

Quasipolynomial density bounds for K -point configurations in $\mathbb{Z}^d$

Andrew Lott

Ákos Magyar

Nagendar Reddy Ponagandla

September 11, 2026

Abstract

Let $d, K, N \in \mathbb{N}$ with $K \geq 3$ and $d \geq 4K + 4$. Let $\Delta \subset \mathbb{Z}^d$ be the vertex set of a nondegenerate $(K - 1)$ -simplex, and let $A \subseteq [N]^d$ contain no nontrivial similar copy of Δ . We prove that

$$|A| \ll_{\Delta, d} N^d \exp\left(-c_{\Delta, d} \sqrt{\log N}\right)$$

improving upon a polylogarithmic bound due to Magyar. We perform a density increment argument using the circle method, and we introduce a “cut operator” method to decouple the weighted exponential sum over the system of quadratic forms describing the simplex. Our proof combines ideas from graph theory, functional analysis, and the geometry of numbers. In the process, we apply Finner’s fractional form of Hölder’s inequality, the analytic large sieve, and Kim’s mean value formula for primitive lattice flags.

Contents

1	Introduction	1
2	Outline of the proof	3
3	Notation and standard definitions	9
4	The counting form	10
5	Cut operators	12
6	The minor arcs	20
7	The major arcs	25
8	The density increment and completion of the proof	36
A	Application to a continuous variant of the problem	39
B	Counting solutions to the matrix equation	40

1 Introduction

A basic problem in additive combinatorics is to determine which finite patterns must occur in dense subsets of integer lattices. At the qualitative level, the multidimensional Szemerédi theorem

of Furstenberg and Katznelson [6] gives a general answer: for every finite set $F \subseteq \mathbb{Z}^d$, every subset of $\mathbb{Z}^d$ of positive upper Banach density contains a homothetic copy $x + qF$, with $x \in \mathbb{Z}^d$ and $q \in \mathbb{N}$. The quantitative bounds obtainable from this general theorem are, however, extremely weak; see [8, 22].

Here we study the corresponding quantitative problem for similarity classes of simplices. Let

$$\Delta = \{v_0, \dots, v_{K-1}\} \subseteq \mathbb{Z}^d$$

be the vertex set of a fixed nondegenerate $(K-1)$ -simplex. We say that $\Delta' = \{x_0, \dots, x_{K-1}\} \subseteq \mathbb{Z}^d$ is a *similar copy* of Δ if

$$|x_i - x_j|^2 = t|v_i - v_j|^2 \quad (0 \leq i < j \leq K-1) \quad (1.1)$$

for some $t > 0$, or equivalently if Δ' is a translated and rotated image of a dilate of Δ . Every homothetic copy is similar, so the qualitative existence of such configurations follows from the preceding theorem; our concern is the quantitative dependence on the density.

Our main result is the following.

Theorem 1.1. *Let $K \geq 3$, and let $\Delta \subseteq \mathbb{Z}^d$ be the vertex set of a fixed nondegenerate integral $(K-1)$ -simplex. If $d \geq 4K+4$, then there are constants $C, c > 0$, depending only on Δ and d , such that every set $A \subseteq [N]^d$ containing no nontrivial similar copy of Δ satisfies*

$$|A| \leq CN^d \exp(-c\sqrt{\log N}).$$

In the common range $d \geq 4K+4$, this improves an earlier result of the second author [17], which states that, when $d \geq 2K+3$, every set $A \subseteq [N]^d$ containing no nontrivial similar copy of Δ satisfies

$$|A| \leq C_{\Delta,d} N^d (\log N)^{-2/(11K)}.$$

To the best of our knowledge, this is the only previous quantitative result formulated for similar copies of an arbitrary nondegenerate integral simplex.

By Chebyshev's bound, theorem 1.1 immediately gives the following quantitative relative result in the prime lattice; compare this to [3, 5, 23]. Let $\mathbb{P}$ denote the set of primes.

Corollary 1.2. *Let $K \geq 3$, and let $\Delta \subseteq \mathbb{Z}^d$ be the vertex set of a fixed nondegenerate integral $(K-1)$ -simplex. If $d \geq 4K+4$, then there are constants $C, c > 0$, depending only on Δ and d , such that every set $A \subseteq (\mathbb{P} \cap [N])^d$ containing no nontrivial similar copy of Δ satisfies*

$$|A| \leq C|\mathbb{P} \cap [N]|^d \exp(-c\sqrt{\log N}).$$

An equivalent form of our main result is that a set $A \subseteq [N]^d$ of density α must contain a similar copy once

$$N \geq \exp(C_{\Delta,d} (1 + \log(1/\alpha))^2). \quad (1.2)$$

Theorem 1.1 also directly implies the following quantitative result in the Euclidean setting.

Corollary 1.3. *Let $K \geq 3$, $d \geq 4K+4$, and let $\Delta \subseteq \mathbb{Z}^d$ be as above. For every measurable $E \subseteq [0, 1]^d$ with $|E| \geq \alpha > 0$, there are $x \in \mathbb{R}^d$, $U \in O(d)$, and*

$$\lambda \geq \exp(-C_{\Delta,d} (1 + \log(1/\alpha))^2)$$

such that $x + \lambda U \Delta \subseteq E$.

A short proof, based on averaging Theorem 1.1 over translates of a sufficiently fine lattice, is given in Appendix A. For comparison, a polylogarithmic bound follows from the proof of Bourgain's simplex theorem in dimensions $d \geq K$ [1, 16]. The cut-operator method also has a direct and simpler implementation in the Euclidean setting, which should permit a lower dimension threshold; we do not pursue this direction here.

Beyond [17], quantitative results in the literature concern special configurations or indirect parametrizations by nonsingular matrix progressions. A notable example is the corners configuration, consisting of axis-parallel isosceles right triangles. Building on ideas originating in the work of Kelley and Meka [11], Jaber, Liu, Lovett, Ostuni and Sawhney [10] recently proved that every corner-free set $A \subseteq [N]^2$ satisfies

$$|A| \leq N^2 \exp(-c(\log N)^{1/600}).$$

For an integral planar triangle, complex multiplication gives a parametrization of its similar copies in the form x , $x + T_1 y$, $x + T_2 y$, where T_1, T_2 , and $T_1 - T_2$ are nonsingular rational matrices. Thus the problem becomes a matrix-valued version of Roth's theorem. Quaternionic constructions, and more generally integral orthogonal pairwise anticommuting Hurwitz–Radon matrices [7], similarly give four-term parametrizations of tetrahedral shapes after passage to a suitable auxiliary lattice. Prendiville's theorem [19] then yields a density bound of the form $(\log \log N)^{-c}$ for these parametrizations.

This Hurwitz–Radon construction extends to arbitrary nondegenerate simplices, and more generally to configurations of fixed affine rank, but it has an unavoidable dimensional cost. The sharp Hurwitz–Radon theorem forces a $(K - 1)$ -simplex represented in this way to use matrices of order $\exp(\Omega(K))$; Geramita and Pullman [7] show that the optimal families may be chosen integral. Prendiville's work also suggests that longer nonsingular matrix progressions should be accessible through higher-order Fourier analysis. Together with the recent quasipolynomial inverse theorem of Leng, Sah and Sawhney [15, 14], this suggests that one might hope for a density bound of the form

$$\exp(-(\log \log N)^{cK})$$

for each fixed nonsingular matrix progression. Such a result is not presently available: a multidimensional localized density-increment argument would still have to be developed. Even if carried out, this route would give a substantially weaker dependence on N than Theorem 1.1 and, through the Hurwitz–Radon parametrization, would require an exponentially large dimension.

