.22), for specific choice of $L_i > 0$. Recall that we defined $f_{L,q} = \mathbf{1}_A * \psi_{L,q}$ where, considering as distribution on $\mathbb{R}^n$,

$$\psi_{L,q} = q^n \delta_{(q\mathbb{Z})^n}\, \psi_L \quad \text{where} \quad \psi_L(x) = L^{-n}\psi(x/L),$$

and $\delta_{(q\mathbb{Z})^n}$ denotes the discrete (counting) measure supported on the lattice $(q\mathbb{Z})^n$. By Poisson summation, if $\phi \in \mathcal{C}^\infty(\mathbb{T}^n)$ then

$$\langle\, \widehat{\delta}_{(q\mathbb{Z})^n}, \phi\,\rangle \;=\; \langle\, \delta_{(q\mathbb{Z})^n}, \widehat{\phi}\,\rangle \;=\; q^{-n} \sum_{l\in\mathbb{Z}^n} \phi(l/q).$$

Thus

$$\widehat{\psi}_{L,q}(\xi) \;=\; q^n \left(\widehat{\delta}_{(q\mathbb{Z})^n} * \widehat{\psi}_L\right)(\xi) \;=\; \sum_{l\in\mathbb{Z}^n} \widehat{\psi}\,(L(\xi - l/q)). \tag{2.2.66}$$

We can now state the main result of this section, given a set $A \subseteq B_R \cap \mathbb{Z}^n$ such that $|A| \geq \varepsilon R^n$, an integral k-dimensional simplex $\triangle \subseteq \mathbb{R}^n$ with $T = T_\triangle$, and a positive integer λ.

Lemma 2.4. *Let $k \geq 2$, $n > 2k+4$, and let $\bar{c} > 0$ be a positive constant. Let $\bar{C} > 0$ and define*

$$L_1 = \bar{C}^{-1}\, e(T)^{-4} \varepsilon^{9(k+1)}\, \lambda^{\frac{1}{2}}\, |T|^{\frac{1}{2}}, \qquad q(\varepsilon) = l.c.m.\,\{q \leq \bar{C}\varepsilon^{-\frac{4(k+1)}{n-2k-4}}\}. \tag{2.2.67}$$

If $\bar{C} = \bar{C}(n, k, \bar{c})$ is large enough and if

$$\lambda \;\geq\; \bar{C}\, q(\varepsilon)^2 \varepsilon^{-18(k+1)}\, e(T)^{\frac{4k(n-k-1)}{n-2k-2}}, \tag{2.2.68}$$

then one has

$$|E_{\lambda T}(\mathbf{1}_A; \mathbf{1}_A, f_{L_1,q(\varepsilon)})| \;\leq\; \bar{c}\, \varepsilon^{k+1}\, R^n\, \det(\lambda T)^{n-k-1}. \tag{2.2.69}$$

Proof. Let $1 \leq i \leq k$ be fixed. Applying inequality (4.6) for $f = f_1 = \mathbf{1}_A$, $f_2 = f_{L_1,q(\varepsilon)}$, one has

$$|E^i_{\lambda T}(\mathbf{1}_A; \mathbf{1}_A, f_{L_1,q(\varepsilon)})| \;\leq\; C\, \|\mathbf{1}_A\|_2^2\, \det(\lambda T_i)^{\frac{(n-k)}{4}} \left(\sup_{\xi\in\mathbb{T}^n} |1 - \widehat{\psi}_{L_1,q(\varepsilon)}(\xi)|^2 \sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\, |\widehat{S}_{\lambda T,\mathbf{m}^i}(\xi)|^2\right)^{\frac{1}{2}}.$$

Since $\|\mathbf{1}_A\|_2^2 = |A| \leq R^n$, it is enough to show that

$$\sup_{\xi\in\mathbb{T}^n} |1 - \widehat{\psi}_{L_1,q(\varepsilon)}(\xi)|\, (\sum_{\mathbf{m}^i} S_{\lambda T_i}(\mathbf{m}^i)\, |\widehat{S}_{\lambda T,\mathbf{m}^i}(\xi)|^2)^{\frac{1}{2}} \;\leq\; c_1\, \varepsilon^{k+1}\, \det(\lambda T)^{\frac{n-k-1}{2}}\, \det(\lambda T_i)^{-\frac{n-k}{4}}, \tag{2.2.70}$$

for some constant $c_1 = c_1(n,k,\bar{c}) > 0$ small enough. By our assumptions, $L_1 > q(\varepsilon)$, hence the supports of the functions $\widehat{\psi}(L_1(\xi - l/q(\varepsilon)))$ are disjoint for different values of $l \in \mathbb{Z}^n$. Thus if there is an l_0 such that: $|\xi - l_0/q(\varepsilon)| \leq C_1^{-1}\,\varepsilon^{k+1} L_1^{-1}$ where C_1 is large enough w.r.t. c_1, then

$$|1 - \sum_{l\in\mathbb{Z}^n} \widehat{\psi}(L_1(\xi - l/q(\varepsilon)))| = |1 - \widehat{\psi}(L_1(\xi - l_0/q(\varepsilon)))| \leq c_1\,\varepsilon^{k+1},$$

using the fact that $|1-\widehat{\psi}(\eta)| \lesssim |\eta|$ for $\eta \in \mathbb{R}^n$, and (2.2.69) follows from (2.2.61) and the assumption (2.2.68).

In the opposite case, for all $l \in \mathbb{Z}^n$ and $1 \leq q \leq \bar{C}\varepsilon^{-\frac{4(k+1)}{n-2k-4}}$, one has by (2.2.68)

$$|\xi - l/q| = |\xi - l'/q(\varepsilon)| > C_1^{-1}\,\varepsilon^{k+1} L_1^{-1} \;\geq\; (\bar{C}/C_1)\,e(T)^4\,\varepsilon^{-8(k+1)}\lambda^{-\frac{1}{2}}\,|T|^{-\frac{1}{2}}.$$

Thus one can apply inequality (2.2.52) with parameters

$$\delta = (C_1/\bar{C})\,e(T)^{-4}\varepsilon^{-8(k+1)} \qquad q_0 = \bar{C}\,\varepsilon^{-\frac{4(k+1)}{n-2k-4}},$$

using the fact that $\lambda \;\geq\; \bar{C}\,e(T)^{\frac{2k(n-k-1)}{n-2k-2}}\,\varepsilon^{-\frac{4(k+1)}{n-2k-2}}$ inequality (2.2.69) follows, if the constant $\bar{C} = \bar{C}(n,k,\bar{c})$ is chosen large enough. □

Proof of Lemma 2.1.

We will proceed as in Section 2.2.2. Assume that for a given $\lambda \in \mathbb{N}$, the simplex $\sqrt{\lambda}\triangle$ cannot be embedded in A, that is

$$N_{\lambda T}(\mathbf{1}_A, \mathbf{1}_A, \ldots, \mathbf{1}_A) = 0. \tag{2.2.71}$$

Choosing $L = C(\lambda|T|)^{1/2}$ such that $R/L \in \mathbb{Z}$ and $q = q(\varepsilon)$ defined in 2.2.1, Lemma 2.3 implies that

$$N_{\lambda T}(\mathbf{1}_A, f_{L,q}, \ldots, f_{L,q}) \geq c_0\;\det(\lambda T)^{\frac{n-k-1}{2}}\,\varepsilon^{k+1} R^n.$$

Assuming that the parameters R, ε and λ satisfy $R > L_2(\lambda,\varepsilon) > L_1(\lambda,\varepsilon) > q(\varepsilon)$, where

$$L_1(\lambda,\varepsilon) = \bar{C}^{-1}\,e(T)^{-4}\varepsilon^{9(k+1)}(\lambda T)^{1/2}, \quad L_2(\lambda,\varepsilon) = \bar{C}\,\varepsilon^{-(k+1)}\,(\lambda T)^{1/2}$$

we have that both (2.2.67) and (2.2.68) is satisfied. Thus by Lemma 2.5,

$$N_{\lambda T}(\mathbf{1}_A, f_{L_1,q}, \ldots, f_{L_1,q}) \leq \frac{c_0}{2}\;\det(\lambda T)^{\frac{n-k-1}{2}}\,\varepsilon^{k+1} R^n,$$

where we wrote $L_1 = L_1\lambda,\varepsilon$ and $q = q(\varepsilon)$ for simplicity of notations. Using Proposition 2.3 with $f = \mathbf{1}_A$, $f_1 = f_{L_1,q}$ and $f_2 = f_{L,q}$ it follows

$$\|f_{L_1,q} - f_{L,q}\|_2^2 = \int_{\mathbb{T}^n} |\hat{\mathbf{1}}_A(\xi)|^2\,|\widehat{\psi}_{L_1,q} - \widehat{\psi}_{L,q}|^2\,d\xi \;\geq c_1\,\varepsilon^{2k+2}\,R^n, \tag{2.2.72}$$

for some constant $0 < c_1 \leq 1$. Note that $\widehat{\psi}_{L_1,q} - \widehat{\psi}_{L,q}$ is supported on $(\frac{1}{q}\mathbb{Z})^n + [-\frac{1}{2L_1}, \frac{1}{2L_1}]^n$. Moreover, if $\xi = \frac{l}{q} + \eta$ with $\eta \in [-\frac{1}{2L_2}, \frac{1}{2L_2}]^n$ for a given $L_2 > C_1\,\varepsilon^{-(k+1)}L$, then

$$|\widehat{\psi}_{L_1,q}(\xi) - \widehat{\psi}_{L,q}(\xi)| = |\widehat{\psi}(L_1\eta) - \widehat{\psi}(L\eta)| \leq C\,L/L_2 \leq \frac{c_1}{2}\,\varepsilon^{k+1},$$

as long as $C_1 \gg c_1^{-1}$. Thus integrating over the complement of the set $\mathbb{T}_{\lambda,\varepsilon} = \mathbb{T}_{(L_1(\lambda,\varepsilon), L_2(\lambda,\varepsilon), q(\varepsilon))}$

$$\int_{\mathbb{T}^n/\mathbb{T}_{\lambda,\varepsilon}} |\hat{\mathbf{1}}_A(\xi)|^2\,|\widehat{\psi}_{L_1,q} - \widehat{\psi}_{L,q}|^2\,d\xi \;\leq \frac{c_1}{4}\,\varepsilon^{2k+2}\,R^n. \tag{2.2.73}$$

Estimates (2.2.72)-(2.2.73) imply estimate (2.2.9) and Lemma 2.1 is proved.

2.2.4 Fourier transforms and Siegel theta functions

In this section we prove Lemma 2.2 using the theory of theta functions. All arguments given here are independent of the rest of the paper, based on the approach in [67, 95] of estimating Fourier coefficients of Siegel modular forms vanishing at cusps. The basic difference is that the above mentioned works dealt with the case $\mathcal{X}=0$ while we need to consider those values of $\mathcal{X}$ which are "away" from rational points P/q ($P\in\mathbb{Z}^{k\times k}$) with small denominator q. The related theta functions are not modular forms, but behave very similarly, at such points $\mathcal{X}$, and hence most arguments of [67] can be adopted to our situation. We start by recalling some of the basic definitions and notions.

Let $\mathbb{H}_k=\{Z=X+iY:\ Z^t=Z,\ Y>0\}$ denote the Siegel upper half-plane of genus k. Following the definition (1.3.2) in [1], let $\theta_k:\mathbb{H}_k\times\mathbb{R}^k\times\mathbb{R}^k\to\mathbb{C}$ be the theta function defined by

$$\theta_k(Z,\xi,\eta)=\sum_{m\in\mathbb{Z}^k} e^{\pi i\,(\,Z(m-\eta)\cdot(m-\eta)+2m\cdot\xi-\,\xi\cdot\eta\,)}. \tag{2.2.74}$$

Note that the above sum converges uniformly on the domain $\{Z:\ Im\ Z>\varepsilon E_k\}$, for every $\varepsilon>0$. Here E_k is the $k\times k$ identity matrix, and by the notation $A\geq B$ we mean that $A-B>0$, that is a positive $k\times k$ matrix. Next, we define the theta functions $\theta_{n,k}:\mathbb{H}_k\times\mathbb{R}^{n\times k}\times\mathbb{R}^{n\times k}\to\mathbb{C}$. Let $\mathcal{X}=(\xi_1,\ldots,\xi_n)$, $\mathcal{E}=(\eta_1,\ldots,\eta_n)$ be $n\times k$ matrices with the i-th row being ξ (resp. η_i) for $1\leq i\leq n$. Define

$$\theta_{n,k}(Z,\mathcal{X},\mathcal{E})=\prod_{i=1}^{n}\theta_k(Z,\xi_i,\eta_i). \tag{2.2.75}$$

Using (2.2.74) , and the fact that $tr(AB)=tr(BA)$ for $A,B\in\mathbb{R}^{n\times k}$, one may also write

$$\theta_{n,k}(Z,\mathcal{X},\mathcal{E})=\sum_{M\in\mathbb{Z}^{n\times k}} e^{\pi i\,tr\big(\,(M-\mathcal{E})Z(M-\mathcal{E})^t+2M^t\mathcal{X}-\,\mathcal{E}^t\mathcal{X}\,\big)}. \tag{2.2.76}$$

These theta functions will play a crucial role. Indeed, one has

Proposition 2.7. *Let $T>0$ be an integral $k\times k$ matrix, and let $\mathcal{X}\in\mathbb{R}^{n\times k}$. Then*

$$|\widehat{S}_T(\mathcal{X})|\lesssim\int_{[0,2]^{\frac{k(k+1)}{2}}}|\theta_{n,k}(X+iT^{-1},-\mathcal{X},0)|\,dX \tag{2.2.77}$$

where $dX=\prod_{1\leq i\leq j\leq k}dx_{ij}$.

Proof. For simplicity of notation, let $I_k=[0,2]^{\frac{k(k+1)}{2}}$. If $M\in\mathbb{Z}^{n\times k}$, then

$$\int_{I_k} e^{\pi i\,tr\,(\,(M^tM\,-T)X)}\,dX=\begin{cases}2^{\frac{k(k+1)}{2}} & ,\text{ if } M^tM=T\\ 0 & ,\text{ otherwise}\end{cases}$$

If $MM^t=T$ then $tr(M^tMT^{-1})=tr(MT^{-1}M^t)=n$, thus

$$\begin{aligned}\widehat{S}_T(\mathcal{X})&=2^{-\frac{k(k+1)}{2}}e^{\pi n}\sum_{M\in\mathbb{Z}^{n\times k}}e^{-\pi\,tr(MT^{-1}M^t)}\int_{I_k}e^{\pi i\,tr(\,(M^tM\,-T)X-2M^t\mathcal{X})}\,dX\\ &=2^{-\frac{k(k+1)}{2}}e^{\pi n}\int_{I_k}e^{-\pi i\,tr(T\mathcal{X})}\sum_{M\in\mathbb{Z}^{n\times k}}e^{\pi i\,tr(\,M(X+iT^{-1})M^t-2M^t\mathcal{X})}\,dX.\end{aligned}$$

Note that the inner sum is: $\theta_{n,k}(X+iT^{-1},-\mathcal{X},0)$, which converges uniformly for $X\in I_k$, and hence the last equality is justified. Taking absolute values in the integral the proposition follows. $\square$

We will use the approach of [67], in estimating the integral in formula (2.2.74), by partitioning the range of integration I_k, and estimating the theta function separately on each part by exploiting its transformation properties. Note that in one dimension, when $k=1$, this leads to the so-called Farey arcs decomposition. Let

$$\Gamma_k = \left\{ \gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix};\ AB^t = BA^t,\ CD^t = DC^t,\ AD^t - BC^t = E_k \right\}. \tag{2.2.78}$$

denote the integral symplectic group. The group Γ_k acts on $\mathbb{H}_k$ as a group of analytic automorphisms, the action being defined by: $\gamma\langle Z\rangle = (AZ+B)(CZ+D)^{-1}$ for $\gamma \in \Gamma_k,\ Z \in \mathbb{H}_k$. Let us recall also the subgroup of integral modular substitutions:

$$\Gamma_{k,\infty} = \left\{ \gamma = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix};\ AB^t = BA^t,\ AD^t = E_k \right\} \tag{2.2.79}$$

It is immediate that writing $U = A^t$ and $S = AB^t$, that $D = U^{-1}$ and $B = SU^{-1}$, moreover S is symmetric, and $U \in GL(k,\mathbb{Z})$, that is: $\det(U) = \pm 1$. The action of such $\gamma \in \Gamma_{k,\infty}$ on $Z \in \mathbb{H}_k$ takes the form:

$$\gamma\langle Z\rangle = U^t A U + S. \tag{2.2.80}$$

we will adopt also the notation $Z[U] = U^t Z U$. The general linear group $GL(k,\mathbb{Z})$ acts on the space $\mathcal{P}_k$ of positive $k \times k$ matrices, via the action: $Y \to Y[U],\ Y \in \mathcal{P}_k$, and let $\mathcal{R}_k$ denote the corresponding so-called Minkowksi domain, see [KL, Definition 1, p12]. A matrix $Y = (y_{ij}) \in \mathcal{R}_k$ is called reduced. We recall that for a reduced matrix Y

$$Y \approx Y_D\ , \qquad y_{11} \le y_{22} \le \ldots \le y_{kk}. \tag{2.2.81}$$

where $Y_D = diag(y_{11},\ldots,y_{kk})$ denotes the diagonal part of Y, and $A \approx B$ means that $A - c_k B > 0$, $B - c_k A > 0$ for some constant $c_k > 0$. For a proof of these facts, see [KL,Lemma 2, p.20]. A fundamental domain $\mathcal{D}_k$ for the action of Γ_k on $\mathbb{H}_k$, called the Siegel domain, consists of all matrices $Z = X + iY$, $(X = (x_{ij}))$, satisfying

$$Y \in \mathcal{R}_k, \quad |x_{ij}| \le 1/2, \qquad |\det(CZ+D)| \ge 1, \quad \forall\ \gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_k. \tag{2.2.82}$$

The second rows of the matrices $\gamma \in \Gamma_k$ are parameterized by the so-called coprime symmetric pairs of integral matrices (C,D), which means that CD^t is symmetric and the matrices GC and GD with a matrix G of order k are both integral only if G is integral, see [1], Lemma 2.1.17.It is clear from definition (2.2.79) that if $\gamma_2 = \gamma\gamma_1$ with second rows (C_2,D_2) and (C_1,D_1) for some $\gamma \in \Gamma_{k,\infty}$, then $(C_2,D_2) = (UC_1,UD_1)$ for some $U \in GL(k,\mathbb{Z})$. On the other hand, if both γ_1 and γ_2 have the same second row (C,D) then $\gamma_2\gamma_1^{-1} \in \Gamma_{k,\infty}$. This gives the parameterization of the group $\Gamma_{k,\infty}\backslash\Gamma_k$ by equivalence classes of coprime symmetric pairs (C,D) via the equivalence relation $(C_2,D_2) \sim (C_1,D_1)$ if $(C_2,D_2) = (UC_1,UD_1)$ for some $U \in GL(k,\mathbb{Z})$, see also [1], p.54. We will use the notation $[\gamma] = [C,D] \in \Gamma_{k,\infty}\backslash\Gamma_k$.

It is clear that if one defines the domain: $\mathbb{F}_k = \cup_{\gamma\in\Gamma_{k,\infty}} \gamma\mathcal{D}_k$, then $\mathbb{H}_k = \bigcup_{[\gamma]\in\Gamma_{k,\infty}\backslash\Gamma_k} \gamma^{-1}\mathbb{F}_k$ is a non-overlapping cover of the Siegel upper half-plane. Correspondingly, for a given matrix $T > 0$ of order k, define the Farey arc dissection of level T, as the cover

$$I_k = \bigcup_{[\gamma]\in\Gamma_{k,\infty}\backslash\Gamma_k} I_T[\gamma], \qquad I_T[\gamma] = \{X \in I_k :\ X + iT^{-1} \in \gamma^{-1}\mathbb{F}_k\} \tag{2.2.83}$$

We will need the following transformation property of the functions $|\theta_{n,k}(Z,\mathcal{X},\mathcal{E})|$ with respect to $\gamma \in \Gamma_k$, which is immediate from Proposition 1.3.2 and Theorem 1.3.6 in [1], see formulas (1.3.7)-(1.3.10) there. Let $\xi, \eta \in \mathbb{R}^k$, $Z \in \mathbb{H}_k$, and $\gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_k$. Then one has

$$|\theta_k(Z,\xi,\eta)| = |\det(CZ+D)|^{-\frac{1}{2}}\ |\theta_k(\gamma\langle Z\rangle,\, A\xi - B\eta - k_\gamma/2,\, C\xi - D\eta - n_\gamma/2)|, \tag{2.2.84}$$

for some vectors $k_\gamma, n_\gamma \in \mathbb{Z}^k$ depending only on the matrix γ. If $\mathcal{X} = (\xi_1, \ldots, \xi_n)$ is a real $n \times k$ matrix with the i-th row being ξ_i, for $1 \leq i \leq n$, then by (2.2.75)

$$|\theta_{n,k}(Z,\mathcal{X},0)| = |\det(CZ+D)|^{-\frac{n}{2}}\ |\theta_{n,k}(\gamma\langle Z\rangle,\, \mathcal{X}A^t - K_\gamma/2,\, \mathcal{X}C^t - N_\gamma/2)|, \tag{2.2.85}$$

for some matrices $K_\gamma, N_\gamma \in \mathbb{Z}^{n\times k}$ depending only on the matrix γ. Let us recall the following quantity associated to a positive matrix $Y \in \mathbb{R}^{k\times k}$.

$$min(Y) = \min_{x\in\mathbb{Z}^n, x\neq 0} Yx\cdot x. \tag{2.2.86}$$

It is clear that $\mu(Y) \leq min(Y)$, and it follows from (5.8) that $\mu(Y) \approx min(Y)$ if Y is reduced.

Proposition 2.8. *Let $\mathcal{X} \in \mathbb{R}^{n\times k}$, $T \in \mathbb{Z}^{k\times k}$ such that $T > 0$, and $\tau > 0$ be given. If (C, D) is a coprime symmetric pair, then for $Z \in I_T[C,D]$ one has*

$$|\theta_{n,k}(Z,\mathcal{X},0)| \lesssim |\det(CZ+D)|^{-\frac{n}{2}}. \tag{2.2.87}$$

Let $q = \det(C)$, $[\gamma] = [C,D]$ and $Y = Im\gamma\langle Z\rangle$. If $q \neq 0$, and for every $P \in \mathbb{Z}^{n\times k}$

$$|\mathcal{X} - P/2q| \geq \tau \tag{2.2.88}$$

then one has

$$|\theta_{n,k}(Z,\mathcal{X},0)| \lesssim |\det(CZ+D)|^{-\frac{n}{2}} \left(e^{-c\min(Y)} + e^{-c\tau^2\mu(C^tYC)}\right), \tag{2.2.89}$$

for some constant $c > 0$ depending only on the dimension k.

Proof. By formula (2.2.84) it is enough to show that

$$|\theta_{n,k}(\gamma\langle Z\rangle,\, \mathcal{X}A^t - K_\gamma/2,\, \mathcal{X}C^t - N_\gamma/2)| \lesssim 1. \tag{2.2.90}$$

Since $\gamma\langle Z\rangle \in \mathbb{F}_k$, there is a $U \in GL(k,\mathbb{Z})$ and a symmetric $S \in \mathbb{Z}^{k\times k}$, such that $\gamma\langle Z\rangle = U^tZ_1U + S$ with $Z_1 \in \mathcal{D}_k$. Taking absolute values in (2.2.76) one obtains, using the notation $A[B] = B^tAB$,

$$\begin{aligned} |\theta_{n,k}(\gamma\langle Z\rangle,\, \mathcal{X}A^t - K_\gamma/2,\, \mathcal{X}C^t - N_\gamma/2)| &\leq \sum_{M\in\mathbb{Z}^{n\times k}} e^{-\pi\, tr(Y[C\mathcal{X}^t - M^t - N_\gamma^t/2])} \\ &= \sum_{M_1\in\mathbb{Z}^{n\times k}} e^{-\pi\, tr(Y_1[C_1\mathcal{X}^t - M_1^t - N_1^t/2])}, \end{aligned} \tag{2.2.91}$$

where $M_1 = MU^t$ runs through $Z^{n\times k}$, $C_1 = UC$, $N_1 = N_\gamma U^t$ and $Y_1 = Im\,Z_1 = U^tYU$. Since $Z_1 \in \mathcal{D}_k$, $Y_1 \geq c_kE_k$ for some constant $c_k > 0$. Let $M_0 \in \mathbb{Z}^{n\times k}$ be such that

$$|\mathcal{X}C_1^t - M_0 - N_1/2| = \min_{M\in\mathbb{Z}^{n\times k}} |\mathcal{X}C_1^t - M - N_1/2|,$$

and write $M_2 = M_1 - M_0$. Since $Z_1 \in \mathcal{D}_k$, one has that $\mu(Y_1) \approx min(Y_1) \gtrsim 1$ see [1], thus the right side of (2.2.91) is further estimated by

$$e^{-c'|\mathcal{X}C_1^t - M_0 - N_1/2|^2} + \sum_{M_2 \neq 0} e^{-c' min(Y_1)\,|M_2|^2} \lesssim 1. \tag{2.2.92}$$

If $q = \det(C) \neq 0$ and one assumes (2.2.89), then $\det(C_1) = \pm q \neq 0$ and $(M_0 + N_1/2)(C_1^t)^{-1} = P/2q$ for some $P \in \mathbb{Z}^{n\times k}$. Thus

$$tr(Y_1[C_1\mathcal{X}^t - M_0^t - N_1^t/2]) = tr((C_1^t Y_1 C_1)\,[\mathcal{X}^t - P^t/2q]) \geq \tau^2 \mu(C_1^t Y_1 C_1).$$

Thus the expression in formula (2.2.92) is bounded by

$$e^{-c\,\tau^2 \mu(C_1^t Y_1 C_1)} + e^{-c\min(Y_1)},$$

for some constant $c > 0$ depending only on k. Since $C_1^t Y_1 C_1 = C^t Y C$ and $min(Y_1) = min(Y)$, the proposition is proved. $\square$

We will estimate below the sum of the integrals

$$J_{T,\mathcal{X}}[C,D] = \int_{I_T[C,D]} |\theta_{n,k}(Z,\mathcal{X},0)|\,dX, \tag{2.2.93}$$

over all coprime symmetric pairs $[C,D]$, using bounds (2.2.87) and (2.2.89). Most of the estimates needed, were done in [67] in the proofs of Propositions 1.4.10 and 1.4.11, which we recall without proofs, however we give detailed proofs of similar estimates not discussed in [67].
To be more precise, define the quantities

$$J_T^0[C,D] = \int_{I_T[C,D]} |\det(CZ+D)|^{-\frac{n}{2}}\,dX. \tag{2.2.94}$$

$$J_T^1[C,D] = \int_{I_T[C,D]} |\det(CZ+D)|^{-\frac{n}{2}}\,e^{-c\,min(Y)}\,dX. \tag{2.2.95}$$

$$J_{T,\tau}^2[C,D] = \int_{I_T[C,D]} |\det(CZ+D)|^{-\frac{n}{2}}\,e^{-c\tau^2\,\mu(C^tYC)}\,dX, \tag{2.2.96}$$

where $Y = Im\,\gamma\langle Z\rangle$ and $\gamma \in \Gamma_k$ such that $[\gamma] = [C,D] \in \Gamma_{k,\infty}\backslash\Gamma$. The following estimates are proved in [K], (see Proposition 1.4.10 together with Lemma 1.4.4. and estimate (39) there)

Proposition 2.9. *Let T be a positive integral matrix, and let $[C,D]$ be a coprime symmetric pair such that* $\det(C) \neq 0$. *Then one has the following estimates*

$$\sum_{S^t=S} J_T^0[C, D+CS] \;\lesssim\; \det(T)^{\frac{n-k-1}{2}}\;|\det(C)|^{-\frac{n}{2}}. \tag{2.2.97}$$

$$\sum_{S^t=S} J_T^1[C, D+CS] \;\lesssim\; \det(T)^{\frac{n-k-1}{2}}\;|\det(C)|^{-k}\;min(T)^{-\frac{n-2k}{4}}, \tag{2.2.98}$$

where the summation is taken over all symmetric integral matrices S.

Using the same argument as in the proof of the above statements given in [67], one obtains

Proposition 2.10. *Let T be a positive integral matrix, let $\tau > 0$, and let $[C, D]$ be a coprime symmetric pair such that* $\det(C) \neq 0$. *Then*

$$\sum_{S^t=S} J^2_{T,\tau}[C, D + CS] \;\lesssim\; \det(T)^{\frac{n-k-1}{2}} \;|\det(C)|^{-\frac{n}{2}}\; (\tau^2\mu(T))^{-\frac{n-2k}{4}}, \tag{2.2.99}$$

where the summation is taken over all symmetric integral matrices S.

Proof. Using the fact that

$$Im\,\gamma\langle Z\rangle = ((CZ + D)(Im\,Z)^{-1}(C\bar{Z} + D)^t)^{-1},$$

it follows for $Z = X + iT^{-1}$ that

$$Y[C] = (T\,[X + C^{-1}D + iT^{-1}])^{-1} = (T[X + C^{-1}D] + T^{-1})^{-1}.$$

Thus by (2.2.96)

$$\begin{aligned}
&\sum_{S^t=S} J^2_{T,\tau}[C, D + CS] \;\lesssim \qquad\qquad (2.2.100)\\
&\lesssim \sum_{S^t=S} \int_{I_k} |\det(C)|^{-\frac{n}{2}}\, |\det(X + C^{-1}D + S + iT^{-1})|^{-\frac{n}{2}}\; e^{-c\tau^2\,\mu(T[X+C^{-1}D+S]+T^{-1})^{-1}}\, dX\\
&\lesssim |\det(C)|^{-\frac{n}{2}} \int_{\mathbb{R}^{\frac{k(k+1)}{2}}} |\det(X + iT^{-1})|^{-\frac{n}{2}}\, e^{-c\tau^2\,\mu((T[X]+iT^{-1})^{-1})}\, dX.
\end{aligned}$$

Let $T^{\frac{1}{2}}$ denote the positive square root of T, and let $X_1 = X[T^{\frac{1}{2}}]$. Then by a change of variables $dX = \det(T)^{-\frac{k+1}{2}} dX_1$, the expression in (2.2.100) takes the form

$$\det(T)^{\frac{n-k-1}{2}}\, |\det(C)|^{-\frac{n}{2}} \int_{\mathbb{R}^{\frac{k(k+1)}{2}}} \det(X_1^2 + E_k)^{-\frac{n}{4}}\, e^{-c\tau^2\,\mu((X_1^2+E_k)^{-1}[T^{\frac{1}{2}}])}. \tag{2.2.101}$$

Note that the above expression depends just on the conjugacy class of the symmetric matrix X_1. Thus writing $X_1 = V^t\, diag(w_1, \ldots, w_k)\, V$ for some orthogonal matrix $V \in O(k)$, with $|w_1| \geq \ldots \geq |w_k|$ being the eigenvalues of the matrix X_1, it follows that

$$\mu(T^{\frac{1}{2}}(X_1^2 + E_k)^{-1}T^{\frac{1}{2}}) \geq (1 + w_1^2)^{-1}\,\mu(T).$$

By the Weyl integral formula:

$$dX_1 = \prod_{1\leq i<j\leq k} |w_i - w_j|\;\; dw_1 \ldots dw_k\, dV \leq \prod_{1\leq i\leq k} (1 + w_i^2)^{\frac{k-1}{2}}\;\; dw_1 \ldots dw_k\, dV.$$

Since $n > 2k$, using the above change of variables, one estimates the integral in (2.2.101) by

$$\int_{\mathbb{R}^k} (1 + w_1^2)^{-\frac{n}{4}+\frac{k-1}{2}}\;\, e^{-c\tau^2\mu(T)\,(1+w_1^2)^{-1}}\, dw_1 \;\lesssim\; (\tau^2\mu(T))^{-\frac{n-2k}{4}}.$$

This proves the proposition. $\square$

The map $[C, D] \to C^{-1}D$ provides a one-one and onto correspondence between the classes of coprime symmetric pairs $[C, D] \in \Gamma_{k,\infty}\backslash\Gamma_k$ and the space of symmetric rational matrices R of order k, see Lemma 1.4.6 in [67]. Note that the pairs $[C, D + CS]$ correspond to the matrices

$R+S$ with symmetric $S\in\mathbb{Z}^{k\times k}$. Thus using Proposition 2.10 9, one needs to estimate the sum of $\sum_{S^t=S} J_{T,\mathcal{X}}[C,D+CS]=J_{T,\mathcal{X}}[R]$ over the space of modulo 1 incongruent symmetric rational matrices, which we will denote by $\mathbb{Q}(1)^{k\times k}$, where $\mathbb{Q}(1)=\mathbb{Q}/\mathbb{Z}$, $\mathbb{Q}$ being the set of rational numbers. Let us introduce the notation: $d(R)=|\det(C)|$ for $R=C^{-1}D$, and recall the following estimate, proved in Lemma 1.4.9 in [67]; for $u>0$ and $s>1$ one has

$$u^{-s}\sum_{1\leq d(R)\leq u} d(R)^{-k}+\sum_{d(R)>u} d(R)^{-k-s}\lesssim (2+\frac{1}{s-1})\,u^{1-s}, \tag{2.2.102}$$

where the summation is taken over $[R]\in\mathbb{Q}(1)^{k\times k}$.

Proposition 2.11. *Let T be a positive integral matrix, let $\tau>0$ and $q_0\in\mathbb{N}$. Let $\mathcal{X}\in\mathbb{R}^{n\times k}$ such that for all $1\leq q\leq q_0$ and $P\in\mathbb{Z}^{n\times k}$*

$$|\mathcal{X}-P/q|\geq\tau. \tag{2.2.103}$$

Then one has

$$\sum_{R\in\mathbb{Q}(1)^{k\times k},\, d(R)\neq 0} J_{T,\mathcal{X}}[R]\ \lesssim\ \det(T)^{\frac{n-k-1}{2}}\left((\tau^2\mu(T))^{-\frac{n-2k-2}{4}}+q_0^{-\frac{n-2k-2}{2}}\right). \tag{2.2.104}$$

Proof. By Propositions 2.8-2.9 one has

$$J_{T,\mathcal{X}}[R]\ \lesssim\ \det(T)^{\frac{n-k-1}{2}}\,d(R)^{-\frac{n}{2}},$$

thus by (2.2.102) applied for $s=n/2-k>1$ and $u=1$

$$\sum_{d(R)\neq 0} J_{T,\mathcal{X}}[R]\ \lesssim\ \det(T)^{\frac{n-k-1}{2}}. \tag{2.2.105}$$

If $\mathcal{X}$ satisfies (2.2.103) then for $1\leq d(R)\leq q_0/2$ one has by (2.2.90) and (2.2.98)-(2.2.99)

$$J_{T,\mathcal{X}}[R]\ \lesssim\ \det(T)^{\frac{n-k-1}{2}}\left(d(R)^{-\frac{n}{2}}(\tau^2\mu(T))^{-\frac{n-2k}{4}}+d(R)^{-k}min(T)^{-\frac{n-2k}{4}}\right).$$

Clearly $|\tau|\lesssim 1$, thus $\tau^2\mu(T)\lesssim min(T)$ so the right side is bounded by

$$J_{T,\mathcal{X}}[R]\ \lesssim\ \det(T)^{\frac{n-k-1}{2}}\,d(R)^{-k}(\tau^2\mu(T))^{-\frac{n-2k}{4}}. \tag{2.2.106}$$

By inequality (2.2.102) applied for $s=n/2-k$, $u=q_0/2$

$$\sum_{d(R)\neq 0} J_{T,\mathcal{X}}[R]\ \lesssim\ \det(T)^{\frac{n-k-1}{2}}\,(q_0\,(\tau^2\mu(T))^{-\frac{n-2k}{4}}+q_0^{-\frac{n-2k-2}{2}}),$$

which is bounded by the right side of formula (2.2.104). □

Next, we estimate the sum $J_{T,\mathcal{X}}[C,D]$ over the classes $[C,D]$ of coprime symmetric pairs for which $\det(C)=0$. We will use the estimate

$$J_{T,\mathcal{X}}[C,D]\ \lesssim\ J^0_T[C,D]=\int_{I_T[C,D]}|\det(CZ+D)|^{-\frac{n}{2}}\,dX.$$

which follows from (2.2.88) and (2.2.94). First we show that one may assume T is reduced in our estimates below.

Proposition 2.12. *Let $T \in \mathbb{Z}^{k\times k}$ such that $T > 0$ and let $T_1 = T[V]$ for some $V \in GL(k,\mathbb{Z})$. Let $0 \le r < k$, and let $rank(C)$ stand for the rank of the matrix C. Then*

$$\sum_{[C,D],\ rank(C)=r} J^0_{T_1}[C,D] = \sum_{[C,D],\ rank(C)=r} J^0_T[C,D] \tag{2.2.107}$$

Proof. Let $U \in GL(k,\mathbb{Z})$ such that $U^{-1} = V^t$. Then $T^{-1} = T_1^{-1}[U^{-1}]$, and writing $Z = X + iT^{-1}$ for $Z \in I_T[C,D]$ a straightforward calculation shows that

$$|\det(CZ+D)| = |\det(C_1Z_1+D_1)|,$$

with $C_1 = C(U^t)^{-1},\ D_1 = DU,\ X_1 = X[U^{-1}]$ and $Z_1 = X_1 + iT_1^{-1}$. Notice that $Z_1 = h\langle Z\rangle$ with $h = \begin{pmatrix} (U^t)^{-1} & 0 \\ 0 & U \end{pmatrix}$, and if $\gamma = \begin{pmatrix} * & * \\ C & D \end{pmatrix}$ then $\gamma \cdot h = \gamma_1$ with $\gamma_1 = \begin{pmatrix} * & * \\ C_1 & D_1 \end{pmatrix}$. It follows $\gamma\langle Z\rangle = \gamma_1\langle Z_1\rangle$, hence $X \in I_T[C,D]$ exactly when $X_1 = I_{T_1}[C_1,D_1]$ and one has

$$\int_{I_T[C,D]} |\det(CZ+D)|^{-\frac{n}{2}}\ dX \ = \ \int_{I_{T_1}[C_1,D_1]} |\det(CZ_1+D_1)|^{-\frac{n}{2}}\ dX_1$$

The map $[C,D] \to [C_1,D_1] = [C(U^t)^{-1}, DU]$ is one-one and onto from the classes of coprime symmetric pairs $[C,D]$ with $rank(C) = r$ to itself, and the proposition is proved. □

Let $T > 0$ be integral, and let $T_1 = T[U]$ be reduced, with $U \in GL(k,\mathbb{Z})$. We recall that $T_1 \approx diag(t_{1,1},\ldots,t_{k,k})$, where $t_{i,i}$ $(1 \le i \le k)$ denote the diagonal entries of the matrix T_1, see (2.2.81). For reduced matrices the estimate of the sum in (2.2.107) goes back to [98], and is given in Lemma 1.4.11 in [67], which we recall without proofs, see formulas (39) and (43)-(44) there.

Proposition 2.13. *Let $T_1 \in \mathbb{Z}^{k\times k}$ be reduced, and let $0 \le r < k$. Then*

$$\sum_{[C,D],\ rank(C)=r} J^0_{T_1}[C,D] \ \lesssim\ (t_{k,k}\cdot\ldots\cdot t_{k-r+1,k-r+1})^{\frac{n-r-1}{2}},$$

where $t_{i,i}$ $(1 \le i \le k)$ denote the diagonal entries of the matrix T_1.

It is easy to see that

$$e(T_1) \lesssim e(T). \tag{2.2.108}$$

Indeed,

$$t_{1,1} = Te_1\cdot e_1 = T(Ue_1)\cdot Ue_1 \ge \mu(T) \qquad \text{and}$$

$$|T| \ \ge\ \sup_{|x|=1} T_1(U^{-1}x)\cdot U^{-1}x \ \gtrsim\ \sup_{|x|=1} t_{k,k}\,(U^{-1}x)_k^2 \ \ge\ t_{k,k},$$

as U^{-1} is integral, where $(U^{-1}x)_k$ denotes the k-th entry of the vector $U^{-1}x$.

Finally, one has $r(n-r-1) \le (k-1)(n-k)$ for $0 \le r \le k-1$, thus Proposition 2.13 and inequality (2.2.108) implies

Corollary 2.1. *Let $T \in \mathbb{Z}^{k\times k}$ such that $T > 0$. Then*

$$\sum_{[C,D],\ \det(C)=0} J^0_T[C,D] \ \lesssim\ |T|^{\frac{(k-1)(n-k)}{2}}.$$

Note that a proof of this corollary is also given in [95], see formulas (25)-(26) there.
Lemma 2.2 follows immediately from Proposition 2.11 and Corollary 2.1.
This finishes to proof of Theorem 2.3.

3 A weak hypergraph regularity lemma and products of simplices in sets of positive density

As described in Section 2 Graham has conjectured that Bourgain's simplex theorem extends to all finite spherical configurations S in Euclidean spaces. The first breakthrough in this direction was obtained in joint work with Lyall [76] where it was shown that configurations of four points forming a 2-dimensional geometric rectangle, and more generally any configuration that is the direct product of two non-degenerate simplices in $\mathbb{R}^n$ are Ramsey. This section is based on the recent joint work [79] where the results in [76] were extended to cover configurations with a a higher dimensional product structure in both the settings of Euclidean spaces and the integer lattice $\mathbb{Z}^d$. Let us state our main results.

By a d-dimensional rectangle in a space $\mathbb{R}^D$ (or in $\mathbb{Z}^D$) we mean a set of the form $\mathcal{R} = \{x_{11}, x_{12}\} \times \cdots \times \{x_{d1}, x_{d2}\}$ with side vectors $v_i = x_{i2} - x_{i1}$ $(i = 1, \ldots, d)$ being pairwise orthogonal.

Theorem 3.1. *Let $\mathcal{R}$ be 2^d points forming the vertices of a fixed d-dimensional rectangle in $\mathbb{R}^{2d}$.*

(i) If $S \subseteq \mathbb{R}^{2d}$ has positive upper Banach density, then there exists a threshold $\lambda_0 = \lambda_0(S, \mathcal{R})$ such that S contains an isometric copy of $\lambda\mathcal{R}$ for all $\lambda \geq \lambda_0$.

(ii) For any $0 < \delta \leq 1$ there exists a constant $c = c(\delta, \mathcal{R}) > 0$ such that any $S \subseteq [0,1]^{2d}$ with $|S| \geq \delta$ is guaranteed to contain an isometric copy of $\lambda\mathcal{R}$ for all λ in some interval of length at least c.

Moreover, if $\mathcal{R}$ has sidelengths given by $t_1, \ldots, t_d$, then the isometric copies of $\lambda\mathcal{R}$ in both (i) *and* (ii) *above can all be realized in the special form $\{x_{11}, x_{12}\} \times \cdots \times \{x_{d1}, x_{d2}\} \subseteq \mathbb{R}^2 \times \cdots \times \mathbb{R}^2$ with each $|x_{j2} - x_{j1}| = \lambda t_j$.*

More generally, we have the following extension of Bourgain's simplex theorem.

Theorem 3.2. *Let $\Delta = \Delta_1 \times \cdots \times \Delta_d \subseteq \mathbb{R}^n$, where $\mathbb{R}^n = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$ and each $\Delta_j \subseteq \mathbb{R}^{n_j}$ is a non-degenerate simplex of n_j points.*

(i) If $S \subseteq \mathbb{R}^n$ has positive upper Banach density, then there exists a threshold $\lambda_0 = \lambda_0(S, \Delta)$ such that S contains an isometric copy of $\lambda\Delta$ for all $\lambda \geq \lambda_0$.

(ii) For any $0 < \delta \leq 1$ there exists a constant $c = c(\delta, \Delta) > 0$ such that any $S \subseteq [0,1]^n$ with $|S| \geq \delta$ is guaranteed to contain an isometric copy of $\lambda\Delta$ for all λ in some interval of length at least c.

Moreover the isometric copies of $\lambda\Delta$ in both (i) *and* (ii) *above can all be realized in the special form $\Delta'_1 \times \cdots \times \Delta'_d$ with each $\Delta'_j \subseteq \mathbb{R}^{n_j}$ an isometric copy of $\lambda\Delta_j$.*

Our main results in this Section are in the context of sets of positive upper density of the integer lattice.

Theorem 3.3. *Let $0 < \delta \leq 1$ and $\mathcal{R}$ be 2^d points forming the vertices of a d-dimensional rectangle in $\mathbb{Z}^{5d}$.*

(i) If $S \subseteq \mathbb{Z}^{5d}$ has upper Banach density at least δ, then there exist integers $q_0 = q_0(\delta, \mathcal{R})$ and $\lambda_0 = \lambda_0(S, \mathcal{R})$ such that S contains an isometric copy of $q_0\lambda\mathcal{R}$ for all $\lambda \in \sqrt{\mathbb{N}}$ with $\lambda \geq \lambda_0$.

(ii) There exists a constant $N(\delta,\mathcal{R})$ such that if $N \geq N(\delta,\mathcal{R})$, then any $S \subseteq \{1,\dots,N\}^{5d}$ with cardinality $|S| \geq \delta N^{5d}$ will necessarily contain an isometric copy of $\lambda\mathcal{R}$ for some $\lambda \in \sqrt{\mathbb{N}}$ with $1 \leq \lambda \leq N$.

If $\mathcal{R}$ has side lengths given by $t_1,\dots,t_d$, then each of the isometric copies in (i) *and* (ii) *above can be realized in the form $\{x_{11},x_{12}\} \times \cdots \times \{x_{d1},x_{d2}\} \subseteq \mathbb{Z}^5 \times \cdots \times \mathbb{Z}^5$ with each $|x_{j2}-x_{j1}| = q_0\lambda t_j$ and λt_j, respectively.*

The above results extend to more general patterns where d-dimensional rectangles are replaced with direct products of non-degenerate simplices.

Theorem 3.4. *Let $0 < \delta \leq 1$ and $\Delta = \Delta_1 \times \cdots \times \Delta_d \subseteq \mathbb{Z}^n$, where $\mathbb{Z}^n = \mathbb{Z}^{2n_1+3} \times \cdots \times \mathbb{Z}^{2n_d+3}$ and each $\Delta_i \subseteq \mathbb{Z}^{2n_i+3}$ is a non-degenerate simplex of n_i points.*

(i) If $S \subseteq \mathbb{Z}^n$ has upper Banach density at least δ, then there exist integers $q_0 = q_0(\delta,\Delta)$ and $\lambda_0 = \lambda_0(S,\Delta)$ such that S contains an isometric copy of $q_0\lambda\Delta$ for all $\lambda \in \sqrt{\mathbb{N}}$ with $\lambda \geq \lambda_0$.

(ii) There exists a constant $N(\delta,\Delta)$ such that if $N \geq N(\delta,\Delta)$, then any $S \subseteq \{1,\dots,N\}^n$ with cardinality $|S| \geq \delta N^n$ will necessarily contain an isometric copy of $\lambda\Delta$ for some $\lambda \in \sqrt{\mathbb{N}}$ with $1 \leq \lambda \leq N$.

Moreover, each of the isometric copies in (i) *and* (ii) *above can be realized in the special form $\Delta_1' \times \cdots \times \Delta_d'$ with each $\Delta_i' \subseteq \mathbb{Z}^{2n_i+3}$ an isometric copy of $q_0\lambda\Delta_j$ and $\lambda\Delta_j$, respectively.*

The constants $N(\delta,\Delta)$ and $q_0(\delta,\Delta)$ can be taken less than $\exp^{(d)}(C_\Delta \delta^{-13n_1\cdots n_d})$ where $\exp^{(k)}(m)$ is a k-fold tower of exponentials defined by $\exp^{(1)}(m) = \exp(m)$ and $\exp^{(k+1)}(m) = \exp(\exp^{(k)}(m))$, for $k \geq 1$.

3.1 Outline and notations.

Our proofs are based on adapting Gowers type box-norms [48] and on developing a weak hypergraph regularity lemma [39, 108] and an associated counting lemma, in the context of Euclidean spaces and the integer lattice. As the notations in the general case are quite cumbersome, in Section 3.2 we introduce our approach in the model case of finite fields and prove an analogue of Theorem 3.1 in this setting. In Section 3.3 we review Theorem 3.2 for a single simplex and ultimately establish the base case of our general inductive approach to Theorem 3.2. The general case which we present in the Section 3.4. The proof of Theorem 3.4 is outlined in Sections 3.5 and 3.6.

We will consider the parameters $d, n_1, \dots, n_d$ fixed and will not indicate the dependence on them. Thus we will write $f = O(g)$ or alternatively $f \ll g$ if $|f| \leq C(n_1,\dots,n_d)g$. If the implicit constants in our estimates depend on additional parameters $\varepsilon, \delta, K, \dots$ the we will write $f = O_{\varepsilon,\delta,K,\dots}(g)$ or $f \lesssim_{\varepsilon,\delta,K,\dots} g$.

Given an $\varepsilon > 0$ and a (finite or infinite) sequence $L_0 \geq L_1 \geq \cdots > 0$, we will say that the sequence is *ε-admissible* if $L_j/L_{j+1} \in \mathbb{N}$ and $L_{j+1} \ll \varepsilon^2 L_j$ for all $j \geq 1$. Moreover, if $q \in \mathbb{N}$ is given and $L_j \in \mathbb{N}$ for all $1 \leq j \leq J$, then we will call the sequence $L_0 \geq L_1 \geq \cdots \geq L_J$ *(ε, q)-admissible* if in addition $L_J/q \in \mathbb{N}$. Such sequences of scales will often appear in our statements both in the continuous and the discrete case.

3.2 Model case: vector spaces over finite fields.

In this section we will illustrate our general method by giving a complete proof of Theorem 3.1 in the model setting of $\mathbb{F}_q^n$ where $\mathbb{F}_q$ denotes the finite field of q elements. We do this as the notation and arguments are more transparent in this setting yet many of the main ideas are still present.
We say that two vectors $u, v \in \mathbb{F}_q^n$ are orthogonal, if $x \cdot y = 0$, where "$\cdot$" stands for the usual dot product. A rectangle in $\mathbb{F}_q^n$ is then a set $\mathcal{R} = \{x_1, y_1\} \times \cdots \times \{x_n, y_n\}$ with side vectors $y_i - x_i$ being pairwise orthogonal.
The finite field analogue of Theorem 3.1 is the following

Proposition 3.1. *For any $0 < \delta \leq 1$ there exists an integer $q_0 = q_0(\delta)$ with the following property: If $q \geq q_0$ and $t_1, \ldots, t_d \in \mathbb{F}_q^*$, then any $S \subseteq \mathbb{F}_q^{2d}$ with $|S| \geq \delta\, q^{2d}$ will contain points*

$$\{x_{11}, x_{12}\} \times \cdots \times \{x_{d1}, x_{d2}\} \subseteq V_1 \times \cdots \times V_d \quad \textit{with} \quad |x_{j2} - x_{j1}|^2 = t_j \textit{ for } 1 \leq j \leq d$$

where we have written $\mathbb{F}_q^{2d} = V_1 \times \cdots \times V_d$ with $V_j \simeq \mathbb{F}_q^2$ pairwise orthogonal coordinate subspaces.