2 Outline of the proof

2.1 Setup

Let $F_0, \dots, F_{K-1} : [N]^d \rightarrow [-1, 1]$, extended by zero to $\mathbb{Z}^d$, and $L := \binom{K}{2}$, and let $w = \mathbf{1}_{[1,2]}$. Fix $R = \lfloor \sigma N \rfloor$ for some sufficiently small constant $0 < \sigma < 1$ depending only on Δ , assume N is sufficiently large depending on Δ and d , and write $w_R(n) = R^{-2}w(n/R^2)$. We count weighted solutions to equation (1.1) for $t = n \in \mathbb{N}$, $n \asymp R^2$, via the multilinear expression

$$\Lambda_N(F_0, \dots, F_{K-1}) = \sum_n w_R(n) \sum_{\substack{z_0, \dots, z_{K-1} \in [N]^d \\ |z_i - z_j|^2 = n |v_i - v_j|^2 \ (i < j)}} \prod_{i=0}^{K-1} F_i(z_i). \quad (2.1)$$

Kitaoka's estimate [13] supplies

$$\Lambda_N(\mathbf{1}_{[N]^d}, \dots, \mathbf{1}_{[N]^d}) \gtrsim N^d R^{(K-1)d-2L}. \quad (2.2)$$

Here $(K-1)d$ is the number of relative variables, while $2L = 2\binom{K}{2}$ is the contribution of the $\binom{K}{2}$ quadratic equations.

Suppose that $A \subset [N]^d$ has density $\alpha > 0$ and contains no nontrivial similar copy of Δ . Then $\Lambda_N(\mathbf{1}_A, \dots, \mathbf{1}_A) = 0$. Thus, writing $\mathbf{1}_A = \alpha \mathbf{1}_{[N]^d} + g$, the pigeonhole principle yields, for some $F_0, \dots, F_{K-1}$,

$$|\Lambda_N(F_0, \dots, F_{K-1})| \gtrsim \alpha^s N^d R^{(K-1)d-2L}, \quad (2.3)$$

with exactly $s > 0$ of the F_i equal to g and the remaining F_i equal to $\mathbf{1}_{[N]^d}$. The rest of the proof is to turn this lower bound into a density increment.

As is customary in the circle method, one expresses the quadratic relations via the integrals

$$\mathbf{1}_{|z_i - z_j|^2 = n|v_i - v_j|^2} = \int_{\mathbb{T}} e(\beta_{ij}|z_i - z_j|^2) e(-n\beta_{ij}|v_i - v_j|^2) d\beta_{ij}, \quad e(t) := \exp(2\pi it).$$

This gives

$$\Lambda_N(F_0, \dots, F_{K-1}) = \int_{\mathbb{T}^L} \widehat{w}_R \left(\sum_{i < j} |v_i - v_j|^2 \beta_{ij} \right) \mathcal{S}_\beta(F_0, \dots, F_{K-1}) d\beta, \quad \|\widehat{w}_R\|_\infty \lesssim 1, \quad (2.4)$$

where $\widehat{w}_R(t) := \sum_n w_R(n) e(-nt)$ and

$$\mathcal{S}_\beta(F_0, \dots, F_{K-1}) = \sum_{z_0, \dots, z_{K-1} \in [N]^d} e \left(\sum_{i < j} \beta_{ij} |z_i - z_j|^2 \right) \prod_{i=0}^{K-1} F_i(z_i). \quad (2.5)$$

2.2 The cut-operator reduction

Write $\mathcal{E}$ for the edge set of the complete graph on K vertices, and note that $L = |\mathcal{E}| = \binom{K}{2}$. We use the convention $\beta_{ji} = \beta_{ij}$ for $i < j$.

Fix a balanced cut $S \mid T$, that is, a partition of the vertex set into two classes whose sizes differ by at most one. The phase $\mathcal{Q}_\beta(z)$ in (2.5) decomposes as

$$\mathcal{Q}_\beta(z) = \mathcal{Q}_\beta^S(z_S) + \mathcal{Q}_\beta^T(z_T) - 2 \sum_{i \in S, j \in T} \beta_{ij} z_i \cdot z_j,$$

where $z_S = (z_i)_{i \in S}$ and $z_T = (z_j)_{j \in T}$. The first two terms depend on only one side of the cut. For fixed $\Theta \in \mathbb{T}^{|S||T|}$, the crossing term naturally leads to the operator

$$\mathcal{K}_{R,d}^{S,T}(\Theta) : \ell^2([R]^d)^T \rightarrow \ell^2([R]^d)^S$$

defined by

$$(\mathcal{K}_{R,d}^{S,T}(\Theta)h)(z_S) := \sum_{z_T \in ([R]^d)^T} e \left(\sum_{i \in S, j \in T} \Theta_{ij} z_i \cdot z_j \right) h(z_T).$$

We call $\mathcal{K}_{R,d}^{S,T}(\Theta)$ a *cut operator*. Because the dot product separates coordinatewise, $\mathcal{K}_{R,d}^{S,T}(\Theta)$ is the d -fold tensor product of the operator $\mathcal{K}_R^{S,T}(\Theta) := \mathcal{K}_{R,1}^{S,T}(\Theta)$. For ease of notation, we work with the normalized operator norm

$$m_R^{S,T}(\Theta) := R^{-K/2} \|\mathcal{K}_R^{S,T}(\Theta)\|_{2 \rightarrow 2}, \quad (0 \leq m_R^{S,T}(\Theta) \leq 1). \quad (2.6)$$

Let $\beta_{S,T} = (\beta_{ij})_{i \in S, j \in T}$ denote the family of crossing frequencies of the cut $S \mid T$. A key observation is to write

$$\mathcal{S}_\beta(F_0, \dots, F_{K-1}) = \langle U_S \mathcal{K}_{N,d}^{S,T}(-2\beta_{S,T}) U_T H_T, H_S \rangle,$$

where

$$H_S(z_S) := \prod_{i \in S} F_i(z_i), \quad H_T(z_T) := \prod_{j \in T} F_j(z_j),$$

and U_S, U_T are the unitary multiplication operators

$$(U_S h)(z_S) := e(\mathcal{Q}_\beta^S(z_S)) h(z_S), \quad (U_T h)(z_T) := e(\mathcal{Q}_\beta^T(z_T)) h(z_T).$$

Since $\|H_S\|_2 = \prod_{i \in S} \|F_i\|_2$ and $\|H_T\|_2 = \prod_{j \in T} \|F_j\|_2$, the Cauchy–Schwarz inequality, unitarity, and the multiplicativity of the $\ell^2 \rightarrow \ell^2$ operator norm with respect to tensor products imply that

$$|\mathcal{S}_\beta(F_0, \dots, F_{K-1})| \leq N^{Kd/2} m_N^{S,T}(-2\beta_{S,T})^d \prod_{i=0}^{K-1} \|F_i\|_2. \quad (2.7)$$

Let $\mathcal{B}_K$ be the set of balanced cuts; we introduce the notation

$$w_K := \frac{L}{|S||T||\mathcal{B}_K|}, \quad \tau := \frac{d|S||T|}{2L}.$$

By symmetry it is easy to see that

$$w_K \# \{S \mid T \in \mathcal{B}_K : e \text{ crosses } S \mid T\} = 1 \quad (e \in \mathcal{E}). \quad (2.8)$$

Since $2\tau w_K = d/|\mathcal{B}_K|$, taking the geometric mean of (2.7) over $S \mid T \in \mathcal{B}_K$ yields

$$|\mathcal{S}_\beta(F_0, \dots, F_{K-1})| \leq N^{Kd/2} \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K} \prod_{i=0}^{K-1} \|F_i\|_2. \quad (2.9)$$

The problem of estimating the exponential sum $\mathcal{S}_\beta$ has thus been reduced to understanding the normalized one-dimensional cut-operator norms.

2.3 The mean-value estimate

The first principal estimate is a high-moment bound for the cut operator norm.