3.2.1 Overview of the proof of Proposition 3.1

Write $\mathbb{F}_q^{2d} = V_1 \times \ldots \times V_d$ with $V_j \simeq \mathbb{F}_q^2$ pairwise orthogonal coordinate subspaces. For any $\underline{t} := (t_1, \ldots, t_d) \in \mathbb{F}_q^*$ and $S \subseteq \mathbb{F}_q^{2d}$ we define

$$\mathcal{N}_{\underline{t}}(1_S) := \mathbb{E}_{\underline{x}_1 \in V_1^2, \ldots, \underline{x}_d \in V_d^2} \prod_{(\ell_1, \ldots, \ell_d) \in \{1,2\}^d} 1_S(x_{1\ell_1}, \ldots, x_{d\ell_d}) \prod_{j=1}^{d} \sigma_{t_j}(x_{j2} - x_{j1})$$

where we used the shorthand notation $\underline{x}_j := (x_{j1}, x_{j2})$ for each $1 \leq j \leq d$ and the averaging notation:

$$\mathbb{E}_{x \in A} f(x) := \frac{1}{|A|} \sum_{x \in A} f(x)$$

for a finite set $A \neq \emptyset$. We have also used the notation

$$\sigma_t(x) = \begin{cases} q & \text{if } |x|^2 = t \\ 0 & \text{otherwise} \end{cases}$$

for each $t \in \mathbb{F}_q^*$. Note that the function σ_t may be viewed as the discrete analogue of the normalized surface area measure on the sphere of radius $\sqrt{t}$. It is well-known, see [62], that

$$\mathbb{E}_{x \in \mathbb{F}_q^2}\ \sigma_t(x) = 1 + O(q^{-1/2})$$

and for all $\xi \neq 0$ one has

$$\hat{\sigma}_t(\xi) := \mathbb{E}_{x \in \mathbb{F}_q^2}\ \sigma_t(x)\, e^{2\pi i \frac{x \cdot \xi}{q}} = O(q^{-1/2}).$$

Note that if $\mathcal{N}_{\underline{t}}(1_S) > 0$, then this implies that S contains a rectangle of the form $\{x_{11}, x_{12}\} \times \cdots \times \{x_{d1}, x_{d2}\}$ with $x_{j1}, x_{j2} \in V_j$ and $|x_{j2} - x_{j1}|^2 = t_j$ for $1 \leq j \leq d$.
Our approach to Proposition 3.1 in fact establishes the following quantitatively stronger result.

Proposition 3.2. *For any $0 < \varepsilon \leq 1$ there exists an integer $q_0 = q_0(\varepsilon)$ with the following property: If $q \geq q_0$, then for any $S \subseteq \mathbb{F}_q^{2d}$ and $t_1, \ldots, t_d \in \mathbb{F}_q^*$ one has*

$$\mathcal{N}_{\underline{t}}(1_S) > \left(\frac{|S|}{q^{2d}}\right)^{2^d} - \varepsilon$$

where we have written $\mathbb{F}_q^{2d} = V_1 \times \ldots \times V_d$ with $V_j \simeq \mathbb{F}_q^2$ pairwise orthogonal coordinate subspaces.

A crucial observation in the proof of Proposition 3.2 is that the averages $\mathcal{N}_{\underline{t}}(1_S)$ can be compared to ones which can be easily estimated from below. We define, for any $S \subseteq \mathbb{F}_q^{2d}$, the (unrestricted) count

$$\mathcal{M}(1_S) := \mathbb{E}_{\underline{x}_1 \in V_1^2, \dots, \underline{x}_d \in V_d^2} \prod_{(\ell_1, \dots, \ell_d) \in \{1,2\}^d} 1_S(x_{1\ell_1}, \dots, x_{d\ell_d}).$$

It is easy to see, by carefully applying Cauchy-Schwarz d times to $\mathbb{E}_{x_{11} \in V_1, \dots, x_{d1} \in V_d} 1_S(x_{11}, \dots, x_{d1})$, that

$$\mathcal{M}(1_S) \geq \left(\frac{|S|}{q^{2d}} \right)^{2^d}. \tag{3.2.1}$$

Our approach to Proposition 3.2 therefore reduces to establishing that for any $\varepsilon > 0$ one has

$$\mathcal{N}_{\underline{t}}(1_S) \;=\; \mathcal{M}(1_S) + O(\varepsilon) + O_\varepsilon(q^{-1/2}). \tag{3.2.2}$$

The validity of (3.2.2) will follow immediately from the $d = k$ case of Proposition 3.3 below. However, before we can state this *counting lemma* we need to introduce some further notation from the theory of hypergraphs, notation that we shall ultimately make use of throughout the paper.

3.2.2 Hypergraph notation and a counting lemma

In order to streamline our notation we will make use the language of hypergraphs. For $J := \{1, \dots, d\}$ and $1 \leq k \leq d$, we let $\mathcal{H}_{d,k} = \{e \subseteq J;\ |e| = k\}$ denote the full k-regular hypergraph on the vertex set J. For $K := \{jl;\ j \in J, l \in \{1,2\}\}$ we define the projection $\pi : K \to J$ as $\pi(jl) := j$ and use this in turn to define the hypergraph bundle

$$\mathcal{H}_{d,k}^{\underline{2}} := \{e \subseteq K;\ |e| = |\pi(e)| = k\}$$

using the shorthand notation $\underline{2} = (2, 2, \dots, 2)$ to indicate that $|\pi^{-1}(j)| = 2$ for all $j \in J$.
Notice when $k = d$ then $\mathcal{H}_{d,d}$ consists of one element, the set $e = \{1, \dots, d\}$, and

$$\mathcal{H}_{d,d}^{\underline{2}} = \{\, \{1l_1, \dots, dl_d\};\ (l_1, \dots, l_d) \in \{1,2\}^d\}.$$

Let $V := \mathbb{F}_q^{2d}$ and $V = V_1 \times \ldots \times V_d$ with $V_j \simeq \mathbb{F}_q^2$ pairwise orthogonal coordinate subspaces. For a given $\underline{x} = (x_{11}, x_{12}, \dots, x_{d1}, x_{d2}) \in V^2$ with $x_{j1}, x_{j2} \in V_j$ and a given edge $e = \{1l_1, \dots, dl_d\}$, we write

$$\underline{x}_e := (x_{1l_1}, \dots, x_{dl_d}).$$

Note that the map $\underline{x} \to \underline{x}_e$ defines a projection $\pi_e : V^2 \to V$. With this notation, we can clearly now write

$$\mathcal{N}_{\underline{t}}(1_S) = \mathbb{E}_{\underline{x} \in V^2} \prod_{e \in \mathcal{H}_{d,d}^{\underline{2}}} 1_S(\underline{x}_e) \prod_{j=1}^{d} \sigma_{t_j}(x_{j2} - x_{j1})$$

$$\mathcal{M}(1_S) = \mathbb{E}_{\underline{x} \in V^2} \prod_{e \in \mathcal{H}_{d,d}^{\underline{2}}} 1_S(\underline{x}_e).$$

Now for any $1 \leq k \leq d$ and any edge $e' \in \mathcal{H}_{d,k}$, i.e. $e' \subseteq \{1, \dots, d\}$, $|e'| = k$, we let $V_{e'} := \prod_{j \in e'} V_j$. For every $\underline{x} \in V^2$ and $e \in \mathcal{H}_{d,k}^{\underline{2}}$, we define $\underline{x}_e := \pi_e(\underline{x})$ where $\pi_e : V^2 \to V_{\pi(e)}$ is the natural projection map.

Our key *counting lemma*, Proposition 3.3 below, which we will establish by induction on $1 \leq k \leq d$ below, is then the statement that given a family of functions $f_e : V_{\pi(e)} \to [-1,1]$, $e \in \mathcal{H}^2_{d,k}$, the averages (generalizing those discussed above) which are defined by

$$\mathcal{N}_{\underline{t}}(f_e;\ e \in \mathcal{H}^2_{d,k}) := \mathbb{E}_{\underline{x} \in V^2} \prod_{e \in \mathcal{H}^2_{d,k}} f_e(\underline{x}_e) \prod_{j=1}^{d} \sigma_{t_j}(x_{j2} - x_{j1}) \tag{3.2.3}$$

$$\mathcal{M}(f_e;\, e \in \mathcal{H}^2_{d,k}) := \mathbb{E}_{\underline{x} \in V^2} \prod_{e \in \mathcal{H}^2_{d,k}} f_e(\underline{x}_e). \tag{3.2.4}$$

are approximately equal. Specifically, one has

Proposition 3.3 (Counting Lemma)**.** *Let $1 \leq k \leq d$ and $0 < \varepsilon \leq 1$. For any collection of functions*

$$f_e : V_{\pi(e)} \to [-1,1] \text{ with } e \in \mathcal{H}^2_{d,k}$$

one has

$$\mathcal{N}_{\underline{t}}(f_e;\ e \in \mathcal{H}^2_{d,k}) \ = \ \mathcal{M}(f_e;\, e \in \mathcal{H}^2_{d,k}) + O(\varepsilon) + O_\varepsilon(q^{-1/2}). \tag{3.2.5}$$

If we apply this Proposition with $d = k$ and $f_e = 1_S$ for all $e \in \mathcal{H}^2_{d,d}$, then Theorem 3.1 clearly follows given the lower bound (3.2.1).

3.2.3 Proof of Proposition 3.3

We will establish Proposition 3.3 by inducting on $1 \leq k \leq d$.
For $k = 1$ the result follows from the basic observation that if $f_1, f_2 : \mathbb{F}_q^2 \to [-1,1]$ and let $t \in \mathbb{F}_q^*$, then

$$\begin{aligned}
\mathbb{E}_{x_1,x_2 \in \mathbb{F}_q^2}\ f_1(x_1) f_2(x_2)\, \sigma_t(x_2 - x_1) &= \sum_{\xi \in \mathbb{F}_q^2} \hat{f}_1(\xi)\hat{f}_2(\xi)\hat{\sigma}_t(\xi) \\
&= \hat{f}_1(0)\hat{f}_2(0) + O(q^{-1/2}) \qquad (3.2.6) \\
&= \mathbb{E}_{x_1,x_2 \in \mathbb{F}_q^2}\ f_1(x_1) f_2(x_2) + O(q^{-1/2})
\end{aligned}$$

by the properties of the function $\hat{\sigma}$ given above.
To see how this implies Proposition 3.3 for $k = 1$ we note that since $\mathcal{H}^2_{d,1} = \{jl :\ 1 \leq j \leq d,\ 1 \leq l \leq 2\}$ it follows that

$$\begin{aligned}
\mathcal{N}_{\underline{t}}(f_e;\ e \in \mathcal{H}^2_{d,1}) &= \prod_{j=1}^{d} \mathbb{E}_{x_{j1},x_{j2} \in \mathbb{F}_q^2}\ f_{j1}(x_{j1}) f_{j2}(x_{j2})\, \sigma_t(x_{j2} - x_{j1}) \\
&= \prod_{j=1}^{d} \mathbb{E}_{x_{j1},x_{j2} \in \mathbb{F}_q^2}\ f_{j1}(x_{j1}) f_{j2}(x_{j2}) + O(q^{-1/2}) = \mathcal{M}(f_e;\, e \in \mathcal{H}^2_{d,1}) + O(q^{-1/2}).
\end{aligned}$$

The induction step has two main ingredients, the first is an estimate of the type which is often referred to as a generalized von-Neumann inequality, namely

Lemma 3.1. *Let $1 \leq k \leq d$. For any collection of functions $f_e : V_{\pi(e)} \to [-1,1]$ with $e \in \mathcal{H}^2_{d,k}$ one has*

$$\mathcal{N}_{\underline{t}}(f_e;\ e \in \mathcal{H}^2_{d,k}) \ \leq \ \min_{e \in \mathcal{H}^2_{d,k}} \|f_e\|_{\square(V_{\pi(e)})} + O(q^{-1/2}) \tag{3.2.7}$$

where for any $e \in \mathcal{H}^2_{d,k}$ *and* $f : V_{\pi(e)} \to [-1,1]$ *we define*

$$\|f\|^{2^k}_{\square(V_{\pi(e)})} := \mathbb{E}_{\underline{x}\in V^2_{\pi(e)}} \prod_{e\in\mathcal{H}^2_{d,k}} f(\underline{x}_e). \tag{3.2.8}$$

The corresponding inequality for the multilinear expression $\mathcal{M}(f_e;\ e \in \mathcal{H}^2_{d,k})$, namely the fact that

$$\mathcal{M}(f_e;\ e \in \mathcal{H}^2_{d,k}) \le \prod_{e\in\mathcal{H}^2_{d,k}} \|f_e\|_{\square(V_{\pi(e)})} \le \min_{e\in\mathcal{H}^2_{d,k}} \|f_e\|_{\square(V_{\pi(e)})}$$

is well-known and is referred to as the Gowers-Cauchy-Schwarz inequality [48].

The second and main ingredient is an approximate decomposition of a graph to simpler ones, and is essentially the so-called weak (hypergraph) regularity lemma of Frieze and Kannan [39]. We choose to state this from a somewhat more abstract/probabilistic point of view, a perspective that will be particularly helpful when we consider our general results in the continuous and discrete settings.

We will first introduce this in the case $d = 2$. A bipartite graph with (finite) vertex sets V_1, V_2 is a set $S \subseteq V_1 \times V_2$ and a function $f : V_1 \times V_2 \to \mathbb{R}$ may be viewed as weighted bipartite graph with weights $f(x_1, x_2)$ on the edges (x_1, x_2). If $\mathbf{P}_1$ and $\mathbf{P}_2$ are partitions of V_1 and V_2 respectively then $\mathbf{P} = \mathbf{P}_1 \times \mathbf{P}_2$ is a partition $V_1 \times V_2$ and we let $\mathbb{E}(f|\mathbf{P})$ denote the function that is constant and equal to $\mathbb{E}_{x\in A} f(x)$ on each atom $A = A_1 \times A_2$ of $\mathbf{P}$. The weak regularity lemma states that for any $\varepsilon > 0$ and for any weighted graph $f : V_1 \times V_2 \to [-1,1]$ there exist partitions $\mathbf{P}_i$ of V_i with $|\mathbf{P}_i| \le 2^{O(\varepsilon^{-2})}$ for $i = 1, 2$, so that

$$|\mathbb{E}_{x_1\in V_1}\mathbb{E}_{x_2\in V_2}(f - \mathbb{E}(f|\mathbf{P}))(x_1, x_2)\ \mathbf{1}_{U_1}(x_1)\mathbf{1}_{U_2}(x_2)| \le \varepsilon \tag{3.2.9}$$

for all $U_1 \subseteq V_1$ and $U_2 \subseteq V_2$. Informally this means that the graph f can be approximated with precision ε with the "low complexity" graph $\mathbb{E}(f, \mathbf{P})$. If we consider the σ-algebras $\mathcal{B}_i$ generated by the partitions $\mathbf{P}_i$ and the σ-algebra $\mathcal{B} = \mathcal{B}_1 \vee \mathcal{B}_2$ generated by $\mathbf{P}_1 \times \mathbf{P}_2$ then we have $\mathbb{E}(f|\mathcal{B})$, the so-called conditional expectation function of f. Moreover it is easy to see, using Cauchy-Schwarz, that estimate (3.2.9) follows from

$$\|f - \mathbb{E}(f|\mathcal{B}_1 \vee \mathcal{B}_2)\|_{\square(V_1\times V_2)} \le \varepsilon. \tag{3.2.10}$$

With this more probabilistic point of view the weak regularity lemma says that the function f can be approximated with precision ε by a low complexity function $\mathbb{E}(f|\mathcal{B}_1 \bigvee \mathcal{B}_2)$, corresponding to σ-algebras $\mathcal{B}_i$ on V_i generated by $O(\varepsilon^{-2})$ sets. This formulation is also referred to as a Koopman-von Neumann type decomposition, see Corollary 6.3 in [109].

We will need a natural extension to k-regular hypergraphs. See [108, 48], and also [26] for extension to sparse hypergraphs. Given an edge $e' \in \mathcal{H}_{d,k}$ of k elements we define its boundary $\partial e' := \{\mathfrak{f}' \in \mathcal{H}_{d,k-1};\ \mathfrak{f}' \subseteq e'\}$. For each $\mathfrak{f}' = e' \backslash \{j\} \in \partial e'$ let $\mathcal{B}'_{\mathfrak{f}}$ be a σ-algebra on $V_{\mathfrak{f}'} := \prod_{j\in\mathfrak{f}'} V_j$ and $\bar{\mathcal{B}}_{\mathfrak{f}'} := \{U \times V_j;\ U \in \mathcal{B}_{\mathfrak{f}'}\}$ denote its pull-back over the space $V_{e'}$. The σ-algebra $\mathcal{B} = \bigvee_{\mathfrak{f}'\in\partial e'} \mathcal{B}_{\mathfrak{f}'}$ is the smallest σ-algebra on $\partial e'$ containing $\bar{\mathcal{B}}_{\mathfrak{f}'}$ for all
$\mathfrak{f}' \in \partial e'$. Note that the atoms of $\mathcal{B}$ are of the form $A = \bigcap_{\mathfrak{f}'\in\partial e'} A_{\mathfrak{f}'}$ where $A_{\mathfrak{f}'}$ is an atom of $\bar{\mathcal{B}}_{\mathfrak{f}'}$. We say that the complexity of a σ-algebra $\mathcal{B}_{\mathfrak{f}'}$ is at most m, and write complex$(\mathcal{B}_{\mathfrak{f}'}) \le m$, if it is generated by m sets.

Lemma 3.2 (Weak hypergraph regularity lemma)**.** *Let* $1 \le k \le d$ *and* $f_e : V_{\pi(e)} \to [-1,1]$ *be a given function for each* $e \in \mathcal{H}^2_{d,k}$. *For any* $\varepsilon > 0$ *there exists* σ*-algebras* $\mathcal{B}_{\mathfrak{f}'}$ *on* $V_{\mathfrak{f}'}$ *for each* $\mathfrak{f}' \in \mathcal{H}_{d,k-1}$ *such that*

$$\mathrm{complex}(\mathcal{B}_{\mathfrak{f}'}) = O(\varepsilon^{-2^{k+1}}) \tag{3.2.11}$$

and

$$\|f_e - \mathbb{E}(f_e| \bigvee_{\mathfrak{f}'\in\partial\pi(e)} \mathcal{B}_{\mathfrak{f}'})\|_{\square(V_{\pi(e)})} \leq \varepsilon \quad \textit{for all } e \in \mathcal{H}^2_{d,k}. \tag{3.2.12}$$

The proof of Lemmas 3.1 and 3.2 are presented in Section 3.2.4 below. We close this subsection by demonstrating how these lemmas can be combined to establish Proposition 3.3.

Proof of Proposition 3.3.
Let $\varepsilon > 0$, $2 \leq k \leq d$ and assume that the lemma holds for $k-1$. It follows from Lemma 3.2 that there exists σ-algebras $\mathcal{B}_{\mathfrak{f}'}$ of complexity $O(\varepsilon^{-2^{k+1}})$ on $V_{\mathfrak{f}'}$ for each $\mathfrak{f}' \in \mathcal{H}_{d,k-1}$ for which (3.2.12) holds for all $e \in \mathcal{H}^2_{d,k}$. For each $e \in \mathcal{H}^2_{d,k}$ we let $\bar{f}_e := \mathbb{E}(f_e| \bigvee_{\mathfrak{f}'\in\partial\pi(e)} \mathcal{B}_{\mathfrak{f}'})$ and write $f_e = \bar{f}_e + h_e$. By Lemma 3.1 and multi-linearity we have that

$$\mathcal{N}_{\underline{t}}(f_e;\ e \in \mathcal{H}^2_{d,k}) = \mathcal{N}_{\underline{t}}(\bar{f}_e;\ e \in \mathcal{H}^2_{d,k}) + O(\varepsilon) + O(q^{-1/2}) \tag{3.2.13}$$

and also by the Gowers-Cauchy-Schwarz inequality

$$\mathcal{M}(f_e;\ e \in \mathcal{H}^2_{d,k}) = \mathcal{M}(\bar{f}_e;\ e \in \mathcal{H}^2_{d,k}) + O(\varepsilon). \tag{3.2.14}$$

The conditional expectation functions $\bar{f}_e$ are linear combinations of the indicator functions 1_{A_e} of the atoms A_e of the σ-algebras $\mathcal{B}_e := \bigvee_{\mathfrak{f}'\in\partial\pi(e)} \mathcal{B}_{\mathfrak{f}'}$. Since the number of terms in this linear combination is at most $2^{C\varepsilon^{-2^{k+1}}}$, with coefficients at most 1 in modulus, plugging these into the multi-linear expressions $\mathcal{N}_{\underline{t}}(\bar{f}_e;\ e \in \mathcal{H}^2_{d,k})$ and $\mathcal{M}(\bar{f}_e;\ e \in \mathcal{H}^2_{d,k})$ one obtains a linear combination of expressions of the form $\mathcal{N}_{\underline{t}}(1_{A_e};\ e \in \mathcal{H}^2_{d,k})$ and $\mathcal{M}(1_{A_e};\ e \in \mathcal{H}^2_{d,k})$ respectively with each A_e being an atoms of $\mathcal{B}_e$ for all $e \in \mathcal{H}^2_{d,k}$.
The key observation is that these expressions are at level $k-1$ instead of k. Indeed, $1_{A_e} = \prod_{\mathfrak{f}'\in\partial\pi(e)} 1_{A_{e\mathfrak{f}'}}$ where $A_{e\mathfrak{f}'} = A'_{e\mathfrak{f}'} \times V_j$, with $A'_{e\mathfrak{f}'}$ being an atom of $\mathcal{B}_{\mathfrak{f}'}$ when $\mathfrak{f}' = \pi(e)\backslash\{j\}$. If $e = (j_1l_1, \ldots, jl, \ldots, j_kl_k)$, let $p_{\mathfrak{f}'}(e) := (j_1l_1, \ldots, j_kl_k) \in \mathcal{H}^2_{d,k-1}$, obtained from e by removing the jl-entry. Then we have $1_{A_{e\mathfrak{f}'}}(\underline{x}_e) = 1_{A'_{e\mathfrak{f}'}}(\underline{x}_{p_{\mathfrak{f}}'(e)})$ since $x_{jl} \in V_j$, and hence

$$1_{A_e}(\underline{x}_e) = \prod_{\mathfrak{f}'\in\partial\pi(e)} 1_{A'_{e\mathfrak{f}'}}(\underline{x}_{p'_{\mathfrak{f}}(e)}).$$

It therefore follows that

$$\begin{aligned}\mathcal{N}_{\underline{t}}(1_{A_e};\ e \in \mathcal{H}^2_{d,k}) &= \mathbb{E}_{\underline{x}\in V^2} \prod_{e\in\mathcal{H}^2_{d,k}} \prod_{\mathfrak{f}'\in\partial\pi(e)} 1_{A'_{e\mathfrak{f}'}}(\underline{x}_{p_{\mathfrak{f}'}(e)}) \prod_{j=1}^d \sigma_{t_j}(x_{j2} - x_{j1}) \\ &= \mathbb{E}_{\underline{x}\in V^2} \prod_{\mathfrak{f}\in\mathcal{H}^2_{d,k-1}} \underbrace{\prod_{\substack{e\in\mathcal{H}^2_{d,k},\, \mathfrak{f}'\in\partial\pi(e) \\ p_{\mathfrak{f}'}(e)=\mathfrak{f}}} 1_{A'_{e\mathfrak{f}'}}(\underline{x}_{p_{\mathfrak{f}'}(e)})}_{=:g_{\mathfrak{f}}} \prod_{j=1}^d \sigma_{t_j}(x_{j2} - x_{j1}) = \mathcal{N}_{\underline{t}}(g_{\mathfrak{f}};\ \mathfrak{f} \in \mathcal{H}^2_{d,k-1})\end{aligned}$$

and similarly that

$$\mathcal{M}(1_{A_e};\ e \in \mathcal{H}^2_{d,k}) = \mathcal{M}(g_{\mathfrak{f}};\ \mathfrak{f} \in \mathcal{H}^2_{d,k-1}).$$

It then follows from the induction hypotheses that

$$\mathcal{N}_{\underline{t}}(1_{A_e};\ e \in \mathcal{H}^2_{d,k}) = \mathcal{M}(1_{A_e};\ e \in \mathcal{H}^2_{d,k}) + O(\varepsilon_1) + O_{\varepsilon_1}(q^{-1/2})$$

for any $\varepsilon_1 > 0$. If we choose $\varepsilon_1 := 2^{-C_1\,\varepsilon^{-2^{k+1}}}$, with $C_1 \gg 1$ sufficiently large, then $\varepsilon_1\, 2^{C\varepsilon^{-2^{k+1}}} = O(\varepsilon)$ and it follows that

$$\mathcal{N}_{\underline{t}}(\bar{f}_e;\ e \in \mathcal{H}^2_{d,k}) = \mathcal{M}(\bar{f}_e;\ e \in \mathcal{H}^2_{d,k}) + O(\varepsilon) + O_\varepsilon(q^{-1/2}).$$

This, together with (3.2.13) and (3.2.14), establishes that (3.2.5) hold for $d = k$ as required. □

3.2.4 Proof of Lemmas 3.1 and 3.2

Proof of Lemma 3.1. We start by observing the following consequence of (3.2.6), namely that

$$\left|\mathbb{E}_{x_1,x_2\in\mathbb{F}_q^2} f_1(x_1)f_2(x_2)\sigma_t(x_2-x_1)\right|^2 \le \mathbb{E}_{x_1,x_2\in\mathbb{F}_q^2} f_1(x_1)f_1(x_2) + O(q^{-1/2}) \tag{3.2.15}$$

for any $f_1, f_2 : \mathbb{F}_q^2 \to [-1,1]$ and $t \in \mathbb{F}_q^*$.

Now, fix an edge, say $e_0 = (11, 21, \ldots, k1)$. Partition the edges $e \in \mathcal{H}^2_{d,k}$ into three groups; the first group consisting of edges e for which $1 \notin \pi(e)$, the second where $11 \in e$ and write $e = (11, e')$ with $e' \in \mathcal{H}^2_{d-1,k-1}$ and the third when $12 \in e$, using the notation $\mathcal{H}^2_{d-1,k-1} := \{(j_2l_2, \ldots, j_kl_k)\}$. Accordingly we can write

$$\mathcal{N}_t(f_e;\ e \in \mathcal{H}^2_{d,k}) = \mathbb{E}_{\underline{x}\in V^2} \prod_{1\notin\pi(e)} f_e(\underline{x}_e) \prod_{e'\in\mathcal{H}^2_{d-1,k-1}} f_{(11,e')}(x_{11},\underline{x}_{e'}) \prod_{e'\in\mathcal{H}^2_{d-1,k-1}} f_{(12,e')}(x_{12},\underline{x}_{e'}) \prod_{j=1}^{d} \sigma_{t_j}(x_{j2}-x_{j1}). \tag{3.2.16}$$

If for given $x \in V_1$ and $\underline{x}' = (x_{21}, x_{22}, \ldots, x_{d1}, x_{d2}) \in V_2^2 \times \ldots \times V_d^2$ we define

$$g_1(x,\underline{x}') := \prod_{e'\in\mathcal{H}^2_{d-1,k-1}} f_{(11,e')}(x,\underline{x}_{e'}) \qquad \text{and} \qquad g_2(x,\underline{x}') := \prod_{e'\in\mathcal{H}^2_{d-1,k-1}} f_{(12,e')}(x,\underline{x}_{e'})$$

then we can write

$$\begin{aligned}\mathcal{N}_t(f_e;\ e \in \mathcal{H}^2_{d,k}) = \mathbb{E}_{x_{21},x_{22},\ldots,x_{d1},x_{d2}} &\prod_{1\notin\pi(e)} f_e(\underline{x}_e) \prod_{j=2}^{d} \sigma_{t_j}(x_{j2}-x_{j1}) \\ &\times \mathbb{E}_{x_{11},x_{12}}\ g_1(x_{11},\underline{x}')g_2(x_{12},\underline{x}')\,\sigma_{t_1}(x_{12}-x_{11}).\end{aligned} \tag{3.2.17}$$

By (3.2.15) we can estimate the inner sum in (3.2.17) by the square root of

$$\mathbb{E}_{x_{11},x_{12}}\ g_1(x_{11},\underline{x}')g_1(x_{12},\underline{x}') + O(q^{-1/2}).$$

Thus by Cauchy-Schwarz, and the fact that $f_e : V_{\pi(e)} \to [-1,1]$ for all $e \in \mathcal{H}^2_{d,k}$, we can conclude that

$$\mathcal{N}_t(f_e;\ e \in \mathcal{H}^2_{d,k})^2 \le \mathbb{E}_{x_{11},x_{12},\ldots,x_{d1},x_{d2}} \prod_{e'\in\mathcal{H}^2_{d-1,k-1}} f_{(11,e')}(x_{11},\underline{x}_{e'})f_{(11,e')}(x_{12},\underline{x}_{e'}) \prod_{j=2}^{d} \sigma_{t_j}(x_{j2}-x_{j2}). \tag{3.2.18}$$

The expression on the right hand side of the inequality above is similar to that in (3.2.16) except for the following changes. The functions f_e for $1 \notin e$ are eliminated i.e. replaced by 1, as well as the factor σ_{t_1}. The functions $f_{(12,e')}$, are replaced by $f_{(11,e')}$ for all $e' \in \mathcal{H}^2_{d-1,k-1}$. Repeating the same procedure for $j = 2, \ldots, k$ one eliminates all the factors σ_{t_j} for $1 \le j \le k$, moreover all the functions f_e for edges e such that $j \notin \pi(e)$ for some $1 \leq j \leq k$, which leaves only the edges e so that $\pi(e) = (1, 2, \ldots, k)$, moreover for such edges the functions f_e are eventually replaced by $f_{e_0} = f_{11,21,\ldots,k1}$. The factors $\sigma_{t_j}(x_{j2} - x_{j1})$ are not changed for $j > k$ however as the function f_{e_0} does not depend on the variables x_{jl} for $j > k$, averaging over these variables gives rise to a factor of $1 + O(q^{-1/2})$. Thus one obtains the following final estimate

$$\mathcal{N}_t(f_e;\ e \in \mathcal{H}^2_{d,k})^{2^k} \leq \mathbb{E}_{x_{11},x_{12},\ldots,x_{k1},x_{k2}} \prod_{\pi(e)=(1,\ldots,k)} f_{e_0}(\underline{x}_e) + O(q^{-1/2}) = \|f_{e_0}\|^{2^k}_{\square(V_{\pi(e_0)})} + O(q^{-1/2}). \tag{3.2.19}$$

This proves the lemma, as it is clear that the above procedure can be applied to any edge in place of $e_0 = (11, 21, \ldots, k1)$. $\square$

Proof of Lemma 3.2. For a function $f_e : V_{\pi(e)} \to [-1, 1]$ and a σ-algebra $\mathcal{B}_{\pi(e)}$ on $V_{\pi(e)}$ define the *energy* of f_e with respect to $\mathcal{B}_{\pi(e)}$ as

$$\mathcal{E}(f_e, \mathcal{B}_{\pi(e)}) := \|\mathbb{E}(f_e|\mathcal{B}_{\pi(e)})\|_2^2 = \mathbb{E}_{x\in V_{\pi(e)}} |\mathbb{E}(f_e|\mathcal{B}_{\pi(e)})(x)|^2,$$

and for a family of functions f_e and σ-algebras $\mathcal{B}_{\pi(e)}$, $e \in \mathcal{H}^2_{d,k}$ its total energy as

$$\mathcal{E}(f_e, \mathcal{B}_{\pi(e)};\, e \in \mathcal{H}^2_{d,k}) := \sum_{e\in\mathcal{H}^2_{d,k}} \mathcal{E}(f_e, \mathcal{B}_{\pi(e)}).$$

We will show that if (3.2.12) does not hold for a family of σ-algebras $\mathcal{B}_{\pi(e)} = \bigvee_{\mathfrak{f}'\in\partial\pi(e)} \mathcal{B}_{\mathfrak{f}'}$, then the σ-algebras $\mathcal{B}_{\mathfrak{f}'}$ can be refined so that the total energy of the system increases by a quantity depending only on ε. Since the functions f_e are bounded the total energy of the system is $O(1)$, the energy increment process must stop in $O_\varepsilon(1)$ steps, and (3.2.12) must hold. The idea of this procedure appears already in the proof of Szemerédi's regularity lemma [104], and have been used since in various places [39, 108, 48].

Initially set $\mathcal{B}_{\mathfrak{f}'} := \{\emptyset, V_{\mathfrak{f}'}\}$ and hence $\mathcal{B}_{\pi(e)} = \{\emptyset, V_{\pi(e)}\}$ to be the trivial σ-algebras. Assume that in general (3.2.12) does not hold for a family of σ-algebras $\mathcal{B}_{\mathfrak{f}'}$, with $\mathfrak{f}' \in \mathcal{H}_{d,k-1}$. Then there exists an edge $e \in \mathcal{H}^2_{d,k}$ so that $\|g_e\|_{\square(V_{\pi(e)})} \geq \varepsilon$, with $g_e := f_e - \mathbb{E}(f_e|\mathcal{B}_{\pi(e)})$. Let $e = (11, \ldots, k1)$ for simplicity of notation, hence $\pi(e) = (1, \ldots, k)$. Then, with notation $\underline{x}' = (x_{12}, \ldots, x_{k2})$, one has

$$\begin{aligned}\varepsilon^{2^k} \leq \|g_e\|^{2^k}_{\square(V_{\pi(e)})} &= \mathbb{E}_{x_{11},x_{12},\ldots,x_{k1},x_{k2}} \prod_{l_1,\ldots,l_k=1,2} g_e(x_{1l_1}, \ldots, x_{kl_k}) \\ &\leq \mathbb{E}_{x_{12},\ldots,x_{k2}} \Big|\mathbb{E}_{x_{11},\ldots,x_{k1}} g_e(x_{11}, \ldots, x_{k1}) \prod_{j=1}^{k} h_{j,\underline{x}'}(x_{11}, \ldots, x_{j-1\,1}, x_{j+1\,1}, \ldots, x_{k1})\Big|\end{aligned}$$

for some functions $h_{j,\underline{x}'}$ that are bounded by 1 in magnitude. Indeed if and edge $e \neq (11, \ldots, k1)$ then x_e does not depend at least one of the variables x_{j1}. Thus there must be an $\underline{x}'$ for which the inner sum in the above expression is at least ε^{2^k}. Fix such an $\underline{x}'$. Decomposing the functions $h_{j,\underline{x}'}$ into their positive and negative parts and then writing them as an average of indicator functions, one obtains that there sets $B_j \subseteq V_{\pi(e)\setminus\{j\}}$ such that

$$\Big|\mathbb{E}_{x_{11},\ldots,x_{k1}} g_e(x_{11}, \ldots, x_{k1}) \prod_{j=1}^{k} 1_{B_j}(x_{11}, \ldots, x_{j-1\,1}, x_{j+1\,1}, \ldots, x_{k1})\Big| \;\geq\; 2^{-k}\,\varepsilon^{2^k}$$

which can be written more succinctly, using the inner product notation, as

$$\Big|\langle f_e - \mathbb{E}(f_e|\mathcal{B}_{\pi(e)}), \prod_{j=1}^{k} 1_{B_j}\rangle\Big| \geq 2^{-k}\,\varepsilon^{2^k}. \tag{3.2.20}$$

For $\mathfrak{f}' = \partial\pi(e)\backslash\{j\}$ let $\mathcal{B}'_{\mathfrak{f}'}$ be the σ-algebra generated by $\mathcal{B}_{\mathfrak{f}'}$ and the set B_j and let $\mathcal{B}'_{\pi(e)} := \bigvee_{\mathfrak{f}'\in\partial\pi(e)} \mathcal{B}'_{\mathfrak{f}'}$. Since the functions 1_{B_j} are measurable with respect to the σ-algebra $\mathcal{B}'_{\pi(e)}$ for all $1 \leq j \leq k$, we have that

$$\langle f_e - \mathbb{E}(f_e|\mathcal{B}'_{\pi(e)}), \prod_{j=1}^{k} 1_{B_j}\rangle = 0 \tag{3.2.21}$$

and hence, by Cauchy-Schwarz, that

$$\|\mathbb{E}(f_e|\mathcal{B}'_{\pi(e)}) - \mathbb{E}(f_e|\mathcal{B}_{\pi(e)})\|_2^2 = \|\mathbb{E}(f_e|\mathcal{B}'_{\pi(e)})\|_2^2 - \|\mathbb{E}(f_e|\mathcal{B}_{\pi(e)})\|_2^2 \geq 2^{-2k}\,\varepsilon^{2^{k+1}}. \tag{3.2.22}$$

Note that the first equality above follows from the fact that conditional expectation function $\mathbb{E}(f|\mathcal{B})$ is the orthogonal projection of f to the subspace of $\mathcal{B}$-measurable functions in L^2. This also implies that energy of a function is always increasing when the underlying σ-algebra is refined, and (3.2.22) tells us that the energy of f_e is increased by at least $c_k\,\varepsilon^{2^{k+1}}$.
For $\mathfrak{f}' \notin \partial\pi(e)$ we set $\mathcal{B}'_{\mathfrak{f}'} := \mathcal{B}_{\mathfrak{f}'}$. Then the total energy of the family f_e with respect to the system $\mathcal{B}'_{\pi(e)} = \bigvee_{\mathfrak{f}'\in\partial\pi(e)} \mathcal{B}'_{\mathfrak{f}'}$, $e \in \mathcal{H}^2_{d,k}$ is also increased by at least $c_k\,\varepsilon^{2^{k+1}}$.
It is clear that the complexity of the σ-algebras $\mathcal{B}_{\mathfrak{f}'}$ are increased by at most 1, hence, as explained above, the lemma follows by applying this energy increment process at most $O(\varepsilon^{-2^{k+1}})$ times. □

3.3 The base case of an inductive strategy to prove Theorem 3.2

In this section we will ultimately establish the base case of our more general inductive argument. We however start by giving a quick review of the proof of Theorem 3.2 when $d = 1$ (which contains both Theorem B and Corollary B), namely the case of a single simplex. This was originally addressed in [16] and revisited in [76] and [77].

3.3.1 A single simplex in $\mathbb{R}^n$

Let $Q \subseteq \mathbb{R}^n$ be a fixed cube and let $l(Q)$ denotes its side length.
Let $\Delta^0 = \{v_1 = 0, v_2, \ldots, v_n\} \subseteq \mathbb{R}^n$ be a fixed non-degenerate simplex and define $t_{kl} := v_k \cdot v_l$ for $2 \leq k, l \leq n$ where "$\cdot$" is the dot product on $\mathbb{R}^n$. Given $\lambda > 0$, a simplex $\Delta = \{x_1 = 0, x_2, \ldots, x_n\} \subseteq \mathbb{R}^n$ is isometric to $\lambda\Delta^0$ if and only if $x_k \cdot x_l = \lambda^2 t_{kl}$ for all $2 \leq k, l \leq n$. Thus the configuration space $S_{\lambda\Delta^0}$ of isometric copies of $\lambda\Delta^0$ is a non-singular real variety given by the above equations. Let $\sigma_{\lambda\Delta^0}$ be natural normalized surface area measure on $S_{\lambda\Delta^0}$, described in [16], [76], and [77]. It is clear that the variable x_1 can be replaced by any of the variables x_i by redefining the constants t_{kl}.
For any family of functions $f_1, \ldots, f_n : Q \to [-1, 1]$ and $0 < \lambda \ll l(Q)$ we define the multi-linear expression

$$\mathcal{N}^1_{\lambda\Delta^0,Q}(f_1, \ldots, f_n) := \fint_{x_1\in Q} \int_{x_2,\ldots,x_n} f_1(x_1)\ldots f_n(x_n)\, d\sigma_{\lambda\Delta_0}(x_2 - x_1, \ldots, x_n - x_1)\, dx_1. \tag{3.3.1}$$

We note that all of our functions are 1-bounded and both integrals, in fact all integrals in this paper, are normalized. Recall that we are using the normalized integral notation $\fint_A f := \frac{1}{|A|}\int_A f$. Since

the normalized measure $\sigma_{\lambda\Delta^0}$ is supported on $S_{\lambda\Delta_0}$ we will not indicate the support of the variables $(x_2, \dots, x_n)$ explicitly.
Note also that if $S \subseteq Q$ is a measurable set and $\mathcal{N}^1_{\lambda\Delta^0,Q}(1_S, \dots, 1_S) > 0$ then S must contain an isometric copy of $\lambda\Delta^0$. The following proposition (with $Q = [0,1]^n$) immediately establishes Theorem 3.2 for $d = 1$.

Proposition 3.4. *For any $0 < \varepsilon \le 1$ there exists an integer $J = O(\varepsilon^{-2} \log \varepsilon^{-1})$ with the following property:*
Given any lacunary sequence $l(Q) \ge \lambda_1 \ge \dots \ge \lambda_J$ and $S \subseteq Q$, there is some $1 \le j < J$ such that

$$\mathcal{N}^1_{\lambda\Delta^0,Q}(1_S, \dots, 1_S) > \left(\frac{|S|}{|Q|}\right)^n - \varepsilon \tag{3.3.2}$$

for all $\lambda \in [\lambda_{j+1}, \lambda_j]$.

Our approach to establishing Proposition 3.4 is to compare the above expressions to simpler ones for which it is easy to obtain lower bounds. Given a scale $0 < \lambda \ll l(Q)$ we define the multi-linear expression

$$\mathcal{M}^1_{\lambda,Q}(f_1, \dots, f_n) := \fint_{t \in Q} \fint_{x_1, x_2, \dots, x_n \in t+Q(\lambda)} f_1(x_1) \dots f_n(x_n)\, dx_1 \dots dx_n\, dt \tag{3.3.3}$$

where $Q(\lambda) = [-\frac{\lambda}{2}, \frac{\lambda}{2}]^n$ and $t + Q(\lambda)$ is the shift of the cube $Q(\lambda)$ by the vector t. Note that if $S \subseteq Q$ is a set of measure $|S| \ge \delta|Q|$ for some $\delta > 0$, then for a given $\varepsilon > 0$, Hölder implies

$$\mathcal{M}^1_{\lambda,Q}(1_S, \dots, 1_S) = \fint_{t \in Q} \left(\fint_{x \in t+Q(\lambda)} 1_S(x)\, dx \right)^n dt \ge \left(\fint_{t \in Q} \fint_{x \in t+Q(\lambda)} 1_S(x)\, dx\, dt \right)^n \ge \delta^n - O(\varepsilon), \tag{3.3.4}$$

for all scales $0 < \lambda \ll \varepsilon\, l(Q)$.
Recall that for any $\varepsilon > 0$ we call a sequence $L_1 \ge \dots \ge L_J$ *ε-admissible* if $L_j / L_{j+1} \in \mathbb{N}$ and $L_{j+1} \ll \varepsilon^2 L_j$ for all $1 \le j < J$. Note that given any lacunary sequence $l(Q) \ge \lambda_1 \ge \dots \ge \lambda_{J'}$ with $J' \gg (\log \varepsilon^{-1})\, J$, one can always finds an ε-admissible sequence of scales $l(Q) \ge L_1 \ge \dots \ge L_J$ such that for each $1 \le j < J$ the interval $[L_{j+1}, L_j]$ contains at least two consecutive elements from the original lacunary sequence.
In light of this observation, and the one above regarding a lower bound for $\mathcal{M}^1_{\lambda,Q}(1_S, \dots, 1_S)$, our proof of Proposition 3.4 reduces to establishing the following “counting lemma”.

Proposition 3.5. *Let $0 < \varepsilon < 1$. There exists an integer $J_1 = O(\varepsilon^{-2})$ such that for any ε-admissible sequence of scales $l(Q) \ge L_1 \ge \dots \ge L_{J_1}$ and $S \subseteq Q$ there is some $1 \le j < J_1$ such that*

$$\mathcal{N}^1_{\lambda\Delta^0,Q}(1_S, \dots, 1_S) = \mathcal{M}^1_{\lambda,Q}(1_S, \dots, 1_S) + O(\varepsilon) \tag{3.3.5}$$

for all $\lambda \in [L_{j+1}, L_j]$.

There are two main ingredients in the proof of Proposition 3.5, this will be typical to all of our arguments. The first ingredient is a result which establishes that the our multi-linear forms $\mathcal{N}^1_{\lambda\Delta^0,Q}(f_1, \dots, f_n)$ are controlled by an appropriate norm which measures the uniformity of distribution of functions $f : Q \to [-1, 1]$ with respect to particular scales L. This is analogous to estimates in additive combinatorics [48], [110] which are often referred to as generalized von-Neumann inequalities.
The result below was proved in [76] for $Q = [0,1]^n$, however a simple scaling of the variables x_i transfers the result to an arbitrary cube Q.

Lemma 3.3 (A Generalized von-Neumann inequality [76]). *Let $\varepsilon > 0$, $0 < \lambda \ll l(Q)$, and $0 < L \ll \varepsilon^6 \lambda$.*
For any collections of functions $f_1, \ldots, f_n : Q \to [-1,1]$ we have

$$|\mathcal{N}^1_{\lambda\Delta^0,Q}(f_1, \ldots, f_n)| \leq \min_{i=1,\ldots,n} \|f_i\|_{U^1_L(Q)} + O(\varepsilon) \tag{3.3.6}$$

where for any $f : Q \to [-1,1]$ we define

$$\|f\|^2_{U^1_L(Q)} := \fint_{t\in Q} \Big| \fint_{x\in t+Q(L)} f(x)\,dx \Big|^2 dt \tag{3.3.7}$$

with $t + Q(L)$ denoting the shift of the cube $Q(L) = [-\frac{L}{2}, \frac{L}{2}]^n$ by the vector t.

The corresponding inequality for the multilinear expression $\mathcal{M}^1_{\lambda,Q}(f_1, \ldots, f_n)$, namely the fact that

$$\mathcal{M}^1_{\lambda,Q}(f_1, \ldots, f_n) \leq \min_{i=1,\ldots,n} \|f_i\|_{U^1_L(Q)} + O(\varepsilon)$$

whenever $0 < L \ll \varepsilon^6\lambda$ follows easily from Cauchy-Schwarz together with the simple observation that

$$\|f\|_{U^1_L(Q)} \leq \|f\|_{U^1_{L'}(Q)} + O(\varepsilon)$$

whenever $L' \ll \varepsilon L$.
The second key ingredient, proved in [77] and generalized in Lemma 3.5 below, is a Koopman-von Neumann type decomposition of functions where the underlying σ-algebras are generated by cubes of a fixed length. To recall it, let $Q \subseteq \mathbb{R}^n$ be a cube, $L > 0$ be scale that divides $l(Q)$, $Q(L) = [-\frac{L}{2}, \frac{L}{2}]^n$, and $\mathcal{G}_{L,Q}$ denote the collection of cubes $t + Q(L)$ partitioning the cube Q and $\Gamma_{L,Q}$ denote the grids corresponding to the centers of the cubes. By a slightly abuse of notation we also write $\mathcal{G}_{L,Q}$ for the σ-algebra generated by the grid. Recall that the conditional expectation function $\mathbb{E}(f|\mathcal{G}_{L,Q})$ is constant and equal to $\fint_A f$ on each cube $A \in \mathcal{G}_{L,Q}$.

Lemma 3.4 (A Koopman-von Neumann type decomposition [77]). *Let $0 < \varepsilon \leq 1$ and $Q \subseteq \mathbb{R}^n$ be a cube.*
There exists an integer $\bar{J}_1 = O(\varepsilon^{-2})$ such that for any ε-admissible sequence $l(Q) \geq L_1 \geq \cdots \geq L_{\bar{J}_1}$ and function $f : Q \to [-1,1]$ there is some $1 \leq j < \bar{J}_1$ such that

$$\|f - \mathbb{E}(f|\mathcal{G}_{L_j,Q})\|_{U^1_{L_{j+1}}(Q)} \leq \varepsilon \tag{3.3.8}$$

Proof of Proposition 3.5. Let $\mathcal{G}_{L_j,Q}$ be the grid obtained from Lemma 3.4 for the functions $f = 1_S$ for some fixed $\varepsilon > 0$. Let $\bar{f} := \mathbb{E}(f|\mathcal{G}_{L_j,Q})$, then by (3.3.6) and multi-linearity, we have

$$\mathcal{N}^1_{\lambda\Delta^0,Q}(f, \ldots, f) = \mathcal{N}^1_{\lambda\Delta^0,Q}(\bar{f}, \ldots, \bar{f}) + O(\varepsilon),$$

and also

$$\mathcal{M}^1_{\lambda,Q}(f, \ldots, f) = \mathcal{M}^1_{\lambda,Q}(\bar{f}, \ldots, \bar{f}) + O(\varepsilon)$$

provided for $\varepsilon^{-6}L_{j+1} \ll \lambda$. Thus in showing (3.5.4) one can replace the functions f with $\bar{f}$. If we make the additional assumption that $\lambda \ll \varepsilon L_j$ then it is easy to see, using the fact that the function $\bar{f}$ is constant on the cubes $Q_t(L_j) \in \mathcal{G}_{L_j,Q}$, that

$$\mathcal{N}^1_{\lambda\Delta^0,Q}(\bar{f}, \ldots, \bar{f}) = \mathcal{M}^1_{\lambda,Q}(\bar{f}, \ldots, \bar{f}) + O(\varepsilon).$$

Since the condition $\varepsilon^{-6}L_{j+1} \ll \lambda \ll \varepsilon L_j$ can be replaced with $L_{j+1} \ll \lambda \ll L_j$ if one passes to a subsequence of scales, for example $L'_j = L_{5j}$, this completes the proof of Proposition 3.5. $\square$

3.3.2 The base case of a general inductive strategy

In this section, as preparation to handle the case of products of simplices, we prove a parametric version of Proposition 3.5, namely Proposition 3.6 below, which will serve as the base case for later inductive arguments.

Let $Q = Q_1 \times \cdots \times Q_d$ with $Q_i \subseteq \mathbb{R}^{n_i}$ be cubes of equal side length $l(Q)$. Let L be a scale dividing $l(Q)$ and for each $\underline{t} = (t_1, \ldots, t_d) \in \Gamma_{L,Q}$ let $Q_{\underline{t}}(L) = \underline{t} + Q(L)$ and $Q_{t_i}(L) = t_i + Q_i(L)$. Note that $Q_{\underline{t}}(L) = Q_{t_1}(L) \times \cdots \times Q_{t_d}(L)$. Here $Q(L) = [-\frac{L}{2}, \frac{L}{2}]^n$ and $Q_i(L) = [-\frac{L}{2}, \frac{L}{2}]^{n_i}$ for each $1 \leq i \leq d$. Let $\Delta_i^0 = \{v_1^i, \ldots, v_{n_i}^i\} \subseteq \mathbb{R}^{n_i}$ be a non-degenerate simplex for each $1 \leq i \leq d$.

Proposition 3.6 (Parametric Counting Lemma on $\mathbb{R}^n$ for Simplices)**.**
Let $0 < \varepsilon \leq 1$ and $R \geq 1$. There exists an integer $J_1 = J_1(\varepsilon, R) = O(R\,\varepsilon^{-4})$ such that for any ε-admissible sequence of scales $L_0 \geq L_1 \geq \cdots \geq L_{J_1}$ with the property that L_0 divides $l(Q)$ and collection of functions

$$f_{k,\underline{t}}^{i,r} : Q_{t_i}(L_0) \to [-1,1] \quad \text{with} \quad 1 \leq i \leq d,\ 1 \leq k \leq n_i,\ 1 \leq r \leq R \text{ and } \underline{t} \in \Gamma_{L_0,Q}$$

there exists $1 \leq j < J_1$ and a set $T_\varepsilon \subseteq \Gamma_{L_0,Q}$ of size $|T_\varepsilon| \leq \varepsilon |\Gamma_{L_0,Q}|$ such that

$$\mathcal{N}^1_{\lambda\Delta_i^0, Q_{t_i}(L_0)}(f_{1,\underline{t}}^{i,r}, \ldots, f_{n_i,\underline{t}}^{i,r}) = \mathcal{M}^1_{\lambda, Q_{t_i}(L_0)}(f_{1,\underline{t}}^{i,r}, \ldots, f_{n_i,\underline{t}}^{i,r}) + O(\varepsilon) \tag{3.3.9}$$

for all $\lambda \in [L_{j+1}, L_j]$ and $\underline{t} \notin T_\varepsilon$ uniformly in $1 \leq i \leq d$ and $1 \leq r \leq R$.