Lemma 2.1. *Let $S \mid T$ be a balanced cut. For every positive integer R ,*

$$\int_{\mathbb{T}^{|S||T|}} m_R^{S,T}(\Theta)^{2K+4} d\Theta \ll_K R^{-2|S||T|}. \quad (2.10)$$

The power $R^{-2|S||T|}$ is the natural saving associated with the $|S||T|$ crossing frequencies. To prove the lemma, we observe first that

$$\|\mathcal{K}_R^{S,T}(\Theta)\|_{2 \rightarrow 2}^{2K+4} = \|(\mathcal{K}_R^{S,T}(\Theta) \mathcal{K}_R^{S,T}(\Theta)^*)^{K+2}\|_{2 \rightarrow 2}$$

and that the latter norm is bounded by the trace of $(\mathcal{K}_R^{S,T}(\Theta) \mathcal{K}_R^{S,T}(\Theta)^*)^{K+2}$. Expanding the trace and integrating in Θ reduces the problem, up to at most R^K choices of base points, to counting bounded integer matrices

$$U \in \mathbb{Z}^{|S| \times (K+1)}, \quad V \in \mathbb{Z}^{|T| \times (K+1)}, \quad UV^\top = 0. \quad (2.11)$$

We follow a geometric route exploiting the special structure of this matrix equation. In particular, after treating $U = 0$ and $V = 0$ separately, the saturated row lattices Γ_U and Γ_V given by U and V

determine the primitive 2-step lattice flag $\Gamma_U \subseteq \Gamma_V^{\perp_Z} \subseteq \mathbb{Z}^{K+1}$, where $\Gamma_V^{\perp_Z}$ is the integral orthogonal complement of Γ_V . Thus we reduce the count of solutions to the matrix equation to counting primitive 2-step lattice flags subject to certain constraints. Indeed, organizing the count by its ranks and covolumes, and combining geometry of numbers with Kim's mean-value formula for primitive lattice flags [12], gives the upper bound

$$\#\{(U, V) : UV^\top = 0, \|U\|_\infty, \|V\|_\infty \ll R\} \ll_K R^{K(K+1)-2|S||T|}.$$

Combining this with the R^K choices of base points gives precisely the bound required in Lemma 2.1 after normalization by $R^{-K(K+2)}$.

Finner's fractional form of Hölder's inequality [4] turns the one-cut estimates into the global estimate used in both the major and minor arc estimates.

Corollary 2.2. *If $\tau \geq K + 2$, then*

$$\int_{\mathbb{T}^L} \prod_{S|T \in \mathcal{B}_K} m_R^{S,T}(-2\theta_{S,T})^{2\tau w_K} d\theta \ll_{K,d} R^{-2L}. \quad (2.12)$$

Indeed, the covering condition in Finner's inequality is exactly (2.8). Since $m_R^{S,T} \leq 1$ and $2\tau \geq 2K + 4$, Lemma 2.1 gives

$$\begin{aligned} \int_{\mathbb{T}^L} \prod_{S|T \in \mathcal{B}_K} m_R^{S,T}(-2\theta_{S,T})^{2\tau w_K} d\theta &\leq \prod_{S|T \in \mathcal{B}_K} \left(\int_{\mathbb{T}^{|S||T|}} m_R^{S,T}(\Theta)^{2\tau} d\Theta \right)^{w_K} \\ &\ll \prod_{S|T \in \mathcal{B}_K} R^{-2|S||T|w_K} = R^{-2L}. \end{aligned}$$

For $K = 4$, the mechanism is especially transparent. There are six edge frequencies and three balanced $2 \mid 2$ cuts,

$$01 \mid 23, \quad 02 \mid 13, \quad 03 \mid 12.$$

Each cut sees four edges, every edge crosses exactly two cuts, and $w_4 = 1/2$. Lemma 2.1 is a twelfth-moment estimate, and Finner's inequality gives the bound

$$(R^{-8})^{1/2} (R^{-8})^{1/2} (R^{-8})^{1/2} = R^{-12} = R^{-2\binom{4}{2}}.$$

It is natural to ask whether the lattice-flag count can be replaced by a general circle-method or modular-form estimate. Because (2.11) is a singular system with low-rank pencils, direct application of the general circle-method estimates in [20, 9] would require taking a trace moment of order proportional to K^2 and would ultimately lead to a dimension hypothesis $d \gg K^2$. Kitaoka's theory of Siegel modular forms treats related positive-definite representation problems [13], but we are unaware of a modular-form estimate that directly supplies the required upper bound for this singular, indefinite matrix equation. The lattice-flag argument exploits its bilinear structure and is what allows the dimension threshold to remain linear in K .

2.4 The major and minor arc estimates

The major arcs.

Fix a suitable parameter $2 \leq Q \leq R$. On a major arc, after passing to a common denominator, write $\beta = a/q + \theta$, where $q \leq 2Q^L$, $(a, q) = 1$, and $\|\theta\|_\infty \leq Q/R^2$. We may approximate $\mathcal{S}_{a/q+\theta}$ at

the cost of a negligible error, by a major arc form, consisting of a complete-sum component and an archimedean component:

$$\mathcal{M}_{q,a,\theta}(G_0, \dots, G_{K-1}) = \mathbb{E}_{r \bmod q} \sum_{u \in (\mathbb{Z}^d)^K} e_q(\mathcal{Q}_a(r)) e(\mathcal{Q}_\theta(u)) \prod_{i=0}^{K-1} G_i(u_i, r_i),$$

where each G_i is obtained from the convolution of F_i with the normalized counting measure μ_q on a q -spaced box, defined below. More precisely, $G_i(u, r) = (\mu_q * F_i)(y)$, where $y \equiv r \pmod{q}$ and $u - y \in \{0, \dots, q-1\}^d$.

Define an operator

$$\mathcal{K}_{q,a,d}^{S,T} : \ell^2((\mathbb{Z}/q\mathbb{Z})^d)^T \longrightarrow \ell^2((\mathbb{Z}/q\mathbb{Z})^d)^S$$

with normalized counting measure on both spaces, given by

$$(\mathcal{K}_{q,a,d}^{S,T} H)(r_S) := \mathbb{E}_{r_T \bmod q} e_q \left(-2 \sum_{i \in S, j \in T} a_{ij} r_i \cdot r_j \right) H(r_T).$$

By a similar cut reduction, for each cut $S \mid T$, $\mathcal{M}_{q,a,\theta}$ is controlled by $\mathcal{K}_{q,a,d}^{S,T} \otimes \mathcal{K}_{P,d}^{S,T}(-2\theta_{S,T})$, where P is a suitably chosen parameter proportional to N . In this way, $\mathcal{M}_{q,a,\theta}$ itself does not factor, but the cut operator is a tensor product and thus its operator norm does factor:

$$\|\mathcal{K}_{q,a,d}^{S,T} \otimes \mathcal{K}_{P,d}^{S,T}(-2\theta_{S,T})\|_{2 \rightarrow 2} = \|\mathcal{K}_{q,a,d}^{S,T}\|_{2 \rightarrow 2} \|\mathcal{K}_{P,d}^{S,T}(-2\theta_{S,T})\|_{2 \rightarrow 2}.$$

Hence, the major arc analysis reduces to understanding the operator norms of $\mathcal{K}_{q,a,d}^{S,T}$ and $\mathcal{K}_{P,d}^{S,T}(-2\theta_{S,T})$.

An elementary orthogonality argument gives

$$\|\mathcal{K}_{q,a,d}^{S,T}\|_{2 \rightarrow 2} = |\text{im}(2a_{S|T})|^{-d/2}, \quad (2.13)$$

where the matrix $2a_{S|T} = (2a_{ij})_{i \in S, j \in T}$ is regarded as a homomorphism $(\mathbb{Z}/q\mathbb{Z})^T \rightarrow (\mathbb{Z}/q\mathbb{Z})^S$. Corollary 2.2 treats $\mathcal{K}_{P,d}^{S,T}$, and we treat $\mathcal{K}_{q,a,d}^{S,T}$ via row and column operations and an elementary counting argument.

The minor arcs.

The minor arcs require an additional pointwise estimate. For $\xi \in \mathbb{T}$, set

$$\mathcal{K}_R(\xi) = (e(\xi r s))_{r,s \in [R]}, \quad m_R(\xi) = R^{-1} \|\mathcal{K}_R(\xi)\|_{2 \rightarrow 2}.$$

After isolating the variables attached to a crossing edge $e = (i, j)$, a simple application of Cauchy–Schwarz in the remaining variables gives

$$m_R^{S,T}(\Theta) \leq m_R(\Theta_{ij}) \quad (i \in S, j \in T). \quad (2.14)$$

Lemma 2.3. *Let $2 \leq Q \leq R$. Suppose that $\xi \in \mathbb{T}$ admits no coprime rational approximation a/q satisfying*

$$1 \leq q \leq Q, \quad \|\xi - a/q\|_{\mathbb{T}} \leq \frac{Q}{qR^2}.$$

Then

$$m_R(\xi)^2 \ll Q^{-1}. \quad (2.15)$$

Lemma 2.3 is a Weyl-type estimate; we prove it using the large sieve inequality. Let $\mathfrak{m}_Q$ be the set of $\beta \in \mathbb{T}^L$ for which at least one edge frequency $-2\beta_e$ satisfies the hypothesis of the lemma at scale R . Since $R \leq N$, it also satisfies that hypothesis at scale N . Hence, if e is such an edge, every cut crossing e satisfies $m_N^{S,T}(-2\beta_{S,T}) \ll Q^{-1/2}$. Under the dimension hypothesis,

$$\gamma := \tau - (K + 2) > 0.$$

We use the mean value estimate on the first $2K + 4$ powers and the pointwise estimate on the remaining 2γ powers. Finner's inequality and (2.8) give

$$\int_{\mathfrak{m}_Q} \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K} d\beta \ll_{K,d} Q^{-\gamma} N^{-2L}. \quad (2.16)$$

2.5 Extracting the density increment

It remains to explain why the major arcs retain useful combinatorial information. Let μ_q be normalized counting measure on a long q -spaced box,

$$\mu_q = N_q^{-d} \mathbf{1}_{q\{0, \dots, N_q-1\}^d}, \quad N_q := \left\lfloor \frac{R}{qQ^{L(L+2)+3}} \right\rfloor.$$

Denote by Λ_N^{maj} the contribution to (2.4) from the major arcs. The major-arc estimate gives, for some $\eta = \eta(K, d) > 1$,

$$\begin{aligned} |\Lambda_N^{\text{maj}}(F_0, \dots, F_{K-1})| &\ll_{\Delta, d} R^{(K-1)d-2L} \sum_{q \leq 2Q^L} q^{-1-\eta} \prod_{i=0}^{K-1} \|\mu_q * F_i\|_{\ell^K(\mathbb{Z}^d)} \\ &\quad + Q^{-2} R^{Kd/2-2L} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \end{aligned} \quad (2.17)$$

The first term is the important one, as $\mu_q * g$ gives information about the density of A on a q -spaced box.

Returning to (2.3), choose Q to be a sufficiently large fixed power of α^{-1} , multiplied by a sufficiently large constant. Unless $N \ll_{\Delta, d} \alpha^{-C}$, the scale conditions above hold. The minor-arc contribution bounded using (2.16) and the error term in (2.17) are then negligible. Since

$$\|\mu_q * \mathbf{1}_{[N]^d}\|_{\ell^K(\mathbb{Z}^d)} \leq N^{d/K}$$

and $\sum_{q \geq 1} q^{-1-\eta} < \infty$, the main term in (2.17) implies that, for some $q \ll_{\Delta, d} \alpha^{-C}$,

$$\sum_{x \in \mathbb{Z}^d} |\mu_q * g(x)|^K \gg_{\Delta, d} \alpha^K N^d.$$

A standard positive-negative part argument, using $\mu_q * g \geq -\alpha$ and $\sum_x \mu_q * g(x) = 0$, gives a point $y \in [1 + q(N_q - 1), N]^d$ for which

$$(\mu_q * \mathbf{1}_A)(y) > (1 + \kappa)\alpha,$$

where $\kappa > 0$ depends only on Δ and d . Thus either $N \ll_{\Delta, d} \alpha^{-C}$, or A has density at least $(1 + \kappa)\alpha$ on a d -dimensional progression of common difference $q \ll \alpha^{-C}$ and side length $N' \gg \alpha^C N$. Translating and rescaling this progression produces another Δ -free set of increased density. The standard iteration has $O(1 + \log(1/\alpha))$ steps, each costing $O(1 + \log(1/\alpha))$ in the logarithm of the scale. Hence

$$\log N \ll_{\Delta, d} (1 + \log(1/\alpha))^2,$$

which is the content of Theorem 1.1.

2.6 Further directions

The cut operator method may offer a route to quantitative results for more complex configurations. A natural test case is a simplex together with its barycenter, or more general barycentric configurations consisting of the vertices of a simplex together with the barycenters of all of its ℓ -dimensional faces. We leave these configurations as a possible direction for future work.

2.7 Organization of the proof

Section 4 introduces the weighted counting form. Section 5 develops the cut reduction and proves the mean value estimate. Sections 6 and 7 establish the minor- and major-arc estimates. Section 8 proves and iterates the density increment. The lattice-flag count underlying Lemma 2.1 is proved in Appendix B.

3 Notation and standard definitions

We use the following notation throughout the paper.

- $\mathbb{N} = \{1, 2, \dots\}$, $[N] = \{1, \dots, N\}$ for a positive integer N , and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. We write $e(x) := e^{2\pi i x}$ and $e_q(x) := e(x/q)$.
- For $x \in \mathbb{R}$, let $\|x\|_{\mathbb{T}} := \min_{n \in \mathbb{Z}} |x - n|$. If $x = (x_1, \dots, x_m)$ lies in $\mathbb{R}^m$, then $\|x\|_{\infty} := \max_j |x_j|$, while for $x \in \mathbb{T}^m$ we put $\|x\|_{\infty} := \max_j \|x_j\|_{\mathbb{T}}$. The notation $\text{diam}_{\infty}(E)$ refers to diameter with respect to the maximum norm.
- For a set E , $\mathbf{1}_E$ denotes its indicator function. If X is finite, $\mathbb{E}_{x \in X}$ denotes normalized averaging over X .
- For functions $\mu, F : G \rightarrow \mathbb{C}$, define their convolution by

$$(\mu * F)(y) := \sum_{x \in G} \mu(x) F(y - x).$$

- If $T : H_1 \rightarrow H_2$ is a bounded linear operator between Hilbert spaces, its adjoint $T^* : H_2 \rightarrow H_1$ is determined by $\langle Tf, g \rangle_{H_2} = \langle f, T^*g \rangle_{H_1}$. We define

$$\|T\|_{2 \rightarrow 2} := \sup_{\substack{f \in H_1 \\ \|f\|_{H_1} = 1}} \|Tf\|_{H_2}.$$

If $(e_j)_j$ is any orthonormal basis of H_1 , then

$$\|T\|_{\text{HS}} := \left(\sum_j \|Te_j\|_{H_2}^2 \right)^{1/2}.$$

This value is independent of the choice of orthonormal basis.

- If $a \in (\mathbb{Z}/q\mathbb{Z})^m$, then $(a, q) = 1$ means that the coordinates of a together with q have no common divisor greater than 1.
- We use $X \ll Y$ and $X = O(Y)$ interchangeably to mean that $|X| \leq CY$ for some constant $C > 0$. We write $X \gg Y$ if $Y \ll X$, and $X \asymp Y$ if both $X \ll Y$ and $Y \ll X$. Subscripts indicate the permitted dependence of the implied constants.

4 The counting form

Recall the definition of Δ from the introduction, and suppose that $A \subseteq [N]^d$ contains no copy similar to Δ . In this section we define a weighted count for copies of Δ , extract a large term involving the balanced function, and then express the count by Fourier inversion.