The proof of Proposition 3.6 will follow from Lemma 3.3 and the following generalization of Lemma 3.4 in which we simultaneously consider a family of functions supported on the subcubes in a partition of an original cube Q.

Lemma 3.5 (A simultaneous Koopman-von Neumann type decomposition)**.**
Let $0 < \varepsilon \leq 1$, $m \geq 1$, and $Q \subseteq \mathbb{R}^n$ be a cube. There exists an integer $\bar{J}_1 = O(m\varepsilon^{-3})$ such that for any ε-admissible sequence $L_0 \geq L_1 \geq \cdots \geq L_{\bar{J}_1}$ with the property that L_0 divides $l(Q)$, and collection of functions

$$f_{1,t}, \ldots, f_{m,t} : Q_t(L_0) \to [-1,1]$$

defined for each $t \in \Gamma_{L_0,Q}$, there is some $1 \leq j < \bar{J}_1$ and a set $T_\varepsilon \subseteq \Gamma_{L_0,Q}$ of size $|T_\varepsilon| \leq \varepsilon|\Gamma_{L_0,Q}|$ such that

$$\|f_{i,t} - \mathbb{E}(f_{i,t}|\mathcal{G}_{L_j,Q_t(L_0)})\|_{U^1_{L_{j+1}}(Q_t(L_0))} \leq \varepsilon \tag{3.3.10}$$

for all $1 \leq i \leq m$ and $t \notin T_\varepsilon$.

Proof of Proposition 3.6. Fix $1 \leq i \leq d$. For $1 \leq k \leq n_i$ and $\underline{t} = (t_1, \ldots, t_d) \in \Gamma_{L_0,Q}$, we will abuse notation and write

$$f_{k,\underline{t}}^{i,r}(x_1, \ldots, x_d) := f_{k,\underline{t}}^{i,r}(x_i)$$

for $(x_1, \ldots, x_d) \in Q_{\underline{t}}(L_0)$.

If we apply Lemma 3.5 to the family of functions $f_{k,\underline{t}}^{i,r}$ on $Q_{\underline{t}}(L_0)$ for $1 \leq i \leq d$, $1 \leq k \leq n_i$, and $1 \leq r \leq R$, with $m = (n_1 + \ldots + n_d)R$, then this produces a grid $\mathcal{G}_{L_j,Q}$ for some $1 \leq j \leq \bar{J}_1 = O(\varepsilon^{-3}R)$, and a set $T_\varepsilon \subseteq \Gamma_{L_0,Q}$ of size $|T_\varepsilon| \leq \varepsilon|\Gamma_{L_0,Q}|$, such that

$$\|f_{k,\underline{t}}^{i,r} - \mathbb{E}(f_{k,\underline{t}}^{i,r}|\mathcal{G}_{L_j,Q})\|_{U^1_{L_{j+1}}(Q_{\underline{t}}(L_0))} \leq \varepsilon$$

uniformly for $1 \leq i \leq d$, $1 \leq k \leq n_i$ and $1 \leq r \leq R$ for $t \notin T_\varepsilon$.

Since $f^{i,r}_{k,\underline{t}}(x_1,\dots,x_d) = f^{i,r}_{k,\underline{t}}(x_i)$ for $(x_1,\dots,x_d) \in Q_{\underline{t}}(L_0)$ it is easy to see that

$$\|f^{i,r}_{k,\underline{t}} - \mathbb{E}(f^{i,r}_{k,\underline{t}}|\mathcal{G}_{L_j,Q})\|_{U^1_{L_{j+1}}(Q_{\underline{t}}(L_0))} = \|f^{i,r}_{k,\underline{t}} - \mathbb{E}(f^{i,r}_{k,\underline{t}}|\mathcal{G}_{L_j,Q_i})\|_{U^1_{L_{j+1}}(Q_{t_i}(L_0))}.$$

Let $\bar{f}^{i,r}_{k,\underline{t}} := \mathbb{E}(f^{i,r}_{k,\underline{t}}|\mathcal{G}_{L_j,Q_i})$, then by Lemma 3.3, one has

$$\mathcal{N}^1_{\lambda\Delta^0_i,Q_{t_i}(L_0)}(f^{i,r}_{1,\underline{t}},\dots,f^{i,r}_{n_i,\underline{t}}) = \mathcal{N}^1_{\lambda\Delta^0_i,Q_{t_i}(L_0)}(\bar{f}^{i,r}_{1,\underline{t}},\dots,\bar{f}^{i,r}_{n_i,\underline{t}}) + O(\varepsilon),$$

and

$$\mathcal{M}^1_{\lambda,Q_{t_i}(L_0)}(f^{i,r}_{1,\underline{t}},\dots,f^{i,r}_{n_i,\underline{t}}) = \mathcal{M}^1_{\lambda,Q_{t_i}(L_0)}(\bar{f}^{i,r}_{1,\underline{t}},\dots,\bar{f}^{i,r}_{n_i,\underline{t}}) + O(\varepsilon)$$

for all $\underline{t} \notin T_\varepsilon$ provided $\varepsilon^{-6}L_{j+1} \ll \lambda$. Finally, if we also have $\lambda \ll \varepsilon L_j$ then it is easy to see that

$$\mathcal{N}^1_{\lambda\Delta^0_i,Q_{t_i}(L_0)}(\bar{f}^{i,r}_{1,\underline{t}},\dots,\bar{f}^{i,r}_{n_i,\underline{t}}) = \mathcal{M}^1_{\lambda,Q_{t_i}(L_0)}(\bar{f}^{i,r}_{1,\underline{t}},\dots,\bar{f}^{i,r}_{n_i,\underline{t}}) + O(\varepsilon)$$

as the functions $\bar{f}^{i,r}_{k,\underline{t}}$ are constant on cubes $Q_{t_i}(L_j)$ of $\mathcal{G}_{L_j,Q_i}$, which are of size $L_j \ll \varepsilon L_0$. Passing first to a subsequence of scales, for example $L'_j = L_{5j}$, the condition $\varepsilon^{-6}L_{j+1} \ll \lambda \ll \varepsilon L_j$ can be replaced with $L_{j+1} \ll \lambda \ll L_j$ so this completes the proof of the Proposition. □

We conclude this section with a sketch of the proof of Lemma 3.5. These arguments are standard, see for example the proof of Lemma 3.4 given in [76].

Proof of Lemma 3.5. First we make an observation about the $U^1_L(Q)$-norm. Suppose $0 < L' \ll \varepsilon^2 L$ with L' dividing L. If $s \in \Gamma_{L',Q}$ and $t \in Q_s(L')$ then $|t-s| = O(L')$ and hence

$$\fint_{x\in Q_t(L)} g(x)\,dx = \fint_{x\in Q_s(L)} g(x)\,dx + O(L'/L)$$

for any function $g : Q \to [-1,1]$. Moreover, since the cube $Q_s(L)$ is partitioned into the smaller cubes $Q_t(L')$, we have by Cauchy-Schwarz

$$\Big|\fint_{x\in Q_s(L)} g(x)\,dx\Big|^2 \le \mathbb{E}_{t\in\Gamma_{L',Q_s(L)}}\Big|\fint_{x\in Q_t(L')} g(x)\,dx\Big|^2.$$

From these observations it is easy to see that

$$\|g\|^2_{U^1_L(Q)} = \fint_{t\in Q}\Big|\fint_{x\in Q_t(L)} g(x)\,dx\Big|^2 dt \le \mathbb{E}_{t\in\Gamma_{L',Q}}\Big|\fint_{x\in Q_t(L')} g(x)\,dx\Big|^2 + O(L'/L)$$

and we note that the right side of the above expression is $\|\mathbb{E}(g|\mathcal{G}_{L',Q})\|^2_{L^2(Q)}$ since the conditional expectation function $\mathbb{E}(g|\mathcal{G}_{L',Q})$ is constant and equal to $\fint_{x\in Q_t(L')} g(x)\,dx$ on the cubes $Q_t(L')$. Suppose that (3.3.10) does not hold for some $1 \le i \le m$ for every t in some set $T_\varepsilon \subseteq \Gamma_{L_0,Q}$ of size $|T_\varepsilon| > \varepsilon\,|\Gamma_{L_0,Q}|$. If we apply the above observation to $g := f_{i,t} - \mathbb{E}(f_{i,t}|\mathcal{G}_{L_j,Q_t(L_0)})$, for every $t \in T_\varepsilon$, we obtain by orthogonality that

$$\sum_{i=1}^m \|\mathbb{E}(f_{i,t}|\mathcal{G}_{L_{j+2},Q_t(L_0)})\|^2_{L^2(Q_t(L_0))} \ge \sum_{i=1}^m \|\mathbb{E}(f_{i,t}|\mathcal{G}_{L_j,Q_t(L_0)})\|^2_{L^2(Q_t(L_0))} + c\varepsilon^2$$

for some constant $c > 0$.

If we now define $f_i : Q \to [-1,1]$ such that $f_i|_{(Q_t(L_0))} = f_{i,t}$, for $1 \le i \le m$, average over $t \in \Gamma_{L_0,Q}$, and use the fact $\|f_i\|^2_{L^2(Q)} = \mathbb{E}_{t\in\Gamma_{L_0,Q}} \|f_{i,t}\|^2_{L^2(Q_t(L_0))}$, we obtain

$$\sum_{i=1}^m \|\mathbb{E}(f_i|\mathcal{G}_{L_{j+2},Q})\|^2_{L^2(Q)} \ge \sum_{i=1}^m \|\mathbb{E}(f_i|\mathcal{G}_{L_j,Q})\|^2_{L^2(Q)} + c\varepsilon^3. \tag{3.3.11}$$

It is clear that the sums in the above expressions are bounded by m for all $j \ge 1$, thus (3.3.11) cannot hold for some $1 \le j \le \bar{J}_1$ for $\bar{J}_1 := C\, m\, \varepsilon^{-3}$. This implies that (3.3.10) must hold for some $1 \le j \le \bar{J}_1$, for all $1 \le i \le m$ and all $t \notin T_\varepsilon$ for a set $T_\varepsilon \subseteq \Gamma_{L_0,Q}$ of size $|T_\varepsilon| \le \varepsilon\, |\Gamma_{L_0,Q}|$. $\square$

3.4 The general case: products of simplices in $\mathbb{R}^d$

After these preparations we will now consider the general case of Theorem 3.2. Let $Q = Q_1 \times \cdots \times Q_d \subseteq \mathbb{R}^n$ with $Q_i \subseteq \mathbb{R}^{n_i}$ cubes of equal side length $l(Q)$ and $\Delta^0 = \Delta_1^0 \times \cdots \times \Delta_d^0$ with each $\Delta_i \subseteq \mathbb{R}^{n_i}$ a non-degenerate simplex of n_i points for $1 \le i \le d$.

We will use a generalized version of the hypergraph terminology introduced in Section 3.2. In particular, for a vertex set $I = \{1, 2, \ldots, d\}$ and set $K = \{il;\ 1 \le i \le d,\, 1 \le l \le n_i\}$ we will let $\pi : K \to I$ denote the projection defined by $\pi(il) := i$. As before we will let $\mathcal{H}_{d,k} := \{e \subseteq I;\ |e| = k\}$ denote the complete k-regular hypergraph with vertex set I, and for the multi-index $\underline{n} = (n_1, \ldots, n_d)$ define the hypergraph bundle

$$\mathcal{H}^{\underline{n}}_{d,k} := \{e \subseteq K;\ |e| = |\pi(e)| = k\}$$

noting that $|\pi^{-1}(i)| = n_i$ for all $i \in I$.

In order to parameterize the vertices of direct products of simplices, i.e. sets of the form $\Delta = \Delta_1 \times \cdots \times \Delta_d$ with $\Delta_i \subseteq Q_i$, we consider points $\underline{x} = (\underline{x}_1, \ldots, \underline{x}_d)$ with $\underline{x}_i = (x_{i1}, \ldots, x_{in_i}) \in Q_i^{n_i}$ for each $i \in I$. Now for any $1 \le k \le d$ and any edge $e' \in \mathcal{H}_{d,k}$ we will write $Q_{e'} := \prod_{i\in e'} Q_i$, and for every $\underline{x} \in Q_1^{n_1} \times \cdots \times Q_d^{n_d}$ and $e \in \mathcal{H}^{\underline{n}}_{d,k}$ we define $\underline{x}_e := \pi_e(\underline{x})$, where $\pi_e : Q_1^{n_1} \times \cdots \times Q_d^{n_d} \to Q_{\pi(e)}$ is the natural projection map. Writing $\Delta_i = \{x_{i1}, \ldots, x_{in_i}\}$ we have that $\Delta_1 \times \cdots \times \Delta_d = \{\underline{x}_e :\ e \in \mathcal{H}^{\underline{n}}_{d,d}\}$ since every edge $\underline{x}_e$ is of the form $(x_{1l_1}, \ldots, x_{dl_d})$. We can therefore identify points $\underline{x}$ with configurations of the form $\Delta_1 \times \cdots \times \Delta_d$.

For any $0 < \lambda \ll l(Q)$ the measures $d\sigma_{\lambda\Delta_i^0}$, introduced in Section 3.3.1, are supported on points $(y_2, \ldots, y_{n_i})$ for which the simplex $\Delta_i = \{0, y_2, \ldots, y_{n_i}\}$ is isometric to $\lambda\Delta_i^0$. For simplicity of notation we will write

$$\int_{\underline{x}_i} f(\underline{x}_i)\, d\sigma_i^\lambda(\underline{x}_i) := \fint_{x_{i1}\in Q_i} \int_{x_{i2},\ldots,x_{in_i}} f(\underline{x}_i)\, d\sigma_{\lambda\Delta_i^0}(x_{i2} - x_{i1}, \ldots, x_{in_i} - x_{i1})\, dx_{i1}$$

Note that the support of the measure $d\sigma_i^\lambda$ is the set of points $\underline{x}_i$ so that the simplex $\Delta_i := \{x_{i1}, \ldots, x_{in_i}\}$ is isometric to $\lambda\Delta_i^0$ and $x_{i1} \in Q_i$, moreover the measure is normalized. Thus if $S \subseteq Q$ is a set then the density of configurations Δ in S of the form $\Delta = \Delta_1 \times \ldots \times \Delta_d$ with each $\Delta_i \subseteq Q_i$ an isometric copy of $\lambda\Delta_i^0$ is given by the expression

$$\mathcal{N}^d_{\lambda\Delta^0,Q}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) := \int_{\underline{x}_1} \cdots \int_{\underline{x}_d} \prod_{e\in\mathcal{H}^{\underline{n}}_{d,d}} 1_S(\underline{x}_e)\ d\sigma_1^\lambda(\underline{x}_1) \ldots d\sigma_d^\lambda(\underline{x}_d). \tag{3.4.1}$$

The proof of Theorem 3.2 reduces to establishing the following stronger quantitative result.

Proposition 3.7. *For any $0 < \varepsilon \ll 1$ there exists an integer $J_d = J_d(\varepsilon)$ with the following property; Given any lacunary sequence $l(Q) \geq \lambda_1 \geq \cdots \geq \lambda_{J_d}$ and $S \subseteq Q$, there is some $1 \leq j < J_d$ such that*

$$\mathcal{N}^d_{\lambda\Delta^0,Q}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) > \left(\frac{|S|}{|Q|}\right)^{n_1\cdots n_d} - \varepsilon \tag{3.4.2}$$

for all $\lambda \in [\lambda_{j+1}, \lambda_j]$.

Quantitative Remark. A careful analysis of our proof reveals that there is a choice of $J_d(\varepsilon)$ which is less than $\exp^{(d)}(\log(C_\Delta \varepsilon^{-3}))$, where $\exp^{(k)} k(m)$ is again the tower-exponential function defined by $\exp^{(1)}(m) = \exp(m)$ and $\exp^{(k+1)}(m) = \exp(\exp^{(k)}(m))$ for $k \geq 1$.

For any $0 < \lambda \ll l(Q)$ and set $S \subseteq Q$ we define the expression:

$$\mathcal{M}^d_{\lambda,Q}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) := \fint_{\underline{t}\in Q} \mathcal{M}^d_{\underline{t}+Q(\lambda)}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d})\, d\underline{t} \tag{3.4.3}$$

where $Q(\lambda) = [-\frac{\lambda}{2}, \frac{\lambda}{2}]^n$ and

$$\mathcal{M}^d_{\widetilde{Q}}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) := \fint_{\underline{x}_1\in\widetilde{Q}_1^{n_1}} \cdots \fint_{\underline{x}_d\in\widetilde{Q}_d^{n_d}} \prod_{e\in\mathcal{H}^{\underline{n}}_{d,d}} 1_S(\underline{x}_e)\, d\underline{x}_1 \ldots d\underline{x}_d \tag{3.4.4}$$

for any cube $\widetilde{Q} \subseteq Q$ of the form $\widetilde{Q} = \widetilde{Q}_1 \times \cdots \times \widetilde{Q}_d$ with $\widetilde{Q}_i \subseteq Q_i$ for $1 \leq i \leq d$. Note that if $S \subseteq Q$ is a set of measure $|S| \geq \delta|Q|$ for some $\delta > 0$, then careful applications of Hölder's inequality give

$$\mathcal{M}^d_{\lambda,Q}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) \geq \fint_{\underline{t}\in Q} \left(\fint_{(x_1,\ldots,x_d)\in\underline{t}+Q(\lambda)} 1_S(x_1,\ldots,x_d)\, dx_1 \ldots dx_d \right)^{n_1\cdots n_d} d\underline{t} \geq \delta^{n_1\cdots n_d} - O(\varepsilon)$$

for all scales $0 < \lambda \ll \varepsilon\, l(Q)$.

In light of the discussion above, and that preceding Proposition 3.5, we see that Proposition 3.7, and hence Theorem 3.2 in general, will follows as a consequence of the following

Proposition 3.8. *Let $0 < \varepsilon \ll 1$. There exists an integer $J_d = J_d(\varepsilon)$ such that for any ε-admissible sequence of scales $l(Q) \geq L_1 \geq \cdots \geq L_{J_d}$ and $S \subseteq Q$ there is some $1 \leq j < J_d$ such that*

$$\mathcal{N}^d_{\lambda\Delta^0,Q}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) = \mathcal{M}^d_{\lambda,Q}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) + O(\varepsilon) \tag{3.4.5}$$

for all $\lambda \in [L_{j+1}, L_j]$.

The validity of Proposition 3.8 will follow immediately from the $d = k$ case of Proposition 3.9 below.

3.4.1 Reduction of Proposition 3.8 to a more general "local" counting lemma

For any given $1 \leq k \leq d$ and collection of functions $f_e : Q_{\pi(e)} \to [-1, 1]$ with $e \in \mathcal{H}^{\underline{n}}_{d,k}$ we define the following multi-linear expressions

$$\mathcal{N}^d_{\lambda\Delta^0,Q}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k}) := \int_{\underline{x}_1} \cdots \int_{\underline{x}_d} \prod_{e\in\mathcal{H}^{\underline{n}}_{d,k}} f_e(\underline{x}_e)\; d\sigma_1^\lambda(\underline{x}_1) \ldots .d\sigma_d^\lambda(\underline{x}_d) \tag{3.4.6}$$

and

$$\mathcal{M}^d_{\lambda,Q}(f_e\,;\, e \in \mathcal{H}^{\underline{n}}_{d,k}) := \fint_{\underline{t}\in Q} \mathcal{M}^d_{\underline{t}+Q(\lambda)}(f_e\,;\, e \in \mathcal{H}^{\underline{n}}_{d,k})\, d\underline{t} \tag{3.4.7}$$

where $Q(\lambda)=[-\frac{\lambda}{2},\frac{\lambda}{2}]^n$ and

$$\mathcal{M}^d_{\widetilde{Q}}(f_e\,;\,e\in\mathcal{H}^{\underline{n}}_{d,k}) := \fint_{\underline{x}_1\in\widetilde{Q}_1^{n_1}}\cdots\fint_{\underline{x}_d\in\widetilde{Q}_d^{n_d}}\prod_{e\in\mathcal{H}^{\underline{n}}_{d,k}} f_e(\underline{x}_e)\,d\underline{x}_1\ldots d\underline{x}_d \tag{3.4.8}$$

for any cube $\widetilde{Q}\subseteq Q$ of the form $\widetilde{Q}=\widetilde{Q}_1\times\cdots\times\widetilde{Q}_d$ with $\widetilde{Q}_i\subseteq Q_i$ for $1\le i\le d$.

Our strategy to proving Proposition 3.8 is the same as illustrated in the finite field settings, that is we would like to compare averages $\mathcal{N}_{\lambda\Delta^0,Q}(f_e;e\in\mathcal{H}^{\underline{n}}_{d,k})$ to those of $\mathcal{M}^d_{\lambda,Q}(f_e\,;\,e\in\mathcal{H}^{\underline{n}}_{d,k})$, at certain scales $\lambda\in[L_{j+1},L_j]$, inductively for $1\le k\le d$. However in the Euclidean case, an extra complication emerges due to the fact the (hypergraph) regularity lemma, the analogue of Lemma 3.2, does not produce σ-algebras $\mathcal{B}_{\mathfrak{f}}$, for $\mathfrak{f}\in\mathcal{H}^{\underline{n}}_{d,k-1}$, on the cubes $Q_{\mathfrak{f}}$. In a similar manner to the case for $d=2$ discussed in the previous section, we will only obtain σ-algebras "local" on cubes $Q_{\underline{t}_{\mathfrak{f}}}(L_0)$ at some scale $L_0>0$. This will have the effect that the functions f_e will be replaced by a family of functions $f_{e,\underline{t}}$, where $\underline{t}$ runs through a grid $\Gamma_{L_0,Q}$.

To be more precise, let $L>0$ be a scale dividing the side-length $l(Q)$. For $\underline{t}\in\Gamma_{L,Q}$ and $e'\in\mathcal{H}_{d,k}$ we will use $\underline{t}_{e'}$ to denote the projection of $\underline{t}$ onto $Q_{e'}$ and $Q_{\underline{t}_{e'}}(L):=\underline{t}_{e'}+Q_{e'}(L)$ to denote the projection of the cube $Q_{\underline{t}}(L)$ centered at $\underline{t}$ onto $Q_{e'}$. It is then easy to see that for any $\varepsilon>0$ we have

$$\mathcal{N}^d_{\lambda\Delta^0,Q}(f_e;e\in\mathcal{H}^{\underline{n}}_{d,k}) = \mathbb{E}_{\underline{t}\in\Gamma_{L,Q}}\,\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{t}}(L)}(f_{e,\underline{t}}\,;e\in\mathcal{H}^{\underline{n}}_{d,k})+O(\varepsilon) \tag{3.4.9}$$

and

$$\mathcal{M}^d_{\lambda,Q}(f_e;e\in\mathcal{H}^{\underline{n}}_{d,k}) = \mathbb{E}_{\underline{t}\in\Gamma_{L,Q}}\,\mathcal{M}^d_{\lambda,Q_{\underline{t}}(L)}(f_{e,\underline{t}}\,;e\in\mathcal{H}^{\underline{n}}_{d,k})+O(\varepsilon) \tag{3.4.10}$$

provided $0<\lambda\ll\varepsilon L$ where $f_{e,\underline{t}}$ denotes the restriction of a function f_e to the cube $Q_{\underline{t}}(L)$.

At this point the proof of Proposition 3.8 reduces to showing that the expressions in (3.4.9) and (3.4.10) only differ by $O(\varepsilon)$ at some scales $\lambda\in[L_{j+1},L_j]$, given an ε-admissible sequence $L_0\ge L_1\ge\cdots\ge L_J$, for any collection of bounded functions $f_{e,\underline{t}}$, $e\in\mathcal{H}^{\underline{n}}_{d,k}$, $\underline{t}\in\Gamma_{L_0,Q}$. Indeed, our crucial result will be the following

Proposition 3.9 (Local Counting Lemma)**.** *Let $0<\varepsilon\ll1$ and $M\ge1$. There exists an integer $J_k=J_k(\varepsilon,M)$ such that for any ε-admissible sequence of scales $L_0\ge L_1\ge\cdots\ge L_{J_k}$ with the property that L_0 divides $l(Q)$, and collection of functions*

$$f^m_{e,\underline{t}}:Q_{t_{\pi(e)}}(L_0):\to[-1,1]\quad\textit{with}\quad e\in\mathcal{H}^{\underline{n}}_{d,k},\ 1\le m\le M\ \textit{and}\ \underline{t}\in\Gamma_{L_0,Q}$$

there exists $1\le j<J_k$ and a set $T_\varepsilon\subseteq\Gamma_{L_0,Q}$ of size $|T_\varepsilon|\le\varepsilon|\Gamma_{L_0,Q}|$ such that

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}};\,e\in\mathcal{H}^{\underline{n}}_{d,k}) = \mathcal{M}_{\lambda,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}};\,e\in\mathcal{H}^{\underline{n}}_{d,k})+O(\varepsilon) \tag{3.4.11}$$

for all $\lambda\in[L_{j+1},L_j]$ and $\underline{t}\notin T_\varepsilon$ uniformly in $e\in\mathcal{H}^{\underline{n}}_{d,k}$ and $1\le m\le M$.

3.4.2 Proof of Proposition 3.9

We will prove Proposition 3.9 by induction on $1\le k\le d$. For $k=1$ this is basically Proposition 3.6. Indeed, in this case for a given $\underline{t}=(t_1,\ldots,t_d)\in\Gamma_{L_0,Q}$ and edge $e\in\mathcal{H}^{\underline{n}}_{d,1}=\{il\,:\,1\le i\le d,\,1\le l\le n_i\}$ we have that $f^m_{e,\underline{t}}(\underline{x}_e)=f^m_{il,\underline{t}}(x_{il})$ with $x_{il}\in Q_{t_i}(L_0)$ and hence both

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}};\,e\in\mathcal{H}^{\underline{n}}_{d,1}) = \prod_{i=1}^d\,\mathcal{N}^1_{\lambda\Delta^0_i,Q_{t_i}(L_0)}(f^m_{i1,\underline{t}},\ldots,f^m_{in_i,\underline{t}})$$

$$\mathcal{M}^d_{\lambda,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}};\, e\in\mathcal{H}^{\underline{n}}_{d,1}) = \prod_{i=1}^{d} \mathcal{M}^1_{\lambda,Q_{t_i}(L_0)}(f^m_{i1,\underline{t}},\dots,f^m_{in_i,\underline{t}}).$$

By Proposition 3.6 there exists an $1\le j<J_1=O(M\varepsilon^{-4})$ and an exceptional set $T_\varepsilon\subseteq\Gamma_{L_0,Q}$ of size $|T_\varepsilon|\le\varepsilon|\Gamma_{L_0,Q}|$, such that uniformly for $\underline{t}\notin T_\varepsilon$ and for $1\le i\le d$, one has

$$\mathcal{N}^1_{\lambda\Delta^0_i,Q_{t_i}(L_0)}(f^m_{i1,\underline{t}},\dots,f^m_{in_i,\underline{t}}) = \mathcal{M}^1_{\lambda,Q_{t_i}(L_0)}(f^m_{i1,\underline{t}},\dots,f^m_{in_i,\underline{t}}) + O(\varepsilon)$$

hence

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}};\, e\in\mathcal{H}^{\underline{n}}_{d,1}) = \mathcal{M}^d_{\lambda,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}};\, e\in\mathcal{H}^{\underline{n}}_{d,1}) + O(\varepsilon)$$

as the all factors are trivially bounded by 1 in magnitude. This implies (3.4.11) for $k=1$.

For the induction step we again need two main ingredients. The first establishes that the our multi-linear forms $\mathcal{N}^d_{\lambda\Delta^0,Q}(f_e;\, e\in\mathcal{H}^{\underline{n}}_{d,k})$ are controlled by an appropriate box-type norm attached to a scale L.

Let $Q=Q_1\times\cdots\times Q_d$ and $1\le k\le d$. For any scale $0<L\ll l(Q)$ and function $f:Q_{e'}\to[-1,1]$ with $e'\in\mathcal{H}_{d,k}$ we define its local box norm at scale L by

$$\|f\|^{2^k}_{\square_L(Q_{e'})} := \fint_{\underline{s}\in Q_{e'}} \|f\|^{2^k}_{\square(\underline{s}+Q(L))}\, d\underline{s} \tag{3.4.12}$$

where

$$\|f\|^{2^k}_{\square(\widetilde{Q})} := \fint_{x_{11},x_{12}\in\widetilde{Q}_1}\cdots\fint_{x_{k1},x_{k2}\in\widetilde{Q}_k} \prod_{(\ell_1,\dots,\ell_k)\in\{1,2\}^k} f(x_{1\ell_1},\dots,x_{k\ell_k})\, dx_{11}\, dx_{12}\dots dx_{k1}\, dx_{k2} \tag{3.4.13}$$

for any cube $\widetilde{Q}$ of the form $\widetilde{Q}=\widetilde{Q}_1\times\cdots\times\widetilde{Q}_k$.

Lemma 3.6 (Generalized von-Neumann inequality)**.** *Let $\varepsilon>0$, $0<\lambda\ll l(Q)$ and let $0<L\ll(\varepsilon^{2^k})^6\lambda$. For any $1\le k\le d$ and collection of functions $f_e:Q_{\pi(e)}\to[-1,1]$ with $e\in\mathcal{H}^{\underline{n}}_{d,k}$, we have both*

$$|\mathcal{N}^d_{\lambda\Delta^0,Q}(f_e;\, e\in\mathcal{H}^{\underline{n}}_{d,k})| \le \min_{e\in\mathcal{H}^{\underline{n}}_{d,k}} \|f_e\|_{\square_L(Q_{\pi(e)})} + O(\varepsilon) \tag{3.4.14}$$

$$|\mathcal{M}^d_{\lambda,Q}(f_e;\, e\in\mathcal{H}^{\underline{n}}_{d,k})| \le \min_{e\in\mathcal{H}^{\underline{n}}_{d,k}} \|f_e\|_{\square_L(Q_{\pi(e)})}. \tag{3.4.15}$$

The crucial ingredient is the following analogue of the weak hypergraph regularity lemma.

Lemma 3.7 (Parametric weak hypergraph regularity lemma for $\mathbb{R}^n$)**.** *Let $0<\varepsilon\ll1$, $M\ge1$, and $1\le k\le d$.*

There exists $\bar{J}_k=O(M\varepsilon^{-2^{k+3}})$ such that for any ε^{2^k}-admissible sequence $L_0\ge L_1\ge\cdots\ge L_{\bar{J}_k}$ with the property that L_0 divides $l(Q)$ and collection of functions

$$f^m_{e,\underline{t}}:Q_{\underline{t}_{\pi(e)}}(L_0)\to[-1,1] \quad \textit{with} \quad e\in\mathcal{H}^{\underline{n}}_{d,k},\ 1\le m\le M,\ \textit{and}\ \underline{t}\in\Gamma_{L_0,Q}$$

there is some $1\le j<\bar{J}_k$ and σ-algebras $\mathcal{B}_{e',\underline{t}}$ of scale L_j on $Q_{\underline{t}_{e'}}(L_0)$ for each $\underline{t}\in\Gamma_{L_0,Q}$ and $e'\in\mathcal{H}_{d,k}$ such that

$$\|f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}})\|_{\square_{L_{j+1}}(Q_{\underline{t}_{\pi(e)}}(L_0))} \le \varepsilon \tag{3.4.16}$$

uniformly for all $t\notin T_\varepsilon$, $e\in\mathcal{H}^{\underline{n}}_{d,k}$, and $1\le m\le M$, where $T_\varepsilon\subseteq\Gamma_{L_0,Q}$ with $|T_\varepsilon|\le\varepsilon|\Gamma_{L_0,Q}|$.

Moreover, the σ-algebras $\mathcal{B}_{e',\underline{t}}$ have the additional local structure that the exist σ-algebras $\mathcal{B}_{e',\mathfrak{f}',\underline{s}}$ on $Q_{\underline{s}_{\mathfrak{f}'}}(L_j)$ with $\mathrm{complex}(\mathcal{B}_{e',\mathfrak{f}',\underline{s}})=O(j)$ *for each $\underline{s}\in\Gamma_{L_j,Q}$, $e'\in\mathcal{H}_{d,k}$, and $\mathfrak{f}'\in\partial e'$ such that if $\underline{s}\in Q_{\underline{t}}(L_0)$, then*

$$\mathcal{B}_{e',\underline{t}}\big|_{Q_{\underline{s}_{e'}}(L_j)} = \bigvee_{\mathfrak{f}'\in\partial e'} \mathcal{B}_{e',\mathfrak{f}',\underline{s}}. \tag{3.4.17}$$

Lemma 3.7 is the parametric and simultaneous version of the extension of Lemma 3.7 to the product of d simplices. The difference is that in the general case one has to deal with a parametric family of functions $f^m_{e,\underline{t}}$ as $\underline{t}$ is running through a grid $\Gamma_{L_0,Q}$. The essential new content of Lemma 3.7 is that one can develop σ-algebras $\mathcal{B}_{e',\underline{t}}$ on the cubes $Q_{\underline{t}}(L_0)$ with respect to the family of functions $f^m_{e,\underline{t}}$ such that the local structure described above and (3.4.16) hold simultaneously for almost all $\underline{t} \in \Gamma_{L_0,Q}$.

Proof of Proposition 3.9. Assume the Proposition holds for $k-1$.

Let $\varepsilon > 0$, $\varepsilon_1 := \exp\left(-C_1\varepsilon^{-2^{k+3}}\right)$ for some large constant $C_1 = C_1(n,k,d) \gg 1$, and $\{L_j\}_{j\geq 1}$ be an ε_1-admissible sequence of scales. Set $F(\varepsilon) := J_{k-1}(\varepsilon_1, M)$ with $M = \varepsilon\,\varepsilon_1^{-1}$.
For $L \in \{L_j\}_{j\geq 1}$ we again write index$(L) = j$ if $L = L_j$. We now choose a subsequence $\{L'_j\} \subseteq \{L_j\}$ so that $L'_0 = L_0$ and index$(L'_{j+1}) \geq$ index$(L'_j) + F(\varepsilon) + 2$. Lemma 3.7 then guarantees the existence of σ-algebras $\mathcal{B}_{e',\underline{t}}$ of scale L'_j on $Q_{\underline{t}_{e'}}(L_0)$ for each $\underline{t} \in \Gamma_{L_0,Q}$ and $e' \in \mathcal{H}_{d,k}$, with the local structure described above, such that

$$\|f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}})\|_{\square_{L'_{j+1}}(Q_{\underline{t}_{\pi(e)}}(L_0))} \leq \varepsilon \tag{3.4.18}$$

uniformly for all $t \notin T'_\varepsilon$, $e \in \mathcal{H}^{\underline{n}}_{d,k}$, and $1 \leq m \leq M$, for some $1 \leq j < \bar{J}_k(\varepsilon, M) = O(M\varepsilon^{-2^{k+3}})$, where $T'_\varepsilon \subseteq \Gamma_{L_0,Q}$ with $|T'_\varepsilon| \leq \varepsilon|\Gamma_{L_0,Q}|$. Let $\bar{f}^m_{e,\underline{t}} := \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}})$ for $\underline{t} \in \Gamma_{L_0,Q}$ and $e \in \mathcal{H}^{\underline{n}}_{d,k}$. If $t \notin T'_\varepsilon$, then by (3.4.14), (3.4.15), and (3.4.16) we have both

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{t}}(L_0)}(\bar{f}^m_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon) \tag{3.4.19}$$

$$\mathcal{M}^d_{\lambda,Q_{\underline{t}}(L_0)}(f^m_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathcal{M}^d_{\lambda,Q_{\underline{t}}(L_0)}(\bar{f}^m_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon). \tag{3.4.20}$$

provided $(\varepsilon^{-2^k})^6 L'_{j+1} \ll \lambda$. For given $\underline{s} \in \Gamma_{L'_j,Q_{\underline{t}}(L_0)}$ one may write $\bar{f}^m_{e,\underline{s}}$ for the restriction of $\bar{f}^m_{e,\underline{t}}$ on the cube $Q_{\underline{s}}(L'_j) \subseteq Q_{\underline{t}}(L_0)$, as $\underline{s}$ uniquely determines $\underline{t}$. By localization, provided $\lambda \ll \varepsilon L'_j$, we then have both

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{t}}(L_0)}(\bar{f}^m_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathbb{E}_{\underline{s}\in\Gamma_{L'_j,Q_{\underline{t}}(L_0)}}\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(\bar{f}^m_{e,\underline{s}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon), \tag{3.4.21}$$

$$\mathcal{M}^d_{\lambda,Q_{\underline{t}}(L_0)}(\bar{f}^m_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathbb{E}_{\underline{s}\in\Gamma_{L'_j,Q_{\underline{t}}(L_0)}}\mathcal{M}^d_{\lambda,Q_{\underline{s}}(L'_j)}(\bar{f}^m_{e,\underline{s}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon). \tag{3.4.22}$$

For a fixed cube $Q_{\underline{s}}(L'_j)$ we have that

$$\bar{f}^m_{e,\underline{s}} = \sum_{r_e=1}^{R_{e,\underline{s}}} \alpha_{\underline{s},r_e,m}\ 1_{A^{r_e}_{\pi(e),\underline{s}}} \tag{3.4.23}$$

where $\{A^{r_e}_{\pi(e),\underline{s}}\}_{1\leq r\leq R_{e,\underline{s}}}$ is the family of atoms of the σ-algebra $\mathcal{B}_{\pi(e),\underline{t}}$ restricted to the cube $Q_{\underline{s}}(L'_j)$. Note that $|\alpha_{\underline{s},r_e}| \leq 1$ and $|R_{e,\underline{s}}| = O(\exp\left(C\varepsilon^{-2^{k+3}}\right))$. By adding the empty set to the collection of atoms one may assume $|R_{e,\underline{s}}| = R := \exp\left(C\varepsilon^{-2^{k+3}}\right)$ for all $e \in \mathcal{H}^{\underline{n}}_{d,k}$ and $\underline{s} \in \Gamma_{L'_j,Q}$. Then, by multi-linearity, using the notations $\underline{r} = (r_e)_{e\in\mathcal{H}^{\underline{n}}_{d,k}}$ and $\alpha_{\underline{r},\underline{s}} = \prod_e \alpha_{\underline{s},r_e}$, one has both

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(\bar{f}^m_{\underline{s},e};\ e \in \mathcal{H}^{\underline{n}}_{d,k}) = \sum_{\underline{r}} \alpha_{\underline{s},\underline{r},m}\ \mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(1_{A^{r_e}_{\pi(e),\underline{s}}};\ e \in \mathcal{H}^{\underline{n}}_{d,k}) \tag{3.4.24}$$

$$\mathcal{M}^d_{\lambda,Q_{\underline{s}}(L'_j)}(\bar{f}^m_{\underline{s},e};\, e\in\mathcal{H}^n_{d,k}) = \sum_{\underline{r}} \alpha_{\underline{s},\underline{r},m}\ \mathcal{M}^d_{\lambda,Q_{\underline{s}}(L'_j)}(1_{A^{re}_{\pi(e),\underline{s}}};\, e\in\mathcal{H}^n_{d,k}). \tag{3.4.25}$$

The key observation is that these expressions in the sum above are all at level $k-1$ instead of k. To see this let $e=(i_1l_1,\ldots,i_ml_m,\ldots,i_kl_k)$ so $e'=\pi(e)=(i_1,\ldots,i_m,\ldots,i_k)$. If $\mathfrak{f}'=e'\backslash\{i_m\}$ then recall that the edge $p_{\mathfrak{f}'}(e)=(i_1l_1,\ldots,i_kl_k)\in\mathcal{H}^n_{d,k-1}$ is obtained from e by removing the i_ml_m-entry. Thus, for any atom $A_{e',\underline{s}}$ of $\mathcal{B}_{\underline{s},e'}(L'_j)$ we have by (3.4.17), that

$$1_{A_{e',\underline{s}}}(\underline{x}_e) = \prod_{\mathfrak{f}'\in\partial e'} 1_{A_{e',\mathfrak{f}',\underline{s},}}(\underline{x}_{p_{\mathfrak{f}'}(e)}) \tag{3.4.26}$$

where $A_{e',\mathfrak{f}',\underline{s}}$ is an atom of the σ-algebra $\mathcal{B}_{e',\mathfrak{f}',\underline{s}}$. Thus

$$\prod_{e\in\mathcal{H}^n_{d,k}} 1_{A^{re}_{\pi(e),\underline{s}}}(\underline{x}_e) = \prod_{\mathfrak{f}\in\mathcal{H}^n_{d,k-1}} \prod_{\substack{e\in\mathcal{H}^n_{d,k},\mathfrak{f}'\in\partial\pi(e)\\ p_{\mathfrak{f}'}(e)=f}} 1_{A^{re}_{\pi(e),\mathfrak{f}',\underline{s}}}(\underline{x}_{\mathfrak{f}}) = \prod_{\mathfrak{f}\in\mathcal{H}^n_{d,k-1}} g^r_{\mathfrak{f},\underline{s}}(\underline{x}_{\mathfrak{f}}). \tag{3.4.27}$$

It follows that

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(1_{A^{re}_{\pi(e),\underline{s}}};\, e\in\mathcal{H}^n_{d,k}) = \mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(g^r_{\mathfrak{f},\underline{s}};\, \mathfrak{f}\in\mathcal{H}^n_{d,k-1}) \tag{3.4.28}$$

and hence that

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(\bar{f}^m_{e,\underline{s}};\, e\in\mathcal{H}^n_{d,k}) = \sum_{\underline{r}} \alpha_{\underline{s},\underline{r},m}\ \mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(g^r_{\mathfrak{f},\underline{s}};\, \mathfrak{f}\in\mathcal{H}^n_{d,k-1}) \tag{3.4.29}$$

and similarly

$$\mathcal{M}^d_{\lambda,Q_{\underline{s}}(L'_j)}(\bar{f}^m_{e,\underline{s}};\, e\in\mathcal{H}^n_{d,k}) = \sum_{\underline{r}} \alpha_{\underline{r},\underline{s},m}\ \mathcal{M}^d_{\lambda,Q_{\underline{s}}(L'_j)}(g^r_{\mathfrak{f},\underline{s}};\, \mathfrak{f}\in\mathcal{H}^n_{d,k-1}). \tag{3.4.30}$$

Note that number of index vectors $\underline{r}=(r_e)_{e\in\mathcal{H}^n_{d,k}}$ is R^D with $D:=|\mathcal{H}^n_{d,k}|$ and hence $R^D\leq M$ if $C_1\gg 1$.
Writing $j':=\text{index}(L'_j)$ and $J':=\text{index}(L'_{j+1})$ it then follows from our inductive hypothesis functions, applied with respect to the ε_1-admissible sequence of scales

$$L_{j'+1}\geq L_{j'+2}\geq\cdots\geq L_{J'-1}$$

which is possible as $J'-j'\gg J_{k-1}(\varepsilon_1,R^D)$, that there is a scale L_j with $j'\leq j<J'$ so that

$$\mathcal{N}_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(g^r_{\underline{s},\mathfrak{f}};\, \mathfrak{f}\in\mathcal{H}^n_{d,k-1}) = \mathcal{M}_{\lambda,Q_{\underline{s}}(L'_j)}(g^r_{\underline{s},\mathfrak{f}};\, \mathfrak{f}\in\mathcal{H}^n_{d,k-1}) + O(\varepsilon_1) \tag{3.4.31}$$

for all $\lambda\in[L_{j+1},L_j]$ uniformly in $\underline{r}$ for $\underline{s}\notin S_{\varepsilon_1}$, where $S_{\varepsilon_1}\subseteq\Gamma_{L'_j,Q}$ is a set of size $|S_{\varepsilon_1}|\leq\varepsilon_1|\Gamma_{L'_j,Q}|$. Since the cubes $Q_{\underline{t}}(L_0)$ form a partition of Q as $\underline{t}$ runs through the grid $\Gamma_{L_0,Q}$ the relative density of the set S_{ε_1} can substantially increase only of a few cubes $Q_{\underline{t}}(L_0)$. Indeed, it is easy to see that $|T''_{\varepsilon_1}|\leq\varepsilon_1^{1/2}|\Gamma_{L_0,Q}|$ for the set

$$T''_{\varepsilon_1}:=\{\underline{t}\in\Gamma_{L_0,Q}:\ |S_{\varepsilon_1}\cap Q_{\underline{t}}(L_0)|\geq\varepsilon_1^{1/2}\,|\Gamma_{L'_j,Q}\cap Q_{\underline{t}}(L_0)|\}.$$

We claim that (3.4.11) holds for $\lambda\in[L_{j+1},L_j]$ uniformly in $t\notin T_\varepsilon:=T'_\varepsilon\cup T''_{\varepsilon_1}$, $e\in\mathcal{H}^n_{d,k}$, and $1\leq m\leq M$. Indeed, from (3.6.17), (3.6.18), and (3.4.31) and the fact that $|\alpha_{\underline{s},\underline{r}}|\leq 1$, it follows

$$\mathcal{N}^d_{\lambda\Delta^0,Q_{\underline{s}}(L'_j)}(\bar{f}_{e,\underline{s}};\, e\in\mathcal{H}^n_{d,k}) = \mathcal{M}^d_{\lambda,Q_{\underline{s}}(L'_j)}(\bar{f}_{e,\underline{s}};\, e\in\mathcal{H}^n_{d,k}) + O(\varepsilon)$$

for $\underline{s} \notin S_{\varepsilon_1} \cap Q_{\underline{t}}(L_0)$ since $R^D \varepsilon_1 \ll \varepsilon$. Finally, the fact that $\underline{t} \notin T''_{\varepsilon_1}$ together with localization, namely (3.4.21) and (3.4.22), ensures that averaging over $\Gamma_{L'_j, Q_{\underline{t}}(L_0)}$ gives

$$\mathcal{N}^d_{\lambda\Delta^0, Q_{\underline{t}}(L_0)}(\bar{f}_{e,\underline{t}};\, e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathcal{M}^d_{\lambda, Q_{\underline{t}}(L_0)}(\bar{f}_{e,\underline{t}};\, e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon) + O(\varepsilon_1^{1/2})$$

which in light of (3.4.19), (3.4.20), and the fact that $\varepsilon_1 \ll \varepsilon^2$ complete the proof. $\square$

3.4.3 Proof of Lemmas 3.6 and 3.7

Proof of Lemma 3.6. The argument is similar to that of Lemma 3.1. Fix an edge, say $e_0 = (11, 12, \ldots, 1k)$, and partition the edges $e \in \mathcal{H}^{\underline{n}}_{d,k}$ in to as follows. Let $\mathcal{H}_0$ be the set of those edges e for which $1 \notin \pi(e)$, and for $l = 1, \ldots, n_1$ let $\mathcal{H}_l$ denote the collection of edges of the form $e = (1l, j_2 l_2, \ldots, j_k l_k)$, in other words $e \in \mathcal{H}_l$ if $e = (1l, e')$ for some edge $e' = (j_2 l_2, \ldots, j_k l_k) \in \mathcal{H}^{\underline{n}}_{d-1,k-1}$. Accordingly write

$$\prod_{e \in \mathcal{H}^{\underline{n}}_{d,k}} f_e(\underline{x}_e) = \prod_{e \in \mathcal{H}_0} f_e(\underline{x}_e) \prod_{l=1}^{n_1} \prod_{e' \in \mathcal{H}^{\underline{n}}_{d-1,k-1}} f_{1l,e'}(x_{1l}, \underline{x}_{e'}).$$

For $x \in Q_1$ and $\underline{x}' = (\underline{x}_2, \ldots, \underline{x}_d)$ with $\underline{x}_i \in Q_i^{n_i}$, define

$$g_l(x, \underline{x}') := \prod_{e' \in \mathcal{H}^{\underline{n}}_{d-1,k-1}} f_{1l,e'}(x_{1l}, \underline{x}_{e'}) \tag{3.4.32}$$

Then one may write

$$\mathcal{N}^d_{\lambda\Delta^0, Q}(f_e;\, e \in \mathcal{H}^{\underline{n}}_{d,k}) = \fint_{\underline{x}_2} \cdots \fint_{\underline{x}_d} \prod_{e \in \mathcal{H}_0} f_e(\underline{x}_e) \left(\fint_{\underline{x}_1} \prod_{l=1}^{n_1} g_l(x_{1l}, \underline{x}')\, d\sigma_1^\lambda(\underline{x}_1) \right) d\sigma_d^\lambda(\underline{x}_d) \ldots d\sigma_2^\lambda(\underline{x}_2). \tag{3.4.33}$$

For the inner integrals we have, using (3.3.6), the estimate

$$\left(\fint_{\underline{x}_1} \prod_{l=1}^{n_1} g_l(x_{1l}, \underline{x}')\, d\sigma_1^\lambda \right)^2 \leq \|g_1\|^2_{U^1_L(Q)} + O(\varepsilon^{2^k}) = \fint_{y_{11}} \int_{y_{12}} g_1(y_{11}) g_1(y_{12}) \psi^1_L(y_{12} - y_{11})\, dy_{11}\, dy_{12} + O(\varepsilon^{2^k}).$$

provided $0 < L \ll (\varepsilon^{2^k})^6 \lambda$, where we use the notation

$$\psi^i_L(y_2 - y_1) = \int_t \chi^i_L(y_1 - t) \chi^i_L(y_2 - t)\, dt$$

with $\chi^i_L := L^{-n_i} 1_{[-L/2, L/2]^{n_i}}$ for $1 \leq i \leq k$. By Cauchy-Schwarz we then have

$$\left| \mathcal{N}^d_{\lambda\Delta^0, Q}(f_e;\, e \in \mathcal{H}^{\underline{n}}_{d,k}) \right|^2 \leq \int_{\underline{y}_1} \fint_{\underline{x}_2} \cdots \fint_{\underline{x}_d} \prod_{e' \in \mathcal{H}^{\underline{n}}_{d-1,k-1}} f_{11,e'}(x_{11}, \underline{x}_{e'}) f_{11,e'}(x_{12}, \underline{x}_{e'})\, d\sigma_d^\lambda \ldots d\sigma_2^\lambda\, d\omega^1_L(\underline{y}_1) + O(\varepsilon^{2^k})$$

where $d\omega^i_L(\underline{y}_i) = |Q_i|^{-1} \psi^i_L(y_{i2} - y_{i1})\, dy_{i1}\, dy_{i2}$ with $\underline{y}_i = (y_{i1}, y_{i2}) \in Q_i^2$ for $1 \leq i \leq k$.
The expression we have obtained above is similar to the one in (3.4.2) except for the following changes. The variable $\underline{x}_1 \in Q_1^{n_1}$ is replaced by $\underline{y}_1 \in Q_1^2$ and the measure $d\sigma_1^\lambda$ by $d\omega^1_L$. The functions $f_{1l,e'}$ are replaced by $f_{11,e'}$, for $1 \leq l \leq n_1$, while the functions f_e for all $e \in \mathcal{H}^{\underline{n}}_{d,k}$ such that $1 \notin \pi(e)$ are eliminated, that is replaced by 1. Repeating the same procedure for $i = 2, \ldots, k$ replaces all

variables $\underline{x}_i$ with variables $\underline{y}_i$ as well as the measures $d\sigma_i^\lambda$ with $d\omega_L^i$. The procedure eliminates all functions f_e when e is an edge such that $i \notin \pi(e)$ for some $1 \le i \le k$; for the remaining edges, when $\pi(e) = (1,\dots,k)$, it replaces the functions f_e with $f_{e_0} = f_{11,21,\dots,1k}$. For $k < i$ the variables $\underline{x}_i$ and the measures $d\sigma_i^\lambda$ are not changed, however integrating in these variables will have no contribution as the measures are normalized. Thus one obtains the following final estimate

$$\left|\mathcal{N}_{\lambda\Delta^0,Q}(f_e;\, e \in \mathcal{H}^n_{d,k}\right|^{2^k} \le \frac{1}{|Q_1|}\int_{\underline{y}_1} \cdots \frac{1}{|Q_k|}\int_{\underline{y}_k} \prod_{e\in\mathcal{H}^2_{k,k}} f_{e_0}(\underline{y}_e) \prod_{i=1}^k \psi_L^i(y_{i2}-y_{i1})\, dy_{i1}\, dy_{i2} + O(\varepsilon^{2^k}) \tag{3.4.34}$$

noting that these integrals are not normalized. Thus, one may write the expression in (3.4.34), using a change of variables $y_{i1} := y_{i1} - t_i$, $y_{i2} := y_{i2} - t_i$, as

$$\frac{1}{|Q_1|}\int_{t_1} \fint_{\underline{y}_1\in t_1+Q_1} \cdots \frac{1}{|Q_k|}\int_{t_k} \fint_{\underline{y}_k\in t_k+Q_k} \prod_{e\in\mathcal{H}^2_{k,k}} f_{e_0}(\underline{y}_e)\, d\underline{y}_1 \dots d\underline{y}_k\, d\underline{t} = \|f_{e_0}\|^{2^k}_{\square_L(Q_{\pi(e_0)})} + O(\varepsilon^{2^k}) \tag{3.4.35}$$

where the last equality follows from the facts that the function f_{e_0} is supported on the cube $Q_{\pi(e_0)}$ and hence the integration in $\underline{t}$ is restricted to the cube $Q + Q(L)$, giving rise an error of $O(L/l(Q))$. Estimate (3.4.14) follows from (3.4.34) and (3.4.35) noting that the above procedure can be applied to any $e \in \mathcal{H}^n_{d,k}$ in place of e_0. Estimate (3.4.15) is established similarly. □

Proof of Lemma 3.7. For $j = 0$ we set $\mathcal{B}_{e',\underline{t}}(L_0) := \{Q_{\underline{t}}(L_0), \emptyset\}$ and $\mathcal{B}_{e',\mathfrak{f}',\underline{s}}(L_0) := \{Q_{\underline{s}_{\mathfrak{f}'}}(L_0), \emptyset\}$ for $e' \in \mathcal{H}_{d,k}$, $\mathfrak{f}' \in \partial e'$, and $\underline{t}, \underline{s} \in \Gamma_{L_0,Q}$. We will develop σ-algebras $\mathcal{B}_{e',\underline{t}}(L_j)$ of scale L_j such that (3.4.17) holds with complex$(\mathcal{B}_{e',\mathfrak{f}',\underline{s}}(L_j)) \le j$.
We define the *total energy* of a family of functions $f^m_{e,\underline{t}}$ with respect to a family of σ-algebras $\mathcal{B}_{e',\underline{t}}(L_j)$ as

$$\mathcal{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j)) := \mathbb{E}_{\underline{t}\in\Gamma_{L_0,Q}} \sum_{m=1}^M \sum_{e\in\mathcal{H}^n_{d,k}} \|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}}(L_j))\|^2_{L^2(Q_{\underline{t}_{\pi(e)}}(L_0))}. \tag{3.4.36}$$

Since $|f^m_{e,\underline{t}}| \le 1$ for all e, m, and $\underline{t}$ it follows that the total energy is bounded by $M\cdot|\mathcal{H}^n_{d,k}| = O(M)$. Our strategy will be to show that if (3.4.16) does not hold then there exist a family of σ-algebras $\mathcal{B}_{e',\underline{t}}(L_{j+2})$ such that the total energy of the family of functions $f^m_{e,\underline{t}}$ is increased by at least $c_k\varepsilon^{2^{k+3}}$ with respect to this new family of σ-algebras, and at the same time ensuring that (3.4.17) remains valid with complex$(\mathcal{B}_{e',\mathfrak{f}',\underline{s}}(L_{j+2})) \le j+2$. This iterative process must stop at some $j = O(M\,\varepsilon^{-2^{k+3}})$ proving the Lemma.
Assume that we have developed σ-algebras $\mathcal{B}_{e',\underline{t}}(L_j)$ and $\mathcal{B}_{e',\mathfrak{f}',\underline{s}}(L_j)$ of scale L_j such that (3.4.17) holds with complex$(\mathcal{B}_{e',\mathfrak{f}',\underline{s}}(L_j)) \le j$. If (3.4.16) does not hold then $|T_\varepsilon| \ge \varepsilon|\Gamma_{L_0,Q}|$ for the set

$$T_\varepsilon := \{\underline{t} \in \Gamma_{L_0,Q} \,:\, \|f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}}(L_j))\|_{\square_{L_{j+1}}(Q_{\underline{t}_{\pi(e)}}(L_0))} \ge \varepsilon \ \text{ for some } e \in \mathcal{H}^n_{d,k} \text{ and } 1 \le m \le M\}.$$

Fix $\underline{t} \in T_\varepsilon$ and let $e \in \mathcal{H}^n_{d,k}$ and $1 \le m \le M$ be such that

$$\|f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}}(L_j))\|_{\square_{L_{j+1}}(Q_{\underline{t}_{\pi(e)}}(L_0))} \ge \varepsilon$$

and write $e' := \pi(e)$. Consider the partition of the cube $Q_{\underline{t}_{e'}}(L_0)$ into small cubes $Q_{\underline{s}_{e'}}(L_{j+2})$ where $\underline{s}_{e'} \in \Gamma_{L_{j+2},Q_{e'}} \cap Q_{\underline{t}_{e'}}(L_0)$. By the localization properties of the $\square_{L_{j+1}}(Q)$-norm, and the fact that $L_{j+2} \ll \varepsilon^{2^k} L_{j+1}$ we have that

$$\|f\|^{2^k}_{\square_{L_{j+1}}(Q_{\underline{t}_{e'}}(L_0))} \le \mathbb{E}_{\underline{s}_{e'}\in\Gamma_{L_{j+2},Q_{\underline{t}_{e'}}(L_0)}}\ \|f\|^{2^k}_{\square(Q_{\underline{s}_{e'}}(L_{j+2}))} + \frac{\varepsilon^{2^k}}{2}$$

for any function $f : Q_{\underline{t}_{e'}}(L_0) \to [-1,1]$. Thus there exists a set $S_{\varepsilon,e,\underline{t}} \subseteq \Gamma_{L_{j+2},Q_{\underline{t}_{e'}}(L_0)}$ of size

$$|S_{\varepsilon,e,\underline{t}}| \geq \frac{\varepsilon^{2^k}}{4} |\Gamma_{L_{j+2},Q_{\underline{t}_{e'}}(L_0)}|$$

such that

$$\|f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j))\|^{2^k}_{\square(Q_{\underline{s}_{e'}}(L_{j+2})} \geq \frac{\varepsilon^{2^k}}{4} \tag{3.4.37}$$

for all $\underline{s}_{e'} \in S_{\varepsilon,e,\underline{t}}$.

For a given cube Q and functions $f, g : Q \to \mathbb{R}$, define the normalized inner product of f and g as

$$\langle f, g\rangle_Q := \fint_Q f(x)g(x)\,dx.$$

Then by the well-known property of the $\square$-norm, see for example [109] or the proof of Lemma 3.2, it follows from (3.4.37) that there exits sets

$$B_{\mathfrak{f}',\underline{s}_{e'},\underline{t}} \subseteq Q_{\underline{s}_{\mathfrak{f}'}}(L_{j+2})$$

for $\mathfrak{f}' \in \partial e'$ such that

$$\Big\langle f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j))\,,\ \prod_{\mathfrak{f}'\in\partial e'} 1_{B_{\mathfrak{f}',\underline{s}_{e'},\underline{t}}}\Big\rangle_{Q_{\underline{s}_{e'}}(L_{j+2})} \geq \frac{\varepsilon^{2^k}}{2^{k+2}}. \tag{3.4.38}$$

If $\underline{s} \in \Gamma_{L_{j+2},Q}$ then there is a unique $\underline{t} = \underline{t}(\underline{s}) \in \Gamma_{L_0,Q}$ such that $\underline{s} \in Q_{\underline{t}}(L_0)$. If $\underline{t} \in T_\varepsilon$ and $\underline{s}_{e'} \in S_{\varepsilon,e,\underline{t}}$ then we define the σ-algebras $\mathcal{B}_{\mathfrak{f}',e',\underline{s}}(L_{j+2})$ on $Q_{\underline{s}_{\mathfrak{f}'}}(L_{j+2})$ as follows. Write $B_{\mathfrak{f}',e',\underline{s}} = B_{\mathfrak{f}',\underline{s}_{e'},\underline{t}}$ where $\underline{t} = \underline{t}(\underline{s})$ and let $\mathcal{B}_{\mathfrak{f}',e',\underline{s}}(L_{j+2})$ be the σ-algebra generated by the set $B_{\mathfrak{f}',e',\underline{s}}$ and the σ-algebra $\mathcal{B}_{\mathfrak{f}',e',\underline{s}'}(L_j)$ restricted to $Q_{\underline{s}_{\mathfrak{f}'}}(L_{j+2})$ where $\underline{s}' \in \Gamma_{L_j,Q}$ is the unique element so that $\underline{s} \in Q_{\underline{s}'}(L_j)$. Note that that the complexity of the σ-algebra $\mathcal{B}_{\mathfrak{f}',e',\underline{s}}(L_{j+2})$ is at most one larger then the complexity of the σ-algebra $\mathcal{B}_{\mathfrak{f}',e',\underline{s}'}(L_j)$ as restricting a σ-algebra to a set does not increase its complexity. If $\underline{t} = \underline{t}(\underline{s}) \notin T_\varepsilon$ or $\underline{s}_{e'} \notin S_{\varepsilon,e,\underline{t}}$ then let $\mathcal{B}_{\mathfrak{f}',e',\underline{s}}(L_{j+2})$ be simply the restriction of $\mathcal{B}_{\mathfrak{f}',e',\underline{s}'}(L_j)$ to the cube $Q_{\underline{s}_{\mathfrak{f}'}}(L_{j+2})$, or equivalently define the sets $B_{\mathfrak{f}',e',\underline{s}} := Q_{\underline{s}_{\mathfrak{f}'}}(L_{j+2})$. Finally, let

$$\mathcal{B}_{e',\underline{s}}(L_{j+2}) := \bigvee_{\mathfrak{f}'\in\partial e'} \mathcal{B}_{\mathfrak{f}',e',\underline{s}}(L_{j+2}) \tag{3.4.39}$$

be the corresponding σ-algebra on the cube $Q_{\underline{s}_{e'}}(L_{j+2})$.

Since the cubes $Q_{\underline{s}_{e'}}(L_{j+2})$ partition the cube $Q_{\underline{t}_{e'}}(L_0)$ as $\underline{s}_{e'}$ runs through the grid $\Gamma_{L_{j+2},Q_{e'}} \cap Q_{\underline{t}_{e'}}(L_0)$, these σ-algebras define a σ-algebra $\mathcal{B}_{e',\underline{t}}(L_{j+2})$ on $Q_{\underline{t}_{e'}}(L_0)$, such that its restriction to the cubes $Q_{\underline{s}_{e'}}(L_{j+2})$ is equal to the σ-algebras $\mathcal{B}_{e',\underline{s}}(L_{j+2})$.

Since the function $\prod_{\mathfrak{f}'\in\partial e'} 1_{B_{\mathfrak{f}',e',\underline{s}}}$ is measurable with respect to the σ-algebra $\mathcal{B}_{e',\underline{t}}(L_{j+2})$ restricted to the cube $Q_{\underline{s}_{e'}}(L_{j+2})$ one clearly has

$$\langle\, f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_{j+2})),\ \prod_{\mathfrak{f}'\in\partial e'} 1_{B_{\mathfrak{f}',e',\underline{s}}}\,\rangle_{Q_{\underline{s}_{e'}}(L_{j+2})} = 0. \tag{3.4.40}$$

and hence, by (3.4.38), that

$$\langle\, \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_{j+2})) - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j)),\ \prod_{\mathfrak{f}'\in\partial e'} 1_{B_{\mathfrak{f}',e',\underline{s}}}\,\rangle_{Q_{\underline{s}_{e'}}(L_{j+2})} \geq \frac{\varepsilon^{2^k}}{2^{k+2}}. \tag{3.4.41}$$

It then follows from Cauchy-Schwarz and orthogonality, using the fact that the σ-algebra $\mathcal{B}_{e',\underline{t}}(L_{j+2}))$ is a refinement of $\mathcal{B}_{e',\underline{t}}(L_{j+2})$, that

$$\begin{aligned}\|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_{j+2}))-\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j))\|^2_{L^2(Q_{\underline{s}_{e'}}(L_{j+2}))} \quad (3.4.42)\\ = \|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_{j+2}))\|^2_{L^2(Q_{\underline{s}_{e'}}(L_{j+2}))} - \|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j))\|^2_{L^2(Q_{\underline{s}_{e'}}(L_{j+2}))}\\ \geq \left(\frac{\varepsilon^{2^k}}{2^{k+2}}\right)^2\end{aligned}$$

for $\underline{s}_{e'} \in S_{\varepsilon,e,\underline{t}}$. Since $|S_{\varepsilon,e,\underline{t}}| \geq \dfrac{\varepsilon^{2^k}}{4}|\Gamma_{L_{j+2},Q_{\underline{t}_{e'}}}(L_0)|$ averaging over $\underline{s}_{e'} \in \Gamma_{L_{j+2},Q_{\underline{t}_{e'}}}(L_0)$ implies

$$\|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_{j+2}))\|^2_{L^2(Q_{\underline{t}_{e'}}(L_0))} \;\geq\; \|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j))\|^2_{L^2(Q_{\underline{t}_{e'}}(L_0))} + \frac{\varepsilon^{2^{k+2}}}{2^{2k+6}}. \quad (3.4.43)$$

At this point we have shown that if $\underline{t} \in T_\varepsilon$ then there exists an edge $e \in \mathcal{H}^n_{d,k}$, $1 \leq m \leq M$, and σ-algebras $\mathcal{B}_{e',\underline{t}}(L_{j+2}))$ of scale L_{j+2} on $Q_{\underline{t}_{e'}}(L_0)$, with $e' = \pi(e)$, such that (3.4.43) holds.
For all $e'' \in \mathcal{H}_{d,k}$ with $e'' \neq e'$ let $\mathcal{B}_{\mathfrak{f}',e'',\underline{s}}(L_{j+2})$ be the restriction of the σ-algebra $\mathcal{B}_{\mathfrak{f}',e'',\underline{s}'}(L_j)$ to the cube $Q_{\underline{s}_{\mathfrak{f}'}}(L_{j+2})$, where $\underline{s}'$ is such that $\underline{s} \in Q_{\underline{s}'}(L_j)$. By (3.4.39) this implies that $\mathcal{B}_{e'',\underline{s}}(L_{j+2})$ is also the restriction of $\mathcal{B}_{e'',\underline{s}'}(L_j)$ to the cube $Q_{\underline{s}_{e''}}(L_{j+2})$, and hence the σ-algebra $\mathcal{B}_{e'',\underline{t}}(L_{j+2})$ is generated by the grid $\mathcal{G}_{L_{j+2},Q_{\underline{t}_{e''}}(L_0)}$ and the σ-algebra $\mathcal{B}_{e'',\underline{t}}(L_j)$.
We have therefore defined a family of the σ-algebras $\mathcal{B}_{e',\underline{t}}(L_{j+2})$ for $e' \in \mathcal{H}_{d,k}$, satisfying

$$\sum_{m=1}^M \sum_{e\in\mathcal{H}^n_{d,k}} \|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}}(L_{j+2}))\|^2_{L^2(Q_{\underline{t}_{\pi(e)}}(L_0))} \geq \sum_{m=1}^M \sum_{e'\in\mathcal{H}^n_{d,k}} \|\mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}}(L_j))\|^2_{L^2(Q_{\underline{t}_{\pi(e)}}(L_0))} + \frac{\varepsilon^{2^{k+2}}}{2^{2k+6}}.$$

Using the fact that $|T_\varepsilon| \geq \varepsilon|\Gamma_{L_0,Q}|$ and averaging over $\underline{t} \in \Gamma_{L_0,Q}$ it follows using the notations of (3.4.36) that

$$\mathcal{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_{j+2})) \;\geq\; \mathcal{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j)) + \frac{\varepsilon^{2^{k+3}}}{2^{2k+6}}.$$

As the total energy $\mathcal{E}(f^m_{e,\underline{t}}|\mathcal{B}_{e',\underline{t}}(L_j))$ is bounded by $O(M)$, the process must stop at a step $j = O(M\,\varepsilon^{-2^{k+3}})$ where (3.4.16) holds for a σ-algebra of "local complexity" at most j, completing the proof of Lemma 3.7. □

3.5 The base case of an inductive strategy to establish Theorem 3.4

In this section we will ultimately establish the base case of our more general inductive argument. We will however start by giving a (new) proof of Theorem B′, namely the case $d = 1$ of Theorem 3.4.

3.5.1 A Single Simplex in $\mathbb{Z}^n$

Let $\Delta^0 = \{v_1 = 0, v_2, \dots, v_{n_1}\}$ be a fixed non-degenerate simplex of n_1 points in $\mathbb{Z}^n$ with $n = 2n_1+3$ and define $t_{kl} := v_k \cdot v_l$ for $2 \leq k, l \leq n_1$. Recall, see [85], that a simplex $\Delta = \{m_1 = 0, \dots, m_{n_1}\} \subseteq \mathbb{Z}^n$ is isometric to $\lambda\Delta^0$ if and only if $m_k \cdot m_l = \lambda^2 t_{kl}$ for all $2 \leq k, l \leq n_1$.

For any positive integer q and $\lambda \in q\sqrt{\mathbb{N}}$ we define $S_{\lambda\Delta^0,q}(m_2,\dots,m_{n_1}) : \mathbb{Z}^{n(n_1-1)} \to \{0,1\}$ be the function whose value is 1 if $m_k \cdot m_l = \lambda^2 t_{kl}$ with both m_k and m_l in $(q\mathbb{Z})^n$ for all $2 \leq k, l \leq n_1$ and is equal to 0 otherwise. It is a well-known fact in number theory, see [67] or [85], that for $n \geq 2n_1 + 1$ we have that

$$\sum_{m_2,\dots,m_{n_1}} S_{\lambda\Delta^0,q}(m_2,\dots,m_{n_1}) = \rho(\Delta^0)\,(\lambda/q)^{(n-n_1)(n_1-1)}(1+O(\lambda^{-\tau}))$$

for some absolute constant $\tau > 0$ and some constant $\rho(\Delta^0) > 0$, the so-called singular series, which can be interpreted as the product of the densities of the solutions of the above system of equations among the p-adics and among the reals. Thus if we define

$$\sigma_{\lambda\Delta^0,q} := \rho(\Delta^0)^{-1}(\lambda/q)^{-(n-n_1)(n_1-1)} S_{\lambda\Delta^0,q}$$

then $\sigma_{\lambda\Delta^0,q}$ is normalized in so much that

$$\sum_{m_2,\dots,m_{n_1}} \sigma_{\lambda\Delta^0,q}(m_2,\dots,m_{n_1}) = 1 + O(\lambda^{-\tau})$$

for some absolute constant $\tau > 0$.

Let $Q \subseteq \mathbb{Z}^n$ be a fixed cube and let $l(Q)$ denotes its side length. For any family of functions

$$f_1,\dots,f_{n_1} : Q \to [-1,1]$$

and $0 < \lambda \ll l(Q)$ we define the following two multi-linear expressions

$$\mathcal{N}^1_{\lambda\Delta^0,q,Q}(f_1,\dots,f_{n_1}) := \mathbb{E}_{m_1\in Q} \sum_{m_2,\dots,m_{n_1}} f_1(m_1)\dots f_{n_1}(m_{n_1})\,\sigma_{\lambda\Delta^0,q}(m_2-m_1,\dots,m_{n_1}-m_1) \quad (3.5.1)$$

and

$$\mathcal{M}^1_{\lambda,q,Q}(f_1,\dots,f_{n_1}) := \mathbb{E}_{t\in Q}\,\mathbb{E}_{m_1,\dots,m_{n_1}\in t+Q(q,\lambda)}\; f_1(m_1)\dots f_{n_1}(m_{n_1}), \quad (3.5.2)$$

where $Q(q,\lambda) := [-\frac{\lambda}{2},\frac{\lambda}{2}]^n \cap (q\mathbb{Z})^n$. Note that if $S \subseteq Q$ and $\mathcal{N}^1_{\lambda\Delta^0,q,Q}(1_S,\dots,1_S) > 0$ then S must contain an isometric copy of $\lambda\Delta^0$, while if $|S| \geq \delta|Q|$ for some $\delta > 0$ then as before Hölder implies that

$$\mathcal{M}^1_{\lambda,q,Q}(1_S,\dots,1_S) \geq \delta^n - O(\varepsilon) \quad (3.5.3)$$

for all scales $\lambda \in q\sqrt{\mathbb{N}}$ with $0 < \lambda \ll \varepsilon\, l(Q)$.

Recall that for any $0 < \varepsilon \ll 1$ and positive integer q we call a sequence $L_1 \geq \cdots \geq L_J$ (ε,q)-*admissible* if $L_j/L_{j+1} \in \mathbb{N}$ and $L_{j+1} \ll \varepsilon^2 L_j$ for all $1 \leq j < J$ and $L_J/q \in \mathbb{N}$. Note that if $\lambda_1 \geq \cdots \geq \lambda_{J'} \geq 1$ is any lacunary sequence in $q\sqrt{\mathbb{N}}$ with $J' \gg (\log\varepsilon^{-1})\,J + \log q$, one can always finds an (ε,q)-admissible sequence of scales $L_1 \geq \cdots \geq L_J$ with the property that for each $1 \leq j < J$ the interval $[L_{j+1},L_j]$ contains at least two consecutive elements from the original lacunary sequence.
In light of these observations we see that the following "counting lemma" ultimately establishes a quantitatively stronger version of Proposition B$'$ and hence immediately establishes Theorem 3.4 for $d = 1$.

Proposition 3.10. *Let $0 < \varepsilon \ll 1$ and $q_j := q_1(\varepsilon)^j$ for $j \geq 1$ with $q_1(\varepsilon) :=$ lcm$\{1 \leq q \leq C\varepsilon^{-10}\}$. There exists $J_1 = O(\varepsilon^{-2})$ such that for any (ε,q_{J_1})-admissible sequence of scales $l(Q) \geq L_1 \geq \cdots \geq L_{J_1}$ and $S \subseteq Q$ there is some $1 \leq j < J_1$ such that*

$$\mathcal{N}^1_{\lambda\Delta^0,q_j,Q}(1_S,\dots,1_S) = \mathcal{M}^1_{\lambda,q_j,Q}(1_S,\dots,1_S) + O(\varepsilon) \quad (3.5.4)$$

for all $\lambda \in q_j\sqrt{\mathbb{N}}$ with $L_{j+1} \leq \lambda \leq L_j$.

As in the continuous setting the proof of Proposition 3.10 has two main ingredients, namely Lemmas 3.8 and 3.9 below. In these lemmas, and for the remainder Sections 3.5 and 3.6, we will continue to use the notation

$$q_1(\varepsilon) := \ lcm\{1 \leq q \leq C\varepsilon^{-10}\}$$

for any given $\varepsilon > 0$.

Lemma 3.8 (A Generalized von Neumann inequality).
Let $0 < \varepsilon \ll 1$, $q, q' \in \mathbb{N}$ with $qq_1(\varepsilon)|q'$, and $\lambda \in q\sqrt{\mathbb{N}}$ with $\lambda \ll l(Q)$ and $1 \ll L \ll \varepsilon^{10}\lambda$. For any collection of functions $f_1, \dots, f_{n_1} : Q \to [-1, 1]$ we have

$$|\mathcal{N}^1_{\lambda\Delta^0,q,Q}(f_1, \dots, f_{n_1})| \leq \min_{1\leq i\leq n_1} \|f_i\|_{U^1_{q',L}(Q)} + O(\varepsilon) \tag{3.5.5}$$

where for any function $f : Q \to [-1, 1]$ we define

$$\|f\|_{U^1_{q,L}(Q)} := \Big(\frac{1}{|Q|}\sum_{t\in Q} |f * \chi_{q,L}(t)|^2\Big)^{1/2} \tag{3.5.6}$$

with $\chi_{q,L}$ denoting the normalized characteristic function of the cubes $Q(q, L) := [-\frac{L}{2}, \frac{L}{2}]^n \cap (q\mathbb{Z})^n$.

For any cube $Q \subseteq \mathbb{Z}^n$ of side length $l(Q)$ and $q, L \in \mathbb{N}$ satisfying $q \ll L$ with L dividing $l(Q)$, we shall now partition Q into cubic grids $Q_t(q, L) = t + ((q\mathbb{Z})^n \cap Q(L))$, with $Q(L) = [-\frac{L}{2}, \frac{L}{2}]^n$ as usual. These grids form the atoms of a σ-algebra $\mathcal{G}_{q,L,Q}$. Note that if $q|q'$ and $L'|L$ then $\mathcal{G}_{q,L,Q} \subseteq \mathcal{G}_{q',L',Q}$.

Lemma 3.9 (A Koopman-von Neumann type decomposition).
Let $0 < \varepsilon \ll 1$ and $q_j := q_1(\varepsilon)^j$ for all $j \geq 1$. There exists an integer $\bar{J}_1 = O(\varepsilon^{-2})$ such that any $(\varepsilon, q_{\bar{J}_1})$-admissible sequence of scales $l(Q) \geq L_1 \geq \cdots \geq L_{\bar{J}_1}$ and function $f : Q \to [-1, 1]$ there is some $1 \leq j < \bar{J}_1$ such that

$$\|f - \mathbb{E}(f|\mathcal{G}_{q_j,L_j,Q})\|_{U^1_{q_{j+1},L_{j+1}}(Q)} \leq \varepsilon. \tag{3.5.7}$$

The reduction of Proposition 3.10 to these two lemmas is essentially identical to the analogous argument in the continuous setting as presented at the end of Section 3.3.1, we choose to omit the details.

Proof of Lemma 3.8. We will rely on some prior exponential sum estimates, specifically Propositions 4.2 and 4.4 in [85]. First we deal with the case $n_1 \geq 3$. By the change of variables $m_1 := m_1$, $m_i := m_i - m_1$ for $2 \leq i \leq n_1$, one may write

$$\mathcal{N}^1_{\lambda\Delta^0,q,Q}(f_1, \dots, f_{n_1}) := \mathbb{E}_{m_1\in Q_N} \sum_{m_2,\dots,m_{n_1}} f_1(m_1)f_2(m_1+m_2)\cdots f_{n_1}(m_1+m_{n_1})\,\sigma_{\lambda\Delta^0,q}(m_2, \dots, m_{n_1}).$$

We now write

$$\sigma_{\lambda\Delta^0,q}(m_2, \dots, m_{n_1}) = \sigma_{\lambda\Delta^{0\prime},q}(m_2, \dots, m_{n_1-1})\,\sigma^{m_2,\dots,m_{n_1-1}}_{\lambda,q}(m_{n_1})$$

where $\Delta^{0\prime} = \{v_1 = 0, v_2, \dots, v_{n_1-1}\}$ and for each $m_2, \dots, m_{n_1-1} \in (q\mathbb{Z})^n$ we are using $\sigma^{m_2,\dots,m_{n_1-1}}_{\lambda,q}(m)$ denote the (essentially) normalized indicator function of the subset of $(q\mathbb{Z})^n$ that contains m if and only if $m \cdot m_k = \lambda^2 t_{kn_1}$ for all $2 \leq k \leq n_1$.
Using the fact that $|f_i| \leq 1$, together with Cauchy-Schwarz and Plancherel, one can then easily see that

$$|\mathcal{N}^1_{\lambda\Delta^0,q,Q}(f_1, \dots, f_{n_1})|^2 \leq |Q|^{-1} \int_{\xi\in\mathbb{T}^n} |\widehat{f}_{n_1}(\xi)|^2 H_{\lambda,q}(\xi)\, d\xi \tag{3.5.8}$$

with
$$H_{\lambda,q}(\xi) = \sum_{m_2,\dots,m_{n_1}} \sigma_{\lambda\Delta^{0\prime},q}(m_2,\dots,m_{n_1-1})\, |\sigma_{\lambda,q}^{\widehat{m_2,\dots,m_{n_1-1}}}(\xi)|^2.$$

It then follows by Propositions 4.2 and 4.4 in [85], with $\delta = \varepsilon^4$ and after rescaling by q, that in addition to being non-negative and uniformly bounded in ξ we in fact have

$$H_{\lambda,q}(\xi) = O(\varepsilon) \quad \text{whenever} \quad \left| q\xi - \frac{l}{q_1(\varepsilon)} \right| \geq \frac{q}{\varepsilon^4 \lambda}, \tag{3.5.9}$$

for all $l \in \mathbb{Z}^n$.
We note that the expression $H_{\lambda,q}(\xi)$ may be interpreted as the Fourier transform of the indicator function of the set of integer points on a certain variety, and estimate (3.5.9) indicates that this concentrates near rational points of small denominator. It is this crucial fact from number theory which makes results like Theorem B′ possible.
Since
$$\widehat{\chi}_{q,L}(\xi) = \frac{q^n}{L^n} \sum_{m \in [-\frac{L}{2},\frac{L}{2})^n,\ q|m} e^{-2\pi i m\cdot\xi}$$

it is easy to see that $\widehat{\chi}_{q,L}(l/q) = 1$ for all $l \in \mathbb{Z}^n$ and that there exists some absolute constant $C > 0$ such that
$$0 \leq 1 - \widehat{\chi}_{q,L}(\xi)^2 \leq C\, L\, |\xi - l/q| \tag{3.5.10}$$
for all $\xi \in \mathbb{T}^n$ and $l \in \mathbb{Z}^n$. It is then easy to see using our assumption that $qq_1(\varepsilon)|q'$ that
$$0 \leq H_{\lambda,q}(\xi)(1 - \widehat{\chi}_{q',L}(\xi)^2) \leq C\varepsilon \tag{3.5.11}$$
for some constant $C > 0$ uniformly in $\xi \in \mathbb{T}^n$ provided $L \ll \varepsilon^5\lambda$. Substituting inequality (3.5.7) into (3.5.8), we obtain

$$\begin{aligned}|\mathcal{N}^1_{\lambda\Delta^0,q,Q}(f_1,\dots,f_{n_1})|^2 &\leq |Q|^{-1} \left(\int |\hat{f}_{n_1}(\xi)|^2 H_\lambda(\xi) \widehat{\chi}_{q',L}(\xi)^2\, d\xi \,+ \int |\hat{f}_{n_1}(\xi)|^2 H_\lambda(\xi)(1 - \widehat{\chi}_{q',L}(\xi)^2)\, d\xi \right) \\ &\leq \|f_{n_1}\|^2_{U^1_{q',L}(Q)} + O(\varepsilon)\end{aligned}$$

provided $L \ll \varepsilon^5\lambda$. This proves Lemma 3.8 for $k \geq 3$, as it is clear that by re-indexing the above estimate holds for any of the functions f_i in place of f_{n_1}. For $n_1 = 2$ an easy modification of arguments in [78], specifically the proof of Lemma 3 therein, establishes that

$$|\mathcal{N}^1_{\lambda\Delta^0,q,Q}(f_1,f_2)|^2 \leq \|f_i\|^2_{U^1_{q',L}(Q)} + O(\varepsilon)$$

for $i = 1, 2$ provided $L \ll \varepsilon^5\lambda$. □

Proof of Lemma 3.9. Let $q, L \in \mathbb{N}$ such that $L|N$, $q|L$. The "modulo q" grids $Q_t(q,L) = t + Q(q,L)$ partition the cube Q with t running through the set $\Gamma_{q,L,Q} = \{1,\dots,q\}^n + \Gamma_{L,Q}$, where $\Gamma_{L,Q}$ denote the centers of the "integer" grids $t + Q(L)$ in an initial partition of Q. Let q', L' be positive integers so that $q|q'$, $L'|L$ and $L' \ll \varepsilon^2 L$. If $s \in \Gamma_{q',L',Q}$ and $t \in Q_s(q',L')$ then $|t - s| = O(L')$ and hence

$$\mathbb{E}_{x\in Q_t(q,L)}\, g(x) = \mathbb{E}_{x\in Q_s(q,L)}\, g(x) + O(L'/L)$$

for any function $g : Q \to [-1,1]$. Moreover, since the cube $Q_s(q,L)$ is partitioned into the smaller cubes $Q_t(q',L')$, we have by Cauchy-Schwarz

$$|\mathbb{E}_{x\in Q_s(q,L)}\, g(x)|^2 \leq \mathbb{E}_{t\in\Gamma_{q',L',Q_s(q,L)}} |\mathbb{E}_{x\in Q_t(q',L')} g(x)|^2.$$

From this it is easy to see that

$$\|g\|^2_{U^1_{q,L}(Q)} = \mathbb{E}_{t\in Q}|\mathbb{E}_{x\in Q_t(q,L)}g(x)|^2 \leq \mathbb{E}_{t\in\Gamma_{q',L',Q}}\ |\mathbb{E}_{x\in Q_t(q',L')}g(x)|^2 + O(L'/L)$$

and we note that the right side of the above expression is $\|\mathbb{E}(g|\mathcal{G}_{q',L',Q})\|^2_{L^2(Q)}$ since the conditional expectation function $\mathbb{E}(g|\mathcal{G}_{q',L',Q})$ is constant and equal to $\mathbb{E}_{x\in Q_t(q',L')}g(x)$ on the cubes $Q_t(q',L')$.
Now suppose (3.5.7) does not hold for some $j\geq 1$, that is

$$\|f-\mathbb{E}(f|\mathcal{G}_{q_j,L_j,Q})\|^2_{U^1_{q_{j+1},L_{j+1}}(Q)} \geq \varepsilon^2.$$

Since $L_{j+2} \ll \varepsilon^2 L_{j+1}$, $L_{j+2}|L_j$, and $q_{j+1}|q_{j+2}$ we can apply the above observations to $g := f - \mathbb{E}(f|\mathcal{G}_{q_j,L_j,Q})$ and obtain, by orthogonality, that

$$\|\mathbb{E}(f|\mathcal{G}_{q_{j+2},L_{j+2},Q})\|^2_{L^2(Q)} \geq \|\mathbb{E}(f|\mathcal{G}_{q_j,L_j,Q})\|^2_{L^2(Q)} + c\varepsilon^2 \tag{3.5.12}$$

for some constant $c>0$. Since the above expressions are clearly bounded by 1, the above procedure must stop in $O(\varepsilon^{-2})$ steps at which (3.5.7) must hold for some $1\leq j\leq \bar{J}_1(\varepsilon)$ with $\bar{J}_1(\varepsilon)=O(\varepsilon^{-2})$. $\square$

3.5.2 The base case of our general inductive strategy

Let $Q=Q_1\times\ldots\times Q_d$ with $Q_i\subseteq\mathbb{Z}^{2n_i+3}$ be cubes of equal side length $l(Q)$ and $\Delta_i^0\subseteq\mathbb{Z}^{2n_i+3}$ be a non-degenerate simplex of n_i points for $1\leq i\leq d$.
We note that for any $q_0\in\mathbb{N}$ and scale L_0 dividing $l(Q)$ if $\underline{t}=(t_1,\ldots,t_d)\in\Gamma_{q_0,L_0,Q}$, then the corresponding grids $Q_{\underline{t}}(q_0,L_0)$ in the partition of Q take the form $Q_{\underline{t}}(q_0,L_0)=Q_{t_1}(q_0,L_0)\times\cdots\times Q_{t_d}(q_0,L_0)$.
As in the continuous setting we will ultimately need a parametric version of Proposition 3.10, namely Proposition 3.11 below.

Proposition 3.11 (Parametric Counting Lemma on $\mathbb{Z}^n$ for Simplices)**.** *Let $0<\varepsilon\leq 1$ and $R\geq 1$. There exists an integer $J_1=J_1(\varepsilon,R)=O(R\,\varepsilon^{-4})$ such that for any (ε,q_{J_1})-admissible sequence of scales $L_0\geq L_1\geq\cdots\geq L_{J_1}$ with L_0 dividing $l(Q)$ and $q_j:=q_0q_1(\varepsilon)^j$ for $0\leq j\leq J_1$ with $q_0\in\mathbb{N}$, and collection of functions*

$$f^{i,r}_{k,\underline{t}}:\ Q_{t_i}(q_0,L_0)\to[-1,1]\quad\textit{with}\ \ 1\leq i\leq d,\ 1\leq k\leq n_i,\ 1\leq r\leq R\ \textit{and}\ \underline{t}\in\Gamma_{q_0,L_0,Q}$$

there exists $1\leq j<J_1$ and a set $T_\varepsilon\subseteq\Gamma_{q_0,L_0,Q}$ of size $|T_\varepsilon|\leq\varepsilon|\Gamma_{q_0,L_0,Q}|$ such that

$$\mathcal{N}^1_{\lambda\Delta_i^0,q_j,Q_{t_i}(q_0,L_0)}(f^{i,r}_{1,\underline{t}},\ldots,f^{i,r}_{n_i,\underline{t}}) = \mathcal{M}^1_{\lambda,q_j,Q_{t_i}(q_0,L_0)}(f^{i,r}_{1,\underline{t}},\ldots,f^{i,r}_{n_i,\underline{t}}) + O(\varepsilon) \tag{3.5.13}$$

for all $\lambda\in q_j\sqrt{\mathbb{N}}$ with $L_{j+1}\leq\lambda\leq L_j$ and $\underline{t}\notin T_\varepsilon$ uniformly in $1\leq i\leq d$ and $1\leq r\leq R$.

This proposition follows, as the analogous result did in the continuous setting, from Lemma 3.8 and the follow parametric version of Lemma 3.9.

Lemma 3.10 (A simultaneous Koopman-von Neumann type decomposition)**.**
Let $0<\varepsilon\ll 1$, $m\geq 1$, and $Q\subseteq\mathbb{Z}^n$ be a cube. There exists an integer $\bar{J}_1=O(m\varepsilon^{-3})$ such that for any $(\varepsilon,q_{\bar{J}_1})$-admissible sequence $L_0\geq L_1\geq\cdots\geq L_{\bar{J}_1}$ with L_0 dividing $l(Q)$ and $q_j:=q_0q_1(\varepsilon)^j$ for $0\leq j\leq\bar{J}_1$ with $q_0\in\mathbb{N}$, and collection of functions

$$f_{1,t},\ldots f_{m,t}:Q_t(q_0,L_0)\to[-1,1]$$

defined for each $t \in \Gamma_{q_0,L_0,Q}$, there is some $1 \leq j < \bar{J}_1$ and a set $T_\varepsilon \subseteq \Gamma_{q_0,L_0,Q}$ of size $|T_\varepsilon| \leq \varepsilon|\Gamma_{q_0,L_0,Q}|$ such that

$$\|f_{i,t} - \mathbb{E}(f_{i,t}|\mathcal{G}_{q_j,L_j,Q_t(q_0,L_0)})\|_{U^1_{q_{j+1},L_{j+1}}(Q_t(q_0,L_0))} \leq \varepsilon \tag{3.5.14}$$

for all $1 \leq i \leq m$ and $t \notin T_\varepsilon$.

Lemma 3.10 above is of course the discrete analogue of Lemma 3.4. Since the proofs of Proposition 3.11 and Lemma 3.10 are almost identical to the arguments presented in Section 3.3.2 we choose to omit these details.

3.6 Proof of Theorem 3.4: products of simplices in $\mathbb{Z}^d$

After the preparations in Section 3.5 we can proceed very similarly as in Section 3.4 to prove our main result in the discrete case, namely Theorem 3.4. The main difference will be that given $0 < \varepsilon \ll 1$ and $1 \leq k \leq d$, we construct a positive integer $q_k(\varepsilon)$ and assume that all our sequences of scales will be $(\varepsilon, q_k(\varepsilon))$-admissible. The cubes $Q_{\underline{t}}(L)$ will be naturally now be replaced by the grids $Q_{\underline{t}}(q, L)$ of the form that already appear in Section 3.5 where we always assume $q|L$.
Let $\Delta^0 = \Delta^0_1 \times \ldots \times \Delta^0_d$ with each $\Delta^0_i \subseteq \mathbb{Z}^{2n_i+3}$ a non-degenerate simplex of n_i points for $1 \leq i \leq d$ and $Q = Q_1 \times \ldots \times Q_d \subseteq \mathbb{Z}^n$ with $Q_i \subseteq \mathbb{Z}^{2n_i+3}$ cubes of equal side length $l(Q)$ (taken much larger than the diameter of Δ^0). We will use the same parameterizations in terms of hypergraph bundles $\mathcal{H}^{\underline{n}}_{d,k}$ and corresponding notations as in Section 3.4 to count the configurations $\Delta = \Delta_1 \times \ldots \times \Delta_d \subseteq Q$ with each $\Delta_i \subseteq Q_i$ an isometric copy of $\lambda\Delta^0_i$ for some $\lambda \in \sqrt{\mathbb{N}}$.
Given any positive integer q and $\lambda \in q\sqrt{\mathbb{N}}$ we will make use of the notation

$$\sum_{\underline{x}_i} f(\underline{x}_i)\, \sigma^i_{\lambda,q}(\underline{x}_i) := \mathbb{E}_{x_{i1} \in Q_i} \sum_{x_{i2},\ldots,x_{in_i}} f(\underline{x}_i)\, \sigma_{\lambda\Delta^0_i,q}(x_{i2} - x_{i1}, \ldots, x_{in_i} - x_{i1})\, dx_{i1} \tag{3.6.1}$$

with $\sigma_{\lambda\Delta^0_i,q}$ as defined in the previous section and $\underline{x}_i = (x_{i1}, \ldots, x_{in_i}) \in Q_i^{n_i}$.
Note that if $S \subseteq Q$ then the density of configurations Δ in S, of the form $\Delta = \Delta_1 \times \ldots \times \Delta_d$ with each $\Delta_i \subseteq Q_i$ an isometric copy of $\lambda\Delta^0_i$ for some $\lambda \in q\sqrt{\mathbb{N}}$ is given by the expression

$$\mathcal{N}^d_{\lambda\Delta^0,q,Q}(1_S\, ;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) := \sum_{\underline{x}_1} \cdots \sum_{\underline{x}_d} \prod_{e \in \mathcal{H}^{\underline{n}}_{d,d}} 1_S(\underline{x}_e)\ \sigma^1_{\lambda,q}(\underline{x}_1) \ldots \sigma^d_{\lambda,q}(\underline{x}_d). \tag{3.6.2}$$

More generally, for any given $1 \leq k \leq d$ and a family of functions $f_e : Q_{\pi(e)} \to [-1, 1]$ with $e \in \mathcal{H}^{\underline{n}}_{d,k}$ we define the multi-linear expression

$$\mathcal{N}^d_{\lambda\Delta^0,q,Q}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k}) := \sum_{\underline{x}_1} \cdots \sum_{\underline{x}_d} \prod_{e \in \mathcal{H}^{\underline{n}}_{d,k}} f_e(\underline{x}_e)\ \sigma^1_{\lambda,q}(\underline{x}_1) \ldots . \sigma^d_{\lambda,q}(\underline{x}_d). \tag{3.6.3}$$

as well as

$$\mathcal{M}^d_{\lambda,q,Q}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k}) := \mathbb{E}_{\underline{t} \in Q}\ \mathcal{M}^d_{\underline{t}+Q(q,L)}\,(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k}) \tag{3.6.4}$$

where $Q(q, L) = Q_1(q, L) \times \cdots \times Q_d(q, L)$ with each $Q_i(q, L) = (q\mathbb{Z} \cap [-\frac{L}{2}, \frac{L}{2}])^{2n_i+3}$ and

$$\mathcal{M}^d_{\widetilde{Q}}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k}) := \mathbb{E}_{\underline{x}_1 \in \widetilde{Q}_1^{n_1}} \cdots\ \mathbb{E}_{\underline{x}_d \in \widetilde{Q}_d^{n_d}} \prod_{e \in \mathcal{H}^{\underline{n}}_{d,k}} f_e(\underline{x}_e) \tag{3.6.5}$$

for any cube $\widetilde{Q} \subseteq Q$ of the form $\widetilde{Q} = \widetilde{Q}_1 \times \cdots \times \widetilde{Q}_d$ with $\widetilde{Q}_i \subseteq Q_i$ for $1 \leq i \leq d$.

We note that it is easy to show, as in the continuous, that if $S \subseteq Q$ with $|S| \geq \delta|Q|$ for some $\delta > 0$ then
$$\mathcal{M}^d_{\lambda,q,Q}(1_S; e \in \mathcal{H}^{\underline{n}}_{d,d}) \geq \delta^{n_1 \cdots n_d} - O(\varepsilon) \tag{3.6.6}$$
for all scales $\lambda \in q\sqrt{\mathbb{N}}$ with $0 < \lambda \ll \varepsilon\, l(Q)$. In light of this observation and the discussion preceding Proposition 3.10 the proof of Theorem 3.4 reduces, as it did in the continuous setting, to the following

Proposition 3.12. *Let $0 < \varepsilon \ll 1$. There exist positive integers $J_d = J_d(\varepsilon)$ and $q_d(\varepsilon)$ such that for any $(\varepsilon, q_d(\varepsilon)^{J_d})$-admissible sequence of scales $l(Q) \geq L_1 \geq \cdots \geq L_{J_1}$ and $S \subseteq Q$ there is some $1 \leq j < J_d$ such that*
$$\mathcal{N}^d_{\lambda\Delta^0,q_j,Q}(1_S\,;\, e \in \mathcal{H}^{\underline{n}}_{d,d}) = \mathcal{M}^d_{\lambda,q_j,Q}(1_S; e \in \mathcal{H}^{\underline{n}}_{d,d}) + O(\varepsilon), \tag{3.6.7}$$
for all $\lambda \in q_j\sqrt{\mathbb{N}}$ with $L_{j+1} \leq \lambda \leq L_j$ with $q_j := q_d(\varepsilon)^j$.

Quantitative Remark. A careful analysis of our proof reveals that there exist choices of $J_d(\varepsilon)$ and $q_d(\varepsilon)$ which are less than $W_d(\log(C_\Delta\varepsilon^{-3}))$ and $W_d(C_\Delta\varepsilon^{-13})$ respectively where $W_k(m)$ is again the tower-exponential function defined by $W_1(m) = \exp(m)$ and $W_{k+1}(m) = \exp(W_k(m))$ for $k \geq 1$.
The proof of Proposition 3.12 follows along the same lines as the analogous result in the continuous setting. As before we will compare the averages $\mathcal{N}^d_{\lambda\Delta^0,q,Q}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k})$ to those of $\mathcal{M}^d_{\lambda,q,Q}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k})$, at certain scales q and $\lambda \in q\sqrt{\mathbb{N}}$ with with $L_{j+1} \leq \lambda \leq L_j$, inductively for $1 \leq k \leq d$. As the arguments closely follow those given in Section 3.4 we will be brief and emphasize mainly just the additional features.

3.6.1 Reduction of Proposition 3.12 to a more general "local" counting lemma

For any given $1 \leq k \leq d$ and a family of functions $f_e : Q_{\pi(e)} \to [-1, 1]$ with $e \in \mathcal{H}^{\underline{n}}_{d,k}$ it is easy to see that for any $\varepsilon > 0$, scale $L_0 > 0$ dividing the side-length $l(Q)$, and $q_0|q$ we have
$$\mathcal{N}^d_{\lambda\Delta^0,q,Q}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathbb{E}_{\underline{t}\in\Gamma_{q_0,L_0,Q}}\, \mathcal{N}^d_{\lambda\Delta^0,q,Q_{\underline{t}}(q_0,L_0)}(f_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon) \tag{3.6.8}$$
and
$$\mathcal{M}^d_{\lambda,q,Q}(f_e; e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathbb{E}_{\underline{t}\in\Gamma_{L,Q}}\, \mathcal{M}^d_{\lambda,q,Q_{\underline{t}}(q_0,L_0)}(f_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon) \tag{3.6.9}$$
provided $0 < \lambda \ll \varepsilon L_0$ where $f_{e,\underline{t}}$ denotes the restriction of a function f_e to the cube $Q_{\underline{t}}(q_0, L_0)$.
Thus the proof of Proposition 3.12 reduces to showing that the expressions in (3.6.8) and (3.6.9) only differ by $O(\varepsilon)$ for all scales $\lambda \in q\sqrt{\mathbb{N}}$ with $L_{j+1} \leq \lambda \leq L_j$, given an (ε, q)-admissible sequence $L_0 \geq L_1 \geq \cdots \geq L_J$, for any collection of bounded functions $f_{e,\underline{t}}$, $e \in \mathcal{H}^{\underline{n}}_{d,k}$, $\underline{t} \in \Gamma_{q_0,L_0,Q}$. Indeed, our crucial result will be the following

Proposition 3.13 (Local Counting Lemma in $\mathbb{Z}^n$)**.** *Let $0 < \varepsilon \ll 1$ and $q_0, M \in \mathbb{N}$.*
There exist positive integers $J_k = J_k(\varepsilon, M)$ and $q_k(\varepsilon)$ such that for any (ε, q_{J_d})-admissible sequence of scales $L_0 \geq L_1 \geq \cdots \geq L_{J_1}$ with L_0 dividing $l(Q)$ and $q_j := q_0\, q_k(\varepsilon)^j$ for $j \geq 1$, and collection of functions
$$f^m_{e,\underline{t}} : Q_{t_{\pi(e)}}(q_0, L_0) :\to [-1, 1] \quad \text{with} \quad e \in \mathcal{H}^{\underline{n}}_{d,k},\ 1 \leq m \leq M \ \text{and}\ \underline{t} \in \Gamma_{q_0,L_0,Q}$$
there exists $1 \leq j < J_k$ and a set $T_\varepsilon \subseteq \Gamma_{q_0,L_0,Q}$ of size $|T_\varepsilon| \leq \varepsilon|\Gamma_{q_0,L_0,Q}|$ such that
$$\mathcal{N}^d_{\lambda\Delta^0,q_j,Q_{\underline{t}}(q_0,L_0)}(f_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) = \mathcal{M}^d_{\lambda,q_j,Q_{\underline{t}}(q_0,L_0)}(f_{e,\underline{t}}; e \in \mathcal{H}^{\underline{n}}_{d,k}) + O(\varepsilon) \tag{3.6.10}$$
for all $\lambda \in q_j\sqrt{\mathbb{N}}$ with $L_{j+1} \leq \lambda \leq L_j$ and $\underline{t} \notin T_\varepsilon$ uniformly in $e \in \mathcal{H}^{\underline{n}}_{d,k}$ and $1 \leq m \leq M$.

Note that if $k=d$, $L_0=l(Q)$, $q_0=M=1$, then $|\Gamma_{q_0,L_0,Q}|=1$, and moreover if $f_{e,\underline{t}}=1_S$ for all $e\in\mathcal{H}^{\underline{n}}_{d,k}$ for a set $S\subseteq Q$, then Proposition 3.13 reduces to precisely Proposition 3.12. In fact, Proposition 3.13 is a parametric, multi-linear and simultaneous extension of Proposition 3.12 which we need in the induction step, i.e. when going from level $k-1$ to level k.