4.1 The simplex equations

Put $L := \binom{K}{2}$. Let

$$\mathcal{E} := \{\{i, j\} : 0 \leq i < j \leq K-1\}, \quad |\mathcal{E}| = L.$$

We count configurations $z = (z_0, \dots, z_{K-1}) \in (\mathbb{Z}^d)^K$ satisfying

$$|z_i - z_j|^2 = n|v_i - v_j|^2 \quad (\forall \{i, j\} \in \mathcal{E}) \quad (4.1)$$

for some integer $n > 0$. These equations say that the points $z_0, \dots, z_{K-1}$ form a similar copy of Δ with dilation factor $\sqrt{n}$. It is enough to consider integer values of the squared dilation factor: a set containing no similar copy of Δ contains, in particular, no such copy with integer squared dilation factor. Moreover, Kitaoka's simplex count, which we apply shortly, counts precisely the copies occurring at these scales.

For an integer $R \geq 1$, define

$$w_R(n) := R^{-2} \mathbf{1}_{\{R^2 \leq n < 2R^2\}}, \quad n \in \mathbb{Z}.$$

Thus w_R restricts the squared dilation factor to $n \asymp R^2$ and $\sum_n w_R(n) = 1$. Put

$$D_\Delta := \max_{\{i, j\} \in \mathcal{E}} |v_i - v_j|, \quad D_0 := 1 + \left\lfloor \sqrt{2} D_\Delta \right\rfloor.$$

Whenever (4.1) holds and $w_R(n) \neq 0$, we have

$$|z_i - z_j| = \sqrt{n} |v_i - v_j| < \sqrt{2} R D_\Delta < D_0 R \quad (0 \leq i < j \leq K-1). \quad (4.2)$$

Fix a constant $0 < \sigma < 1/(4D_0)$, depending only on Δ . Choose a fixed threshold $N_0 = N_0(\Delta, d) \geq 2/\sigma$, large enough for the representation estimate (Lemma 4.1) used below. After enlarging the theorem's implicit constant, we may assume $N \geq N_0$; otherwise the conclusion is immediate. Put

$$R = \lfloor \sigma N \rfloor, \quad \alpha = \frac{|A|}{N^d}. \quad (4.3)$$

4.2 The weighted count and the balanced function

For finitely supported functions $F_0, \dots, F_{K-1} : \mathbb{Z}^d \rightarrow \mathbb{C}$, define the simplex counting form

$$\Lambda_N(F_0, \dots, F_{K-1}) := \sum_{z_0, \dots, z_{K-1} \in \mathbb{Z}^d} \sum_{n \in \mathbb{Z}} w_R(n) \mathbf{1}_{|z_i - z_j|^2 = n|v_i - v_j|^2} (\forall \{i, j\} \in \mathcal{E}) \prod_{i=0}^{K-1} F_i(z_i). \quad (4.4)$$

Lemma 4.1 (Kitaoka's simplex count). *If $d \geq 4K + 4$ and $N \geq N_0$, then*

$$\Lambda_N(\mathbf{1}_{[N]^d}, \dots, \mathbf{1}_{[N]^d}) \gg_{\Delta, d} N^d R^{(K-1)d-2L}. \quad (4.5)$$

Proof. Put $m_0 = 0$, and let

$$\mathcal{I}_N := \{x \in [N]^d : 1 + D_0 R \leq x_j \leq N - D_0 R \text{ for } 1 \leq j \leq d\}.$$

Since $D_0 R < N/4$, we have $|\mathcal{I}_N| \gg_d N^d$.

Fix $x \in \mathcal{I}_N$, and write $z_i = x + m_i$. Since $z_i - z_j = m_i - m_j$, the configuration equations are

$$|m_i - m_j|^2 = n|v_i - v_j|^2 \quad (\{i, j\} \in \mathcal{E}).$$

Because $m_0 = 0$, simple rearranging shows that these equations are equivalent to

$$m_i \cdot m_j = n(v_i - v_0) \cdot (v_j - v_0) \quad (1 \leq i, j \leq K-1).$$

Kitaoka's estimate [13, Theorems A–C], in the form recorded in [17, (2.16) and (3.10)], therefore gives, whenever $R^2 \leq n < 2R^2$,

$$\# \left\{ (m_1, \dots, m_{K-1}) \in (\mathbb{Z}^d)^{K-1} : \begin{array}{l} |m_i - m_j|^2 = n|v_i - v_j|^2 \text{ for every } \{i, j\} \in \mathcal{E} \end{array} \right\} \gg_{\Delta, d} n^{(K-1)(d-K)/2}.$$

By (4.2), every such m_i satisfies $|m_i| < D_0 R$, so $x + m_i \in [N]^d$ for every $x \in \mathcal{I}_N$. There are R^2 integers n in this range, and $w_R(n) = R^{-2}$ for each of them. Hence

$$\Lambda_N(\mathbf{1}_{[N]^d}, \dots, \mathbf{1}_{[N]^d}) \gg_{\Delta, d} N^d \cdot R^2 \cdot R^{-2} \cdot R^{(K-1)(d-K)} = N^d R^{(K-1)d-2L} = N^d R^{(K-1)d-2L}.$$

□

Since A contains no similar copy of Δ , we immediately have

$$\Lambda_N(\mathbf{1}_A, \dots, \mathbf{1}_A) = 0. \quad (4.6)$$

We now extract an important consequence of (4.6). There is nothing to prove if $A = \emptyset$, so assume that $\alpha > 0$ and write $g := \mathbf{1}_A - \alpha \mathbf{1}_{[N]^d}$. Then

$$\sum_{x \in \mathbb{Z}^d} g(x) = 0, \quad |g| \leq 1, \quad \|g\|_{\ell^2(\mathbb{Z}^d)}^2 = N^d \alpha(1 - \alpha) \leq \alpha N^d. \quad (4.7)$$

For each $\mathcal{C} \subseteq \{0, \dots, K-1\}$, let $F_i^{\mathcal{C}} = g$ if $i \in \mathcal{C}$ and $F_i^{\mathcal{C}} = \mathbf{1}_{[N]^d}$ otherwise. Multilinearity and (4.6) give the exact expansion

$$0 = \sum_{\mathcal{C} \subseteq \{0, \dots, K-1\}} \alpha^{K-|\mathcal{C}|} \Lambda_N(F_0^{\mathcal{C}}, \dots, F_{K-1}^{\mathcal{C}}). \quad (4.8)$$

The term with $\mathcal{C} = \emptyset$ is $\alpha^K \Lambda_N(\mathbf{1}_{[N]^d}, \dots, \mathbf{1}_{[N]^d})$. Moving this term to the other side, taking absolute values, and using (4.5), we obtain

$$\sum_{\substack{\mathcal{C} \subseteq \{0, \dots, K-1\} \\ \mathcal{C} \neq \emptyset}} \alpha^{K-|\mathcal{C}|} |\Lambda_N(F_0^{\mathcal{C}}, \dots, F_{K-1}^{\mathcal{C}})| \gg_{\Delta, d} \alpha^K N^d R^{(K-1)d-2L}.$$

There are only $2^K - 1$ terms in this sum. Hence some nonempty $\mathcal{B} \subseteq \{0, \dots, K-1\}$, with $s := |\mathcal{B}|$ satisfying $1 \leq s \leq K$, has the property that, writing $F_i = F_i^{\mathcal{B}}$,

$$|\Lambda_N(F_0, \dots, F_{K-1})| \gg_{\Delta, d} \alpha^s N^d R^{(K-1)d-2L}, \quad (4.9)$$

where $F_i = g$ for $i \in \mathcal{B}$ and $F_i = \mathbf{1}_{[N]^d}$ otherwise.

4.3 Applying the orthogonality relation

For $\beta = (\beta_{ij})_{\{i,j\} \in \mathcal{E}} \in \mathbb{T}^L$, we use the convention $\beta_{ji} = \beta_{ij}$. For a tuple $u = (u_0, \dots, u_{K-1})$, set

$$\mathcal{Q}_\beta(u) := \sum_{\{i,j\} \in \mathcal{E}} \beta_{ij} |u_i - u_j|^2.$$

We apply the orthogonality relation to detect solutions to the L equations in the definition of Λ_N , writing each equation as $|z_i - z_j|^2 - n|v_i - v_j|^2 = 0$. If $z = (z_0, \dots, z_{K-1})$ and $v = (v_0, \dots, v_{K-1})$, the resulting phase is

$$\sum_{\{i,j\} \in \mathcal{E}} \beta_{ij} \left(|z_i - z_j|^2 - n|v_i - v_j|^2 \right) = \mathcal{Q}_\beta(z) - n\mathcal{Q}_\beta(v).$$

Thus

$$\mathbf{1}_{|z_i - z_j|^2 = n|v_i - v_j|^2 \ (\forall \{i,j\} \in \mathcal{E})} = \int_{\mathbb{T}^L} e(\mathcal{Q}_\beta(z) - n\mathcal{Q}_\beta(v)) d\beta.$$

Substituting this identity into the counting form gives

$$\Lambda_N(F_0, \dots, F_{K-1}) = \int_{\mathbb{T}^L} \sum_{n \in \mathbb{Z}} w_R(n) e(-n\mathcal{Q}_\beta(v)) \sum_{z_0, \dots, z_{K-1} \in \mathbb{Z}^d} e(\mathcal{Q}_\beta(z)) \prod_{i=0}^{K-1} F_i(z_i) d\beta. \quad (4.10)$$

Observe that the sum over n in (4.10) is $\hat{w}_R(\mathcal{Q}_\beta(v))$. For $\beta \in \mathbb{T}^L$, write

$$\mathcal{S}_\beta(F_0, \dots, F_{K-1}) := \sum_{z_0, \dots, z_{K-1} \in \mathbb{Z}^d} e(\mathcal{Q}_\beta(z)) \prod_{i=0}^{K-1} F_i(z_i). \quad (4.11)$$

We write $\mathcal{S}_\beta$ when the inputs are clear. Since every F_i is supported on $[N]^d$, the sum in (4.11) is finite and may equivalently be taken over $z_0, \dots, z_{K-1} \in [N]^d$. For a measurable set $\mathfrak{B} \subseteq \mathbb{T}^L$, define

$$\Lambda_N^{\mathfrak{B}}(F_0, \dots, F_{K-1}) := \int_{\mathfrak{B}} \hat{w}_R(\mathcal{Q}_\beta(v)) \mathcal{S}_\beta d\beta. \quad (4.12)$$

Taking $\mathfrak{B} = \mathbb{T}^L$ gives $\Lambda_N = \Lambda_N^{\mathbb{T}^L}$.

Since w_R is nonnegative and $\sum_n w_R(n) = 1$,

$$\|\hat{w}_R\|_{L^\infty(\mathbb{T})} \leq 1. \quad (4.13)$$

5 Cut operators

By (4.12), both the major arc and minor arc estimates reduce to bounds for the exponential sum

$$\mathcal{S}_\beta(F_0, \dots, F_{K-1}) = \sum_{z_0, \dots, z_{K-1} \in [N]^d} e(\mathcal{Q}_\beta(z)) \prod_{i=0}^{K-1} F_i(z_i),$$

where the phase is

$$\mathcal{Q}_\beta(z) = \sum_{\{i,j\} \in \mathcal{E}} \beta_{ij} |z_i - z_j|^2.$$

We estimate this sum by partitioning the variables into two blocks and treating the part of the phase coupling those blocks as an operator kernel.

5.1 The cut operator reduction

A *cut* $S \mid T$ is a partition of $\{0, \dots, K-1\}$ into two nonempty sets. Fix such a cut, and for $z = (z_0, \dots, z_{K-1})$, write

$$z_S := (z_i)_{i \in S}, \quad z_T := (z_j)_{j \in T}.$$

An edge contained entirely in S or entirely in T contributes to $\mathcal{Q}_\beta(z)$ a term depending on only one of these two blocks. For a crossing edge, with $i \in S$ and $j \in T$, we have

$$\beta_{ij}|z_i - z_j|^2 = \beta_{ij}|z_i|^2 - 2\beta_{ij}z_i \cdot z_j + \beta_{ij}|z_j|^2.$$

Collecting the terms according to the variables on which they depend, define

$$\begin{aligned} \mathcal{Q}_\beta^S(z_S) &:= \sum_{\substack{\{i,i'\} \in \mathcal{E} \\ i,i' \in S}} \beta_{ii'}|z_i - z_{i'}|^2 + \sum_{i \in S, j \in T} \beta_{ij}|z_i|^2, \\ \mathcal{Q}_\beta^{S,T}(z_S, z_T) &:= -2 \sum_{i \in S, j \in T} \beta_{ij}z_i \cdot z_j, \\ \mathcal{Q}_\beta^T(z_T) &:= \sum_{\substack{\{j,j'\} \in \mathcal{E} \\ j,j' \in T}} \beta_{jj'}|z_j - z_{j'}|^2 + \sum_{i \in S, j \in T} \beta_{ij}|z_j|^2. \end{aligned}$$

Then

$$\mathcal{Q}_\beta(z) = \mathcal{Q}_\beta^S(z_S) + \mathcal{Q}_\beta^{S,T}(z_S, z_T) + \mathcal{Q}_\beta^T(z_T). \quad (5.1)$$

The first and third terms in (5.1) depend on variables from only one side of the cut, while the middle term is the only one involving variables from both sides.

We view $e(\mathcal{Q}_\beta^{S,T}(z_S, z_T))$ as the kernel of an operator taking functions of z_T to functions of z_S . The cut-reduction lemma below shows that the two one-sided phases $e(\mathcal{Q}_\beta^S(z_S))$ and $e(\mathcal{Q}_\beta^T(z_T))$ in (5.1) act by unitary multiplication and therefore do not affect the relevant $2 \rightarrow 2$ operator norm.

The cut operator reduction will be applied in two different measure-space settings. On the minor arcs, $(X_i, \mu_i) = ([N]^d, \#)$, where $\#$ denotes counting measure. On the major arcs, $X_i = B_i \times (\mathbb{Z}/q\mathbb{Z})^d$, where $B_i \subseteq \mathbb{Z}^d$ is finite, and we equip X_i with the measure μ_i determined by

$$\int_{X_i} f d\mu_i = \mathbb{E}_{r_i \bmod q} \sum_{u_i \in B_i} f(u_i, r_i).$$

Therefore, we formulate the reduction for arbitrary finite measure spaces out of convenience.

For $0 \leq i \leq K-1$, let (X_i, μ_i) be a finite measure space and put

$$X_S := \prod_{i \in S} X_i, \quad X_T := \prod_{j \in T} X_j, \quad \mu_S := \bigotimes_{i \in S} \mu_i, \quad \mu_T := \bigotimes_{j \in T} \mu_j.$$

Given a measurable function $\mathcal{Q}^{S,T} : X_S \times X_T \rightarrow \mathbb{T}$, define

$$\begin{aligned} \mathcal{K}^{S,T} : L^2(X_T, \mu_T) &\rightarrow L^2(X_S, \mu_S), \\ (\mathcal{K}^{S,T} h)(x_S) &:= \int_{X_T} e\left(\mathcal{Q}^{S,T}(x_S, x_T)\right) h(x_T) d\mu_T(x_T). \end{aligned}$$

We call every operator of this form a *cut operator*.