3.6.2 Proof of Proposition 3.13

We will prove Proposition 3.13 by induction on $1\leq k\leq d$.
For $k=1$ this is basically Proposition 3.11, exactly as it was in the base case of the proof of Proposition 3.9.
For the induction step we will again need two main ingredients. The first establishes that the our multi-linear forms $\mathcal{N}^d_{\lambda\Delta^0,q,Q}(f_e;e\in\mathcal{H}^{\underline{n}}_{d,k})$ are controlled by a box-type norm attached to scales q' and L.
Let $Q=Q_1\times\ldots\times Q_d$ with $Q_i\subseteq\mathbb{Z}^{2n_i+3}$ be cubes of equal side length $l(Q)$ and $1\leq k\leq d$. For any scale $0<L\ll l(Q)$ and function $f:Q_{e'}\to[-1,1]$ with $e'\in\mathcal{H}_{d,k}$ we define its local box norm at scales q' and L by

$$\|f\|^{2^k}_{\square_{q',L}(Q_{e'})}:=\mathbb{E}_{\underline{s}\in Q_{e'}}\|f\|^{2^k}_{\square(Q_{\underline{s}}(q',L))} \tag{3.6.11}$$

where

$$\|f\|^{2^k}_{\square(\widetilde{Q})}:=\mathbb{E}_{x_{11},x_{12}\in\widetilde{Q}_1}\cdots\ \mathbb{E}_{x_{k1},x_{k2}\in\widetilde{Q}_k}\prod_{(\ell_1,\ldots,\ell_k)\in\{1,2\}^k}f(x_{1\ell_1},\ldots,x_{k\ell_k}) \tag{3.6.12}$$

for any cube $\widetilde{Q}$ of the form $\widetilde{Q}=\widetilde{Q}_1\times\cdots\times\widetilde{Q}_k$. We note that (3.6.4) and (3.6.5) are special cases of (3.6.11) and (3.6.12) with $k=d$, $\underline{n}=(2,\ldots,2)$, and $f_e=f$ for all $e\in\mathcal{H}^{\underline{n}}_{d,d}$.

Lemma 3.11 (A Generalized von-Neumann inequality on $\mathbb{Z}^n$)**.** *Let $1\leq k\leq d$.*
Let $0<\varepsilon\ll 1$, $q,q'\in\mathbb{N}$ with $qq_1(\varepsilon)|q'$, and $\lambda\in q\sqrt{\mathbb{N}}$ with $\lambda\ll l(Q)$ and $1\ll L\ll(\varepsilon^{2^k})^{10}\lambda$. For any collection of functions $f_e:Q_{\pi(e)}\to[-1,1]$ with $e\in\mathcal{H}^{\underline{n}}_{d,k}$ we have both

$$|\mathcal{N}^d_{\lambda\Delta^0,q,Q}(f_e;e\in\mathcal{H}^{\underline{n}}_{d,k})|\leq\min_{e\in\mathcal{H}^{\underline{n}}_{d,k}}\|f_e\|_{\square_{q',L'}(Q_{\pi(e)})}+O(\varepsilon) \tag{3.6.13}$$

and

$$|\mathcal{M}^d_{\lambda,q,Q}(f_e;e\in\mathcal{H}^{\underline{n}}_{d,k})|\leq\min_{e\in\mathcal{H}^{\underline{n}}_{d,k}}\|f_e\|_{\square_{q',L'}(Q_{\pi(e)})}. \tag{3.6.14}$$

The proof of inequalities (3.6.13) and (3.6.14) follow exactly as in the continuous case, see Lemma 3.6, using Lemma 3.8 in place of Lemma 3.3. We omit the details.

The crucial ingredient is again a parametric weak hypergraph regularity lemma, i.e. Lemma 3.7 adapted to the discrete settings. The proof is essentially the same as in the continuous case, with exception that the $\square_{L_j}$-norms are replaced by $\square_{q_j,L_j}$-norms where $q_j=q_0q^j$ is a given sequence of positive integers and $L_0\geq L_1\geq\cdots\geq L_J$ is an (ε,q_J)-admissible sequence of scales. To state it we say that a σ-algebra $\mathcal{B}$ on a cube Q is of *scale* (q,L) if it is refinement of the grid $\mathcal{G}_{q,L,Q}$, i.e. if its atoms partition each cube $Q_{\underline{t}}(q,L)$ of the grid. We will always assume that $q|L$ and $L|l(Q)$. Recall also that we say the complexity of a σ-algebra $\mathcal{B}$ is at most m, and write complex$(\mathcal{B})\leq m$, if it is generated by m sets.

Lemma 3.12 (Parametric weak hypergraph regularity lemma for $\mathbb{Z}^n$)**.**

Let $0 < \varepsilon \ll 1$, $1 \le k \le d$, $q_0, q, L_0, M \in \mathbb{N}$, and let $q_j := q_0 q^j$ for $j \ge 1$. There exists $\bar{J}_k = O(M\varepsilon^{-2^{k+3}})$ such that for any $(\varepsilon^{2^k}, q_{\bar{J}_k})$-admissible sequence $L_0 \ge L_1 \ge \cdots \ge L_{\bar{J}_k}$ with the property that L_0 divides $l(Q)$ and collection of functions

$$f^m_{e,\underline{t}} : Q_{\underline{t}_{\pi(e)}}(q_0, L_0) \to [-1,1] \quad \textit{with} \quad e \in \mathcal{H}^{\underline{n}}_{d,k},\ 1 \le m \le M,\ \textit{and}\ \underline{t} \in \Gamma_{q_0,L_0,Q}$$

there is some $1 \le j < \bar{J}_k$ and σ-algebras $\mathcal{B}_{e',\underline{t}}$ of scale (q_j, L_j) on $Q_{\underline{t}_{e'}}(q_0, L_0)$ for each $\underline{t} \in \Gamma_{q_0,L_0,Q}$ and $e' \in \mathcal{H}_{d,k}$ such that

$$\|f^m_{e,\underline{t}} - \mathbb{E}(f^m_{e,\underline{t}}|\mathcal{B}_{\pi(e),\underline{t}})\|_{\square_{q_{j+1},L_{j+1}}(Q_{\underline{t}_{\pi(e)}}(L_0))} \le \varepsilon \tag{3.6.15}$$

uniformly for all $t \notin T_\varepsilon$, $e \in \mathcal{H}^{\underline{n}}_{d,k}$, and $1 \le m \le M$, where $T_\varepsilon \subseteq \Gamma_{q_0,L_0,Q}$ with $|T_\varepsilon| \le \varepsilon|\Gamma_{q_0,L_0,Q}|$. Moreover, the σ-algebras $\mathcal{B}_{e',\underline{t}}$ have the additional local structure that the exist σ-algebras $\mathcal{B}_{e',\mathfrak{f}',\underline{s}}$ on $Q_{\underline{s}_{\mathfrak{f}'}}(q_j, L_j)$ with complex$(\mathcal{B}_{e',\mathfrak{f}',\underline{s}}) = O(j)$ *for each $\underline{s} \in \Gamma_{q_j,L_j,Q}$, $e' \in \mathcal{H}_{d,k}$, and $\mathfrak{f}' \in \partial e'$ such that if $\underline{s} \in Q_{\underline{t}}(q_0, L_0)$, then*

$$\mathcal{B}_{e',\underline{t}}\big|_{Q_{\underline{s}_{e'}}(q_j,L_j)} = \bigvee_{\mathfrak{f}' \in \partial e'} \mathcal{B}_{e',\mathfrak{f}',\underline{s}}. \tag{3.6.16}$$

The proof of Lemma 3.12 follows exactly as the corresponding proof of Lemma 3.7 in the continuous setting, so we will omit the details. We will however provide some details of how one deduces Proposition 3.13, from Lemmas 3.11 and 3.12. The arguments are again very similar to those in the continuous setting, however one needs to make a careful choice of the integers $q_k(\varepsilon)$, appearing in the statement of the Proposition.

Proof of Proposition 3.13. Let $2 \le k \le d$ and assume that the lemma holds for $k-1$.
Let $0 < \varepsilon \ll 1$ and $\varepsilon_1 := \exp(-C_1 \varepsilon^{-2^{k+3}})$ for some large constant $C_1 = C_1(n,k,d) \gg 1$.
We then define $q_k(\varepsilon) := q_{k-1}(\varepsilon_1)$ recalling that $q_1(\varepsilon) := \ lcm\{1 \le q \le C\varepsilon^{-10}\}$ and note that it is easy to see by induction that $q_k(\varepsilon)|q_k(\varepsilon')$ for $0 < \varepsilon' \le \varepsilon$ and $q_{k-1}(\varepsilon)|q_k(\varepsilon)$. We further define the function $F(\varepsilon) := J_{k-1}(\varepsilon_1, M)$ with $M = \varepsilon\,\varepsilon_1^{-1}$ and recall that $q_j := q_0\, q_k(\varepsilon)^j$ for $j \ge 1$.
We now proceed exactly as in the proof of Proposition 3.9 but with $\{L_j\}_{j\ge1}$ being a $(\varepsilon_1, q_{\widetilde{J}})$-admissible sequence of scales, with $\widetilde{J} \gg F(\varepsilon)\,\bar{J}_k(\varepsilon, M)$. We again choose a subsequence $\{L'_j\} \subseteq \{L_j\}$ so that $L'_0 = L_0$ and $\text{index}(L'_{j+1}) \ge \text{index}(L'_j) + F(\varepsilon) + 2$, but also now set $q'_j = q_{j'}$, where $j' := \text{index}(L'_j)$.
Lemma 3.12 then guarantees the existence of σ-algebras $\mathcal{B}_{e',\underline{t}}$ of scale (q'_j, L'_j) on $Q_{\underline{t}_{e'}}(q_0, L_0)$ for each $\underline{t} \in \Gamma_{q_0,L_0,Q}$ and $e' \in \mathcal{H}_{d,k}$, with the local structure described above, such that (3.6.15) holds uniformly for all $t \notin T'_\varepsilon$, $e \in \mathcal{H}^{\underline{n}}_{d,k}$, and $1 \le m \le M$, for some $1 \le j < \bar{J}_k(\varepsilon, M) = O(M\varepsilon^{-2^{k+3}})$, where $T'_\varepsilon \subseteq \Gamma_{q_0,L_0,Q}$ with $|T'_\varepsilon| \le \varepsilon|\Gamma_{q_0,L_0,Q}|$.
Arguing as in the proof of Proposition 3.9 we can conclude from this that for each $j' \le l < J'$ we have

$$\mathcal{N}^d_{\lambda\Delta^0,q_l,Q_{\underline{s}}(q'_j,L'_j)}(f^m_{e,\underline{s}};\, e \in \mathcal{H}^{\underline{n}}_{d,k}) = \sum_{\underline{r}} \alpha_{\underline{s},\underline{r},m}\, \mathcal{N}^d_{\lambda\Delta^0,q_l,Q_{\underline{s}}(q'_j,L'_j)}(g^{\underline{r}}_{\mathfrak{f},\underline{s}};\, \mathfrak{f} \in \mathcal{H}^{\underline{n}}_{d,k-1}) + O(\varepsilon) \tag{3.6.17}$$

and

$$\mathcal{M}^d_{\lambda,q_l,Q_{\underline{s}}(q'_j,L'_j)}(f^m_{e,\underline{s}};\, e \in \mathcal{H}^{\underline{n}}_{d,k}) = \sum_{\underline{r}} \alpha_{\underline{r},\underline{s},m}\, \mathcal{M}^d_{\lambda,q_l,Q_{\underline{s}}(q'_j,L'_j)}(g^{\underline{r}}_{\mathfrak{f},\underline{s}};\, \mathfrak{f} \in \mathcal{H}^{\underline{n}}_{d,k-1}) + O(\varepsilon) \tag{3.6.18}$$

provided $(\varepsilon^{-2^k})^{10} L'_{j+1} \ll \lambda$ with $\lambda \in q_l\sqrt{\mathbb{N}}$, where each $|\alpha_{\underline{s},r_e}| \le 1$ and number of index vectors $\underline{r} = (r_e)_{e\in\mathcal{H}^{\underline{n}}_{d,k}}$ is R^D with $D := |\mathcal{H}^{\underline{n}}_{d,k}|$ and hence $R^D \le M$ if $C_1 \gg 1$.

By induction, we apply Proposition 3.13 to the sequence of scales $L'_j = L_{j'} \geq L_{j'+1} \geq \cdots \geq L_{J'} = L'_{j+1}$ with $\varepsilon_1 > 0$ and for $q_l := q'_j\, q_k(\varepsilon)^{l-j'} = q_{j'}\, q_{k-1}(\varepsilon_1)^{l-j'}$ where $j' \leq l \leq J'$ with respect to the family of functions $g^r_{\underline{s},\mathfrak{f}} : Q_{\underline{s}_\mathfrak{f}}(q'_j, L'_j) \to [-1,1]$. This is possible as $J' - j' \gg J_{k-1}(\varepsilon_1, R^D)$ and our sequence of scales is $(\varepsilon_1, q_{J'})$-admissible. Thus there exists an index $j' \leq l < J'$ such that for all $\lambda \in q_l\sqrt{\mathbb{N}}$ with $L_{l+1} \leq \lambda \leq L_l$ we have

$$\mathcal{N}^d_{\lambda\Delta^0, q_l, Q_{\underline{s}}(q'_j, L'_j)}\,(g^r_{\mathfrak{f},\underline{s}};\, \mathfrak{f} \in \mathcal{H}^n_{d,k-1}) = \mathcal{M}^d_{\lambda, q_l, Q_{\underline{s}}(q'_j, L'_j)}\,(g^r_{\mathfrak{f},\underline{s}};\, \mathfrak{f} \in \mathcal{H}^n_{d,k-1}) + O(\varepsilon_1) \tag{3.6.19}$$

uniformly in $\underline{r}$ for $\underline{s} \notin S_{\varepsilon_1}$, where $S_{\varepsilon_1} \subseteq \Gamma_{q'_j, L'_j, Q}$ is a set of size $|S_{\varepsilon_1}| \leq \varepsilon_1 |\Gamma_{q'_j, L'_j, Q}|$.
The remainder of the proof follows as just as it did for Proposition 3.9. $\square$

As explained above, this implies Proposition 3.12 and that finishes the proof of our main result, namely Theorem 3.4.

4 Diophantine equations in the primes

Let $\mathfrak{p} = (\mathfrak{p}_1, ..., \mathfrak{p}_r)$ be a system of r_i polynomials with integer coefficients of degree exactly d in the n variables $\mathbf{x} = (x_1, ..., x_n)$. Our primary concern is finding solutions with each coordinate prime, which we call prime points, to the system of equations $\mathfrak{p}(\mathbf{x}) = \mathbf{s}$. Here $\mathbf{s} \in \mathbb{Z}^r$ is a fixed element. The notation $V_{\mathfrak{p},\mathbf{s}}$ is used to denote the complex affine variety defined by this equation. Some properties of the system only depend on the highest part of the polynomials.To each system of polynomials is attached a system of homogenous integral forms $\mathfrak{f}$, which is comprised of the largest degree homogeneous parts of each of the polynomials forming $\mathfrak{p}$.

If $\mathfrak{p}$ is composed entirely of linear forms the known results may be split into a classical regime and a modern one. With $\mathfrak{p}$ is a system of r linear forms, define the rank of $\mathfrak{p}$ to be the minimum number of nonzero coefficients in a non-trivial linear combination

$$\lambda_1\mathfrak{p}_1 + ... + \lambda_r\mathfrak{p}_r,$$

and denote this quantity by $\mathcal{B}_1(\mathfrak{p})$. The classical results on the large scale distribution of prime points on $V_{\mathfrak{p},\mathbf{s}}$ are conditional on the rank being sufficiently large in terms of r (for example, $2r + 1$ follows from what is shown here). In this realm are many well known results such as the ones due Vinogradov [111] and more recently Balog [4]. The modern results are mostly summed up in the work of Green and Tao [52], where the large scale distribution of prime points on $V_{\mathfrak{p},\mathbf{s}}$ is determined only on the condition that $\mathcal{B}_1(\mathfrak{p})$ is at least 3, a quantity independent of r_1. These results cover all scenarios that do not reduce to a binary problem. However, extending the already stunning results of Goldston, Pintz, and Yildirim [49], it has been shown by Maynard and Tao [86] that one of the equations $x_1 - x_2 = 2i$, $i = 1, ..., 123$ does have infinitely many prime solutions [94].

The scenario for systems involving higher degree forms in certainly less clean cut, and even the study of integral points on $V_{\mathfrak{p},\mathbf{s}}$ is a non-trivial problem. General results for the large scale distribution of integral points are provided by Birch [11] and Schmidt [97], which again require the system to be large with respect to certain notions of rank (which is again with respect to the number of forms, but also with respect to the degrees of the forms involved). Working within the limitations of these results, one should expect to be able to understand the large scale distribution of prime points as well. For systems of forms which are additive, for instance the single form $a_1x_1^d + ... + a_nx_n^d$, this is something that has been done and the primary result here is due to Hua [57]. On the opposite end, if the system of forms is a bilinear system, or even contains a large bilinear piece, one can also provide similar results, a particular instance of which is given by Liu [74] for a quadratic form.

However, the main results given in this sections constitute the first instance where it was shown that asymptotic formulas may be obtained for the number of prime solutions to general systems of diophantine equations, provided the rank of the system is sufficiently large with respect to the number and degree of the equations. Very recently the quantitative aspects of our results have been improved bringing the rank condition to among the primes to be similar that those of for the integer solutions, see [115, 75].

4.1 Main results

For a fixed system of polynomials $\mathfrak{p}$, let us define for each prime p the quantity

$$\mu_p = \lim_{t\to\infty} \frac{(p^t)^R M(p^t)}{\phi^n(p^t)},$$

provided the limit exists, where $M(p^t)$ represents the number of solutions to the equation $\mathfrak{p}(\mathbf{x}) = \mathbf{s}$ in the multiplicative group $U_{p^t}^n$. A general heuristic argument suggests that we should have

$$\begin{aligned} \mathcal{M}_{\mathfrak{p},\mathbf{s}}(N) &:= \sum_{\mathbf{x}\in[N]^n} \Lambda(\mathbf{x})\mathbf{1}_{V_{\mathfrak{p},\mathbf{s}}}(\mathbf{x}) \\ &\approx \prod_{p<\infty} \mu_p(\mathbf{s})\mu_\infty(N,\mathbf{s})\, N^{n-D} \end{aligned} \tag{4.1.1}$$

for an appropriate singular integral $\mu_\infty(N,\mathbf{s})$. More precisely, $\mu_\infty(N,\mathbf{s})$ should coincide with the singular series that appears in the study of integral points on $V_{\mathfrak{f},\mathbf{s}}$. Here we have that Λ denotes the von Mangoldt function and $\Lambda(\mathbf{x}) = \Lambda(x_1)...\Lambda(x_n)$.

What is actually shown here is a precise result of this form for systems of polynomials of common degree provided that the system has large rank in the sense of Birch for the nonlinear forms and in the sense described above for linear forms. Let us be given a system of forms $\mathfrak{f} = (\mathfrak{f}^{(1)},...,\mathfrak{f}^{(d)})$ where for $1 \le i \le d$ we have $\mathfrak{f}^{(i)} = (\mathfrak{f}_1^{(i)},...,\mathfrak{f}_{r_i}^{(i)})$ be a system of r_i polynomials with integer coefficients of degree exactly i. Define the Birch rank independently on each level of $\mathfrak{f}$, i.e., for each system $\mathfrak{f}^{(i)}$. Define the singular variety, over $\mathbb{C}^n$, associated to the forms $\mathfrak{f}^{(i)}$, $i \ge 2$, to be the collection of $\mathbf{x}$ such that the Jacobian of $\mathfrak{f}^{(i)}$ at $\mathbf{x}$, given by the matrix of partial derivatives

$$Jac_{\mathfrak{f}^{(i)}}(\mathbf{x}) = \left[\frac{\partial \mathfrak{f}_k^{(i)}}{\partial x_j}(\mathbf{x})\right]_{k=1,j=1}^{r,n},$$

has rank strictly less than r_i. This collection is labeled as $V_{\mathfrak{f}^{(i)}}^*$. The Birch rank $\mathcal{B}_i(\mathfrak{f})$ is defined to be $codim(V_{\mathfrak{f}^{(i)}}^*)$ provided that $r_i \neq 0$, in which case we simply assign the value ∞. This notion is extended to a general polynomial system $\mathfrak{p}$ by defining the rank by $\mathcal{B}_i(\mathfrak{p}) = \mathcal{B}_i(\mathfrak{f})$, where $\mathfrak{f}$ is the system of forms associated to $\mathfrak{p}$.

The main result that is shown here is the following.

Theorem 4.1. *For a given positive integer d, there exists constants $\chi(r,d)$ such that the following holds:*

Let $\mathfrak{p} = \mathfrak{p}^{(d)}$ be a given system of integral polynomials with r polynomials of degree d in n variables, and set $D = dr$. If we have $\mathcal{B}_d(\mathfrak{p}) \ge \chi(R,d)$ then for the equation $\mathfrak{p}(\mathbf{x}) = \mathbf{s}$ we have an asymptotic of the form

$$\mathcal{M}_{\mathfrak{p},\mathbf{s}}(N) \sim \prod_p \mu_p(\mathbf{s})\, \mu_\infty(N,\mathbf{s})\, N^{n-D}$$

as $N \to \infty$. Moreover, if $\mathfrak{p}(\mathbf{x}) = \mathbf{s}$ has a nonsingular solution in $\mathbb{U}_p$, the p-adic integer units, for all primes p, then

$$\prod_p \mu_p(\mathbf{s}) > 0.$$

The quantitative aspects of the constants $\chi(r,d)$ are in general very poor. The terms $\chi_1(r,d)$ may be taken to be $2r+1$. The case for quadratic forms is still somewhat reasonable, for systems of quadratics one can achieve something of the shape $\chi(r,2) \le 2^{2^{Cr^2}}$ (to be compared to $r(r+1)$ for the integral analogue). However, the constants $\chi(1,d)$ already exhibit tower type behaviour in d (to be compared to $d2^d$ for the integral analogue), and the situation quickly worsens from there. For recent results with exponential type rank conditions see[115, 75].

4.2 Overview and notations

The primary technique used in the proof of Theorem 4.1 is the circle method, and the argument is an adaptation of the following mean value approach. If a single integral form $\mathfrak{F}$ of degree d in n variables takes the shape

$$\mathfrak{F}(\mathbf{x}) = x_1^d + \mathfrak{F}_1(\mathbf{y}) + \mathfrak{F}_2(\mathbf{z}),$$

where $\mathbf{x} = (x_1, \mathbf{y}, \mathbf{z})$, then we have the representation

$$\mathcal{M}_{\mathfrak{F},0} = \int_0^1 \left(\sum_{x_1 \in [N]} \Lambda(x_1) e(\alpha x_1^d) \right) \left(\sum_{\mathbf{y} \in [N]^m} \Lambda(\mathbf{y}) e(\alpha \mathfrak{F}_1(\mathbf{y})) \right) \left(\sum_{\mathbf{z} \in [N]^{n-1-m}} \Lambda(\mathbf{z}) e(\alpha \mathfrak{F}_2(\mathbf{z})) \right) d\alpha$$

$$:= \int_0^1 S_0(\alpha) S_1(\alpha) S_2(\alpha) d\alpha.$$

An application of the Cauchy-Schwarz inequality then gives

$$\mathcal{M}_{\mathfrak{F},0}^2 \leq ||S_0||_\infty^2 ||S_1||_2^2 ||S_2||_2^2 \leq ||S_0||_\infty^2 (\log\, N)^{2n-2} Y(N) Z(N),$$

where $Y(N)$ is the number of solutions to the equation $\mathfrak{F}_1(\mathbf{y}) = \mathfrak{F}_1(\mathbf{y}')$ with $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $Z(N)$ is the number of solutions to the equation $\mathfrak{F}_2(\mathbf{z}) = \mathfrak{F}_2(\mathbf{z}')$ with $\mathbf{z}, \mathbf{z}' \in [N]^{n-1-m}$. If $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are assumed to have large rank, then $Y(N)Z(N) = O(N^{2m-d}) O(N^{2(n-1-m)-d}) = O(N^{2n-2-2d}))$. More generally, for any measurable subset $\mathfrak{u} \subset [0,1]$ we have

$$\int_{\mathfrak{u}} S_0(\alpha) S_1(\alpha) S_2(\alpha) d\alpha = O(||S_0||_{\infty(\mathfrak{u})} (\log\, N)^{n-1} N^{n-1-d}), \tag{4.2.1}$$

where $||S_0||_{\infty(\mathfrak{u})}$ denotes the supremum of $|S_0(\alpha)|$ for $\alpha \in \mathfrak{u}$.
This partition into major and minor arcs becomes useful due to the following.

Lemma 4.1. *Given $c > 0$, there exists a C such that $||S_0||_{\infty(\mathfrak{m}(C))} \leq N(\log\, N)^{-c}$.*

This, together with equation 4.2.1, in turn gives the bound

$$\int_{\mathfrak{m}(C)} S_0(\alpha) S_1(\alpha) S_2(\alpha) d\alpha = O((\log\, N)^{-1} N^{n-d}). \tag{4.2.2}$$

Thus one is left with the task of approximating

$$\int_{\mathfrak{M}(C)} \sum_{\mathbf{x} \in [N]^n} e(\alpha \mathfrak{F}(\mathbf{x})) d\alpha,$$

which is in general susceptible to the usual methods.
Now let us at the case of a general form $\mathfrak{F}$ of degree 2. If we introduce a splitting of the variables $\mathbf{x} = (x_1, \mathbf{y}, \mathbf{z})$, we induce a decomposition of the shape

$$\mathfrak{F}(\mathbf{x}) = a x_1^2 + \mathfrak{g}^{(1)}(\mathbf{y}, \mathbf{z}) x_1 + \mathfrak{F}_1(\mathbf{y}) + \mathfrak{F}_2(\mathbf{z}) + \mathfrak{g}^{(2)}(\mathbf{y}, \mathbf{z})$$

for a form $\mathfrak{g}^{(2)}$ which is bilinear in $\mathbf{y}$ and $\mathbf{z}$, and a linear form $\mathfrak{g}^{(1)}$. There are two possible approaches to adapting the above argument to this case.

The first involves a dichotomized argument based on the rank of $\mathfrak{g}^{(2)}$. If we have that $\mathfrak{g}^{(2)}$ has large rank (at least five is shown in [74], one can obtain good bounds on the exponential sum

$$\sum_{\mathbf{y}\in[N]^m}\sum_{\mathbf{z}\in[N]^{n-1-m}}\Lambda(\mathbf{y})\Lambda(\mathbf{z})e(\alpha(\mathfrak{F}_1(\mathbf{y})+\mathfrak{F}_2(\mathbf{z})+\mathfrak{g}^{(2)}(\mathbf{y},\mathbf{z}))),$$

by simply removing the contribution of the von Mangoldt function with two applications of the Cauchy-Schwarz inequality. In this case the methods of Birch are applicable (and the rank bounds are comparable). If $\mathfrak{g}^{(2)}$ has small rank, then it must be the case that $\mathfrak{F}_1$ and $\mathfrak{F}_2$ each have large rank for appropriately chosen splitting of the variables. Write $\mathfrak{g}^{(2)}(\mathbf{y},\mathbf{z})=\langle\mathbf{y},B\mathbf{z}\rangle$ for an appropriately sized matrix B whose rank is small, and split $\mathfrak{g}^{(1)}(\mathbf{y},\mathbf{z})=l_1(\mathbf{y})+l_2(\mathbf{z})$. The above argument can then be run on the intersection of the level sets of $l_1(\mathbf{y})$, $l_2(\mathbf{z})$, and $B\mathbf{z}$, as both $\mathfrak{F}_1$ and $\mathfrak{F}_2$ have large rank on this small codimension linear space. On such an intersection we get an extra power gain, which is equivalent to the codimension. This extra compensates for the loss of originally applying the Cauchy-Schwarz inequality on each linear space. Thus summing over all such level sets gives an appropriate bound.

The second approach, to be fair, is simply a streamlined version of the first which removes the need for a dichotomized approach, and this is the one we shall follow. The main requirement here is an appropriate decomposition of $\mathfrak{F}$ in the form

$$\mathfrak{F}(\mathbf{x})=ax_1^2+\mathfrak{g}^{(1)}(\mathbf{y},\mathbf{z})x_1+\mathfrak{F}_1(\mathbf{y})+\mathfrak{F}_2(\mathbf{z})+\mathfrak{g}^{(2)}(\mathbf{y},\mathbf{z})$$

such the rank of $\mathfrak{F}_1+\mathfrak{g}^{(2)}$ is large with respect to $\mathfrak{g}^{(1)}$, the number of variables of composing $\mathbf{y}$, m, is small, and the rank of $\mathfrak{F}_2$ is large with respect to m. As before, we wish to fix $l_1(\mathbf{y})$, $l_2(\mathbf{z})$, and $B\mathbf{z}$. The difference is that we have no assumption on the rank of B. However, by controlling the value of m provides, we have a way to control the number of linear equations is $\mathbf{z}$. Running the argument as before and summing over the linear spaces leads reduces our problem to providing an appropriate bound for the number of solutions to

$$\begin{aligned}\mathfrak{F}_1(\mathbf{y})+\mathfrak{g}^{(2)}(\mathbf{y},\mathbf{z}) &= \mathfrak{F}_1(\mathbf{y}')+\mathfrak{g}^{(2)}(\mathbf{y}',\mathbf{z})\\ l_1(\mathbf{y}) &= l_1(\mathbf{y}')\\ \mathfrak{F}_2(\mathbf{z}) &= \mathfrak{F}_2(\mathbf{z}')\\ l_2(\mathbf{z}) &= l_2(\mathbf{z}')\\ B\mathbf{z} &= B\mathbf{z}'\end{aligned}\tag{4.2.3}$$

with $\mathbf{y},\mathbf{y}'\in[N]^m$ and $\mathbf{z},\mathbf{z}'\in[N]^{n-1-m}$. This achieved by the rank assumptions of $\mathfrak{F}_1+\mathfrak{g}^{(2)}$ and $\mathfrak{F}_2$ in the original decomposition.

The strategy for forms of higher degree by a similar decomposition of the form

$$\mathfrak{F}(\mathbf{x})=ax_1^d+\mathfrak{g}^{(1)}(\mathbf{y},\mathbf{z})x_1^{d-1}+...+\mathfrak{g}^{(d-1)}(\mathbf{y},\mathbf{z})x_1+\mathfrak{F}_1(\mathbf{y})+\mathfrak{F}_2(\mathbf{z})+\mathfrak{g}^{(d)}(\mathbf{y},\mathbf{z}),$$

where the $\mathfrak{g}^{(i)}$ forms of degree i, and $\mathfrak{F}_1(\mathbf{y})$, $\mathfrak{F}_2(\mathbf{z})$, $\mathfrak{g}^{(d)}$ are forms of degree d. Again we require that the rank of $\mathfrak{F}_1+\mathfrak{g}^{(d)}$ is large with respect to $\mathfrak{g}^{(i)}$ for each $i<d$, the number of variables composing $\mathbf{y}$ is small, and the rank of $\mathfrak{F}_2$ is large with respect to m. That such a decomposition is possible is the subject of section 4. Then we view each form $\mathfrak{g}^{(i)}$, $1\le i\le d$ as a sum of forms in $\mathbf{y}$ with coefficients that are forms in $\mathbf{z}$, and the number of these coefficients is bounded in terms of m. On each of the level sets of this new system of forms in $\mathbf{z}$ we have a system of forms in $\mathbf{y}$, the number of

which is bounded in terms of d. Now passing to the further level of sets of these forms $\mathbf{y}$ provides a place to carry out the simple Cauchy-Schwarz argument at the beginning of this section. Summing back over the level sets then provides a system analogous to the one above.

The only problem with this so far is that the system we end up contains at least a portion of each form $\mathfrak{g}^{(i)}(\mathbf{y}, \mathbf{z})$ for $i = 1, ..., d$, and we have no control on the rank of these forms at all and therefore have no way of dealing with the terminal system. The solution to this problem is found in the work of Schmidt. His results provide a way of partitioning the level sets of a form by the level sets of a system of forms that does have high rank in each degree. Section 3 is dedicated to this. Working with this more regular system as opposed the $\mathfrak{g}^{(i)}$'s does provide a manageable terminal system, and allows for a bound on the minor arc integral.

Extending this method to systems of forms is relatively straightforward at this point, and is of course carried out below. The major annoyance here is the need to isolate larger number of suitable variables $x_1, ..., x_K$ to get a the logarithmic gain on the minor arcs, as opposed to a randomly chosen single variable x_1.

4.2.1 Outline and notation

The outline for the rest of the paper is as follows. Sections 3 and 4 are as described above. The completion of the bound for the integral on the minor arcs is going to be carried out in Section 5. The major arcs are dealt with in Section 6, where the asymptotic formula is shown. Section 7 is dedicated to the proof of Theorem. The final section concludes the work with a few further remarks.

Remarks on notation The symbols $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ denote the integers, the rational numbers, the real numbers, and the complex numbers, respectively. The r-dimensional flat torus $\mathbb{R}^r/\mathbb{Z}^r$ is denoted by $\mathbb{T}^r$. The p-adic integers are denoted by $\mathbb{Z}_p$, and the units of $\mathbb{Z}_p$ are denoted by $\mathbb{U}_p$. The symbol Z_N represents shorthand for the groups $\mathbb{Z}/N\mathbb{Z}$. Also, the shorthand for the multiplicative group Z_N^* is U_N.

For a given measurable set $X \subseteq \mathbb{T}^r$ we shall use the notation $||f||_{p(X)}$ to denote the L^p norm of the function $\mathbf{1}_X f$ with the normalized Lebesgue measure on the r-dimensional flat torus. If X is omitted it is assumed that $X = \mathbb{T}^r$. Here, and in general, $\mathbf{1}_X$ denotes a characteristic function for X in a specified ambient space, and, on occasion, the set X is replaced by a conditional statement which defines it.The Landau o and O notation is used throughout the work. The notation $f \lesssim g$ is sometimes used to replace $f = O(g)$.

4.3 A Regularity lemma for systems of polynomials

In [97], Schmidt provides an alternative definition of rank for a form. For a single form $\mathfrak{F}$ of degree at least 2 defined over a field k, define the *Schmidt rank* $h_k(\mathfrak{F})$ to be the minimum value of l such that there exist a decomposition

$$\mathfrak{F} = \sum_{i=1}^{l} U_i V_i$$

where U_i and V_i are forms defined over k of degree at least one. For a system $\mathfrak{f}^{(d)} = (\mathfrak{f}_1^{(d)}, ..., \mathfrak{f}_{r_d}^{(d)})$ of forms of degree d we define $h_k(\mathfrak{f})$ to be

$$\min\{h_k(\lambda_1 \mathfrak{f}_1^{(d)} + ... + \lambda_{r_d} \mathfrak{f}_{r_d}^{(d)}) \,:\, \lambda_i \neq 0 \, for\, some\, i\}.$$

A few basic properties of the Schmidt rank are important.

- If $\mathfrak{f}$ is defined over a field k, and k' is an extension of k, then $h_{k'}(\mathfrak{f}) \leq h_k(\mathfrak{f})$.
- The Schmidt rank is invariant under invertible linear transformations of k, i.e. $h_k(\mathfrak{f} \circ A) = h_k(\mathfrak{f})$ for $A \in GL_n(k)$.
- If $\mathfrak{f}'(x_2, ..., x_n) = \mathfrak{f}(0, x_2, ..., x_n)$, then $h_k(\mathfrak{F}') \geq h_k(\mathfrak{F}) - r$.

The first two are clear from the definition, and the third simply follows from the fact that $\mathfrak{f}(\mathbf{x}) - \mathfrak{f}'(\mathbf{x})$ is of the form $x_1 \mathfrak{g}(\mathbf{x})$ for some $d-1$ degree system of forms $\mathfrak{g}$. Also, the second and third imply that the rank cannot drop on a codimension j subspace by more than jr.

As observed by Schmidt, the Birch rank $\mathcal{B}_d(\mathfrak{F})$ and the complex Schmidt rank $h_{\mathbb{C}}(\mathfrak{F})$ are essentially equivalent for a form of degree d, being bounded in terms of each other. Of course the same is true for systems as well. The rational Schmidt rank $h_{\mathbb{Q}}$ on the other hand is not equivalent and we need the following, which is a weakened version of a main result in [97].

Definition 4.1. *Let $\mathfrak{f} = (f^{(d)}, ..., \mathfrak{f}^{(1)})$ be a graded system of forms with rational coefficients. Assume that $\mathfrak{f}^{(i)}$ consists of r_i forms for each $1 \leq i \leq d$, and set $R = \sum_i r_i$ and $D = \sum_i i r_i$. The system $\mathfrak{f}$ is said to be regular if $|V_{\mathfrak{f},\mathbf{s}} \cap [N]^n| = O(N^{n-D})$ as $N \to \infty$ holds uniformly for $\mathbf{s} \in \mathbb{Z}^R$.*

Theorem A (Schmidt '86)**.** *For a given positive integers R and d, there exists constants $\rho_i(R, d)$ for $2 \leq i \leq d$ such that the following holds:*

Let $\mathfrak{f} = (\mathfrak{f}^{(d)}, ..., \mathfrak{f}^{(2)})$ be given system of rational forms with r_i forms of degree i composing each subsystem $\mathfrak{f}^{(i)}$, $R = r_2 + ... + r_d$ the total number of forms, and $D = 2r_2 + ... + dr_d$. If we have $h_{\mathbb{Q}}(\mathfrak{f}^{(i)}) \geq \rho_i(R, d)$ for each i, then the system $\mathfrak{f}$ is regular.

One of the key observations of Schmidt is that his definition of rank has a very nice reductive quality with respect to the degree, in the sense that forms of small rank may be replaced by a small number of forms of lesser degree. The next result captures this idea.

Proposition 4.1 (Regularity lemma)**.** *Let $d > 1$ be a fixed integer, and let F be any collection of functions $F_i(R)$ for $i = 2, ..., d$ mapping the nonnegative integers into themselves. For a collection of non-negative integers $r_1, ..., r_d$, there exists constants*

$$C_1(r_1, ..., r_d, F), ..., C_d(r_1, ..., r_d, F)$$

such that the following holds:

Given a system of integral forms $\mathfrak{f} = (\mathfrak{f}^{(d)}, \mathfrak{f}^{(d-1)}, ..., \mathfrak{f}^{(1)})$, where each of the $\mathfrak{f}^{(i)}$ is a system of r_i forms of degree i, there exists a system of rational forms $\mathfrak{g} = (\mathfrak{g}^{(d)}, \mathfrak{g}^{(d-1)}, ..., \mathfrak{g}^{(1)})$ satisfying:

1. *The level sets of $\mathfrak{g}$ partition those of $\mathfrak{f}$.*

2. *The number of forms in each subsystem of $\mathfrak{g}^{(i)}$, say r_i', is at most $C_i(r_1, ..., r_d, F)$ for each $1 \leq i \leq d$.*

3. *Set H to be the linear subspace defined by $\mathfrak{g}^{(1)} = 0$, and R' to be the total number of forms in $\mathfrak{g}$ of degree at least 2. The system $((\mathfrak{g}^{(d)}, \mathfrak{g}^{(d-1)}, ..., \mathfrak{g}^{(2)}))$ has $h_{\mathbb{Q}}(\mathfrak{g}^{(i)}|_H) \geq F_i(R')$ for each $2 \leq i \leq d$. Here $\mathfrak{g}^{(i)}|_H$ denotes the restriction of $\mathfrak{g}^{(i)}$ to the subspace H.*

Proof. The proof is carried out by a double induction on the parameters. First for a fixed d we show that the case r_d with any choice of $r_{d-1}, ..., r_1$ implies the similar scenario for the case $r_d + 1$. Then the induction on d is carried out.

The initial case we need to consider is $d = 2$ with a given function $F_2(R)$. Take a system of forms $\mathfrak{f} = (\mathfrak{f}^{(2)}, \mathfrak{f}^{(1)})$ with $r_2 = 1$ and any value of r_1. If $h_{\mathbb{Q}}(\mathfrak{f}^{(2)}) \geq F_2(r_2) + r_1$ then we may simply take $\mathfrak{f} = \mathfrak{g}$, as restricting $\mathfrak{f}^{(2)}$ to the subspace defined by $\mathfrak{f}^{(1)} = 0$ can only drop the rank by at most r_1. Otherwise $\mathfrak{f}^{(2)} = \sum_{i=1}^{l} U_i V_i$ for some rational linear forms U_i and V_i where $l < F_2(r_2) + r_1$. We may then adjoin the linear forms $U_1, ..., V_l$ to the system $\mathfrak{f}^{(1)}$ to obtain the system $\mathfrak{g} = \mathfrak{g}^{(1)}$. Properties (1) and (2) are easily verified for this system, and property (3) is vacuous.

Now for a fixed value of d assume that the result holds for all systems with maximal degree d for any given collection of functions F when $r_d = j$ and the $r_{d-1}, ..., r_1$ are arbitrary. Consider now a fixed collection of functions F and a system $\mathfrak{f} = (\mathfrak{f}^{(d)}, ..., \mathfrak{f}^{(1)})$ with $r_d = j + 1$ and $r_1, ..., r_{d-1}$ arbitrary. Let $\mathfrak{f}'$ be the system $(\mathfrak{f}^{(d-1)}, ..., \mathfrak{f}^{(1)})$. By the induction hypothesis, there is a system $\mathfrak{g}'$ of rational forms which is a regularization of $\mathfrak{f}'$ with respect to $F_i'(R) := F_i(R+(j+1))$ for $i = 2, ..., (d-1)$.

Now let $\tilde{\mathfrak{g}}' = (\mathfrak{f}^{(d)}, \mathfrak{g}')$. If $\tilde{\mathfrak{g}}'$ fails to be the regularization of $\mathfrak{f}$, then it must happen that $h_{\mathbb{Q}}(\mathfrak{f}^{(d)}) < F_d(R_{\tilde{\mathfrak{g}}'}) + (j+1)r_1(\tilde{\mathfrak{g}}')$, where $R_{\tilde{\mathfrak{g}}'}$ is the number of forms of degree at least two in the system $\tilde{\mathfrak{g}}'$ and $r_1(\tilde{\mathfrak{g}}')$ is the number of linear forms in $\mathfrak{g}'$. As before, in this case there must exist homogeneous rational polynomials U_i and V_i, $i = 1, ..., l < F_d(R_{\tilde{\mathfrak{g}}'}) + r_1$, such that

$$\lambda_1 \mathfrak{f}_1^{(d)} + ... \lambda_{j+1} \mathfrak{f}_{j+1}^{(d)} = \sum_{i \leq l} U_i V_i,$$

where without loss of generality we have $\lambda_{j+1} \neq 0$. Now let $\mathfrak{g}''$ be $\mathfrak{g}'$ adjoined with the those forms U_i and V_i which are not linear combinations of forms already in $\mathfrak{g}'$, and set

$$\tilde{\mathfrak{g}}'' = ((\mathfrak{f}_1^{(d)}, ..., \mathfrak{f}_j^{(d)}), \mathfrak{g}'').$$

By the induction hypothesis there is a system $\mathfrak{g}$ which is the regularization of $\tilde{\mathfrak{g}}''$ with respect to initial collection of functions F. As the number of forms in $\mathfrak{g}''$ is expressible in terms of $r_1, ..., r_d$, and d, the system $\mathfrak{g}$ is the regularization of $\mathfrak{f}$.

The induction argument to go from d to $d + 1$ is simply the above argument carried out with $j = 0$. □

Apply Proposition 4.1 with the functions being given by the values of the Schmidt constants $\rho_i(R, d)$ then provides the following .

Corollary 4.1. *Let $\mathfrak{f} = (\mathfrak{f}^{(d)}, ..., \mathfrak{f}^{(2)})$ be given system of rational forms with r_i forms of degree i composing each subsystem $\mathfrak{f}^{(i)}$. There exists a regular system of forms $\mathfrak{g}$ satisfying conclusions 1) and 2) of Proposition 1.*

4.4 A Decomposition of forms

For $I \subset [n]$, let $\mathbf{y} = (y_i)_{i=1,\dots,n}$ be the vector with components $y_i = x_i$ for $i \in I$ and 0 otherwise. Also let $\mathbf{z}$ be the vector defined similarly for the set $[n]\backslash I$. Note that $\mathbf{y} + \mathbf{z} = \mathbf{x}$. In this section we prove the following decomposition result.

Proposition 4.2. *Let positive integers C_1 and C_2 be given. Let $\mathfrak{f}$ be given system of r rational forms with $\mathcal{B}_d(\mathfrak{f})$ sufficiently large with respect to C_1 and C_2. There exists an $I \subset [n]$ such that $|I| \le C_1 r$ and the associated decomposition*

$$\mathfrak{f}(\mathbf{x}) = \mathfrak{f}_1(\mathbf{y}) + \mathfrak{f}_2(\mathbf{z}) + \mathfrak{g}(\mathbf{y}, \mathbf{z})$$

satisfies $\mathcal{B}_d(\mathfrak{f}_1 + \mathfrak{g}) \ge C_1$ and $\mathcal{B}_d(\mathfrak{f}_2) \ge C_2$.

The proof of this result is carried by dealing directly with the Jacobian matrices. Some notation is helpful. Let $M = M(\mathbf{x})$ be a $i \times j$ matrix whose entries depend on $\mathbf{x}$. The notation $M \propto M'$ is used to imply that M is a submatrix of M' obtained by the deletion of columns, so that M is an $i \times j'$ matrix with $j' \le j$. Let V_M^* be the collection of $\mathbf{x}$ where M has rank strictly less than i. Clearly one has that if $M \propto M'$, then $V_{M'}^* \subseteq V_M^*$.

Lemma 4.2. *If $\mathfrak{f}^{(d)}$ is a system of r integral forms of degree d in n variables, then the restriction of $\mathfrak{f}$ to the hyperplane defined by $x_n = 0$ has rank at least $\mathcal{B}_d(\mathfrak{f}) - r - 1$.*

Proof. Denote the restriction of $\mathfrak{f}$ to the subspace defined by $x_n = 0$ as $\mathcal{F}$. The matrix $Jac_{\mathfrak{F}}$ is then the matrix $Jac_{\mathfrak{f}}$ with the last column deleted and restricted to the space $x_n = 0$. It follows that $V_{\mathcal{F}}^* \cap H \subseteq V_{\mathfrak{f}}^* \cap \{x_n = 0\}$, where H denotes the variety where the last column has all entries equal to zero. As H is defined by at most r equations, it has co-dimension at most r. In turn it follows from the definition that $\mathcal{B}_d(\mathfrak{F}) \ge \mathcal{B}_d(\mathfrak{f}) - r - 1$. □

As the rank of a non-homogeneous quadratic system is defined to be the rank of the homogeneous part, it follows, simply by noting that the Birch rank is invariant under invertible linear transformations[1], that this result extends to general affine linear spaces, i.e., cosets of linear subspaces.

Corollary 4.2. *If H is an affine linear space of co-dimension m, then the restriction of $\mathfrak{f}$ to H has rank at least $\mathcal{B}_d(\mathfrak{f}) - m(r+1)$.*

Now define $\mathcal{C}_{\mathfrak{f}}(k)$ to be the minimal value of m such that there exists an $M \propto Jac_{\mathfrak{f}}$ of size $r \times m$ such that V_M^* has dimension at most $n - k$. This is defined to be infinite if no such value exists.

Lemma 4.3. *For a system $\mathfrak{f}$ of r forms of degree d, one has that*

$$\mathcal{C}_{\mathfrak{f}}(k) \le kr, \ \textit{if} \quad (k-1)((d-1)r)^{k-1} < \mathcal{B}(\mathcal{Q}).$$

Proof. Write the singular variety $V_{\mathfrak{f}}^*$ as an intersection of varieties V_I, where V_I is a the zero set of the characteristic polynomial for the $r \times r$ minor coming from the selecting the columns $I = \{i_1, \dots, i_r\} \subset [n]$. Proceeding inductively, assume we have selected $V^{(l)} = \cap_{j=1}^{l} V_{I_j}$ such that $\dim(V^{(l)}) \le n - l$. The degree of each V_I is at most $r(d-1)$, so that $V^{(l)}$ has at most $((d-1)r)^k$ components. Label the components with degree precisely $n - l$ as $Y_1, \dots, Y_j$, where $j \le ((d-1)r)^k$. For each Y_i, set $N(i)$ to be the set of j's such that V_{I_j} has Y_i as a component, where all of the other elements of $\binom{[n]}{r}$ have also been enumerated. If there is a j such that $j \notin \cup N(i)$,

[1] This fact is essentially the multivariate chain rule.

then $\dim(V_{I_j} \cap V^{(l)}) = n - l - 1$. Otherwise it follows that $\cup N(i) = [n]$. In this case, look at $V^* = \cap_{i=1}^{l}(\cap_{j\in N(i)} V_{I_j})$. In turn it follows that $codim(V^*) \leq \sum_i codim(\cap_{j\in N(i)} V_{I_j}) \leq lr^l$, which cannot happen if $l((d-1)r)^l < \mathcal{B}_d(\mathfrak{f})$. Thus one can choose a $V_{I_{l+1}}$ such that $V^{(l+1)}$ has dimension $l+1$. From this it is also clear that $\mathcal{C}_{\mathfrak{f}}(l+1) \leq \mathcal{C}_{\mathfrak{f}}(l) + r$, which gives the result as $C_{\mathfrak{f}}(1) \leq r$. □

Proof of Proposition 2. We begin by considering the case $\mathfrak{f} = \mathfrak{f}^{(d)}$. Start by applying Lemma 4.3 with $k = C_1$, valid by the assumed lower bound on $codim V_{\mathfrak{f}}^*$. Then there are at most $C_1 r$ columns of $Jac_{\mathfrak{f}}$ providing an associated singular set of dimension at most $n - C_1$. Call this sub-matrix M, and let I denote the collection of the indices of these columns, noting that $m = |I| \leq C_1 r$. The associated decomposition into $\mathfrak{f}_1, \mathfrak{f}_2$, and $\mathfrak{f}_I$ follows. It is easily seen that $M \propto Jac_{\mathfrak{f}_1+\mathfrak{g}}$, and it follows that $V_{\mathfrak{f}_1+\mathfrak{g}}^* \subseteq V_M^*$, and hence $\mathcal{B}(\mathfrak{f}_1 + \mathfrak{g}) \geq C_1$.
Now look at the the matrix W obtained by deleting the columns of M from $Jac_{\mathfrak{f}}$. One now has $Jac_{\mathfrak{f}}(\mathbf{z}) = Jac_{\mathfrak{f}_2+\mathfrak{g}}(0,\mathbf{z})$. There are at most $C_1 r^2$ non-zero entries of $Jac_{\mathfrak{g}}$ which are dependent on $\mathbf{z}$. Then there is a of co-dimension at most $C_1 r^2$, say H, such that $V_{\mathfrak{f}_2} \cap H \subseteq V_Q^* \cap \{\mathbf{x} : \mathbf{y} = 0\}$. This gives the inequality

$$dim(V_{\mathfrak{f}_2}^*) - C_1 r^2 \leq dim(V_{\mathfrak{f}}^*) \leq n - C_1 r(r+1) - C_2.$$

The dimension of $V_{\mathfrak{f}_2}^*$ on the subspace given by $\mathbf{y} = 0$ is then at most $n - C_1 r - C_2 \leq (n-m) - C_2$, i.e., $\mathcal{B}(\mathfrak{f}_2) \geq C_2$. □

4.5 Minor arcs estimates

Assume now throughout this section that we have a fixed system of integral polynomials $\mathfrak{p} = (\mathfrak{p}_1, ..., \mathfrak{p}_r)$, where each $\mathfrak{p}_i$ is of degree d. The system $\mathfrak{f}$ is again the highest degree homogeneous parts of $\mathfrak{p}$.
For a given value of C and an integer $q \leq (\log N)^C$, define a major arc

$$\mathfrak{M}_{\mathbf{a},q}(C) = \{\alpha \in [0,1] : \max_{1\leq i\leq r} |\alpha_i - a_i/q| \leq N^{-d}(\log N)^C\}$$

for each $\mathbf{a} = (a_1, ..., a_r) \in U_q^r$. When $q = 1$ it is to be understood that $U_1 = \{0\}$. These arcs are disjoint, and the union

$$\bigcup_{q\leq(\log N)^C} \bigcup_{\mathbf{a}\in U_q^r} \mathfrak{M}_{\mathbf{a},q}(C)$$

defines the major arcs $\mathfrak{M}(C)$. The minor arcs are then given by

$$\mathfrak{m}(C) = [0,1]\backslash\mathfrak{M}(C).$$

The main result in this section is to deal with the integral representation on the minor arcs.