Lemma 5.1 (Cut-operator reduction). *Let $S \mid T$, (X_i, μ_i) , $\mathcal{Q}^{S,T}$, and $\mathcal{K}^{S,T}$ be as above. Let*

$$\mathcal{Q}^S : X_S \rightarrow \mathbb{T}, \quad \mathcal{Q}^T : X_T \rightarrow \mathbb{T}$$

be measurable. Then, for $f_i \in L^2(X_i, \mu_i)$,

$$\begin{aligned} & \left| \int_{X_S} \int_{X_T} e\left(\mathcal{Q}^S(x_S) + \mathcal{Q}^{S,T}(x_S, x_T) + \mathcal{Q}^T(x_T)\right) \prod_{i=0}^{K-1} f_i(x_i) d\mu_T(x_T) d\mu_S(x_S) \right| \\ & \leq \|\mathcal{K}^{S,T}\|_{2 \rightarrow 2} \prod_{i=0}^{K-1} \|f_i\|_{L^2(X_i, \mu_i)}. \end{aligned} \tag{5.2}$$

Proof. Define multiplication operators by

$$(U_S h)(x_S) := e\left(\mathcal{Q}^S(x_S)\right) h(x_S), \quad (U_T h)(x_T) := e\left(\mathcal{Q}^T(x_T)\right) h(x_T).$$

Both operators are unitary.

For every $h \in L^2(X_T, \mu_T)$,

$$\begin{aligned} (U_S \mathcal{K}^{S,T} U_T h)(x_S) &= e\left(\mathcal{Q}^S(x_S)\right) \int_{X_T} e\left(\mathcal{Q}^{S,T}(x_S, x_T)\right) e\left(\mathcal{Q}^T(x_T)\right) h(x_T) d\mu_T(x_T) \\ &= \int_{X_T} e\left(\mathcal{Q}^S(x_S) + \mathcal{Q}^{S,T}(x_S, x_T) + \mathcal{Q}^T(x_T)\right) h(x_T) d\mu_T(x_T). \end{aligned}$$

Thus $U_S \mathcal{K}^{S,T} U_T$ is exactly the operator with the full phase. Since U_S and U_T are unitary,

$$\|U_S \mathcal{K}^{S,T} U_T\|_{2 \rightarrow 2} = \|\mathcal{K}^{S,T}\|_{2 \rightarrow 2}.$$

The integral in the statement can now be written as

$$\int_{X_S} \left(\prod_{i \in S} f_i(x_i) \right) \left(U_S \mathcal{K}^{S,T} U_T \left[\prod_{j \in T} f_j \right] \right) (x_S) d\mu_S(x_S).$$

Applying Cauchy–Schwarz gives

$$\|\mathcal{K}^{S,T}\|_{2 \rightarrow 2} \left\| \prod_{i \in S} f_i \right\|_{L^2(X_S, \mu_S)} \left\| \prod_{j \in T} f_j \right\|_{L^2(X_T, \mu_T)}.$$

The product measures give

$$\left\| \prod_{i \in S} f_i \right\|_{L^2(X_S, \mu_S)} = \prod_{i \in S} \|f_i\|_{L^2(X_i, \mu_i)}, \quad \left\| \prod_{j \in T} f_j \right\|_{L^2(X_T, \mu_T)} = \prod_{j \in T} \|f_j\|_{L^2(X_j, \mu_j)},$$

which proves the claim. $\square$

5.2 Tensorization

We now consider a cut operator which naturally appears in the present problem. Take $X_i = [R]^d$ with counting measure, and let $\Theta = (\Theta_{ij})_{i \in S, j \in T} \in \mathbb{T}^{|S||T|}$. For $z = (z_0, \dots, z_{K-1}) \in ([R]^d)^K$, set

$$\mathcal{Q}^{S,T}(z_S, z_T) := \sum_{i \in S, j \in T} \Theta_{ij} z_i \cdot z_j.$$

The resulting cut operator is

$$\begin{aligned} \mathcal{K}_{R,d}^{S,T}(\Theta) : \ell^2([R]^d)^T &\longrightarrow \ell^2([R]^d)^S, \\ (\mathcal{K}_{R,d}^{S,T}(\Theta)h)(z_S) &:= \sum_{z_T \in ([R]^d)^T} e\left(\sum_{i \in S, j \in T} \Theta_{ij} z_i \cdot z_j\right) h(z_T). \end{aligned}$$

We identify this d -dimensional cut operator with an $R^{d|S|} \times R^{d|T|}$ matrix:

$$\mathcal{K}_{R,d}^{S,T}(\Theta) = \left(e\left(\sum_{i \in S, j \in T} \Theta_{ij} z_i \cdot z_j\right) \right)_{\substack{z_S \in ([R]^d)^S \\ z_T \in ([R]^d)^T}}.$$

For the one-dimensional case, we write $\mathcal{K}_R^{S,T}(\Theta) := \mathcal{K}_{R,1}^{S,T}(\Theta)$ and define the normalized cut norm by

$$m_R^{S,T}(\Theta) := R^{-K/2} \|\mathcal{K}_R^{S,T}(\Theta)\|_{2 \rightarrow 2}. \quad (5.3)$$

By Cauchy–Schwarz, the operator norm of a finite matrix is at most its Hilbert–Schmidt norm. Since the R^K entries of $\mathcal{K}_R^{S,T}(\Theta)$ all have modulus one,

$$\|\mathcal{K}_R^{S,T}(\Theta)\|_{2 \rightarrow 2} \leq \|\mathcal{K}_R^{S,T}(\Theta)\|_{\text{HS}} = R^{K/2},$$

and therefore $0 \leq m_R^{S,T}(\Theta) \leq 1$.

The following lemma gives a relationship between the operator norm of $\mathcal{K}_R^{S,T}$ and $\mathcal{K}_{R,d}^{S,T}$. This reduction plays a key role throughout the proof.

Lemma 5.2. *For every cut $S \mid T$,*

$$\|\mathcal{K}_{R,d}^{S,T}(\Theta)\|_{2 \rightarrow 2} = \|\mathcal{K}_R^{S,T}(\Theta)\|_{2 \rightarrow 2}^d = R^{Kd/2} m_R^{S,T}(\Theta)^d.$$

Proof. Write $z_i = (z_{i,1}, \dots, z_{i,d})$ and put

$$z_{S,\ell} := (z_{i,\ell})_{i \in S} \in [R]^S, \quad z_{T,\ell} := (z_{j,\ell})_{j \in T} \in [R]^T \quad (1 \leq \ell \leq d).$$

Thus the row index z_S is the same data as $(z_{S,1}, \dots, z_{S,d})$, and the column index z_T is the same data as $(z_{T,1}, \dots, z_{T,d})$. Under this identification, the (z_S, z_T) -entry of $\mathcal{K}_{R,d}^{S,T}(\Theta)$ is

$$\begin{aligned} e\left(\sum_{i \in S, j \in T} \Theta_{ij} z_i \cdot z_j\right) &= e\left(\sum_{\ell=1}^d \sum_{i \in S, j \in T} \Theta_{ij} z_{i,\ell} z_{j,\ell}\right) \\ &= \prod_{\ell=1}^d e\left(\sum_{i \in S, j \in T} \Theta_{ij} z_{i,\ell} z_{j,\ell}\right). \end{aligned}$$

For each ℓ , the corresponding factor is the $(z_{S,\ell}, z_{T,\ell})$ -entry of $\mathcal{K}_R^{S,T}(\Theta)$. This is precisely the definition of a d -fold tensor product:

$$\mathcal{K}_{R,d}^{S,T}(\Theta) = (\mathcal{K}_R^{S,T}(\Theta))^{\otimes d}.$$

The standard multiplicativity of the $2 \rightarrow 2$ operator norm under tensor products now gives

$$\|\mathcal{K}_{R,d}^{S,T}(\Theta)\|_{2 \rightarrow 2} = \|\mathcal{K}_R^{S,T}(\Theta)\|_{2 \rightarrow 2}^d.$$

Finally, by the definition of $m_R^{S,T}(\Theta)$,

$$\|\mathcal{K}_R^{S,T}(\Theta)\|_{2 \rightarrow 2}^d = R^{Kd/2} m_R^{S,T}(\Theta)^d.$$

□

The cut operators $\mathcal{K}_R^{S,T}$ appear in both the major and minor arc estimates. On the minor arcs, Lemma 5.1 and Lemma 5.2 reduce the analysis to understanding $m_N^{S,T}(-2\beta_{S,T})$. On the major arcs, the same reductions lead us to $m_N^{S,T}(-2\theta_{S,T})$ paired with the complete-sum factor $|\text{im}(2a_{S,T})|^{-d/2}$. Here we used the following useful notation: for the coefficient vector $b = (b_e)_{e \in \mathcal{E}}$, put $b_{ij} = b_{ji} := b_{\{i,j\}}$ and write $b_{S,T} := (b_{ij})_{i \in S, j \in T}$ for its matrix of crossing coefficients.