Lemma 4.4. *There exists constant $\chi(r,d)$ such that if we have $\mathcal{B}_d(\mathfrak{p}) \geq \chi(r,d)$, then there exists a C such that*

$$\int_{\mathfrak{m}(C)} e(-\mathbf{s}\cdot\alpha) \sum_{\mathbf{x}\in[N]^n} \Lambda(\mathbf{x}) e(\mathfrak{p}(\mathbf{x})\cdot\alpha)\, d\alpha = O(N^{n-D}(\log N)^{-c}). \tag{4.5.1}$$

holds for any prescribed c with an implied constant independent of $\mathbf{s}$.

Another set of minor arcs is also required for an exponential sum estimate. For each $1 \leq i \leq d$, define for $a^{(i)} \in U_q$ the major arc

$$\mathfrak{N}_{a^{(i)},q}^{(i)}(C) = \{\xi^{(i)} \in \mathbb{T} : |\xi^{(i)} - \frac{a^{(i)}}{q}| \leq N^{-i}(\log N)^C\}.$$

Set
$$\mathfrak{N}_{\mathbf{a},q}(C) = \mathfrak{N}^{(d)}_{a^{(d)},q}(C) \times ... \times \mathfrak{N}^{(1)}_{a^{(1)},q}(C),$$
where $\mathbf{a} = a^{(d)} \times ... \times a^{(1)}$. The major arcs are now
$$\mathfrak{N}(C) = \bigcup_{q \leq (\log\ N)^C} \bigcup_{\mathbf{a} \in U_q^d} \mathfrak{N}_{\mathbf{a},q}(C).$$

Let $\mathfrak{n}(C)$ denote a set of minor arcs $\mathfrak{n}(C) = [0,1] \backslash \mathfrak{N}(C)$.
Define the exponential sum
$$S_0(\beta) = \sum_{x \in [N]} \Lambda(x) e(\beta_d x^d + ... + \beta_1 x),$$

for $(\beta_d, ..., \beta_1) \in \mathbb{T}^d$.

Lemma 4.5. *Given $c > 0$, there exists a C such that $||S_0||_{\infty(\mathfrak{n}(C))} = O(N(\log\ N)^{-c})$.*

For a proof the reader is referred to [57] (Ch. 10, §5, Lemma 10.8).

Proof of Lemma 4.4. Our first goal is to pick an appropriate splitting of the variables $x = (x_1, ..., x_K, \mathbf{y}, \mathbf{z})$ which induces a decomposition
$$\mathfrak{p}(\mathbf{x}) = \mathfrak{p}_0(x_1, ..., x_K, \mathbf{y}, \mathbf{z}) + \mathfrak{p}_1(\mathbf{y}) + \mathfrak{g}(\mathbf{y}, \mathbf{z}) + \mathfrak{p}_2(\mathbf{z})$$

such that the choice of $x_1, ..., x_K$ are useful for applying Lemma 5, and $\mathbf{y}$ consisting of m variables is chosen so that $\mathfrak{p}_1 + \mathfrak{g}$ has large rank with respect to K. If we assume that the rank of $\mathfrak{p}$ is initially large, then $\mathfrak{p}_2$ also has large rank in terms of K and m.

To select the variables $x_1, ..., x_K$, we first consider associated system of forms $\mathfrak{f}$. We collect the r coefficients of each term $x_{i_1}...x_{i_d}$ into a vector $\mathbf{b}_{i_1,...,i_d}$ We select r of these which are linearly independent. The total number of indices involved is our value of K, in this choice is at most dr, and we assume that the corresponding variables to be the first $1 \leq i \leq K$. The variables $x_1, ..., x_K$ have now been selected.

For any choice of $\mathbf{y}$ and $\mathbf{z}$ we have some decomposition of the shape
$$\begin{aligned}
\mathfrak{p}(x_1, ..., x_K, \mathbf{y}, \mathbf{z}) \quad = \quad & \mathfrak{p}(x_1, ..., x_K, 0, ..., 0) + \sum_{j=1}^{d-1} \sum_{1 \leq i_1 < .... < i_j \leq K} \left(\sum_{k=1}^{d-j} \mathfrak{G}^{(k)}_{i_1,...,i_j}(\mathbf{y}, \mathbf{z}) \right) x_{i_1} ... x_{i_j} \\
& + \sum_{\kappa=1}^{d-1} \sum_{1 \leq \iota_1 < .... < \iota_\kappa \leq m} \left(\sum_{k=1}^{d-\kappa} \mathfrak{H}^{(k)}_{0;\iota_1,...,\iota_\kappa}(\mathbf{z}) \right) y_{\iota_1} ... y_{\iota_\kappa} \\
& + \mathfrak{p}_1(0, ..., 0, \mathbf{y}, 0) + \mathfrak{p}_2(0, ..., 0, 0, \mathbf{z}),
\end{aligned} \tag{4.5.2}$$

where for each appropriate set of indices the $\mathfrak{G}^{(k)}_{i_1,...,i_j}$ and the $\mathfrak{H}^{(k)}_{0;\iota_1,...,\iota_\kappa}$ are systems of at most r integral forms of degree k in the appropriate variables.
Let $\mathfrak{G}$ be the collection of all the forms $\mathfrak{G}^{(k)}_{i_1,...,i_j}$, of which there are crudely at most $R_{\mathfrak{G}} \leq d^2 K^d r \leq d^{d+2} r^{d+1}$. Set $\mathfrak{g}_{\mathfrak{G}}$ to be the regularization of $\mathfrak{G}$ with respect to the functions $F_i(R) = \rho_i(R + r, d)$ for $i = 1, ..., d-1$, noting that the number of forms of degree i in $\mathfrak{g}_{\mathfrak{G}}$ is bounded in terms of r and d

by Proposition 4.1

The variables $\mathbf{y}$ are now selected by Proposition 4.2 by the choice of $C_1 = \rho_i(R_{\mathfrak{g}_{\mathfrak{G}}} + r, d)$ so that the forms

$$\mathfrak{f}(0, ..., 0, \mathbf{y}, \mathbf{z}) = \mathfrak{f}_1(\mathbf{y}) + \mathfrak{g}(\mathbf{y}, \mathbf{z}) + \mathfrak{f}_2(z)$$

have $\mathcal{B}_d(\mathfrak{f}_1(\mathbf{y}) + \mathfrak{g}(\mathbf{y}, \mathbf{z})) \geq C_1$ with the number of $\mathbf{y}$ variables, m, being at most $C_1 r$. With this choice we have that system obtained by adjoining the systems $\mathfrak{g}_{\mathfrak{G}}$ and $\mathfrak{f}_1 + \mathfrak{g}$ is in fact a regular system by Theorem A.
We breakdown further the forms of $\mathfrak{g}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z})$ by separating the $\mathbf{y}$ and $\mathbf{z}$ parts:

$$(\mathfrak{g}_{\mathfrak{G}})_i^{(l)}(\mathbf{y}, \mathbf{z}) = \sum_{\kappa=0}^{l} \sum_{1 \leq \iota_1 < .... < \iota_\kappa \leq m} \mathfrak{H}_{i;\iota_1,...,\iota_\kappa}^{(k;l)}(\mathbf{z}) y_{\iota_1} ... y_{\iota_\kappa}. \tag{4.5.3}$$

Note that the right hand side introduces at most $lm^l \leq dm^d$ forms in $\mathbf{z}$ for the ith form of degree l in $\mathfrak{g}_{\mathfrak{G}}$. We collect the forms $\mathfrak{H}_{0;\iota_1,...,\iota_\kappa}^{(k)}$ and $\mathfrak{H}_{i;\iota_1,...,\iota_\kappa}^{(k;l)}$ into a system $\mathfrak{H}$. Then the number of forms $R_{\mathfrak{H}}$ of $\mathfrak{H}$ is at most $R_{\mathfrak{g}_{\mathfrak{G}}} dm^d + rd^2 m^d$. Now we regularize the system $\mathfrak{H}$ with respect to the functions $F_i(R) = \rho_i(R + r, d)$ for $1 \leq i \leq d - 1$ and call the resultant system $\mathfrak{g}_{\mathfrak{H}}$.

If the system $\mathfrak{f}_2(\mathbf{z})$ has rank at least as large as $C_2 = \rho_d(R_{\mathfrak{g}_{\mathfrak{H}}} + r, d)$, then system $\mathfrak{f}_2$ adjoined to the system $\mathfrak{g}_{\mathfrak{H}}$ is a regular system by Theorem A. This can be guaranteed as long the rank of $\mathfrak{f}(0, ..., 0, \mathbf{y}, \mathbf{z})$ is sufficiently large with respect to our choices of C_1 and C_2 by appealing to Proposition 4.2. As loosing the first K variables can drop the rank by at most $K(r+1) \leq r(r+1)d$, and C_1 and C_2 are dependent only on d and r, this is our choice of $\chi(r, d)$. We now define the following sets:

$$W_z(H) = \{\mathbf{z} \in [N]^{n-K-m} : \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) = H\},$$

$$W_y(G; H) = \{\mathbf{y} \in [N]^m : \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}, W_z(H)) = G\},$$

The number of H required is $N^{D_{\mathfrak{g}_{\mathfrak{H}}}}$. The image of $[N]^{n-K}$ under $\mathfrak{g}_{\mathfrak{G}}$ is $O(N^{D_{\mathfrak{g}_{\mathfrak{G}}}})$, and this is an upper bound of the number of G's for any fixed H, where the implied constant does not depend on H.

For any choice of $\mathbf{z} \in W_z(H)$ and $\mathbf{y} \in W_y(G; H)$, the polynomials $\mathfrak{p}$ now take the shape

$$\begin{aligned}
\mathfrak{p}(x_1, ..., x_K, \mathbf{y}, \mathbf{z}) &= \mathfrak{p}(x_1, ..., x_K, 0, ..., 0) + \sum_{j=1}^{d-1} \sum_{1 \leq i_1 < .... < i_j \leq K} \mathbf{c}_{i_1,...,i_j}^{(d)}(G, H) x_{i_1} ... x_{i_j} \\
&\quad + \sum_{\kappa=1}^{d-1} \sum_{1 \leq \iota_1 < .... < \iota_\kappa \leq m} \mathbf{c}_{0;i_1,...,i_j}^{(d)}(H) y_{\iota_1} ... y_{\iota_\kappa} \\
&\quad + \mathfrak{p}_1(0, ..., 0, \mathbf{y}, 0) + \mathfrak{p}_2(0, ..., 0, 0, \mathbf{z}) \\
&:= \mathfrak{P}_0(x_1, ..., x_K, G, H) + \mathfrak{P}_1(\mathbf{y}, H) + \mathfrak{p}_2(\mathbf{z}),
\end{aligned} \tag{4.5.4}$$

which are diagonal.

Define the exponential sums

$$S_0(\alpha, G, H) = \sum_{x_1,...,x_K \in [N]} \Lambda(x_1) ... \Lambda(x_K) e(\alpha \cdot \mathfrak{P}_0(x_1, ..., x_k, G, H)).$$

$$S_1(\alpha, G, H) = \sum_{\mathbf{y}\in W_y(G;H)} \Lambda(\mathbf{y})e(\alpha \cdot \mathfrak{P}_1(\mathbf{y}, H)).$$

$$S_2(\alpha, H) = \sum_{\mathbf{z}\in W_z(H)} \Lambda(\mathbf{z})e(\alpha \cdot \mathfrak{p}_2(\mathbf{z})).$$

Now we have to bound the expression

$$E_C(N) = \sum_H \sum_G \int_{\mathfrak{m}(C)} S_0(\alpha, G, H)S_1(\alpha, G, H)S_2(\alpha, H)e(-\mathbf{s}\cdot\alpha)d\alpha.$$

Proceeding as in Section 4.2, we obtain

$$(E_C(N))^2 \le O\left((\log\ N)^{2n} N^{D_{\mathfrak{g}_{\mathfrak{G}}}+D_{\mathfrak{g}_{\mathfrak{H}}}} \sup_{H,G} ||S_0(\cdot, G, H)||^2_{\infty(\mathfrak{m}(C))}\right) \sum_H \sum_G ||S_1(\cdot, G, H)||_2^2 ||S_2(\cdot, H)||_2^2. \tag{4.5.5}$$

The summands on the right hand side can be expressed as the number of solutions of $\mathfrak{P}_1(\mathbf{y}, H) = \mathfrak{P}_1(\mathbf{y}', H)$ for $\mathbf{y}, \mathbf{y}' \in W_y(G; H)$ times the number of solutions to $\mathfrak{p}_2(\mathbf{z}) = \mathfrak{p}_2(\mathbf{z}')$ for $\mathbf{z}, \mathbf{z}' \in W_z(H)$. The conditions $\mathbf{z}, \mathbf{z}' \in W_1(H)$ may be replaced by the conditions $\mathbf{z}, \mathbf{z}' \in [N]^{(n-K-m)}$ and $\mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) = \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}') = H$. The conditions $\mathbf{y}, \mathbf{y}' \in W_2(G; H)$ may be replaced by the conditions $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $\mathfrak{g}_{\mathfrak{G}}(\mathbf{y}, H) = \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}', H) = G$.

In short, we are summing over all G and H the number of solutions to the system

$$\begin{aligned} \mathfrak{P}_1(\mathbf{y}, H) &= \mathfrak{P}_1(\mathbf{y}', H) \\ \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}, H) &= \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}', H) = G \\ \mathfrak{P}_2(\mathbf{z}) &= \mathfrak{P}_2(\mathbf{z}') \\ \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) &= \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}') = H \end{aligned}$$

for $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $\mathbf{z}, \mathbf{z}' \in [N]^{(n-K-m)}$. With a little rearrangement this becomes

$$\begin{aligned} \mathfrak{P}_1(\mathbf{y}, \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) &= \mathfrak{P}_1(\mathbf{y}', \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) \\ \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}, \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) &= \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}', \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) = G \\ \mathfrak{P}_2(\mathbf{z}) &= \mathfrak{P}_2(\mathbf{z}') \\ \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) &= \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}') = H, \end{aligned}$$

and summing over G and H now simply removes the awkward looking equalities here. And after doing so, by removing the abuse in notation with $\mathfrak{g}_{\mathfrak{H}}$ as an argument puts us in the final form

$$\begin{aligned} \mathfrak{p}_1(\mathbf{y}) + \mathfrak{p}_3(\mathbf{y}, \mathbf{z}) &= \mathfrak{p}_1(\mathbf{y}') + \mathfrak{p}_3(\mathbf{y}', \mathbf{z}) \\ \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z}) &= \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}', \mathbf{z}) \\ \mathfrak{p}_2(\mathbf{z}) &= \mathfrak{p}_2(\mathbf{z}') \\ \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) &= \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}'), \end{aligned} \tag{4.5.6}$$

for $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $\mathbf{z}, \mathbf{z}' \in [N]^{(n-K-m)}$.
Let us call the number of solutions to the system (4.5.6) $\mathcal{W}$. Then (4.5.5) takes the form

$$(E_C(N))^2 \le O\left((\log\ N)^{2n} \mathcal{W} N^{D_{\mathfrak{g}_{\mathfrak{G}}}+D_{\mathfrak{g}_{\mathfrak{H}}}} \sup_{H,G} ||S_0(\cdot, G, H)||^2_{\infty(\mathfrak{m}(C))}\right). \tag{4.5.7}$$

The result is the lemma is immediate from the following two claims.

Claim 1: *Given $c > 0$, there is a C such that the bound*

$$||S_0(\cdot, G, H)||_{\infty(\mathfrak{m}(C))} = O((\log N)^{-c} N^K) \tag{4.5.8}$$

holds uniformly in G and H.

Claim 2: *With the rank of $\mathfrak{p}_1(\mathbf{y}) + \mathfrak{p}_3(\mathbf{y}, \mathbf{z})$ and $\mathfrak{p}_2(\mathbf{z})$ sufficiently large, the bound*

$$\mathcal{W} = O(N^{2(n-K)-2D_{\mathfrak{p}}-D_{\mathfrak{g}_{\mathfrak{H}}}-D_{\mathfrak{g}_{\mathfrak{G}}}}) \tag{4.5.9}$$

holds.

Let us start with claim 1. We look at $\alpha \cdot \mathfrak{P}_0(x_1, ..., x_K, G, H)$, focusing on the coefficients of terms of the form $x_{i_1}...x_{i_d}$ for $1 \leq i_j \leq K$. From our choice of $x_1, ..., x_K$, there is a collection of indices, say $(i_1^\kappa, ..., i_d^\kappa)$ for each $1 \leq \kappa \leq d$, such that the collection $\{\mathbf{b}_{i_1^\kappa, ..., i_d^\kappa}\}$ is linearly independent. Let M denote the $d \times d$ matrix of these coefficient vectors as rows. The coefficient of $x_{i_1^\kappa}...x_{i_d^\kappa}$ in $\alpha \cdot \mathfrak{P}_0(x_1, ..., x_K, G, H)$ is $(M\alpha)_\kappa$. Because M is linearly independent, there is some term of the form $x_{i_1^\kappa}...x_{i_d^\kappa}$ with a coefficient β where $\beta \in \mathfrak{m}(C')$ for some slightly larger C'.

If it happens to be the case that the indices $i_1^\kappa, ..., i_d^\kappa$ are equal, say all 1, then the bound follows directly from the bound in Lemma 4.5 for the x_1 summation, and claim 1 follows by treating the other sums trivially. Otherwise we assume that $x_{i_1^\kappa}...x_{i_d^\kappa} = x_1^{\gamma_1}..., x_l^{\gamma_l}$ where $\sum \gamma_i = d$ and $l < d$. Now look at the sum S_0 in the form

$$\sum_{x_{l+1},...,x_K \in [N]} \Lambda(x_{l+1})...\Lambda(x_K) \sum_{x_1,...,x_l \in [N]} \Lambda(x_1)...\Lambda(x_l) e(\beta x_1^{\gamma_1}...x_l^{\gamma_l} + Q(x_1, ..., x_K, G, H))$$

where $Q(x_1, ..., x_K, G, H)$ is viewed as a polynomial in $x_1, ..., x_l$ of degree less than d with coefficients in the other x_i and the G and the H. Apply the Cauchy-Schwarz inequality γ_i times to the inner sum for each of the variables x_i gives the upper bound

$$(\log N)^{d2^d} N^{2^d l-d-l} \sum_{x_1,...,x_l} \sum_{w_1^1,...,w_{\gamma_1}^1} ... \sum_{w_1^l...w_{\gamma_l}^1} \left(\prod_{i=1}^{l} \Delta_{w_1^i,...,w_{\gamma_l}^i} \mathbf{1}_{x_i \in [N]} \right) e(\beta w_1^1...w_{\gamma_l}^l),$$

where $\Delta_w f(x) = f(x+w)f(x)$ is the multiplicative differencing operator, and $\Delta_{w_1,w_2} = \Delta_{w_2}(\Delta_{w_1})$, and so on. The logarithmic gain on this latter sum now follows form the Weyl method as $\beta \in \mathfrak{m}(C')$ for C' large enough in terms of $d2^d c$, and then claim 1 follows by taking the 2^dth root and summing trivially in $x_{l+1}, ..., x_K$.

To get the bound on $\mathcal{W}$, note that the number of solutions to (5.6) is the product of the number of solutions to

$$\begin{aligned} \mathfrak{p}_1(\mathbf{y}) + \mathfrak{p}_3(\mathbf{y}, \mathbf{z}) &= \mathfrak{p}_1(\mathbf{y}') + \mathfrak{p}_3(\mathbf{y}', \mathbf{z}) \\ \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z}) &= \mathfrak{g}_{\mathfrak{G}}(\mathbf{y}', \mathbf{z}) \end{aligned}$$

with the number to

$$\begin{aligned} \mathfrak{p}_2(\mathbf{z}) &= \mathfrak{p}_2(\mathbf{z}') \\ \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) &= \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}'), \end{aligned}$$

which follows as the two systems can not possibly have any terms that appear in common. As it is the case that all polynomials of $\mathfrak{p}$ have the same degree we see that the $\mathfrak{p}_1 + \mathfrak{p}_3$ and $\mathfrak{g}_{\mathfrak{G}}$ can be treated separately because ranks have been defined independently on each degree. The same goes for $\mathfrak{p}_2$ and $\mathfrak{g}_{\mathfrak{H}}$, and so the result of Schmidt finishes the proof. $\square$

4.6 Major arcs asymptotics

The major arcs are a union of boxes of the form $\mathfrak{M}_{\mathbf{a},q}(C)$, where $q \leq (\log\, N)^C$ and C is now a fixed constant chosen large enough so that lemma 4.4 holds with $c = 1$. For a fixed $\mathbf{a}$ and q, the small size of the associated major arc means that the exponential sum

$$T_{\mathfrak{p}}(\alpha) = \sum_{\mathbf{x} \in [N]^n} \Lambda(\mathbf{x}) e(\mathfrak{p}(\mathbf{x}) \cdot \alpha)$$

can be replaced by any approximation that has a sufficiently large logarithmic power gain in the error.
Upon the actual fixing of a $q \leq (\log\, N)^C$ and an $\mathbf{a} \in U_q^n$, one has[2]

$$\begin{aligned} T_{\mathfrak{p}}(\alpha) &= \sum_{\mathbf{x} \in [N]^n} \Lambda(\mathbf{x}) e(\mathfrak{p}(\mathbf{x}) \cdot \alpha) \\ &= \sum_{\mathbf{g} \in Z_q^n} \sum_{\mathbf{x} \in [N]^n} \mathbf{1}_{\mathbf{x} \equiv \mathbf{g}\,(q)} \Lambda(\mathbf{x}) e(\mathbf{a} \cdot \mathfrak{p}(\mathbf{g})/q) e(\mathfrak{p}(\mathbf{x}) \cdot \tau) \qquad (4.6.1) \\ &= \sum_{\mathbf{g} \in Z_q^n} e(\mathfrak{p}(\mathbf{g}) \cdot \mathbf{a}/q) \int_{\mathbf{X} \in N\mathcal{J}} e(\mathfrak{p}(\mathbf{X}) \cdot \tau) d\psi_{\mathbf{g}}(\mathbf{X}), \end{aligned}$$

where the notations introduced here are $\tau_i = \alpha_i - a_i/q$, and $\psi_{\mathbf{g}}(\mathbf{X}) = \psi_{g_1}(X_1)...\psi_{g_n}(X_n)$ with

$$\psi_l(v) = \sum_{t \leq v,\, t \equiv l\,(q)} \Lambda(t),$$

and $\mathcal{J}$ is the unit cube $[0,1]^n \subset \mathbb{R}^n$.

Lemma 4.6. *For any given a constant c, the estimate*

$$\int_{\mathbf{X} \in N\mathcal{J}} e(\mathfrak{p}(\mathbf{X}) \cdot \tau) d\psi_{\mathbf{g}}(\mathbf{X}) = \mathbf{1}_{\mathbf{g} \in U_q^n} \phi(q)^{-n} \int_{\mathbf{z} \in N\mathcal{J}} e(\mathfrak{p}(\mathbf{z}) \cdot \tau) d\mathbf{z} + O((\log\, N)^{-c} N^n), \qquad (4.6.2)$$

holds on each major arc $\mathfrak{M}_{a,q}(C)$ provided that C is sufficiently large.

Proof. Define for a fixed l the one dimensional signed measure $d\nu_l = d\psi_l - d\omega_l$, where $d\omega_l$ is the Lebesgue measure multipled by the reciprocal of the totient of q if $l \in U_q$, and is zero otherwise. For a continuous function f one then has

$$\int_0^N f(X) d\nu_l(X) = \sum_{x \in [N],\, x \equiv l\,(q)} \Lambda(x) f(x) - \phi(q)^{-1} \int_0^N f(z) dz.$$

Now set $d|\nu_l| = d\omega_l + d\psi_l$, so that

$$\int_{\mathbf{X} \in N\mathcal{J}} e(\mathfrak{p}(\mathbf{X}) \cdot \tau) d\psi_{\mathbf{g}}(\mathbf{X}) = \int_{\mathbf{X} \in N\mathcal{J}} e(\mathfrak{p}(\mathbf{X}) \cdot \tau) \prod_{i=1}^n \left(d\nu_{g_i}(X_i) + d\omega_{g_i}(X_i) \right).$$

[2]There is some ambiguity in the case where N is a prime power, however, there is no harm in assuming that this is not so due to the fact that the prime powers are sparse.

Expanding out the product in the last integral gives the form

$$\int_{\mathbf{X}\in N\mathcal{J}} e(\mathfrak{p}(\mathbf{X})\cdot\tau)d\omega_{\mathbf{g}}(\mathbf{X}) + \sum_{i=1}^{2^n-1}\int_{\mathbf{X}\in N\mathcal{J}} e(\mathfrak{p}(\mathbf{X})\cdot\tau)d\mu_{i,\mathbf{g}}(\mathbf{X}),$$

where $d\mu_{i,\mathbf{g}}$ runs over all of the corresponding product measures, barring the $d\omega_{\mathbf{g}}(\mathbf{X})$ term. Consider

$$\int_{\mathbf{X}\in N\mathcal{J}} e(\mathfrak{p}(\mathbf{X})\cdot\tau)d\mu_{i,\mathbf{g}}(\mathbf{X}),$$

for some fixed i. Assume without loss of generality that $d\mu_{i,\mathbf{y}}$ is of the form

$$d\nu_{g_1}(X_1)d\sigma_{\mathbf{g}}(X_2,...,X_n),$$

where $d\sigma_{\mathbf{g}}$ may be signed in some variables and is independent of g_1. The range of integration for the X_1 variable is a copy of the continuous interval $[0,N]$, and is to be split into smaller disjoint intervals of size $N^1(\log N)^{-c'}$. Here c' is chosen to be between $(c+C)$ and $2(c+C)$ such that $(\log N)^{c'}$ is an integer, say B. The equality $[0,N]=\bigcup_{j=1}^{B} I_j$ follows. Also set $\mathcal{J}_j'=I_j\times[0,N]^{n-1}$, which absorbs the factor of N.

Now, for a fixed interval I_j, select some $t\in I_j$. Then write

$$\begin{aligned}\int_{\mathbf{X}\in\mathcal{J}_j'} e(\mathfrak{p}(\mathbf{X})\cdot\tau)d\mu_{i,\mathbf{g}} \quad &= \quad \int_{\mathbf{X}\in\mathcal{J}_j'} e(\mathfrak{p}(t,X_2,...,X_n)\cdot\tau)d\nu_{g_1}(X_1)d\sigma_{\mathbf{g}}(X_2,...,X_n)\\ &\quad +\int_{\mathbf{X}\in\mathcal{B}_|}(e(\mathfrak{p}(X_1,...,X_n)\cdot\tau)-e(\mathfrak{p}(t,X_2,...,X_n)\cdot\tau))\\ &\quad \times d\nu_{g_1}(X_1)d\sigma_{\mathbf{g}}(X_2,...,X_n)\\ &:= \quad E_1+E_2\end{aligned}$$

The first error term satisfies

$$|E_1|\le\int_{X_2,...,X_n\in[0,N]}|\int_{I_j} d\nu_{g_1}(X_1)|\ d|\sigma_{\mathbf{g}}|(X_2,...,X_n)=O(N^ne^{-c_0\sqrt{\log N}})$$

for some positive constant c_0 by the Siegel-Walfisz theorem, as $q\le(\log N)^C$. To bound E_2, note that on I_j the integrand is

$$O(|\mathfrak{p}(X_1,...,X_n)-\mathfrak{p}(t,X_2,...,X_n))\cdot\tau|)=O((\log N)^{C-c'}).$$

In turn,

$$|E_2|=O((\log N)^{C-c'}))\int_{\mathbf{X}\in\mathcal{J}_j'} d|\nu_{g_1}|(X_1)d|\sigma_{\mathbf{g}}|(X_2,...,X_n)=O(N^n(\log N)^{C-2c'})).$$

There are 2^n-1 error terms on each interval, so summing over the $(\log N)^{c'}$ intervals completes the proof. □

The integral appearing in the last result has a quick reduction, namely

$$\int_{N\mathcal{J}} e(\mathfrak{p}(\mathbf{X})\tau)d\mathbf{X}=\int_{N\mathcal{J}} e(\mathfrak{f}(\mathbf{X})\tau)d\mathbf{X}+O(N^{n-1+\epsilon}),$$

recalling that $\mathfrak{f}$ is the highest degree part of $\mathfrak{p}$. Following along with the work of Birch,

$$\int_{N\mathcal{J}} e(\mathfrak{f}(\mathbf{X})\tau)d\mathbf{X} = N^n \int_{\zeta\in\mathcal{J}} e(\mathfrak{f}(\zeta)\cdot N^2\tau)d\zeta,$$

is denoted by $N^n\mathcal{I}(\mathcal{J}, N^2\tau)$ in [11]. This function is independent of $\mathbf{a}$ and q. Thus the integral over any major arc yields the common integral

$$\int_{|\tau|\le(\log\, N)^C} \mathcal{I}(\mathcal{J}, N^2\tau)e(-\mathbf{s}\cdot\tau)d\tau.$$

With $\mu = N^{-2}\mathbf{s}$, set

$$J(\mu;\Phi) = \int_{|\tau|\le\Phi} \mathcal{I}(\mathcal{J},\tau)e(-\mu\cdot\tau)d\tau,$$

and

$$J(\mu) = \lim_{\Phi\to\infty} J(\mu).$$

The following is Lemma 5.3 in [11].

Lemma 4.7. *The function $J(\mu)$ is continuous and uniformly bounded in μ. Moreover,*

$$|J(\mu) - J(\mu,\Phi)| \lesssim \Phi^{-\frac{1}{2}}$$

holds uniformly in μ.

By defining

$$W_{\mathbf{a},q} = \sum_{\mathbf{g}\in U_q^n} e(\mathfrak{p}(\mathbf{g})\cdot\mathbf{a}/q),$$

one has

Lemma 4.8. *For any given $c > 0$, the estimate*

$$\int_{\mathfrak{M}_{\mathbf{a},q}(C)} T_{\mathfrak{p}}(\alpha)e(-\mathbf{s}\cdot\alpha)d\alpha = N^{n-dr}\phi(q)^{-n}W_{\mathbf{a},q}e(-\mathbf{s}\cdot\mathbf{a}/q)J(\mu) + O(N^{n-dr}(\log\, N)^{-c}),$$

where $\mu = N^{-2}\mathbf{s}$, holds on each major arc $\mathfrak{M}_{\mathbf{a},q}(C)$.

The measure of the major arcs is easily at most $N^{-dr}(\log\, N)^K$ for some constant K. By defining

$$B(\mathbf{s},q) = \sum_{\mathbf{a}\in U_q^r} \phi(q)^{-n}W_{\mathbf{a},q}e(-\mathbf{s}\cdot\mathbf{a}/q)$$
$$\mathfrak{S}(\mathbf{s},N) = \sum_{q\le(\log\, N)^C} B(\mathbf{s},q),$$

it then follows that

Lemma 4.9. *The estimate*

$$\mathcal{M}_{\mathfrak{p},\mathbf{s}}(N) = \mathfrak{S}(\mathbf{s},N)J(\mu)N^{n-dr} + O((\log\, N)^{-c}N^{n-dr}) \tag{4.6.3}$$

holds for any chosen value of c.

4.7 The singular series

Following the outline of Hua ([57], chapter VIII, §2, Lemma 8.1), one can show that $B(\mathbf{s}, q)$ is multiplicative as a function of q. This leads to consideration of the formal identity

$$\mathfrak{S}(\mathbf{s}) := \lim_{N\to\infty} \mathfrak{S}(\mathbf{s}, N) = \prod_{p<\infty} (1 + \sum_{t=1}^{\infty} B(\mathbf{s}, p^t))). \tag{4.7.1}$$

Lemma 4.10. *If $q = p^t$ is a prime power, then*

$$B(\mathbf{s}, q) = O(q^{r-\mathcal{B}(\mathfrak{p})/((d-1)2^d r)+\epsilon}) \tag{4.7.2}$$

holds uniformly in $\mathbf{s}$ as $t \to \infty$. The implied constants can be made independent of p.

Proof. It is shown here that

$$W_{\mathbf{a},q} = O(q^{n-\mathcal{B}(\mathfrak{p})/((d-1)2^d r)+\epsilon}),$$

uniformly for $\mathbf{a} \in U_q^n$, which clearly implies the result by the definition of $B(\mathbf{s}, q)$ and the fact that $q^n/\phi(q)^n \leq 2^n$ independent of p.
The inclusion-exclusion principle is used to bound $W_{\mathbf{a},q}$ when $q = p^t$ when $t \leq d$. Let such a t be fixed, and note that the characteristic function of U_{p^t} decomposes as

$$\mathbf{1}_{U_{p^t}}(x) = 1 - \sum_{h\in Z_{p^{t-1}}} \mathbf{1}_{x=hp}.$$

Applying this in the definition gives

$$\begin{aligned} W_{\mathbf{a},q} &= \sum_{\mathbf{g}\in U_q^n} e(\mathfrak{p}(\mathbf{g})\cdot \mathbf{a}/q) \\ &= \sum_{\mathbf{g}\in Z_q^n} \prod_{i=1}^{n} (1 - \sum_{h_i\in Z_{p^{t-1}}} \mathbf{1}_{g_i=h_i p}) e(\mathfrak{p}(\mathbf{g})\cdot \mathbf{a}/q) \\ &= \sum_{I\subseteq[n]} (-1)^{|I|} \sum_{\mathbf{h}\in Z_{p^{t-1}}^{|I|}} \sum_{\mathbf{g}\in Z_q^n} F_I(\mathbf{g};\mathbf{h}) e(\mathfrak{p}(\mathbf{g})\cdot \mathbf{a}/q), \end{aligned} \tag{4.7.3}$$

where

$$F_I(\mathbf{g};\mathbf{h}) = \prod_{i\in I} \mathbf{1}_{g_i=ph_i}$$

for $\mathbf{h} \in Z_p^{|I|}$. In other words, F_I is the characteristic function of the set $H_{I,\mathbf{h}} = \{\mathbf{g} : g_i = ph_i \,\forall\, i \in I\}$. The sets $I \subseteq [n]$ divided into two categories according to whether $|I| \leq \mathcal{B}(\mathfrak{p})/(r+1)$ or not. If I is a set fitting into the latter category, then the trivial estimate is

$$\left| \sum_{\mathbf{h}\in Z_{p^{t-1}}^{|I|}} \sum_{\mathbf{g}\in Z_q^n} F_I(\mathbf{g};\mathbf{h}) e(\mathfrak{p}(\mathbf{g})\cdot \mathbf{a}/q) \right| = p^{(t-1)|I|}(p^t)^{n-|I|} = (p^t)^{n-|I|/t} \leq q^{n-\mathcal{B}(\mathfrak{p})/(tr+t)} \leq q^{n-\mathcal{B}(\mathfrak{p})/((d-1)2^d r)}.$$

Now let I be a fixed subset of $[n]$ with $|I| \leq \mathcal{B}(\mathfrak{p})/(r+1)$. For each $\mathbf{h}$ the restriction of $\mathfrak{p}$ to the set $H_{I,\mathbf{h}}$ has Birch rank at least $\mathcal{B}(\mathfrak{p}) - |I|(r+1)$ by corollary 4.2. By the work of Birch ([11], Lemma

5.4) it follows that

$$\sum_{\mathbf{h}\in Z_p^{(t-1)|I|}} \left| \sum_{\mathbf{g}\in Z_q^n} F_I(\mathbf{g};\mathbf{h}) e(\mathfrak{p}(\mathbf{g})\cdot \mathbf{a}/q) \right| \lesssim q^{(t-1)|I|/t} q^{n-|I|-(\mathcal{B}(\mathfrak{p})-|I|(r+1))/((d-1)2^{d-1}r)+\epsilon}$$
$$\lesssim q^{n-\mathcal{B}(\mathfrak{p})/((d-1)2^d r)+\epsilon}.$$

Summing over all I yields the bound.
Now let $q=p^t$ for $t>d$. Going back to the definition gives

$$\begin{aligned} W_{\mathbf{a},q} &= \sum_{\mathbf{g}\in U_q^n} e(\mathfrak{p}(\mathbf{g})\cdot \mathbf{a}/q) \\ &= \sum_{\mathbf{g}\in U_p^n} \sum_{\mathbf{h}\in Z_{p^{t-1}}^n} e(\mathfrak{p}(\mathbf{g}+p\mathbf{h})\cdot \mathbf{a}/q). \end{aligned}$$

The system of forms in the exponent can be expanded for each fixed $\mathbf{g}$ as

$$\mathfrak{p}(\mathbf{g}+p\mathbf{h}) = p^d\mathfrak{p}(\mathbf{h}) + f_{\mathbf{g}}(\mathbf{h})$$

for some polynomial $f_{\mathbf{g}}$ of degree at most $d-1$. Then it follows that

$$|W_{\mathbf{a},q}| \leq \sum_{\mathbf{g}\in U_p^n} \left| \sum_{\mathbf{h}\in Z_{p^{t-1}}^n} e(f_{\mathbf{g}}(\mathbf{h}) + p^d\mathfrak{p}(\mathbf{h}))\cdot \mathbf{a}/q) \right|. \tag{4.7.4}$$

The inner sum is now bounded uniformly in $\mathbf{g}$ by an application of the exponential sum estimates in [11] as follows.
Set $P=p^{t-1}$ and $q_1=p^{t-d}$. Then, for each $i=1,...,r$,

$$2|q'a_i - a_i'q_1| \leq P^{-(d-1)+(d-1)r\theta}$$

and

$$1\leq q' \leq P^{(d-1)r\theta}$$

cannot be satisfied if $\theta < 1/(d-1)r$. Then, by Lemma 4.3 of [11],

$$\sum_{\mathbf{h}\in Z_{p^{t-1}}^n} e((p^d\mathfrak{p}(\mathbf{h}))\cdot \mathbf{a}/q + f(\mathbf{h})) = O(P^{n-\mathcal{B}(\mathfrak{p})/((d-1)2^d r)+\epsilon})$$

for any polynomial $f(\mathbf{h})$ of degree strictly less than d. In turn,

$$|W_{\mathbf{a},q}| \leq \sum_{\mathbf{g}\in U_p^n} O(P^{n-B(\mathfrak{p})/((d-1)2^d r)+\epsilon}) = O(q^{n-B(\mathfrak{p})/((d-1)2^d r)+\epsilon}),$$

which is what is needed to complete the proof in this last and final case. □

Now define the local factor for a finite prime p as

$$\mu_p = 1 + \sum_{t=1}^{\infty} B(\mathbf{s},p^t)), \tag{4.7.5}$$

which is well defined as the series is absolutely convergent provided that $\mathcal{B}(\mathfrak{p}) > (d-1)2^d r(r+1)$.
The following result is again an straight forward extension of the results for a single form.

Lemma 4.11. *For each finite prime p, the local factor may be represented as*

$$\mu_p = \lim_{t\to\infty} \frac{(p^t)^R M(p^t)}{\phi^n(p^t)}, \tag{4.7.6}$$

where $M(p^t)$ represents the number of solutions to equation 4.2.1 in the multiplicative group U_{p^t}.

At our disposal now is the fact that the μ_p are positive, which then easily gives the following.

Lemma 4.12. *If $\mathcal{B}(\mathfrak{p}) > (d-1)2^d r(r+1)$ then the local factor for each finite prime satisfies the estimate*

$$\mu_p = 1 + O(p^{-(1+\delta)})$$

for some positive δ, and therefore the product in equation 4.7.1 is absolutely convergent and thusly is in fact well defined.

The observation that

$$|\mathfrak{S}(\mathbf{s}, N) - \mathfrak{S}(\mathbf{s})| = o(1)$$

gives the final form of the asymptotic for $\mathcal{M}_{\mathcal{Q},\mathbf{s}}$.

Theorem 4.2. *The estimate*

$$\mathcal{M}_{\mathfrak{p},\mathbf{s}}(N) = \mathfrak{S}(\mathbf{s})J(\mu)N^{n-dr} + O((\log N)^{-c}N^{n-dr}) \tag{4.7.7}$$

holds for any chosen value of $c > 0$, with the implicit constant in the error term depending on c.

5 Polynomial ergodic theorems for nilpotent group actions

In this section we discuss some of our recent joint work [58] on averages along polynomial sequences in discrete nilpotent groups of step 2. Our main results include boundedness of associated maximal functions and singular integrals operators, an almost everywhere pointwise convergence theorem for ergodic averages along polynomial sequences, the first instance where Bourgain's polynomial ergodic theorem [15, 13, 14] is extended to the non-commutative settings. The last section is about obtaining an asymptotic formula for the umber of solutions to a diophantine systems which is a natural extension of the so-called Waring-Vinogradov system of equations to step-2 nilpotent groups.

Our proofs are based on analytical, number theoretic tools such as a nilpotent Weyl inequality that we obtained earlier in [61], and on complex almost-orthogonality arguments that are designed to replace Fourier transform tools which are not available in the non-commutative nilpotent setting. In particular, we present what we call a *nilpotent circle method* that allows us to adapt some of the ideas of the classical circle method to the setting of nilpotent groups. For the sake of readability and limitations of space (the paper [58] is 118 pages) we do not include all the technical details, however we include the major theorems and lemmas and sketch the crucial ideas of the proof. Our presentation follows [59].

5.1 The Furstenberg–Bergelson–Leibman conjecture

Discrete averages, both of the maximal and singular type, have been considered motivated mainly by open problems in ergodic theory. A fundamental problem in ergodic theory is to establish convergence in norm and pointwise almost everywhere for the polynomial ergodic averages as in

(5.1.2) as $N \to \infty$ for functions $f \in L^p(X)$, $1 \le p \le \infty$. The problem goes back to at least the early 1930's with von Neumann's mean ergodic theorem [112] and Birkhoff's pointwise ergodic theorem [12] and led to profound extensions such as Bourgain's polynomial pointwise ergodic theorem [13, 14, 15] and Furstenberg's ergodic proof [42] of Szemerédi's theorem [105] in particular. Furstenberg's proof was also the starting point of ergodic Ramsey theory, which resulted in many natural generalizations of Szemerédi's theorem, including a polynomial Szemerédi theorem of Bergelson and Leibman [9]. This motivates the following far reaching conjecture known as the Furstenberg–Bergelson–Leibman conjecture [10, Section 5.5, p. 468].

Conjeture 5.1. *Assume that $d, k \ge 1$ are integers, $(X, \mathcal{B}(X), \mu)$ is a probability space, and assume that $T_1, \ldots, T_d : X \to X$ is a given family of invertible measure-preserving transformations on the space $(X, \mathcal{B}(X), \mu)$ that generates a nilpotent group of step k. Assume that $m \ge 1$ is an integer and $P_{1,1}, \ldots, P_{i,j}, \ldots, P_{d,m} : \mathbb{Z} \to \mathbb{Z}$ are polynomial maps with integer coefficients such that $P_{i,j}(0) = 0$. Then for any $f_1, \ldots, f_m \in L^\infty(X)$, the non-conventional multilinear polynomial averages*

$$A_{N;X,T_1,\ldots,T_d}^{P_{1,1},\ldots,P_{d,m}}(f_1, \ldots, f_m)(x) = \frac{1}{2N+1} \sum_{n \in [-N,N] \cap \mathbb{Z}} \prod_{j=1}^{m} f_j(T_1^{P_{1,j}(n)} \cdots T_d^{P_{d,j}(n)} x) \tag{5.1.1}$$

converge for μ-almost every $x \in X$ as $N \to \infty$.

Conjecture 5.1 is a major open problem in ergodic theory that was promoted in person by Furstenberg, see [2, p. 6662], before being published in [10]. Our main result Theorem 5.1 (ii) proves this conjecture in the linear case $m = 1$, provided that the family of transformations $T_1, \ldots, T_d : X \to X$ generates a nilpotent group of step $k = 2$. We call a sequence $A : \mathbb{Z} \to \mathbb{G}$ a polynomial sequence if $D^k A(n) \equiv 1$ for some $k \in \mathbb{N}$, where $DA(n) := A(n)^{-1} A(n+1)$ is the multiplicative differencing operator. It can be shown that this is equivalent of writing $A(n) = g_1^{P_1(n)} \cdots g_d^{P_d(n)}$ for some $g_1, \ldots, g_d \in \mathbb{G}$ and integral polynomials $P_1, \ldots, P_d$. In particular, if $\mathbb{G}$ is a discrete nilpotent group generated by measure preserving transformations $T_1, \ldots, T_d$ then $A(n) = T_1^{P_1} \cdots T_d^{P_d(n)}$ is a polynomial sequence.

Theorem 5.1 (Main result)**.** *Assume that $\mathbb{G}$ is a discrete nilpotent group $\mathbb{G}$ of step 2 and $A : \mathbb{Z} \to \mathbb{G}$ is a polynomial sequence. Then:*
(i) (ℓ^p boundedness of maximal averages) Assume $f : \mathbb{G} \to \mathbb{C}$ is a function and let

$$\mathcal{M} f(g) := \sup_{N \ge 0} \frac{1}{2N+1} \sum_{|n| \le N} |f(A^{-1}(n) \cdot g)|, \qquad g \in \mathbb{G}.$$

Then, for any $p \in (1, \infty]$,

$$\|\mathcal{M} f\|_{\ell^p(\mathbb{G})} \lesssim_p \|f\|_{\ell^p(\mathbb{G})}.$$

(ii) (L^p pointwise ergodic theorems) Assume $\mathbb{G}$ acts by measure-preserving transformations on a σ-finite measure space X, $f \in L^p(X)$, $p \in (1, \infty)$, and let

$$A_N f(x) := \frac{1}{2N+1} \sum_{|n| \le N} f(A^{-1}(n) \cdot x), \qquad x \in X. \tag{5.1.2}$$

Then the sequence $A_N f$ converges pointwise almost everywhere and in the L^p norm as $N \to \infty$.

5.1.1 Earlier pointwise ergodic theorems

The basic linear case $m = d = k = 1$ with $P_{1,1}(n) = n$ follows from Birkhoff's original ergodic theorem [12]. On the other hand, the commutative case $m = d = k = 1$ with an arbitrary polynomial $P = P_{1,1}$ with integer coefficients was a famous open problem of Bellow [5] and Furstenberg [43], solved by Bourgain in his breakthrough papers [13, 14, 15].
Some particular examples of averages (5.1.1) with $m = 1$ and polynomial mappings with degree at most two in the step two nilpotent setting were studied in [60, 83].
The multilinear theory $m \geq 2$, in contrast to the linear theory, is widely open even in the commutative case $k = 1$. Only a few results in the bilinear $m = 2$ and commutative $d = k = 1$ setting are known. Bourgain [17] proved pointwise convergence when $P_{1,1}(n) = an$ and $P_{1,2}(n) = bn$, $a, b \in \mathbb{Z}$. More recently, Krause–Mirek–Tao [66] established pointwise convergence for the polynomial Furstenberg–Weiss averages [44, 45] corresponding to $P_{1,1}(n) = n$ and $P_{1,2}(n) = P(n)$, $\deg P \geq 2$.

5.1.2 Norm convergence

Except for these few cases, there are no other results concerning pointwise convergence for the averages (5.1.1). The situation is completely different, however, for the question of norm convergence, which is much better understood.
A breakthrough paper of Walsh [113] (see also [2]) gives a complete picture of $L^2(X)$ norm convergence of the averages (5.1.1) for any $T_1, \ldots, T_d \in \mathbb{G}$ where $\mathbb{G}$ is a nilpotent group of transformations of a probability space. Prior to this, there was an extensive body of research towards establishing $L^2(X)$ norm convergence, including groundbreaking works of Host–Kra [53], Ziegler [116], Bergelson [6], and Leibman [72]. See also [3, 25, 37, 54, 106] and the survey articles [7, 8, 36] for more details and references, including a comprehensive historical background.

5.1.3 Additional remarks

Bergelson–Leibman [10] showed that convergence may fail if the transformations $T_1, \ldots, T_d$ generate a solvable group, so the nilpotent setting is probably the appropriate setting for Conjecture 5.1. The restriction $p > 1$ is necessary in the case of nonlinear polynomials as was shown in [20, 71].
If $(X, \mathcal{B}(X), \mu)$ is a probability space and the family of measure preserving transformations $(T_1, \ldots, T_{d_1})$ is *totally ergodic*, then Theorem 5.1(ii) implies that

$$\lim_{N\to\infty} A_{N;X}^{P_1,\ldots,P_{d_1}}(f)(x) = \int_X f(y) d\mu(y) \tag{5.1.3}$$

μ-almost everywhere on X. We recall that a family of measure preserving transformations $(T_1, \ldots, T_{d_1})$ is called ergodic on X if $T_j^{-1}(B) = B$ for all $j \in \{1, \ldots, d_1\}$ implies $\mu(B) = 0$ or $\mu(B) = 1$ and is called totally ergodic if the family $(T_1^n, \ldots, T_{d_1}^n)$ is ergodic for all $n \in \mathbb{Z}_+$.

5.2 The universal step-two group $\mathbb{G}_0$

The proof of Theorem 5.1 will follow from our second main result, Theorem 5.2 below, for averages on universal nilpotent groups of step two. We start with some definitions. For integers $d \geq 1$, we define

$$Y_d := \{(l_1, l_2) \in \mathbb{Z} \times \mathbb{Z} : 0 \leq l_2 < l_1 \leq d\}$$

and the "universal" step-two nilpotent Lie groups $\mathbb{G}_0^{\#} = \mathbb{G}_0^{\#}(d)$

$$\mathbb{G}_0^{\#} := \{(x_{l_1 l_2})_{(l_1,l_2)\in Y_d} : x_{l_1 l_2} \in \mathbb{R}\}, \tag{5.2.1}$$

with the group multiplication law

$$[x \cdot y]_{l_1 l_2} := \begin{cases} x_{l_1 0} + y_{l_1 0} & \text{if } l_1 \in \{1, \dots, d\} \text{ and } l_2 = 0, \\ x_{l_1 l_2} + y_{l_1 l_2} + x_{l_1 0} y_{l_2 0} & \text{if } l_1 \in \{1, \dots, d\} \text{ and } l_2 \in \{1, \dots, l_1 - 1\}. \end{cases} \tag{5.2.2}$$

Alternatively, we can also define the group $\mathbb{G}_0^{\#}$ as the set of elements

$$g = (g^{(1)}, g^{(2)}), \qquad g^{(1)} = (g_{l_1 0})_{l_1 \in \{1, \dots, d\}} \in \mathbb{R}^d, \qquad g^{(2)} = (g_{l_1 l_2})_{(l_1, l_2) \in Y_d'} \in \mathbb{R}^{d'}, \tag{5.2.3}$$

where $d' := d(d-1)/2$ and $Y_d' := \{(l_1, l_2) \in Y_d : \, l_2 \geq 1\}$. Letting

$$R_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d'} \quad \text{ denote the bilinear form } \quad [R_0(x, y)]_{l_1 l_2} := x_{l_1 0} y_{l_2 0}, \tag{5.2.4}$$

we notice that the product rule in the group $\mathbb{G}_0^{\#}$ is given by

$$[g \cdot h]^{(1)} := g^{(1)} + h^{(1)}, \qquad [g \cdot h]^{(2)} := g^{(2)} + h^{(2)} + R_0(g^{(1)}, h^{(1)}) \tag{5.2.5}$$

if $g = (g^{(1)}, g^{(2)})$ and $h = (h^{(1)}, h^{(2)})$. For any $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^{\#}$, its inverse is given by

$$g^{-1} = \big(- g^{(1)}, -g^{(2)} + R_0(g^{(1)}, g^{(1)}) \big).$$

The second variable of $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^{\#}$ is called the central variable. Based on the product structure (5.2.5) of the group $\mathbb{G}_0^{\#}$, it is not difficult to see that $g \cdot h = h \cdot g$ for any $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^{\#}$ and $h = (0, h^{(2)}) \in \mathbb{G}_0^{\#}$.
Let $\mathbb{G}_0 = \mathbb{G}_0(d)$ denote the discrete subgroup

$$\mathbb{G}_0 := \mathbb{G}_0^{\#} \cap \mathbb{Z}^{|Y_d|}. \tag{5.2.6}$$

Let $A_0 : \mathbb{R} \to \mathbb{G}_0^{\#}$ denote the canonical polynomial map (or the moment curve on $\mathbb{G}_0^{\#}$)

$$[A_0(x)]_{l_1 l_2} := \begin{cases} x^{l_1} & \text{if } l_2 = 0, \\ 0 & \text{if } l_2 \neq 0, \end{cases} \tag{5.2.7}$$

and notice that $A_0(\mathbb{Z}) \subseteq \mathbb{G}_0$. For $x = (x_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{G}_0^{\#}$ and $\Lambda \in (0, \infty)$, we define

$$\Lambda \circ x := (\Lambda^{l_1 + l_2} x_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{G}_0^{\#}. \tag{5.2.8}$$

Notice that the dilations $\Lambda \circ$ are group homomorphisms on the group $\mathbb{G}_0$ that are compatible with the map A_0, i.e. $\Lambda \circ A_0(x) = A_0(\Lambda x)$.
Let $\chi : \mathbb{R} \to [0, 1]$ be a smooth function supported on the interval $[-2, 2]$. Given any real number $N \geq 1$ and a function $f : \mathbb{G}_0 \to \mathbb{C}$, we can define a smoothed average along the moment curve A_0 by the formula

$$M_N^{\chi}(f)(x) := \sum_{n \in \mathbb{Z}} N^{-1} \chi(N^{-1} n) f(A_0(n)^{-1} \cdot x), \qquad x \in \mathbb{G}_0. \tag{5.2.9}$$

The main advantage of working on the group $\mathbb{G}_0$ with the polynomial map A_0 is the presence of the compatible dilations $\Lambda \circ$ defined in (5.2.8), which lead to a natural family of associated balls. This can be efficiently exploited by noting that M_N^{χ} is a convolution operator on $\mathbb{G}_0$.

The convolution of functions on the group $\mathbb{G}_0$ is defined by the formula

$$(f * g)(x) := \sum_{y\in\mathbb{G}_0} f(y^{-1}\cdot x)g(y) = \sum_{z\in\mathbb{G}_0} f(z)g(x\cdot z^{-1}). \tag{5.2.10}$$

Then it is not difficult to see that $M_N^\chi(f)(x) = f * G_N^\chi(x)$, where

$$G_N^\chi(x) := \sum_{n\in\mathbb{Z}} N^{-1}\chi(N^{-1}n)\mathbf{1}_{\{A_0(n)\}}(x), \qquad x\in\mathbb{G}_0. \tag{5.2.11}$$

We are now ready to state our second main result.

Theorem 5.2 (Boundedness on $\mathbb{G}_0$)**.** *Let $\mathbb{G}_0 = \mathbb{G}_0(d)$, $d\geq 1$, be the discrete nilpotent group defined in* (5.2.6) *and A_0 the polynomial sequence defined in* (5.2.7)*. Then*

(i) (Maximal estimates) If $1 < p \leq \infty$ and $f\in \ell^p(\mathbb{G}_0)$ then

$$\big\| \sup_{N\geq 1} |M_N^\chi(f)|\big\|_{\ell^p(\mathbb{G}_0)} \lesssim_p \|f\|_{\ell^p(\mathbb{G}_0)}, \tag{5.2.12}$$

where M_N^χ is defined as in (5.2.9)*.*

(ii) (Long variational estimates) If $1 < p < \infty$ and $\rho > \max\{p, \frac{p}{p-1}\}$, and $\tau\in(1,2]$ then

$$\big\|V^\rho\big(M_N^\chi(f) : N\in\mathbb{D}_\tau\big)\big\|_{\ell^p(\mathbb{G}_0)} \lesssim_{p,\rho,\tau} \|f\|_{\ell^p(\mathbb{G}_0)}, \tag{5.2.13}$$

where $\mathbb{D}_\tau := \{\tau^n : n\in\mathbb{N}\}$. See (5.3.1) *for the definition of the ρ-variation seminorms V^ρ.*

5.3 Remarks and overview of the proof

We discuss now some of the main ideas in the proofs of Theorems 5.1 and 5.2.

5.3.1 The Calderón transference principle

One can show that Theorem 5.1 is a consequence of Theorem 5.2 upon performing lifting arguments and adapting the Calderón transference principle. Indeed, if $\mathbb{G}^\#$ is a connected and simply connected nilpotent Lie group of step 2, with Lie algebra $\mathcal{G}$, then one can choose so-called exponential coordinates of the second kind associated to a Malcev basis of the Lie algebra $\mathcal{G}$ (see [31], Sec. 1.2) in such a way that

$$\mathbb{G}^\# \simeq \{(x,y)\in\mathbb{R}^{b_1}\times\mathbb{R}^{b_2} : (x,y)\cdot(x',y') = (x+x', y+y'+R(x,x'))\},$$

where $b_1, b_2\in\mathbb{Z}_+$ depend on the Lie algebra $\mathcal{G}$ and $R:\mathbb{R}^{b_1}\times\mathbb{R}^{b_1}\to\mathbb{R}^{b_2}$ is a bilinear form.
Moreover, if $\mathbb{G}\leq\mathbb{G}^\#$ is a discrete co-compact subgroup, then one can choose the Malcev basis such that the discrete subgroup $\mathbb{G}$ is identified with the integer lattice $\mathbb{Z}^b = \mathbb{Z}^{b_1}\times\mathbb{Z}^{b_2}$ (see [31], Thm. 5.1.6 and Prop. 5.3.2). Recall that $A:\mathbb{Z}\to\mathbb{G}$ is a polynomial sequence satisfying $A(0) = 1$. The main point is that one can choose d sufficiently large and a group morphism $T:\mathbb{G}_0\to\mathbb{G}^\#$ such that

$$A(n) = T(A_0(n)) \qquad \text{for any } n\in\mathbb{Z}.$$

Then one can use this group morphism to transfer bounds on operators on the universal group $\mathbb{G}_0$ to bounds on operators on the group $\mathbb{G}$. Theorem 5.1 is thus a consequence of Theorem 5.2 and our main goal therefore is to prove Theorem 5.2.

5.3.2 The variation spaces V^ρ

For any family $(a_t : t \in \mathbb{I})$ of elements of $\mathbb{C}$ indexed by a totally ordered set $\mathbb{I}$, and any exponent $1 \le \rho < \infty$, the ρ-variation seminorm is defined by

$$V^\rho(a_t : t \in \mathbb{I}) := \sup_{J \in \mathbb{Z}_+} \sup_{\substack{t_0 < \cdots < t_J \\ t_j \in \mathbb{I}}} \Big(\sum_{j=0}^{J-1} |a(t_{j+1}) - a(t_j)|^\rho \Big)^{1/\rho}, \tag{5.3.1}$$

where the supremum is taken over all finite increasing sequences in $\mathbb{I}$. It is easy to see that $\rho \mapsto V^\rho$ is non-increasing, and for every $t_0 \in \mathbb{I}$ one has

$$\sup_{t \in \mathbb{I}} |a_t| \le |a_{t_0}| + V^\rho(a_t : t \in \mathbb{I}) \le \sup_{t \in \mathbb{I}} |a_t| + V^\rho(a_t : t \in \mathbb{I}). \tag{5.3.2}$$

In particular, the maximal estimate (5.2.12) follows from the variational estimate (5.2.13). The main point of proving stronger variational estimates such as (5.2.13), with general parameters $\tau \in (1, 2]$, is that it gives an elegant path to deriving pointwise ergodic theorems (which would not follow directly just from maximal estimates such as (5.2.12)). At the same time, the analysis of variational inequalities has many similarities with the analysis of maximal inequalities, and is not substantially more difficult. This is due in large part to the Rademacher–Menshov inequality (see [89, Lemma 2.5]): for any $2 \le \rho < \infty$ and $j_0, m \in \mathbb{N}$ so that $j_0 < 2^m$ and any sequence of complex numbers $(\mathfrak{a}_k : k \in \mathbb{N})$ we have

$$V^\rho(\mathfrak{a}_j : j_0 \le j \le 2^m) \le \sqrt{2} \sum_{i=0}^{m} \Bigg(\sum_{j \in [j_0 2^{-i}, 2^{m-i} - 1] \cap \mathbb{Z}} \Big| \mathfrak{a}_{(j+1)2^i} - \mathfrak{a}_{j2^i} \Big|^2 \Bigg)^{1/2}. \tag{5.3.3}$$

5.3.3 ℓ^p theory

The problem of passing from ℓ^2 estimates to ℓ^p estimates in the context of discrete polynomial averages has been investigated extensively in recent years (see, for example, [87] and the references therein).

The full $\ell^p(\mathbb{G}_0)$ bounds in Theorem 5.2 rely on first proving $\ell^2(\mathbb{G}_0)$ bounds. In fact, we first establish (5.2.13) for $p = 2$ and $\rho > 2$. Then we use the positivity of the operators M_N^χ (i.e. $M_N^\chi(f) \ge 0$ if $f \ge 0$) to prove the maximal operator bounds (5.2.12) for all $p \in (1, \infty]$. Finally, we use vector-valued interpolation between the bounds (5.2.13) with $p = 2$ and $\rho > 2$ and (5.2.12) with $p \in (1, \infty]$ to complete the proof of Theorem 5.2.

5.3.4 Some technical remarks

Theorem 5.2 (i) and (ii) extends the results of [87, 90] to the non-commutative, nilpotent setting. Its conclusions remain true for rough averages, i.e. when $\chi = \mathbf{1}_{[-1,1]}$ in (5.2.9), but it is more convenient to work with smooth averages.

The restriction $p > 1$ in Theorem 5.2 (i) and (ii) is sharp due to [20, 71]. However, the range of $\rho > \max\{p, \frac{p}{p-1}\}$ is only sharp when $p = 2$ due to Lépingle's inequality [73]. One could hope to improve this to the full range $\rho > 2$, but we do not address this here since the limited range $\rho > \max\{p, \frac{p}{p-1}\}$ is already sufficient for us to establish Theorem 5.1.

The restriction $p = 2$ in the singular integral bounds in part (ii) is probably not necessary. In the commutative case one can prove boundedness in the full range $p \in (1, \infty)$ (see [63]), but the proof depends on exploiting certain Fourier multipliers and we do not know at this time if a similar definitive result holds in the nilpotent case.

5.3.5 The main difficulty and a nilpotent circle method

Bourgain's seminal papers [13, 14, 15] generated a large amount of research and progress in the field. Many other discrete operators have been analyzed by many authors motivated by problems in Analysis and Ergodic Theory. See, for example, [20, 60, 63, 65, 66, 71, 83, 87, 89, 90, 92, 93, 103] for some results of this type and more references. A common feature of all of these results, which plays a crucial role in the proofs, is that one can use Fourier analysis techniques, in particular, the powerful framework of the classical circle method, to perform the analysis.
Our situation in Theorem 5.2 is different. The main conceptual issue is that there is no good Fourier transform on nilpotent groups, compatible with the structure of the underlying convolution operators and at the level of analytical precision of the classical circle method. At a more technical level, there is no good resolution of the delta function compatible with the group multiplication on the group $\mathfrak{g}_0$. This prevents us from using a naive implementation of the circle method. The classical delta function resolution

$$\mathbf{1}_{\{0\}}(x^{-1} \cdot y) = \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \mathfrak{e}((y^{(1)} - x^{(1)}).\theta^{(1)})\mathfrak{e}((y^{(2)} - x^{(2)}).\theta^{(2)})\, d\theta^{(1)} d\theta^{(2)},$$

does not detect the group multiplication correctly.

These issues lead to very significant difficulties in the proof and require substantial new ideas. Our main new construction in [58] is what we call a *nilpotent circle method*, an iterative procedure, starting from the center of the group and moving down along its central series. At every stage we identify "minor arcs", and bound their contributions using Weyl's inequalities (the classical Weyl inequality as well as a nilpotent Weyl inequality which was proved in [61]). The final stage involves "major arcs" analysis, which relies on a combination of continuous harmonic analysis on groups $\mathfrak{g}_0^\#$ and arithmetic harmonic analysis over finite integer rings modulo $Q \in \mathbb{Z}_+$. We outline this procedure in Section 5.5 below.

At the implementation level, classical Fourier techniques are replaced with almost orthogonality methods based on exploiting high order T^*T arguments for operators defined on the discrete group $\mathfrak{g}_0$. Investigating high powers of T^*T (i.e. $(T^*T)^r$ for a large $r \in \mathbb{Z}_+$) is consistent with a general heuristic lying behind the proof of Waring-type problems, which says that the more variables that occur in Waring-type equations, the easier it is to find solutions, and we are able to make this heuristic rigorous in our problem. Manipulating the parameter r, by taking r to be very large, we can always decide how many variables we have at our disposal, making our operators "smoother and smoother".

5.3.6 General discrete nilpotent groups

The primary goal is, of course, to remove the restriction that the discrete nilpotent groups $\mathbb{G}$ in Theorem 5.1 are of step 2, and thus establish the full Conjecture 5.1 in the linear $m = 1$ case for arbitrary invertible measure-preserving transformations $T_1, \ldots, T_d$ that generate a nilpotent group of any step $k \geq 2$. The iterative argument we outline in Section 5.5 below could, in principle, be extended to higher step groups. First, matters can be reduced to universal (or free) discrete step-k nilpotent groups $\mathbb{G}_{d,k}$ with generators $g_1, \ldots, g_d$ and to the moment curve $A_0(n) = g_1 g_2^n \cdots g_d^{n^d}$, for higher step groups as well. Using this "universal"-type structure, one could try to go down along the central series of the group and prove minor arcs and transition estimates at every stage.

Indeed, the l^2-theory seems to carry out for step-3 and step-4 groups, however, continuing this way is only possible if one can prove suitable analogues of the nilpotent Weyl's inequalities in Proposition 5.1 on general nilpotent groups of step $k \geq 5$. The point is to have a small (not necessarily optimal, but nontrivial) gain for bounds on oscillatory sums over many variables, corresponding to the kernels of high power $(T^*T)^r$ operators, whenever frequencies are restricted to the minor arcs. In our case, the formulas are explicit, see the identities (5.4.10), and we can use ideas of Davenport [32] and Birch [11] for Diophantine forms in many variables to control the induced oscillatory sums, but the analysis seems to be more complicated for the higher step nilpotent groups.

This is an interesting problem in its own right, corresponding to Waring-type problems on nilpotent groups. A qualitative variant of the Waring problem on nilpotent groups was recently investigated in [55, 56], see also the references given there. We prove a quantitative version on our nilpotent group $\mathbb{G}_0$ in Theorem 5.4 below.

5.3.7 Organization

The rest of this paper is organized as follows: in section 2 we present several nilpotent Weyl estimates proved in [61], which play a key role in the analysis of minor arcs. In section 3 we outline our main new method, the nilpotent circle method, developed in [58] to prove maximal and variational estimates on nilpotent groups. In section 4 we prove a new Waring-type theorem on the nilpotent group $\mathfrak{g}_0$, as an application of the nilpotent Weyl estimates discussed earlier.

5.4 A nilpotent Weyl inequality on the group $\mathbb{G}_0$

In this section we derive explicit formulas used in high order T^*T arguments and discuss a key ingredient in our analysis, namely Weyl inequalities on the group $\mathbb{G}_0$.

5.4.1 High order T^*T arguments and product kernels

Many of our $\ell^2(\mathbb{G}_0)$ estimates will be based on high order T^*T arguments. Assume that

$$S_1, T_1, \ldots, S_r, T_r : \ell^2(\mathbb{G}_0) \to \ell^2(G_0)$$

are convolution operators defined by some $\ell^1(\mathbb{G}_0)$ kernels $L_1, K_1, \ldots, L_r, K_r : \mathbb{G}_0 \to \mathbb{C}$, i.e. $S_j f = f * L_j$ and $T_j f = f * K_j$ for $j \in \{1, \ldots, r\}$. Then the adjoint operators $S_1^*, \ldots, S_r^*$ are also convolution operators, defined by the kernels $L_1^*, \ldots, L_r^*$ given by

$$L_j^*(g) := \overline{L_j(g^{-1})}.$$

Moreover, using (5.2.10), for any $f \in \ell^2(\mathbb{G}_0)$ and $x \in \mathbb{G}_0$, we have

$$(S_1^* T_1 \ldots S_r^* T_r f)(x) = \sum_{h_1, g_1, \ldots, h_r, g_r \in \mathbb{G}_0} \Big\{ \prod_{j=1}^{r} L_j^*(h_j) K_j(g_j) \Big\} f(g_r^{-1} \cdot h_r^{-1} \cdot \ldots \cdot g_1^{-1} \cdot h_1^{-1} \cdot x). \quad (5.4.1)$$

In other words $(S_1^*T_1\dots S_r^*T_rf)(x)=(f*A^r)(x)$, where the kernel A^r is given by

$$A^r(y):=\sum_{h_1,g_1,\dots,h_r,g_r\in\mathbb{G}_0}\Big\{\prod_{j=1}^r\overline{L_j(h_j)}K_j(g_j)\Big\}\mathbf{1}_{\{0\}}(g_r^{-1}\cdot h_r\cdot\ldots\cdot g_1^{-1}\cdot h_1\cdot y). \tag{5.4.2}$$

To use these formulas we decompose $h_j=(h_j^{(1)},h_j^{(2)})$, $g_j=(g_j^{(1)},g_j^{(2)})$ as in (5.2.3). Then

$$[h_1^{-1}\cdot g_1\cdot\ldots\cdot h_r^{-1}\cdot g_r]^{(1)}=\sum_{1\le j\le r}(-h_j^{(1)}+g_j^{(1)}), \tag{5.4.3}$$

$$\begin{aligned}[h_1^{-1}\cdot g_1\cdot\ldots\cdot h_r^{-1}\cdot g_r]^{(2)}&=\sum_{1\le j\le r}\big\{-(h_j^{(2)}-g_j^{(2)})+R_0(h_j^{(1)},h_j^{(1)}-g_j^{(1)})\big\}\\&\quad+\sum_{1\le l<j\le r}R_0(-h_l^{(1)}+g_l^{(1)},-h_j^{(1)}+g_j^{(1)}),\end{aligned} \tag{5.4.4}$$

as a consequence of applying (5.2.5) inductively.

In many of our applications the operators $S_1,T_1,\dots,S_r,T_r$ are equal and, more importantly, are defined by a kernel K that has product structure, i.e.

$$\begin{aligned}&S_1f=T_1f=\ldots=S_rf=T_rf=f*K,\\&K(g)=K(g^{(1)},g^{(2)})=K^{(1)}(g^{(1)})K^{(2)}(g^{(2)}).\end{aligned} \tag{5.4.5}$$

In this case we can derive an additional formula for the kernel A^r. We use the identity

$$\mathbf{1}_{\{0\}}(x^{-1}\cdot y)=\int_{\mathbb{T}^d\times\mathbb{T}^{d'}}\mathfrak{e}((y^{(1)}-x^{(1)}).\theta^{(1)})\mathfrak{e}((y^{(2)}-x^{(2)}).\theta^{(2)})\,d\theta^{(1)}d\theta^{(2)},$$

where $\mathfrak{e}(z):=e^{2\pi iz}$. The formula (5.4.2) shows that

$$A^r(y)=\int_{\mathbb{T}^d\times\mathbb{T}^{d'}}\mathfrak{e}\big(y^{(1)}.\theta^{(1)}\big)\mathfrak{e}\big(y^{(2)}.\theta^{(2)}\big)\Sigma^r\big(\theta^{(1)},\theta^{(2)}\big)\,d\theta^{(1)}d\theta^{(2)}, \tag{5.4.6}$$

where

$$\Sigma^r\big(\theta^{(1)},\theta^{(2)}\big):=\sum_{h_j,g_j\in\mathbb{G}_0}\Big\{\prod_{j=1}^r\overline{K(h_j)}K(g_j)\Big\}\prod_{i=1}^2\mathfrak{e}\big(-[h_1^{-1}\cdot g_1\cdot\ldots\cdot h_r^{-1}\cdot g_r]^{(i)}.\theta^{(i)}\big).$$

Recalling the product formula (5.4.5) we can write

$$\Sigma^r\big(\theta^{(1)},\theta^{(2)}\big)=\Pi^r\big(\theta^{(1)},\theta^{(2)}\big)\Omega^r\big(\theta^{(2)}\big), \tag{5.4.7}$$

for any $(\theta^{(1)},\theta^{(2)})\in\mathbb{T}^d\times\mathbb{T}^{d'}$, where

$$\begin{aligned}\Pi^r\big(\theta^{(1)},\theta^{(2)}\big):=&\sum_{h_j^{(1)},g_j^{(1)}\in\mathbb{Z}^d}\Big\{\prod_{j=1}^r\overline{K^{(1)}(h_j^{(1)})}K^{(1)}(g_j^{(1)})\Big\}\mathfrak{e}\big(\theta^{(1)}.\sum_{1\le j\le r}(h_j^{(1)}-g_j^{(1)})\big)\\&\times\mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le j\le r}R_0(h_j^{(1)},h_j^{(1)}-g_j^{(1)})+\sum_{1\le l<j\le r}R_0(-h_l^{(1)}+g_l^{(1)},-h_j^{(1)}+g_j^{(1)})\Big\}\Big)\end{aligned} \tag{5.4.8}$$

and

$$\begin{aligned}\Omega^r\big(\theta^{(2)}\big):=&\sum_{h_j^{(2)},g_j^{(2)}\in\mathbb{Z}^{d'}}\Big\{\prod_{j=1}^r\overline{K^{(2)}(h_j^{(2)})}K^{(2)}(g_j^{(2)})\Big\}\mathfrak{e}\big(\theta^{(2)}.\sum_{1\le j\le r}(h_j^{(2)}-g_j^{(2)})\big)\\=&\Big|\sum_{g^{(2)}\in\mathbb{Z}^{d'}}K^{(2)}(g^{(2)})\mathfrak{e}\big(-\theta^{(2)}.g^{(2)}\big)\Big|^{2r}.\end{aligned} \tag{5.4.9}$$

5.4.2 Weyl estimates

After applying high order T^*T arguments we often need to estimate exponential sums and oscillatory integrals involving polynomial phases. With the notation in Section 5.2, for $r \geq 1$ let $D, \widetilde{D} : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{G}_0^{\#}$ be defined by

$$\begin{aligned} D((n_1, \ldots, n_r), (m_1, \ldots, m_r)) &:= A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \ldots \cdot A_0(n_r)^{-1} \cdot A_0(m_r), \\ \widetilde{D}((n_1, \ldots, n_r), (m_1, \ldots, m_r)) &:= A_0(n_1) \cdot A_0(m_1)^{-1} \cdot \ldots \cdot A_0(n_r) \cdot A_0(m_r)^{-1}. \end{aligned} \tag{5.4.10}$$

By definition, we have

$$[A_0(n)]_{l_1 l_2} = \begin{cases} n^{l_1} & \text{if } l_2 = 0, \\ 0 & \text{if } l_2 \geq 1, \end{cases} \qquad [A_0(n)^{-1}]_{l_1 l_2} = \begin{cases} -n^{l_1} & \text{if } l_2 = 0, \\ n^{l_1 + l_2} & \text{if } l_2 \geq 1. \end{cases}$$

Thus, using (5.4.3) and (5.4.4), for $x = (x_1, \ldots, x_r) \in \mathbb{R}^r$ and $y = (y_1, \ldots, y_r) \in \mathbb{R}^r$ one has

$$[D(x,y)]_{l_1 l_2} = \begin{cases} \sum_{j=1}^{r} (y_j^{l_1} - x_j^{l_1}) & \text{if } l_2 = 0, \\ \sum_{1 \leq j_1 < j_2 \leq r} (y_{j_1}^{l_1} - x_{j_1}^{l_1})(y_{j_2}^{l_2} - x_{j_2}^{l_2}) + \sum_{j=1}^{r} (x_j^{l_1 + l_2} - x_j^{l_1} y_j^{l_2}) & \text{if } l_2 \geq 1, \end{cases} \tag{5.4.11}$$

and

$$[\widetilde{D}(x,y)]_{l_1 l_2} = \begin{cases} \sum_{j=1}^{r} (x_j^{l_1} - y_j^{l_1}) & \text{if } l_2 = 0, \\ \sum_{1 \leq j_1 < j_2 \leq r} (x_{j_1}^{l_1} - y_{j_1}^{l_1})(x_{j_2}^{l_2} - y_{j_2}^{l_2}) + \sum_{j=1}^{r} (y_j^{l_1 + l_2} - x_j^{l_1} y_j^{l_2}) & \text{if } l_2 \geq 1. \end{cases} \tag{5.4.12}$$

For $P \in \mathbb{Z}_+$ assume $\phi_P^{(j)}, \psi_P^{(j)} : \mathbb{R} \to \mathbb{R}$, $j \in \{1, \ldots, r\}$, are $C^1(\mathbb{R})$ functions with the properties

$$\sup_{1 \leq j \leq r} \big[|\phi_P^{(j)}| + |\psi_P^{(j)}| \big] \leq \mathbf{1}_{[-P,P]}, \qquad \sup_{1 \leq j \leq r} \int_{\mathbb{R}} \big| [\phi_P^{(j)}]'(x) \big| + \big| [\psi_P^{(j)}]'(x) \big| \, dx \leq 1. \tag{5.4.13}$$

For $\theta = (\theta_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{R}^{|Y_d|}$, $r \in \mathbb{Z}_+$, and $P \in \mathbb{Z}_+$ let

$$S_{P,r}(\theta) = \sum_{n,m \in \mathbb{Z}^r} \mathfrak{e}(-D(n,m).\theta) \Big\{ \prod_{j=1}^{r} \phi_P^{(j)}(n_j) \psi_P^{(j)}(m_j) \Big\}$$

and

$$\widetilde{S}_{P,r}(\theta) = \sum_{n,m \in \mathbb{Z}^r} \mathfrak{e}(-\widetilde{D}(n,m).\theta) \Big\{ \prod_{j=1}^{r} \phi_P^{(j)}(n_j) \psi_P^{(j)}(m_j) \Big\},$$

where D and $\widetilde{D}$ are defined as in (5.4.11)–(5.4.12).
The following key estimates are proved in [61, Proposition 5.1 and Lemma 3.1]:

Proposition 5.1. *(i) (Nilpotent Weyl estimate) For any $\varepsilon > 0$ there is $r = r(\varepsilon, d) \in \mathbb{Z}_+$ sufficiently large such that for all $P \in \mathbb{Z}_+$ we have*

$$|S_{P,r}(\theta)| + |\widetilde{S}_{P,r}(\theta)| \lesssim_\varepsilon P^{2r} P^{-1/\varepsilon}, \tag{5.4.14}$$

provided that there is $(l_1, l_2) \in Y_d$ *and an irreducible fraction* $a/q \in \mathbb{Q}$, $q \in \mathbb{Z}_+$, *such that*

$$|\theta_{l_1 l_2} - a/q| \le 1/q^2 \text{ and } q \in [P^\varepsilon, P^{l_1+l_2-\varepsilon}]. \tag{5.4.15}$$

(ii) (Nilpotent Gauss sums) For any irreducible fraction $a/q \in \mathbb{Q}$, $a = (a_{l_1 l_2})_{(l_1,l_2)\in Y_d} \in \mathbb{Z}^{|Y_d|}$, $q \in \mathbb{Z}_+$, *we define the arithmetic coefficients*

$$G(a/q) := q^{-2r} \sum_{v,w\in\mathbb{Z}_q^r} \mathfrak{e}\big(-D(v,w).(a/q)\big), \qquad \widetilde{G}(a/q) := q^{-2r} \sum_{v,w\in\mathbb{Z}_q^r} \mathfrak{e}\big(-\widetilde{D}(v,w).(a/q)\big). \tag{5.4.16}$$

Then for any $\varepsilon > 0$ *there is* $r = r(\varepsilon, d) \in \mathbb{Z}_+$ *sufficiently large such that*

$$|G(a/q)| + |\widetilde{G}(a/q)| \lesssim_\varepsilon q^{-1/\varepsilon}. \tag{5.4.17}$$

We also need a related integral estimate, see Lemma 5.4 in [61]:

Proposition 5.2. *Given* $\varepsilon > 0$ *there is* $r = r(\varepsilon, d)$ *sufficiently large as in Proposition 5.1 such that*

$$\begin{aligned} &\Big|\int_{\mathbb{R}^r\times\mathbb{R}^r} \Big\{\prod_{j=1}^r \phi_j(x_j)\psi_j(y_j)\Big\}\mathfrak{e}\big(-D(x,y).\beta\big)\,dxdy\Big| \lesssim \langle\beta\rangle^{-1/\varepsilon},\\ &\Big|\int_{\mathbb{R}^r\times\mathbb{R}^r} \Big\{\prod_{j=1}^r \phi_j(x_j)\psi_j(y_j)\Big\}\mathfrak{e}\big(-\widetilde{D}(x,y).\beta\big)\,dxdy\Big| \lesssim \langle\beta\rangle^{-1/\varepsilon}, \end{aligned} \tag{5.4.18}$$

for any $\beta \in \mathbb{R}^{|Y_d|}$ *and for any* $C^1(\mathbb{R})$ *functions* $\phi_1, \psi_1, \ldots, \phi_r, \psi_r : \mathbb{R} \to \mathbb{C}$ *satisfying, for any* $j \in \{1, \ldots, r\}$, *the bounds*

$$|\phi_j| + |\psi_j| \le \mathbf{1}_{[-1,1]}(x), \qquad \int_{\mathbb{R}} \big[|\partial_x\phi_j(x)| + |\partial_x\psi_j(x)|\big]\,dx \le 1.$$

These statements should be compared with classical Weyl-type estimates, which are proved for example in [103, Proposition 1]:

Proposition 5.3. *(i) Assume that* $P \ge 1$ *is an integer and* $\phi_P : \mathbb{R} \to \mathbb{R}$ *is a* $C^1(\mathbb{R})$ *function satisfying*

$$|\phi_P| \le \mathbf{1}_{[-P,P]}, \qquad \int_{\mathbb{R}} \big|\phi_P'(x)\big|\,dx \le 1. \tag{5.4.19}$$

Assume that $\varepsilon > 0$ *and* $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$ *has the property that there is* $l \in \{1, \ldots, d\}$ *and an irreducible fraction* $a/q \in \mathbb{Q}$ *with* $q \in \mathbb{Z}_+$, *such that*

$$|\theta_l - a/q| \le 1/q^2 \ \text{ and } \ q \in [P^\varepsilon, P^{l-\varepsilon}]. \tag{5.4.20}$$

Then there is a constant $\overline{C} = \overline{C}_d \ge 1$ *such that*

$$\Big|\sum_{n\in\mathbb{Z}} \phi_P(n)\mathfrak{e}\big(-(\theta_1 n + \ldots + \theta_d n^d)\big)\Big| \lesssim_\varepsilon P^{1-\varepsilon/\overline{C}}. \tag{5.4.21}$$

(ii) For any irreducible fraction $\theta = a/q \in (\mathbb{Z}/q)^d$, $a = (a_1, \ldots, a_d) \in \mathbb{Z}^d$, $q \in \mathbb{Z}_+$, *we have*

$$\Big|q^{-1} \sum_{n\in\mathbb{Z}_q} \mathfrak{e}\big(-(\theta_1 n + \ldots + \theta_d n^d)\big)\Big| \lesssim q^{-1/\overline{C}}. \tag{5.4.22}$$

Notice that formal similarity of Propositions 5.1 and 5.3. They both involve a small but non-trivial gain of a power of P as soon as one of the coefficients of the relevant polynomials is far from rational numbers with small denominators. These estimates can therefore be used efficiently to estimate minor arcs contributions.

We note, however, that the proof of the nilpotent Weyl estimates in Proposition 5.1 is much more involved than the proof of Proposition 5.3. It relies on some classical ideas of Davenport [32] and Birch [11] on treating polynomials in many variables, but one has to identify and exploit suitable non-degeneracy properties of the explicit (but complicated) polynomials D and $\widetilde{D}$ in (5.4.11)–(5.4.12) to make the proof work. All the details of the proof are provided in [61, Section 5]

5.5 A nilpotent circle method

To illustrate our main method, we focus on a particular case of Theorem 5.2, namely on proving boundedness of the maximal function M_N^χ on $\ell^2(\mathbb{G}_0)$. For simplicity of notation, for $k \in \mathbb{N}$ and $x \in \mathbb{G}_0$, let

$$\begin{aligned} \mathcal{M}_k f(x) &:= M^\chi_{2^k} f(x) = \sum_{n\in\mathbb{Z}} 2^{-k}\chi(2^{-k}n) f(A_0(n))^{-1}\cdot x) = (f * K_k)(x), \\ K_k(x) &:= G^\chi_{2^k}(x) = \sum_{n\in\mathbb{Z}} 2^{-k}\chi(2^{-k}n)\mathbf{1}_{\{A_0(n)\}}(x), \end{aligned} \tag{5.5.1}$$

see (5.2.9) and (5.2.11) for the definitions M_N^χ and G_N^χ respectively. With this new notation, our main goal is to prove the following:

Theorem 5.3. *For any $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\big\| \sup_{k\geq 0} |\mathcal{M}_k f| \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{5.5.2}$$

In the rest of this section we outline the proof of this theorem. Our main new construction is an iterative procedure, starting from the center of the group and moving down along its central series, that allows us to use some of the ideas of the classical circle method recursively at every stage. In our case of nilpotent groups of step two, the procedure consists of two basic stages and one additional step corresponding to "major arcs".
Notice that the kernels K_k have product structure

$$K_k(g) := L_k(g^{(1)})\mathbf{1}_{\{0\}}(g^{(2)}), \qquad L_k(g^{(1)}) := \sum_{n\in\mathbb{Z}} 2^{-k}\chi(2^{-k}n)\mathbf{1}_{\{0\}}(g^{(1)} - A_0^{(1)}(n)), \tag{5.5.3}$$

where $A_0^{(1)}(n) := (n, \dots, n^d) \in \mathbb{Z}^d$ and $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0$ as in (5.2.3).

5.5.1 First stage reduction

We first decompose the singular kernel $\mathbf{1}_{\{0\}}(g^{(2)})$ in the central variable $g^{(2)}$ into smoother kernels. For any $s \in \mathcal{N}$ and $m \in \mathbb{Z}_+$ we define the set of rational fractions

$$\mathcal{R}^m_s := \{a/q : a = (a_1, \dots, a_m) \in \mathbb{Z}^m,\ q \in [2^s, 2^{s+1}) \cap \mathbb{Z},\ \gcd(a_1, \dots, a_m, q) = 1\}. \tag{5.5.4}$$

We define also $\mathcal{R}^m_{\leq a} := \bigcup_{0\leq s\leq a} \mathcal{R}^m_s$. For $x^{(1)} = (x^{(1)}_{l_1 0})_{l_1\in\{1,\dots,d\}} \in \mathbb{R}^d$, $x^{(2)} = (x^{(2)}_{l_1 l_2})_{(l_1,l_2)\in Y'_d} \in \mathbb{R}^{d'}$ and $\Lambda \in (0,\infty)$ we define the partial dilations

$$\Lambda \circ x^{(1)} = (\Lambda^{l_1} x^{(1)}_{l_1 0})_{l_1\in\{1,\dots,d\}} \in \mathbb{R}^d, \qquad \Lambda \circ x^{(2)} = (\Lambda^{l_1+l_2} x^{(2)}_{l_1 l_2})_{(l_1,l_2)\in Y'_d} \in \mathbb{R}^{d'}, \tag{5.5.5}$$

which are induced by the group-dilations defined in (5.2.8).
We fix $\eta_0 : \mathbb{R} \to [0,1]$ a smooth even function such that $\mathbf{1}_{[-1,1]} \le \eta_0 \le \mathbf{1}_{[-2,2]}$. For $t \in \mathbb{R}$ and integers $j \ge 1$ we define

$$\eta_j(t) := \eta_0(2^{-j}t) - \eta_0(2^{-j+1}t), \qquad 1 = \sum_{j=0}^{\infty} \eta_j. \tag{5.5.6}$$

For any $A \in [0,\infty)$ we define

$$\eta_{\le A} := \sum_{j\in[0,A]\cap\mathbb{Z}} \eta_j. \tag{5.5.7}$$

By a slight abuse of notation we also let η_j and $\eta_{\le A}$ denote the smooth radial functions on $\mathbb{R}^m$, $m \ge 1$, defined by $\eta_j(x) = \eta_j(|x|)$ and $\eta_{\le A}(x) = \eta_{\le A}(|x|)$. We fix also two small constants $\delta = \delta(d) \ll \delta' = \delta'(d)$ such that $\delta' \in (0,(10d)^{-10}]$ and $\delta \in (0,(\delta')^4]$, and a large constant $D = D(d) \gg \delta^{-8}$, which depend on arithmetic properties of the polynomial sequence A_0 (more precisely on the structural constants in Propositions 5.1–5.2) such that

$$1 \ll 1/\delta' \ll 1/\delta \ll r = r(\delta', d) \ll D. \tag{5.5.8}$$

For $k \ge D^2$ we fix two cutoff functions $\phi_k^{(1)} : \mathbb{R}^d \to [0,1]$, $\phi_k^{(2)} : \mathbb{R}^{d'} \to [0,1]$, such that

$$\phi_k^{(1)}(g^{(1)}) := \eta_{\le\delta k}(2^{-k}\circ g^{(1)}), \qquad \phi_k^{(2)}(g^{(2)}) := \eta_{\le\delta k}(2^{-k}\circ g^{(2)}). \tag{5.5.9}$$

For $k \in \mathcal{N}$ so that $k \ge D^2$ and for any 1-periodic sets of rationals $\mathcal{A} \subseteq \mathbb{Q}^d$, $\mathcal{B} \subseteq \mathbb{Q}^{d'}$ we define the periodic Fourier multipliers by

$$\begin{aligned} \Psi_{k,\mathcal{A}}(\xi^{(1)}) &:= \sum_{a/q\in\mathcal{A}} \eta_{\le\delta' k}(2^k\circ(\xi^{(1)} - a/q)), \qquad \xi^{(1)} \in \mathbb{T}^d, \\ \Xi_{k,\mathcal{B}}(\xi^{(2)}) &:= \sum_{b/q\in\mathcal{B}} \eta_{\le\delta k}(2^k\circ(\xi^{(2)} - b/q)), \qquad \xi^{(2)} \in \mathbb{T}^{d'}. \end{aligned} \tag{5.5.10}$$

For $k \ge D^2$ and $s \in [0,\delta k]\cap\mathbb{Z}$ we define the periodic Fourier multipliers $\Xi_{k,s} : \mathbb{R}^{d'} \to [0,1]$,

$$\Xi_{k,s}(\xi^{(2)}) := \Xi_{k,\mathcal{R}_s^{d'}}(\xi^{(2)}) = \sum_{a/q\in\mathcal{R}_s^{d'}} \eta_{\le\delta k}(2^k\circ(\xi^{(2)} - a/q)). \tag{5.5.11}$$

For $k \ge D^2$ we write

$$\begin{aligned} \mathbf{1}_{\{0\}}(g^{(2)}) &= \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,d\xi^{(2)} \\ &= \sum_{s\in[0,\delta k]\cap\mathbb{Z}} \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi_{k,s}(\xi^{(2)})\,d\xi^{(2)} + \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi_k^c(\xi^{(2)})\,d\xi^{(2)}, \end{aligned} \tag{5.5.12}$$

where $g^{(2)}.\xi^{(2)}$ denotes the usual scalar product of vectors in $\mathbb{R}^{d'}$ and

$$\Xi_k^c := 1 - \sum_{s\in[0,\delta k]\cap\mathbb{Z}} \Xi_{k,s}. \tag{5.5.13}$$

Then we decompose $K_k = K_k^c + \sum_{s\in[0,\delta k]\cap\mathbb{Z}} K_{k,s}$, where, with the notation in (5.5.3), we have

$$K_{k,s}(g) := L_k(g^{(1)})N_{k,s}(g^{(2)}), \qquad K_k^c(g) := L_k(g^{(1)})N_k^c(g^{(2)}), \tag{5.5.14}$$

and

$$\begin{aligned} N_{k,s}(g^{(2)}) &:= \phi_k^{(2)}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi_{k,s}(\xi^{(2)})\, d\xi^{(2)}, \\ N_k^c(g^{(2)}) &:= \phi_k^{(2)}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi_k^c(\xi^{(2)})\, d\xi^{(2)}. \end{aligned} \tag{5.5.15}$$

We first show that we can bound the contributions of the minor arcs in the central variables:

Lemma 5.1. *For any integer $k \geq D^2$ and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\|f * K_k^c\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-k/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{5.5.16}$$

Then we prove our first transition estimate, i.e. we show that we can bound the contributions of the kernels $K_{k,s}$ corresponding to scales $k \geq 0$ not very large. More precisely, for any $s \geq 0$ we define

$$\kappa_s := 2^{2D(s+1)^2}. \tag{5.5.17}$$

Lemma 5.2. *For any integer $s \geq 0$ and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\big\| \sup_{\max(D^2, s/\delta) \leq k < 2\kappa_s} |f * K_{k,s}| \big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{5.5.18}$$

In the commutative setting, minor arcs estimates such as (5.5.16) follow using Weyl estimates and the Plancherel theorem. As we do not have a useful Fourier transform on the group $\mathbb{G}_0$, our main tool to prove the bounds (5.5.16) is a high order T^*T argument. More precisely, we analyze the kernel of the convolution operator $\{(\mathcal{K}_k^c)^*\mathcal{K}_k^c\}^r$, where $\mathcal{K}_k^c f := f * K_k^c$ and r is sufficiently large, and show that its $\ell^1(\mathbb{G}_0)$ norm is $\lesssim 2^{-k}$. The main ingredient in this proof is the non-commutative Weyl estimate in Proposition 5.1 (i).
To prove the transition estimates (5.5.18), we use the Rademacher-Menshov inequality and Khintchine's inequality (leading to logarithmic losses) to reduce to proving the bounds

$$\Big\| \sum_{k \in [J, 2J]} \varkappa_k (f * H_{k,s}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-4s/D^2} \big\|f\big\|_{\ell^2(\mathbb{G}_0)} \tag{5.5.19}$$

for any $J \geq \max(D^2, s/\delta)$ and any coefficients $\varkappa_k \in [-1, 1]$, where $H_{k,s} := K_{k+1,s} - K_{k,s}$. For this, we use a high order version of the Cotlar–Stein lemma, which relies again on precise analysis of the kernel of the convolution operator $\{(\mathcal{H}_{k,s})^*\mathcal{H}_{k,s}\}^r$, where $\mathcal{H}_{k,s} f := f * H_{k,s}$ and r is sufficiently large. The key exponential gain of $2^{-4s/D^2}$ in (5.5.19) is due to the non-commutative Gauss sums estimate, see Proposition 5.1 (ii).

5.5.2 Second stage reduction

In view of Lemmas 5.1–5.2 it remains to prove that

$$\big\| \sup_{k \geq \kappa_s} |f * K_{k,s}| \big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)} \tag{5.5.20}$$

for any fixed integer $s \geq 0$. The kernels $K_{k,s}$ are now reasonably well adapted to a natural family of non-isotropic balls in the central variables, at least when $2^s \approx 1$, and we need to start decomposing in the non-central variables.

We examine the kernels $L_k(g^{(1)})$ defined in (5.5.3), and rewrite them in the form

$$\begin{aligned} L_k(g^{(1)}) &= \sum_{n\in\mathbb{Z}} 2^{-k}\chi(2^{-k}n)\mathbf{1}_{\{0\}}(-A_0^{(1)}(n)+g^{(1)}) \\ &= \phi_k^{(1)}(g^{(1)})\int_{\mathbb{T}^d}\mathfrak{e}(g^{(1)}.\xi^{(1)})S_k(\xi^{(1)})\,d\xi^{(1)}, \end{aligned} \tag{5.5.21}$$

where $g^{(1)}.\xi^{(1)}$ denotes the usual scalar product of vectors in $\mathbb{R}^d$, and

$$S_k(\xi^{(1)}) := \sum_{n\in\mathbb{Z}} 2^{-k}\chi(2^{-k}n)\mathfrak{e}(-A_0^{(1)}(n).\xi^{(1)}). \tag{5.5.22}$$

For any integers $Q\in\mathbb{Z}_+$ and $m\in\mathbb{Z}_+$ we define the set of fractions

$$\widetilde{\mathcal{R}}_Q^m := \{a/Q:\ a=(a_1,\ldots,a_m)\in\mathbb{Z}^m\}. \tag{5.5.23}$$

For any integer $s\geq 0$ we fix a large denominator

$$Q_s := \big(\lfloor 2^{D(s+1)}\rfloor\big)! = 1\cdot 2\cdot\ldots\cdot\lfloor 2^{D(s+1)}\rfloor, \tag{5.5.24}$$

and using (5.5.10) define the periodic multipliers

$$\begin{aligned} \Psi_{k,s}^{\mathrm{low}}(\xi^{(1)}) &:= \Psi_{k,\widetilde{\mathcal{R}}^d_{Q_s}}(\xi^{(1)}) = \sum_{a/q\in\widetilde{\mathcal{R}}^d_{Q_s}} \eta_{\leq\delta' k}(2^k\circ(\xi^{(1)}-a/q)), \\ \Psi_{k,s,t}(\xi^{(1)}) &:= \Psi_{k,\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s}}(\xi^{(1)}) = \sum_{a/q\in\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s}} \eta_{\leq\delta' k}(2^k\circ(\xi^{(1)}-a/q)), \\ \Psi_k^c(\xi^{(1)}) &:= 1-\Psi_{k,s}^{\mathrm{low}}(\xi^{(1)}) - \sum_{t\in[0,\delta' k]\cap\mathbb{Z}}\Psi_{k,s,t}(\xi^{(1)}) \ = 1-\sum_{a/q\in\mathcal{R}^d_{\leq\delta' k}}\eta_{\leq\delta' k}(\tau^k\circ(\xi^{(1)}-a/q)). \end{aligned} \tag{5.5.25}$$

Since $k\geq\kappa_s = 2^{2D(s+1)^2}$ we see that $Q_s\leq 2^{\delta^2 k}$. Therefore the supports of the cutoff functions $\eta_{\leq\delta' k}(2^k\circ(\xi^{(1)}-a/q))$ are all disjoint and the multipliers $\Psi_{k,s}^{\mathrm{low}},\Psi_{k,s,t},\Psi_k^c$ take values in the interval $[0,1]$. Notice also that $\Psi_{k,s,t}\equiv 0$ unless $t\geq D(s+1)$, and that the cutoffs used in these definitions depend on $\delta' k$ not on δk as in the case of the central variables.

We examine the formula (5.5.21) and define the kernels $L_{k,s}^{\mathrm{low}}, L_{k,s,t}, L_k^c:\mathbb{Z}^d\to\mathbb{C}$ by

$$L_*(g^{(1)}) = \phi_k^{(1)}(g^{(1)})\int_{\mathbb{T}^d}\mathfrak{e}(g^{(1)}.\xi^{(1)})S_k(\xi^{(1)})\Psi_*(\xi^{(1)})\,d\xi^{(1)}, \tag{5.5.26}$$

where $(L_*,\Psi_*)\in\{(L_{k,s}^{\mathrm{low}},\Psi_{k,s}^{\mathrm{low}}),(L_{k,s,t},\Psi_{k,s,t}),(L_k^c,\Psi_k^c)\}$. For any $k\geq\kappa_s$ we obtain $K_{k,s}=G_{k,s}^{\mathrm{low}}+\sum_{t\leq\delta' k}G_{k,s,t}+G_{k,s}^c$, where the kernels $G_{k,s}^{\mathrm{low}},G_{k,s,t},G_{k,s}^c:\mathbb{Z}^{|Y_d|}\to\mathbb{C}$ are defined by

$$\begin{aligned} G_{k,s}^{\mathrm{low}}(g) &:= L_{k,s}^{\mathrm{low}}(g^{(1)})N_{k,s}(g^{(2)}), \\ G_{k,s,t}(g) &:= L_{k,s,t}(g^{(1)})N_{k,s}(g^{(2)}), \\ G_{k,s}^c(g) &:= L_k^c(g^{(1)})N_{k,s}(g^{(2)}). \end{aligned} \tag{5.5.27}$$

Our next step is to show that the contributions of the minor arcs corresponding to the kernels $G_{k,s}^c$ can be suitably bounded:

Lemma 5.3. *For any integers $s \geq 0$ and $k \geq \kappa_s$, and for any $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\|f * G^c_{k,s}\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-k/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{5.5.28}$$

Then we prove our second transition estimate, bounding the contributions of the operators defined by the kernels $G_{k,s,t}$ for intermediate values of k.

Lemma 5.4. *For any integers $s \geq 0$, and $t \geq D(s+1)$, and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\Big\| \sup_{\max(\kappa_s, t/\delta') \leq k < 2\kappa_t} |f * G_{k,s,t}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}, \tag{5.5.29}$$

where $\kappa_t = 2^{2D(t+1)^2}$ as in (5.5.17).

The proofs of these estimates are similar to the proofs of the corresponding first stage estimates (5.5.16)–(5.5.18), using high order T^*T arguments. However, instead of using the nilpotent oscillatory sums estimates in Proposition 5.1, we use the classical estimates from Proposition 5.3 here. We emphasize, however, that the underlying nilpotent structure is very important and that these estimates are only possible after performing the two reductions in the first stage, namely, the restriction to major arcs corresponding to denominators $\approx 2^s$ and the restriction to parameters $k \geq \kappa_s$.
We finally remark that the circle method could not have been applied simultaneously to both central and non-central variables, as we would not have been able control efficiently the phase functions arising in the corresponding exponential sums and oscillatory integrals, especially on major arcs.

5.5.3 Final stage: major arcs contributions

After these reductions, it remains to bound the contributions of the "major arcs" in both the central and the non-central variables. More precisely, we prove the following bounds:

Lemma 5.5. *(i) For any integer $s \geq 0$ and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\Big\| \sup_{k \geq \kappa_s} |f * G^{\mathrm{low}}_{k,s}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{5.5.30}$$

(ii) For any integers $s \geq 0$, $t \geq D(s+1)$, and $f \in \ell^2(\mathbb{G}_0)$ we have

$$\Big\| \sup_{k \geq \kappa_t} |f * G_{k,s,t}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{5.5.31}$$

The main idea here is different: we write the kernels $G^{\mathrm{low}}_{k,s}$ and $G_{k,s,t}$ as tensor products of two components up to acceptable errors. One of these components is essentially a maximal average operator on a continuous group, which can be analyzed using the classical method of Christ [22]. The other component is an arithmetic operator-valued analogue of the classical Gauss sums, which leads to the key factors $2^{-s/D^2}$ and $2^{-t/D^2}$ in (5.5.30) and (5.5.31).
More precisely, for any integer $Q \geq 1$ we define the subgroup

$$\mathcal{H}_Q := \{h = (Qh_{l_1 l_2})_{(l_1,l_2) \in Y_d} \in \mathbb{G}_0 : \, h_{l_1,l_2} \in \mathbb{Z}\}. \tag{5.5.32}$$

Clearly $\mathcal{H}_Q \subseteq \mathbb{G}_0$ is a normal subgroup. Let $\mathbb{J}_Q$ denote the coset

$$\mathbb{J}_Q := \{b = (b_{l_1 l_2})_{(l_1,l_2) \in Y_d} \in \mathbb{G}_0 : \, b_{l_1,l_2} \in \mathbb{Z} \cap [0, Q-1]\}, \tag{5.5.33}$$

with the natural induced group structure. Notice that

$$\text{the map } (b,h) \mapsto b \cdot h \text{ defines a bijection from } \mathbb{J}_Q \times \mathcal{H}_Q \text{ to } \mathbb{G}_0. \tag{5.5.34}$$

Assume that $Q \geq 1$ and $2^k \geq Q$. For any $a \in \mathbb{Z}^d$ and $\xi \in \mathbb{R}^d$ let

$$\begin{aligned} J_k(\xi) &:= 2^{-k} \int_{\mathbb{R}} \chi(2^{-k}x)\mathfrak{e}[-A_0^{(1)}(x).\xi]\,dx = \int_{\mathbb{R}} \chi(y)\mathfrak{e}[-A_0^{(1)}(y).(2^k \circ \xi)]\,dy, \\ S(a/Q) &:= Q^{-1} \sum_{n \in \mathbb{Z}_Q} \mathfrak{e}[-A_0^{(1)}(n).a/Q]. \end{aligned} \tag{5.5.35}$$

The point is that the kernels $G_{k,s}^{\mathrm{low}}$ and $G_{k,s,t}$ can be decomposed as tensor products. Indeed, to decompose $G_{k,s,t}$ (the harder case) we set $Q := Q_t = \big(\lfloor 2^{D(t+1)} \rfloor\big)!$ as in (5.5.24). Then we show that if $k \geq \kappa_t$ (so $2^k \gg Q_t^4$), $h \in \mathcal{H}_{Q_t}$ and $b_1, b_2 \in \mathbb{G}_0$ satisfy $|b_1| + |b_2| \leq Q^4$ then

$$G_{k,s,t}(b_1 \cdot h \cdot b_2) \approx W_{k,Q_t}(h) V_{\mathcal{R}_t^d \setminus \widetilde{\mathcal{R}}_{Q_s}^d, \mathcal{R}_s^{d'}, Q_t}(b_1 \cdot b_2), \tag{5.5.36}$$

up to acceptable summable errors. Here

$$W_{k,Q_t}(h) := Q_t^{d+d'} \phi_k(h) \int_{\mathbb{R}^d \times \mathbb{R}^{d'}} \eta_{\leq \delta' k}(2^k \circ \xi) \eta_{\leq \delta k}(2^k \circ \theta) \mathfrak{e}(h.(\xi,\theta)) J_k(\xi)\,d\xi d\theta,$$

$$V_{\mathcal{A},\mathcal{B},Q}(b) := Q^{-d-d'} \Big\{ \sum_{\sigma^{(1)} \in \mathcal{A} \cap [0,1)^d} S(\sigma^{(1)}) \mathfrak{e}[b^{(1)}.(\sigma^{(1)})] \Big\} \Big\{ \sum_{\sigma^{(2)} \in \mathcal{B} \cap [0,1)^{d'}} \mathfrak{e}[b^{(2)}.(\sigma^{(2)})] \Big\},$$

and $\phi_k(h) := \phi_k^{(1)}(h^{(1)}) \phi_k^{(2)}(h^{(2)})$, $h = (h^{(1)}, h^{(2)}) \in \mathcal{H}_{Q_t}$, $b = (b^{(1)}, b^{(2)}) \in \mathbb{G}_0$, and the functions J_k and S are defined in (5.5.35).

Finally, we show that the kernels $V_{s,t} := V_{\mathcal{R}_t^d \setminus \widetilde{\mathcal{R}}_{Q_s}^d, \mathcal{R}_s^{d'}, Q_t}$ (which can be interpreted as an operator-valued Gauss sums) define bounded operators on $\ell^2(\mathbb{J}_{Q_t})$,

$$\|f *_{\mathbb{J}_{Q_t}} V_{s,t}\|_{\ell^2(\mathbb{J}_{Q_t})} \lesssim 2^{-t/D} \|f\|_{\ell^2(\mathbb{J}_{Q_t})}.$$

Moreover, the kernels W_{k,Q_t} are close to classical maximal operators and one can show that

$$\big\| \sup_{k \geq \kappa_t} |f *_{\mathcal{H}_{Q_t}} W_{k,Q_t}| \big\|_{\ell^2(\mathcal{H}_{Q_t})} \lesssim \|f\|_{\ell^2(\mathcal{H}_{Q_t})}.$$

The desired bounds (5.5.31) follow using the approximation formula (5.5.36).

5.6 A nilpotent Waring theorem on the group $\mathbb{G}_0$

The classical Waring problem, solved by Hilbert in 1909, concerns the possibility of writing any positive integer as a sum of finitely many p powers: for any integer $p \geq 1$ there is $r = r(p)$ such that any integer $y \in \mathbb{Z}_+$ can be written in the form

$$y = \sum_{i=1}^{r} m_i^p, \qquad \text{for some non-negative integers } m_1, \ldots, m_r. \tag{5.6.1}$$

There is a vast amount of literature on this problem and its many possible extensions. In particular, the symmetric system of equations

$$\sum_{j=1}^{r} (m_j^s - n_j^s) = 0 \quad (1 \leq s \leq d), \tag{5.6.2}$$

first studied by Vinogradov [111] in relation to the Waring problem, have been the focus of intense recent research, see [114] for some breakthrough results. We are interested here in understanding the analogous question on our discrete nilpotent Lie group $\mathfrak{g}_0$ and for our given polynomial sequence A_0: can one represent elements $g \in \mathfrak{g}_0$ in the form

$$g = A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \ldots \cdot A_0(n_r)^{-1} \cdot A_0(m_r), \tag{5.6.3}$$

for some integers $n_1, m_1, \ldots, n_r, m_r$, provided that r is large enough? We are, in fact, interested in proving a quantitative statement on the number of such representations, for integers $n_1, m_1, \ldots, n_r, m_r \in [N] := [-N, N] \cap \mathbb{Z}$.

We remark that many group elements g cannot be written in the form (5.6.3), due to local obstructions; for instance, if g can be represented in the form (5.6.3) then necessarily $g_{10} \equiv g_{20} \equiv \ldots \equiv g_{d0} \pmod 2$, $g_{10} = g_{30} = \ldots \pmod 3$ etc. For integers $r, N \geq 1$ and $g \in \mathfrak{g}_0$ let

$$S_{r,N}(g) := \big|\big\{(m,n) \in [N]^{2r} : A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \ldots \cdot A_0(n_r)^{-1} \cdot A_0(m_r) = g\big\}\big|. \tag{5.6.4}$$

Our main result in this section is the following:

Theorem 5.4. *(i) There is an integer $r_0(d) \geq 1$ such that if $r \geq r_0(d)$ is sufficiently large and $g \in \mathfrak{g}_0$ then*

$$S_{r,N}(g) = N^{2r}\Big(\prod_{(l_1,l_2)\in Y_d} N^{-|l_1|-|l_2|}\Big)\Big[\mathfrak{S}(g)\int_{\mathbb{R}^{d+d'}} \Phi(\zeta)\mathfrak{e}(-(N^{-1}\circ g).\zeta)\, d\zeta + O_r(N^{-1/2})\Big], \tag{5.6.5}$$

uniformly in $N \in \mathbb{N}$. Here the singular series $\mathfrak{S}$ is defined by

$$\mathfrak{S}(g) := \sum_{a/q \in \mathcal{R}_\infty^{d+d'}} \overline{G(a/q)}\mathfrak{e}(-g.a/q) \tag{5.6.6}$$

and the singular integral Φ is defined by

$$\Phi(\xi) = \int_{[-1,1]^{2r}} \mathfrak{e}\big(D(z,w).\xi\big)\, dz\, dw, \qquad \xi \in \mathbb{R}^{d+d'}. \tag{5.6.7}$$

In particular, all elements $g \in \mathfrak{g}_0$ cannot be represented in the form (5.6.3) *more than a constant times the expected number of representations, i.e.*

$$S_{r,N}(g) \lesssim_r N^{2r}\Big(\prod_{(l_1,l_2)\in Y_d} N^{-|l_1|-|l_2|}\Big) \qquad \text{for any } g \in \mathfrak{g}_0. \tag{5.6.8}$$

(ii) For $r \geq r_0(d)$ as above there is a sufficiently large integer $Q = Q(r)$ such that

$$S_{r,N}(g) = N^{2r}\Big(\prod_{(l_1,l_2)\in Y_d} N^{-|l_1|-|l_2|}\Big)\Big[c_r(g) + O_{r,g}(N^{-1/2})\Big], \tag{5.6.9}$$

for any $g \in \mathcal{H}_Q$ (see definition (5.5.32)*), where $c_r(g) \approx_r 1$ uniformly in g.*

Sketch of the proof: Observe that $D(n,m) = A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \ldots \cdot A_0(n_r)^{-1} \cdot A_0(m_r)$. Using the classical delta function we can write

$$S_{r,N}(g) = \sum_{m,n\in[N]^r} \int_{\mathbb{T}^{d+d'}} \mathfrak{e}\big(D(n,m).\xi\big)\mathfrak{e}(-g.\xi)\, d\xi. \tag{5.6.10}$$

Step 1. We start by decomposing the integration in ξ into major and minor arcs. For any integer $m \geq 1$ and any positive number $M > 0$, we define the set of rational fractions

$$\mathcal{R}^m_{\leq M} := \{a/q :\ a = (a_1, \dots, a_m) \in \mathbb{Z}^m,\ q \in [1, M] \cap \mathbb{Z},\ \gcd(a_1, \dots, a_m, q) = 1\}. \tag{5.6.11}$$

We fix a small constant $\delta = \delta(d) \ll 1$ and a smooth radial function $\eta_0 \colon \mathbb{R}^{|Y_d|} \to [0, 1]$ such that index $|x| \leq 1 \leq \eta_0(x) \leq$ index $|x| \leq 2$, $x \in \mathbb{R}^{|Y_d|}$. For $A > 0$ let $\eta_{\leq A}(x) := \eta_0\big(A^{-1}x\big)$, $x \in \mathbb{R}^{|Y_d|}$. Then we introduce the projections

$$\Xi_N(\xi) := \sum_{a/q \in \mathcal{R}^{d+d'}_{\leq N^\delta}} \eta_{\leq N^\delta}\big(N \circ (\xi - a/q)\big), \qquad \xi \in \mathbb{T}^{d+d'}, \quad N \in \mathbb{N}, \tag{5.6.12}$$

and decompose the integration in (5.6.10) into major and minor arcs, i.e. we define

$$S_{r,N,\mathrm{maj}}(g) := \sum_{m,n \in [N]^r} \int_{\mathbb{T}^{d+d'}} \mathfrak{e}\big(D(n,m).\xi\big)\mathfrak{e}(-g.\xi)\Xi_N(\xi)\, d\xi \tag{5.6.13}$$

$$S_{r,N,\min}(g) := \sum_{m,n \in [N]^r} \int_{\mathbb{T}^{d+d'}} \mathfrak{e}\big(D(n,m).\xi\big)\mathfrak{e}(-g.\xi)\big(1 - \Xi_N(\xi)\big)\, d\xi, \tag{5.6.14}$$

Notice that $S_{r,N}(g) = S_{r,N,\min}(g) + S_{r,N,\mathrm{maj}}(g)$. Moreover

$$|S_{r,N,\min}(g)| \lesssim_r N^{2r-1}\Big(\prod_{(l_1,l_2) \in Y_d} N^{-|l_1|-|l_2|}\Big), \qquad N \in \mathbb{N}, \quad g \in \mathfrak{g}_0 \tag{5.6.15}$$

provided that r is sufficiently large, as a consequence of Proposition 5.1 (i) and the Dirichlet principle. Therefore the contribution of the minor arcs $S_{r,N,\min}(g)$ can be absorbed by the error term in (5.6.5).

Step 2. Next, we deal with the major arcs contributions. Notice that

$$S_{r,N,\mathrm{maj}}(g) = \sum_{a/q \in \mathcal{R}^{d+d'}_{\leq N^\delta} \cap [0,1)^{d+d'}} \mathfrak{e}(-g.a/q) \int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}\big(N \circ \xi\big) I_{r,N,a/q}(\xi)\mathfrak{e}(-g.\xi)\, d\xi, \tag{5.6.16}$$

where

$$I_{r,N,a/q}(\xi) = \sum_{m,n \in [N]^r} \mathfrak{e}\big(D(n,m).(a/q)\big)\mathfrak{e}\big(D(n,m).\xi\big). \tag{5.6.17}$$

Observe that for $a/q \in \mathcal{R}^{d+d'}_{\leq N^\delta} \cap [0,1)^{d+d'}$ and $|N \circ \xi| \lesssim N^\delta$ we have

$$\begin{aligned} I_{r,N,a/q}(\xi) &= \sum_{m,n \in [N/q]^r} \sum_{u,v \in \mathbb{Z}_q^r} \mathfrak{e}\big(D(v,w).(a/q)\big)\mathfrak{e}\big(D(qn,qm).\xi\big) + O(qN^{2r-1+\delta}) \\ &= N^{2r}\overline{G(a/q)}\Phi(N \circ \xi) + O(qN^{2r-1+\delta}), \end{aligned}$$

where $G(a/q)$ is defined in (5.4.16) and Φ is defined in (5.6.7).
Therefore, if $\delta \leq (10d)^{-4}$ then we have

$$\begin{aligned} S_{r,N,\mathrm{maj}}(g) &= N^{2r}\Big(\prod_{(l_1,l_2) \in Y_d} N^{-|l_1|-|l_2|}\Big) \\ &\times \Big[\sum_{a/q \in \mathcal{R}^{d+d'}_{\leq N^\delta}} \overline{G(a/q)}\mathfrak{e}(-g.a/q) \int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}\big(\xi\big)\Phi(\xi)\mathfrak{e}(-g.(N^{-1} \circ \xi))\, d\xi + O_r(N^{-1/2})\Big]. \end{aligned} \tag{5.6.18}$$

It follows from Proposition 5.1 (ii) and Proposition 5.2 that

$$|G(a/q)| \lesssim_r q^{-1/\delta^2}, \qquad (a,q)=1, \tag{5.6.19}$$

and

$$|\Phi(\zeta)| \lesssim_r \langle\zeta\rangle^{-1/\delta^2}, \qquad \zeta\in\mathbb{R}^{d+d'}, \tag{5.6.20}$$

provided that r is sufficiently large. Therefore, recalling the definition (5.6.6),

$$\begin{aligned}&|\mathfrak{S}(g)| \lesssim_r 1,\\ &|[|\Big]\mathfrak{S}(g) - \sum_{a/q\in\mathcal{R}^{d+d'}_{\leq N^\delta}} \overline{G(a/q)}\mathfrak{e}(-g.a/q) \lesssim_r \sum_{q\geq N^\delta} q^{d+d'-1/\delta^2} \lesssim_r N^{-1/(2\delta)}.\end{aligned} \tag{5.6.21}$$

Moreover

$$\begin{aligned}&\Big|\int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}\big(\xi\big)\Phi(\xi)\mathfrak{e}(-g.(N^{-1}\circ\xi))\,d\xi\Big| \lesssim_r 1,\\ &\Big|\int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}\big(\xi\big)\Phi(\xi)\mathfrak{e}(-g.(N^{-1}\circ\xi))\,d\xi - \int_{\mathbb{R}^{d+d'}} \Phi(\xi)\mathfrak{e}(-g.(N^{-1}\circ\xi))\,d\xi\Big| \lesssim_r N^{-1/(2\delta)}.\end{aligned} \tag{5.6.22}$$

It follows from (5.6.18), (5.6.21), and (5.6.22) that

$$S_{r,N,\mathrm{maj}}(g) = N^{2r}\Big(\prod_{(l_1,l_2)\in Y_d} N^{-|l_1|-|l_2|}\Big)\Big[\mathfrak{S}(g)\int_{\mathbb{R}^{d+d'}}\Phi(\xi)\mathfrak{e}(-g.(N^{-1}\circ\xi))\,d\xi + O_r(N^{-1/2})\Big]. \tag{5.6.23}$$

The desired conclusion (5.6.5) follows using also (5.6.15). This completes the proof of part (i) of the theorem.

Step 3. We analyze now the singular series $\mathfrak{S}$ defined in (5.6.6). Observe that

$$\mathfrak{S}(h) = \sum_{q\geq 1} A(q,h), \qquad A(q,h) := \sum_{(a,q)=1} \overline{G(a/q)}\mathfrak{e}(-h.a/q), \tag{5.6.24}$$

for any $h\in\mathfrak{g}_0$. Notice that $A(q,h)$ is multiplicative in the sense that $A(q_1q_2,h) = A(q_1,h)A(q_2,h)$ provided that $(q_1,q_2)=1$ and $h\in\mathfrak{g}_0$. Therefore, letting $\mathbf{P}$ denote the set of primes,

$$\mathfrak{S}(h) = \prod_{p\in\mathbf{P}} B(p,h), \qquad B(p,h) := 1 + \sum_{n\geq 1} A(p^n,h). \tag{5.6.25}$$

For $h\in\mathfrak{g}_0$ and $q\geq 1$ let

$$M(q,h) := \Big|\big\{(m,n)\in\mathbb{Z}_q^{2r} : D(n,m) = h \mod q\big\}\Big|. \tag{5.6.26}$$

We prove that for any $h\in\mathfrak{g}_0$, $p\in\mathbf{P}$ and integer $n\geq 1$ we have

$$1 + \sum_{v=1}^{n} A(p^v,h) = \frac{M(p^n,h)}{p^{n(2r-d-d')}}. \tag{5.6.27}$$

Indeed, for any integer $q \geq 1$ we have

$$\begin{aligned}
M(q,h) = q^{-d-d'} &\sum_{t\in\mathbb{Z}_q^{d+d'}} \sum_{m,n\in\mathbb{Z}_q^r} \mathfrak{e}\big((D(n,m)-h).(t/q)\big)\\
&= q^{-d-d'}\sum_{q_1|q} \sum_{w\in\mathbb{Z}_{q/q_1}^{d+d'},\,(w,q/q_1)=1} \sum_{m,n\in\mathbb{Z}_q^r} \mathfrak{e}\big((D(n,m)-h).(wq_1/q)\big)\\
&= q^{-d-d'}\sum_{q_2|q} \sum_{w\in\mathbb{Z}_{q_2}^{d+d'},\,(w,q_2)=1} \sum_{m,n\in\mathbb{Z}_q^r} \mathfrak{e}\big((D(n,m)-h).(w/q_2)\big)\\
&= q^{-d-d'}\sum_{q_2|q} \sum_{w\in\mathbb{Z}_{q_2}^{d+d'},\,(w,q_2)=1} q^{2r}\overline{G(w/q_2)}\mathfrak{e}\big(-h.(w/q_2)\big)\\
&= q^{2r-d-d'}\sum_{q_2|q} A(q_2,h).
\end{aligned}$$

The identity (5.6.27) follows by applying this with $q=p^n$, $p\in\mathbf{P}$. In particular $\mathfrak{S}(h)$ and $B(p,h)$ are real non-negative numbers,

$$\mathfrak{S}(h),\, B(p,h)\in[0,\infty) \qquad \text{for any } h\in\mathfrak{g}_0,\, p\in\mathbf{P}. \tag{5.6.28}$$

Moreover, using the formulas (5.6.25) and (5.6.27),

$$B(p,h) = \frac{M(p^n,h)}{p^{n(2r-d-d')}} + O_r(2^{-n/\delta})$$

We would like to show now that $\mathfrak{S}(h)\gtrsim_r 1$ for all elements $h\in\mathbb{H}_Q$, in order to be able to exploit the expansion (5.6.5). We notice first that for any integer r sufficiently large there is $p_0(r)\in\mathbf{P}$ such that

$$1/2 \leq \prod_{p\in\mathbf{P},\,p\geq p_0(r)} B(p,h) \leq 3/2, \tag{5.6.29}$$

for any $h\in\mathfrak{g}_0$, due to the rapid decay of the coefficients $G(a/q)$ in (5.6.19).
By Lemma 5.6 there is a point $a_0=(z_0,w_0)$ such that $D(a_0)=0$ and there is a $(d+d')\times(d+d')$ minor $J_D(a_0)\neq 0$. By re-indexing the variables we may assume that this minor is $J_D(a_0) = det\big(\frac{\partial_i D}{\partial x_j}(a_0)\big)_{i,j=K+1}^{N}$, writing $N=2r$, $K=2r-d-d'$ and $x=(m,n)\in\mathbb{Z}^{2r}$. In other words, we may assume that the minor corresponding to the last $d+d'$ columns of the Jacobian matrix of D is non-singular. For a given prime $p\leq p_0(r)$ let $\gamma_p\in\mathcal{N}$ be such that $J_D(a_0)=p^{\gamma_p}u$ with $u\in\mathcal{N}$ and $p\nmid u$. Define $Q=Q(r):=\prod_{p\in\mathbf{P},p\leq p_0(r)} p^{2\gamma_p+1}$.
For $h\in\mathcal{H}_Q$, we have that $D(a_0)=h \mod p^{2\gamma_p+1}$, but $J_D(a_0)\neq 0 \mod p^{\gamma_p+1}$ thus we are in the position to apply Hensel's lemma again with $N=2r$ and $K=2r-d-d'$. Then by Corollary 5.1 we have that $M(p^n,h)\geq p^{(n-2\gamma_p-1)K}$ and hence by (5.6.28) we have $B(p^n,h)\geq p^{-(2\gamma_p+1)K-1}$ for all $n>2\gamma_p,\ n>n(r)$. This proves that

$$\mathfrak{S}(h)\gtrsim_r 1 \quad \text{uniformly for} \quad h\in\mathcal{H}_Q. \tag{5.6.30}$$

Step 4. Finally we analyze the contribution of the singular integral. Since

$$\int_{\mathbb{R}^{d+d'}} \Phi(\zeta)\mathfrak{e}(-(N^{-1}\circ g).\zeta)\,d\zeta = \int_{\mathbb{R}^{d+d'}} \Phi(\zeta)\,d\zeta + O_{r,g}(N^{-1}),$$

due to (5.6.20), to prove the approximate identity (5.6.9) it suffices to prove that

$$\int_{\mathbb{R}^{d+d'}} \Phi(\zeta)\, d\zeta \gtrsim_r 1. \tag{5.6.31}$$

We fix a smooth function $\chi : \mathbb{R}^{d+d'} \to [0,1]$, satisfying $\chi(x) = 1$ if $|x| \le 1/2$, $\chi(x) = 0$ if $|x| \ge 2$, and $\int_{\mathbb{R}^{d+d'}} \chi(x)\, dx = 1$. For $\epsilon \le \epsilon(r)$ sufficiently small we write

$$\int_{\mathbb{R}^{d+d'}} \Phi(\zeta)\widehat{\chi}(\epsilon\zeta)\, d\zeta = \int_{[-1,1]^{2r}} \epsilon^{-(d+d')}\chi(D(z,w)/\epsilon)\, dzdw, \tag{5.6.32}$$

using the definition (5.6.7). In particular, by letting $\epsilon \to 0$, $\int_{\mathbb{R}^{d+d'}} \Phi(\zeta)\, d\zeta$ is a real non-negative number. Moreover, the lower bound (5.6.31) follows from (5.6.32) provided that we can show that there is a point $(z_0, w_0) \in [-1,1]^{2r}$ such that

$$D(z_0, w_0) = 0 \qquad \text{and} \qquad \text{rank}[\nabla_{z,w} D(z_0, w_0)] = d + d'. \tag{5.6.33}$$

□

Note (5.6.33), with r replaced by $2r$, follows easily from the following

Lemma 5.6. *Let $r \ge r_0(d)$. Then there exists $(n,m) \in \mathbb{Z}^{2r}$ such that*

$$\text{rank}[\nabla_{x,y} D(n,m] = d + d'. \tag{5.6.34}$$

Indeed, writing $D_r(x,y) = D(x,y) : \mathbb{R}^{2r} \to \mathbb{G}_0^{\#}$, we have that $D_r(x,y)D_r(n,m)^{-1} = D_{2r}((x,m'),(y,n'))$ with $n' = (n_r, \ldots, n_1)$, $m' = (m_r, \ldots, m_1)$. Assuming (5.6.34) it is clear that the map $D_{2r}((x,m'),(y,n'))$ has maximal rank at $z_0 = (n,m')$, $w_0 = (m,n')$ and (5.6.33) follows.
The proof of Lemma 5.6 is based on counting points $(n,m) \in [N]^{2r}$ at which the rank of the map $\nabla_{x,y} D$ drops. This was also crucial in obtaining the nilpotent Wey estimate (5.4.14).

Proof. Let N be sufficiently large w.r.t. r, d. It is enough to show that

$$|\{n \in [N]^r :\ rank[\nabla_x D(n,m)] < d + d'\}| \lesssim_{d,r} N^{(r+1)/2}, \tag{5.6.35}$$

holds uniformly for $m \in [N]^r$. Fix $m \in [N]^r$. If rank$[\nabla_x D(n,m)] < d + d'$ then by Cramer's rule there exists $b_{l_1 l_2} \in \mathbb{Z}$, $|b_{l_1 l_2}| \lesssim N^{C_d}$ with $b_{l_1 l_2} \ne 0$ for at least one $0 \le l_2 < l_1 \le d$, such that

$$\sum_{0 \le l_2 < l_1 \le d} b_{l_1 l_2}\, \partial_j D_{l_1 l_2}(n,m) = 0, \qquad \textit{for all} \quad 1 \le j \le r. \tag{5.6.36}$$

From (5.4.12) we have that $\partial_j D_{l_1 0}(n,m) = l_1 n_j^{l_1 - 1}$, and for $1 \le l_2$,

$$\partial_j D_{l_1 l_2}(n,m) = l_1 n_j^{l_1-1} \sum_{k>j} (n_k^{l_2} - m_k^{l_2}) + l_2 n_j^{l_2-1} \sum_{k<j} (n_k^{l_1} - m_k^{l_1}) - l_1 n_j^{l_1-1} m_j^{l_2}. \tag{5.6.37}$$

We want to only include terms $k \le j$ and to achieve that we introduce the parameters

$$T_l = T_l(n,m) = \sum_{k=1}^{r} (n_k^{l_2} - m_k^{l_2}), \quad \textit{for} \qquad 1 \le l < d.$$

Note that $T_l \in [-rN^{d-1}, rN^{d-1}]$. For fixed $T = (T_l)_{1 \le l < d}$, write

$$\sum_{k>j} (n_k^{l_2} - m_k^{l_2}) = T_{l_2}(n,m) - \sum_{k \le j} (n_k^{l_2} - m_k^{l_2}),$$

Substituting into (5.6.37), we obtain, up to lower degree terms in the variables $n = (n_1, \dots, n_r)$,

$$\partial_j D_{l_1 l_2}(n,m) = -l_1 \sum_{k \le j} n_j^{l_1-1} n_k^{l_2} + l_2 \sum_{k<j} n_j^{l_2-1} n_k^{l_1}, \qquad \textit{for} \quad 1 \le l_2 < d. \tag{5.6.38}$$

Thus the system in (5.6.36) takes the form

$$\sum_{0 \le l_2 < l_1 \le d} b_{l_1 l_2}\, P^{j,T,m}_{l_1 l_2}(n_1, \dots, n_j) = 0, \qquad (1 \le j \le r). \tag{5.6.39}$$

Notice that for fixed $n_1, \dots, n_{2j-2}$ with $j \le r/2$, the left side of (5.6.39) contains the monomials $b_{l_1 0}\, n_{2j}^{l_1-1}$ and $b_{l_1 l_2}\, n_{2j-1}^{l_1-1} n_{2j}^{l_2}$, and hence is nonvanishing in the variables n_{2j-1}, n_{2j}. This implies that number of solutions to (5.6.39) is at most $2dN$ in the variables n_{2j-1}, n_{2j}.
As the number of choices for parameters $b = (b_{l_1 l_2})_{0 \le l_2 < l_1 \le d}$ and $T = (T_l)_{\le l < d}$ is $\lesssim_{r,d} N^{C_d}$ (with, say $C_d = 2d(d+d')^2$), (5.6.34) follows. □

We remark that (5.6.34) together with the argument proving (5.6.33) also implies that the map $D_{2r} : \mathbb{R}^{4r} \to \mathbb{G}_0^{\#}$ is surjective. Indeed, the image of the map D_r must contain an open ball $B(g,\delta)$ thus the image of D_{2r} must contain an open ball $B(0,\delta') \subseteq B(g,\delta)B(g,\delta)^{-1}$ centered at the origin, then by homogeneity the whole space $\mathbb{G}_0^{\#}$.

Lemma 5.6 together with Hensel's lemma also crucial to show the non-vanishing of the singular series $\mathfrak{S}(h)$ for $h \in \mathcal{H}_Q$. Recall that, given a prime p, the ring of p-adic integers $\widehat{\mathbb{Z}}_p$ is defined as the completion of $\mathbb{Z}$ with respect to the p-adic metric $|m|_p = e^{-k}$, if $m = p^k u$ with $u \in \mathbb{Z},\ p \nmid u$. Then $\widehat{\mathbb{Z}}_p$ is a so-called complete valuation ring with a unique maximal ideal $I_p = p\,\widehat{\mathbb{Z}}_p$ and will write $x = 0 \mod p^k$ if $x \in p^k\,\widehat{\mathbb{Z}}_p$. We have $|x-y|_p \le max\{|x-z|_p, |y-z|_p\}$, hence a sequence $(x_j)_{j \in \mathbb{N}}$ is Cauchy if $|x_{j+1} - x_j|_p \to 0$ as $j \to \infty$.

It follows that any formal power series $g(x)$ converges at x whenever $x = 0 \mod p$. For a vector $x = (x_1, \dots, x_N) \in \widehat{\mathbb{Z}}_p^N$ we say that $x = 0 \mod p^k$ if $x_j = 0 \mod p^k$, for all $1 \le j \le N$. Then any power series $g(x) = g(x_1, \dots, x_N)$ in N variables also converges whenever $x = 0 \mod p$. Moreover, one has the inverse and implicit function theorems for power series maps $g(x) = (g_1(x), \dots, g_N(x)) : \widehat{\mathbb{Z}}_p^N \to \widehat{\mathbb{Z}}_p^N$ without constant terms. Namely, if the Jacobian of the system at origin $J_g(0) \notin I_P$ i.e. is a unit, then g has an inverse power series map $h(x) = (h_1(x), \dots, h_N(x))$, in the sense that $h(g(x)) = g(h(x)) = x$, see Proposition 5.19 in [51]. One also has a corresponding version of the implicit function theorem; for a map $g(x) = (g_{K+1}(x), \dots, g_N(x)) : \widehat{\mathbb{Z}}_p^N \to \widehat{\mathbb{Z}}_p^{N-K}$ such that $det(\frac{\partial g_i}{\partial x_j}(0))_{i,j=K+1}^N \notin I_p$, the inverse image $V_g := g^{-1}(0)$ can be parameterized as $V_g = (t_1, \dots, t_K, h_{K+1}(t), \dots, h_N(t))$ with $t = (t_1, \dots, t_K)$, which can be seen from the inverse function theorem by extending the map with $g_i(x) = x_i$ for $i = 1, \dots, K$. The following extension of the implicit function is often used to show the non-vanishing of the singular series associated to diophantine systems.

Theorem 5.5. *[Hensel's Lemma] Let $f = (f_{K+1}, \dots, f_N) : \widehat{\mathbb{Z}}_p^N \to \widehat{\mathbb{Z}}_p^{N-K}$ be a family of polynomials. Assume there exists an $a \in \widehat{\mathbb{Z}}_p^N$ and an integer $\gamma > 0$, such that*

$$f(a) = 0 \mod p^{2\gamma+1}, \tag{5.6.40}$$

moreover

$$J_f(a) = p^\gamma u,\ u \ne 0 \mod p, \tag{5.6.41}$$

where $J_f(a)$ *is the Jacobian,*

$$J_f(a) = det \left(\frac{\partial g_i}{\partial x_j}(a) \right)_{i,j=K+1}^{N}. \tag{5.6.42}$$

Then there exist power series $h = (h_{K+1}, \ldots, h_N)$ *with* $h_j(0) = 0$*, such that for all* $t = (t_1, \ldots, t_K) = 0 \mod p$*, one has that*

$$f(a + (p^{2d}t, p^d h(t)) = 0. \tag{5.6.43}$$

This means that for each $j = K+1, \ldots, N$, one has

$$f_j(a_1 + p^{2d}t_1, \ldots, a_K + p^{2d}t_K,\, a_{K+1} + p^d h_{K+1}(t), \ldots, a_N + p^d h_N(t)) = 0.$$

This is proved in [51], see Lemma 5.21 and Note 5.22 there, in fact it is shown that all $b \in \widehat{\mathbb{Z}}_p^N$ such that $b = a \mod p^{\gamma+1}$ and $f(b) = 0$ can be parameterized this way. We will use it to obtain the following lower bound, assuming conditions (5.6.40)-(5.6.41) hold.

Corollary 5.1. *Let* $n > 2\gamma$*. Then*

$$|\{b \in \mathbb{Z}_{p^n}^N : \ f(b) = 0 \mod p^n\}| \geq p^{(n-2\gamma-1)K} \tag{5.6.44}$$

Proof. First note that if $t_1 \neq t_2 \mod p^{n-2\gamma}$, then $c_1 \neq c_2 \mod p^n$, where $c_i = a + (p^{2\gamma}t_i, p^\gamma h(t_i))$ for $i = 1, 2$. There are $p^{(n-2\gamma-1)K}$ values of $t \in \mathbb{Z}^K$ such that $t = 0 \mod p$ which fall into different residue classes mod $p^{n-2\gamma}$, thus by (5.6.43) we have at least this many solutions to $f(c) = 0$ in $\widehat{\mathbb{Z}}_p^N$ which fall into different residue classes mod p^n. For each such c let $b \in \mathbb{Z}^N$ such that $b = c \mod p^n$, Then clearly $f(b) = 0 \mod p^n$ and all such $b's$ are distinct mod p^n. □

References

References

[1] A.N.Andrianov Quadratic Forms and Hecke Operators Grundlehren der mathematischen Wissenschaften, Springer-Verlag (1987)

[2] T. Austin. A proof of Walsh's convergence theorem using couplings. Int. Math. Res. Not. IMRN **15** (2015), 6661–6674.

[3] T. Austin. On the norm convergence of non-conventional ergodic averages. Ergodic Theory Dynam. Systems **30** (2010), 321–338.

[4] A. Balog Linear equations in primes Mathematika 39.2 (1992): 367-378

[5] A. Bellow. Measure Theory Oberwolfach 1981. Proceedings of the Conference held at Oberwolfach, June 21–27, 1981. Lecture Notes in Mathematics **945**, editors D. Kölzow and D. Maharam-Stone. Springer-Verlag Berlin Heidelberg (1982). Section: Two problems submitted by A. Bellow, pp. 429–431.

[6] V. Bergelson. Weakly mixing PET. Ergodic Theory Dynam. Systems **7** (1987), no. 3, 337–349.

[7] V. Bergelson. Ergodic Ramsey Theory – an update, Ergodic Theory of $\mathbb{Z}^d$-actions (edited by M. Pollicott and K. Schmidt). London Math. Soc. Lecture Note Series **228** (1996), pp. 1–61.

[8] V. Bergelson. Combinatorial and Diophantine Applications of Ergodic Theory (with appendices by A. Leibman and by A. Quas and M. Wierdl). Handbook of Dynamical Systems, Vol. 1B, B. Hasselblatt and A. Katok, eds., Elsevier, (2006), pp. 745–841.

[9] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden and Szemerédi's theorems. J. Amer. Math. Soc. **9** (1996), 725–753.

[10] V. Bergelson, A. Leibman. A nilpotent Roth theorem. Invent. Math. **147** (2002), 429–470.

[11] B.J. Birch. Forms in many variables. Proc. R. Soc. Lond. A **265** (1962), 245–263.

[12] G. Birkhoff. Proof of the ergodic theorem. Proc. Natl. Acad. Sci. USA **17** (1931), no. 12, 656–660.

[13] J. Bourgain. On the maximal ergodic theorem for certain subsets of the integers. Israel J. Math. **61** (1988), 39–72.

[14] J. Bourgain. On the pointwise ergodic theorem on L^p for arithmetic sets. Israel J. Math. **61** (1988), 73–84.

[15] J. Bourgain. Pointwise ergodic theorems for arithmetic sets, with an appendix by the author, H. Furstenberg, Y. Katznelson and D.S. Ornstein. Inst. Hautes Études Sci. Publ. Math. **69** (1989), 5–45.

[16] J. Bourgain, A Szemerédi type theorem for sets of positive density in $\mathbb{R}^k$, Israel J. Math. 54 (1986), no. 3, 307–316.

[17] J. Bourgain. Double recurrence and almost sure convergence. J. Reine Angew. Math. **404** (1990), 140–161.

[18] J. Bourgain, Exponential sums and nonlinear Schrödinger equations. Geometric and Functional Analysis GAFA 3 (1993): 157-178.

[19] J. Bourgain, Global solutions of nonlinear Schrödinger equations, American Mathematical Soc. Vol. 46 (1999)

[20] Z. Buczolich, R.D. Mauldin. Divergent square averages. Ann. Math. **171** (2010), no. 3, 1479–1530.

[21] A. Calderón. Ergodic theory and translation invariant operators. Proc. Natl. Acad. Sci. USA **59** (1968), 349–353.

[22] M. Christ. Hilbert transforms along curves: I. nilpotent groups. Ann. Math. **122/3** (1985), 575–596.

[23] M. Christ. A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral. Colloq. Math. **60/61** (1990), 601–628.

[24] M. Christ, A. Nagel, E.M. Stein, S. Wainger. Singular and maximal Radon transforms: analysis and geometry. Ann. of Math. (2) **150** (1999), no. 2, 489–577.

[25] Q. Chu, N. Frantzikinakis, B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. Proc. London. Math. Soc. **102** (2011), no. 5, 801–842.

[26] D. Conlon, J. Fox, Y. Zhao, A relative Szemeredi theorem Geometric and Functional Analysis 25.3 (2015): 733-762.

[27] B. Cook, Á. Magyar. Diophantine equations in the primes. Inventiones Mathematicae 198, (2014):701-37.

[28] B. Cook, Á. Magyar, M. Pramanik. A Roth-type theorem for dense subsets of $\mathbb{R}^d$ Bull. Lond. Math. Soc. 49 (2017), no. 4, 676-689.

[29] M. Christ, A. Nagel, E.M. Stein, S. Wainger. Singular and maximal Radon transforms: analysis and geometry Annals of Mathematics, 150(2), (1999): 489-577.

[30] P. Durcik, V. Kovač, Boxes, extended boxes, and sets of positive upper density in the Euclidean space Mathematical Proceedings of the Cambridge Philosophical Society 171/3, (2021): 481-501)

[31] L. J. Corwin, F. P. Greenleaf. Representations of nilpotent Lie groups and their applications. Part I. Basic theory and examples. Cambridge Studies in Advanced Mathematics **18**, Cambridge University Press, Cambridge (1990).

[32] H. Davenport. Cubic forms in thirty-two variables. Phil. Trans. R. Soc. Lond. A **251** (1959), 193–232.

[33] H. Davenport. Analytic methods for Diophantine equations and Diophantine inequalities. Cambridge University Press (1959)

[34] P. Erdős, R.L. Graham, P. Montgomery, B.L. Rothschild, J. Spencer, E. G. Straus. Euclidean Ramsey theorems. I, J. Combinatorial Theory Ser. A **14** (1973), 341–363.

[35] P. Erdős, R.L. Graham, P. Montgomery, B.L. Rothschild, J. Spencer, E. G. Straus. Euclidean Ramsey theorems. II, North-Holland, Amsterdam, Colloq. Math. Soc. János Bolyai, Vol. 10. (1975), 529–557.

[36] N. Frantzikinakis. Some open problems on multiple ergodic averages. Bull. Hellenic Math. Soc. **60** (2016), 41–90.

[37] N. Frantzikinakis, B. Kra. Polynomial averages converge to the product of integrals. Israel J. Math. **148** (2005), 267–276.

[38] H. Furstenberg, Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations J. Analyse Math. 31 (1978), 275-291.

[39] A. Frieze, R. Kannan. The regularity lemma and approximation schemes for dense problems. In Foundations of Computer Science (1996) Proc. 37th Annual Symp. IEEE., 12-20.

[40] H. Furstenberg, Y. Katznelson and B. Weiss, Ergodic theory and configurations in sets of positive density Mathematics of Ramsey theory, 184–198, Algorithms Combin., 5, Springer, Berlin, (199).

[41] P. Frankl and V. Rödl. A partition property of simplices in Euclidean space. J. Amer. Math. Soc. **3** no.1 (1990), 1–7.

[42] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemeredi on arithmetic progressions. J. Anal. Math. **31** (1977), 204–256.

[43] H. Furstenberg. Problems Session, Conference on Ergodic Theory and Applications University of New Hampshire, Durham, NH, June 1982.

[44] H. Furstenberg. Nonconventional ergodic averages. The legacy of John von Neumann (Hempstead, NY, 1988), Proc. Sympos. Pure Math., **50**, Amer. Math. Soc., Providence, RI, (1990), 43–56.

[45] H. Furstenberg, B. Weiss. A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. Convergence in ergodic theory and probability, Ohio State Univ. Math. Res. Inst. Publ., **5**, de Gruyter, Berlin, (1996), 193–227

[46] W. T. Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four Geom. Funct. Anal. **8** no. 3 (1998): 3529–551.

[47] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.

[48] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem Annals of Mathematics (2007), 897-946.

[49] D. Goldston, J. Pintz, Yildirim Primes in tuples I Ann. of Math. 170 (2009): 819-862

[50] R. L. Graham. Recent trends in Euclidean Ramsey theory Discrete Math. **136** no. 1-3, (1994): 119–127.

[51] M. J. Greenberg. Lectures on forms in many variables.

[52] B. Green, T. Tao Linear equations in the primes Ann. of Math.(2) 171.3 (2010): 1753-1850.

[53] B. Host, B. Kra. Non-conventional ergodic averages and nilmanifolds. Ann. Math. **161** (2005), 397–488.

[54] B. Host, B. Kra. Convergence of polynomial ergodic averages. Israel J. Math. **149** (2005), 1–19.

[55] Y.-Q. Hu Polynomial maps and polynomial sequences in groups. Available at arXiv:2105.08000.

[56] Y.-Q. Hu Waring's problem for locally nilpotent groups: the case of discrete Heisenberg groups. Available at arXiv:2011.06683.

[57] L.K. Hua. Additive theory of prime numbers Translations of Mathematical Monographs Vol. 13." Am. Math. Soc., Providence (1965)

[58] A. Ionescu, Á. Magyar, M. Mirek, T. Z. Szarek. Polynomial averages and pointwise ergodic theorems on nilpotent groups. Inventiones Mathematicae 231.3 (2023): 1023-1140.

[59] A. Ionescu, Á. Magyar, M. Mirek, T. Z. Szarek. Polynomial sequences in discrete nilpotent groups of step 2. Advanced Nonlinear Studies 23, no. 1 (2023): 20230085.

[60] A. Ionescu, Á. Magyar, E.M. Stein, S. Wainger. Discrete Radon transforms and applications to ergodic theory. Acta Math. **198** (2007), 231–298.

[61] A. Ionescu, Á. Magyar, S. Wainger. Averages along polynomial sequences in discrete nilpotent Lie groups: Singular Radon transforms.
In Advances in analysis: the legacy of Elias M. Stein, 146–188, Princeton Math. Ser. 50, Princeton Univ. Press, Princeton, NJ, 2014.

[62] A. Iosevich, M. Rudnev Erdős distance problem in vector spaces over finite fields. Trans. Amer. Math. Soc. 359/12 (2007): 6127–6142.

[63] A.D. Ionescu, S. Wainger. L^p boundedness of discrete singular Radon transforms. J. Amer. Math. Soc. **19** (2005), no. 2, 357–383.

[64] R.L. Jones, A. Seeger, J. Wright. Strong variational and jump inequalities in harmonic analysis. Trans. Amer. Math. Soc. **360** (2008), 6711–6742.

[65] B. Krause. Discrete analogoues in harmonic analysis: maximally monomially modulated singular integrals related to Carleson's theorem. Available at arXiv:1803.09431.

[66] B. Krause, M. Mirek, T. Tao. Pointwise ergodic theorems for non-conventional bilinear polynomial averages. Ann. Math. **195** (2022), no. 3, 997–1109.

[67] Y. Kitaoka. Siegel modular forms and representation by quadratic forms.
Lectures on Mathematics and Physics, Tata Institute of Fundamental Research, Springer-Verlag, (1986)

[68] H. Klingen Introductory lectures on Siegel modular forms Cambridge Studies of Advanced Mathematics 20, Cambridge Univ. Press, (1990)

[69] I. Kříž. Permutation groups in Euclidean Ramsey theory. Proc. Amer. Math. Soc. **112/3** (1991): 899-907.

[70] I. Kříž. All trapezoids are Ramsey. Discrete Math. **108** 1-3, (1992): 59-62. Topological, algebraical and combinatorial structures. Frolík's memorial volume.

[71] P. LaVictoire. Universally L^1-Bad Arithmetic Sequences. J. Anal. Math. **113** (2011), no. 1, 241–263.

[72] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. Israel J. Math. **146** (2005), 303–315.

[73] D. Lépingle. La variation d'ordre p des semi-martingales. Z. Wahrscheinlichkeitstheorie Verw. Gebiete **36** (1976), 295–316.

[74] J. Liu Integral points on quadrics with prime coordinates Monats. Math. 164.4 (2011): 439-465

[75] J. Liu, L. Zhao. On forms in prime variables. Trans. Amer. Math. Soc. 376, no. 12 (2023): 8621-8656.

[76] N. Lyall and Á. Magyar, Product of simplices and sets of positive upper density in $\mathbb{R}^d$ Math. Proc. of the Cambridge Philos. Soc. 165. no. 1. (2018), 25-51.

[77] N. Lyall and Á. Magyar, Distance Graphs and sets of positive upper density in $\mathbb{R}^d$ Analysis PDE 13.3 (2020): 685-700.

[78] N. Lyall and Á. Magyar, Distances and trees in dense subsets of $\mathbb{Z}^d$ Israel Journal of Mathematics 240.2 (2020): 769-790.

[79] N. Lyall, Á. Magyar Weak hypergraph regularity and applications to geometric Ramsey theory. Transactions of the Amer. Math. Soc., Ser. B, 9(5) (2022): 160-207.

[80] N. Lyall, Á. Magyar, H. Parshall, Spherical configurations over finite fields Amer. J. Math. 142.2 (2020): 373-404.

[81] Á. Magyar Diophantine equations and Ergodic theorems Amer. J. Math. v.124, (2002): 921-953

[82] Á. Magyar, E.M. Stein, S. Wainger. Discrete analogues in harmonic analysis: spherical averages. Ann. Math. **155** (2002), 189–208.

[83] Á. Magyar, E.M. Stein, S. Wainger. Maximal operators associated to discrete subgroups of nilpotent Lie groups. J. Anal. Math. **101** (2007), 257–312.

[84] Á. Magyar Distance sets of large sets of integer points Israel J. Math., v (2008) pp. 251-263.

[85] Á. Magyar, *k-point configurations in sets of positive density of* Z^n Duke Math. J., v 146/1, (2009) pp. 1-34.

[86] J. Maynard. Small gaps between primes. Annals of mathematics (2015): 383-413.

[87] M. Mirek, E.M. Stein, B. Trojan. $\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: Variational estimates. Invent. Math. **209** (2017), no. 3, 665–748.

[88] M. Mirek, E.M. Stein, P. Zorin-Kranich. Jump inequalities via real interpolation. Math. Ann. **376** (2020), no. 1-2, 797–819.

[89] M. Mirek, E.M. Stein, P. Zorin-Kranich. A bootstrapping approach to jump inequalities and their applications. Anal. PDE **13** (2020), no. 2, 527–558.

[90] M. Mirek, E.M. Stein, P. Zorin-Kranich. Jump inequalities for translation-invariant operators of Radon type on $\mathbb{Z}^d$. Adv. Math. **365** (2020), 107065, pp. 57.

[91] C. Muscalu, W. Schlag. Classical and multilinear harmonic analysis. Vol. II. Cambridge Studies in Advanced Mathematics **138**, Cambridge University Press, Cambridge, 2013.

[92] L. Pierce. Discrete fractional Radon transforms and quadratic forms. Duke Math. J. **161** (2012), 69–106.

[93] L. Pierce, P.-L Yung. A polynomial Carleson operator along the paraboloid. Rev. Mat. Iberoam. **35** (2019), 339–422.

[94] Polymath, D. H. J. A new bound for gaps between primes. Preprint (2013).

[95] S. Raghavan. Modular forms of degree n and representation by quadratic forms Annals of Math. 70, pp. 446-477 (1959)

[96] F. Ricci, E. M. Stein. Harmonic analysis on nilpotent groups and singular integrals I. Oscillatory integrals. J. Funct. Anal. **73** (1987), 179–194.

[97] W. M. Schmidt The density of integer points on homogeneous varieties. Acta Mathematica 154(3-4), (1985):243-96

[98] C. L. Siegel, *On the theory of indefinite quadratic forms* Ann. of Math. (2) 45 (1944), 577-622.

[99] C. Sogge Fourier integrals in classical analysis. Cambridge University Press, (1993)

[100] C. Sogge, E.M. Stein Averages of functions over hypersurfaces in $\mathbb{R}^n$ Invent. Math. 82, 543-556, (1985).

[101] E. M. Stein. Maximal functions: spherical means. Proceedings of the National Academy of Sciences, 73(7), (1976), 2174-2175.

[102] E.M. Stein. Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals. With the assistance of Timothy S. Murphy. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, NJ, 1993.

[103] E.M. Stein, S. Wainger. Discrete analogues in harmonic analysis, I: ℓ^2 estimates for singular Radon transforms. Amer. J. Math. **121** (1999), 1291–1336.

[104] E., Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 199-245.

[105] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. Acta Arith. **27** (1975), 199–245.

[106] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. Ergodic Theory Dynam. Systems **28** (2008), 657–688.

[107] T. Tao, Szemerédi's regularity lemma revisited, arXiv preprint math/0504472 (2005)

[108] T. Tao, A variant of the hypergraph removal lemma Journal of Combinatorial Theory, Series A 113.7 (2006): 1257-1280.

[109] T. Tao, The ergodic and combinatorial approaches to Szemerédi's theorem, Additive combinatorics, 145–193, CRM Proc. Lecture Notes, 43, Amer. Math. Soc., Providence, RI, (2007)

[110] T. Tao and V. Vu Additive combinatorics Cambridge Studies in Advanced Mathematics, 105 (2006)

[111] I.M. Vinogradov Representations of an odd integer as a sum of three primes Goldbach Conjecture 4 (2002): 61.

[112] J. von Neumann. Proof of the quasi-ergodic hypothesis. Proc. Natl. Acad. Sci. USA **18** (1932), 70–82.

[113] M. Walsh. Norm convergence of nilpotent ergodic averages. Ann. Math. **175** (2012), no. 3, 1667–1688.

[114] T.D. Wooley. Nested efficient congruencing and relatives of Vinogradov's mean value theorem. Proceedings of the London Mathematical Society **118.4** (2019), 942-1016.

[115] S. Yamagishi Diophantine equations in primes: Density of prime points on affine hypersurfaces. Duke Mathematical Journal 171, no. 4 (2022): 831-884.

[116] T. Ziegler. Universal characteristic factors and Furstenberg averages. J. Amer. Math. Soc. **20** (2007), 53–97.

[117] T. Ziegler, *Nilfactors of* $\mathbb{R}^m$*-actions and configurations in sets of positive upper density in* $\mathbb{R}^m$ J. Anal. Math. 99 (2006), 249-266.