5.3 Reduction to one crossing edge

We next reduce the normalized cut norm to the single-edge norm associated with any one crossing edge. For $\theta \in \mathbb{T}$, define the single-edge matrix and its normalized norm by

$$\mathcal{K}_R(\theta) := (e(\theta rs))_{r,s \in [R]}, \quad m_R(\theta) := R^{-1} \|\mathcal{K}_R(\theta)\|_{2 \rightarrow 2}.$$

This is the cut operator associated with a cut having one vertex on each side. Thus $\mathcal{K}_R(\theta)$ is an $R \times R$ matrix acting on $\ell^2([R])$, and its r th output coordinate is $\sum_{s \in [R]} e(\theta rs) h(s)$. Its Hilbert–Schmidt norm is R , so $0 \leq m_R(\theta) \leq 1$, and complex conjugation gives $m_R(-\theta) = m_R(\theta)$.

We shall use the standard block form of Schur’s test: if $M = (M_{ab})$ is a $p \times q$ matrix of operators, then

$$\|M\|_{2 \rightarrow 2} \leq \sqrt{pq} \max_{a,b} \|M_{ab}\|_{2 \rightarrow 2}. \quad (5.4)$$

Lemma 5.3 (Single-edge bound). *Let $\Theta \in \mathbb{T}^{|S||T|}$ and let $i \in S, j \in T$. Then $m_R^{S,T}(\Theta) \leq m_R(\Theta_{ij})$.*

Proof. Abbreviate $\mathcal{K} := \mathcal{K}_R^{S,T}(\Theta)$ and separate the two coordinates belonging to the selected edge:

$$\widehat{z}_S := (z_k)_{k \in S \setminus \{i\}}, \quad \widehat{z}_T := (z_\ell)_{\ell \in T \setminus \{j\}}.$$

For fixed $\widehat{z}_S$ and $\widehat{z}_T$, define

$$\mathcal{K}[\widehat{z}_S, \widehat{z}_T] := \left(e\left(\mathcal{Q}^{S,T}(z_S, z_T) \right) \right)_{z_i, z_j \in [R]},$$

where the coordinates contained in $\widehat{z}_S$ and $\widehat{z}_T$ are held fixed. This is the $R \times R$ block of $\mathcal{K}$ in which z_i indexes the rows and z_j indexes the columns.

Every row is specified by $(\widehat{z}_S, z_i)$ and every column by $(\widehat{z}_T, z_j)$. Grouping them by the hatted tuples therefore gives the block matrix

$$\mathcal{K} = (\mathcal{K}[\widehat{z}_S, \widehat{z}_T])_{\substack{\widehat{z}_S \in [R]^{S \setminus \{i\}} \\ \widehat{z}_T \in [R]^{T \setminus \{j\}}}}.$$

Thus there are $R^{|S|-1}$ block rows and $R^{|T|-1}$ block columns, and every block is $R \times R$.

The definition of $\mathcal{Q}^{S,T}$ gives

$$\mathcal{Q}^{S,T}(z_S, z_T) = \Theta_{ij} z_i z_j + z_i \sum_{\ell \in T \setminus \{j\}} \Theta_{i\ell} z_\ell + z_j \sum_{k \in S \setminus \{i\}} \Theta_{kj} z_k + \sum_{\substack{k \in S \setminus \{i\} \\ \ell \in T \setminus \{j\}}} \Theta_{k\ell} z_k z_\ell. \quad (5.5)$$

For $t \in \mathbb{T}$, set

$$D_R(t) := \begin{pmatrix} e(t) & & 0 \\ & \ddots & \\ 0 & & e(Rt) \end{pmatrix}.$$

Consequently the block has the factorization

$$\mathcal{K}[\widehat{z}_S, \widehat{z}_T] = e \left(\sum_{\substack{k \in S \setminus \{i\} \\ \ell \in T \setminus \{j\}}} \Theta_{k\ell} z_k z_\ell \right) D_R \left(\sum_{\ell \in T \setminus \{j\}} \Theta_{i\ell} z_\ell \right) \mathcal{K}_R(\Theta_{ij}) D_R \left(\sum_{k \in S \setminus \{i\}} \Theta_{kj} z_k \right).$$

The scalar factor supplies the last term in 5.5. The left and right diagonal factors supply the terms linear in z_i and z_j , respectively, while $\mathcal{K}_R(\Theta_{ij})$ supplies the term $\Theta_{ij} z_i z_j$. The scalar has modulus one and the two diagonal matrices are unitary, so every block satisfies

$$\|\mathcal{K}[\widehat{z}_S, \widehat{z}_T]\|_{2 \rightarrow 2} = \|\mathcal{K}_R(\Theta_{ij})\|_{2 \rightarrow 2}.$$

Applying (5.4) to the $R^{|S|-1} \times R^{|T|-1}$ block matrix yields

$$\|\mathcal{K}\|_{2 \rightarrow 2} \leq R^{(|S|-1)/2} R^{(|T|-1)/2} \|\mathcal{K}_R(\Theta_{ij})\|_{2 \rightarrow 2} = R^{(K-2)/2} \|\mathcal{K}_R(\Theta_{ij})\|_{2 \rightarrow 2}.$$

Using the two normalizations now gives

$$m_R^{S,T}(\Theta) = R^{-K/2} \|\mathcal{K}\|_{2 \rightarrow 2} \leq R^{-1} \|\mathcal{K}_R(\Theta_{ij})\|_{2 \rightarrow 2} = m_R(\Theta_{ij}).$$

□

5.4 A high-moment bound

We now prove the one-cut mean-value estimate stated in Lemma 2.1. Its arithmetic input is the matrix count from Appendix B, which is derived from Kim's mean-value formula for primitive lattice flags.

Proof of Lemma 2.1. Abbreviate $\mathcal{K} := \mathcal{K}_R^{S,T}(\Theta)$. The matrix $\mathcal{K}\mathcal{K}^*$ is Hermitian and positive semidefinite. If its eigenvalues are $\lambda_1, \dots, \lambda_{R^{|S|}} \geq 0$, then

$$\|\mathcal{K}\|_{2 \rightarrow 2}^{2K+4} = \left(\max_a \lambda_a \right)^{K+2} \leq \sum_a \lambda_a^{K+2} = \text{tr}((\mathcal{K}\mathcal{K}^*)^{K+2}). \quad (5.6)$$

In this way we reduce the proof to analyzing $\text{tr}((\mathcal{K}\mathcal{K}^*)^{K+2})$.

For $z_S, \tilde{z}_S \in [R]^S$, ordinary matrix multiplication gives the $(z_S, \tilde{z}_S)$ -entry of $\mathcal{K}\mathcal{K}^*$ as

$$\begin{aligned} (\mathcal{K}\mathcal{K}^*)_{z_S, \tilde{z}_S} &= \sum_{z_T \in [R]^T} e \left(\sum_{i \in S, j \in T} \Theta_{ij} z_i z_j \right) \overline{e \left(\sum_{i \in S, j \in T} \Theta_{ij} \tilde{z}_i z_j \right)} \\ &= \sum_{z_T \in [R]^T} e \left(\sum_{i \in S, j \in T} \Theta_{ij} (z_i - \tilde{z}_i) z_j \right), \end{aligned} \quad (5.7)$$

and expanding the trace gives

$$\mathrm{tr}((\mathcal{K}\mathcal{K}^*)^{K+2}) = \sum_{z_S^{(1)}, \dots, z_S^{(K+2)} \in [R]^S} (\mathcal{K}\mathcal{K}^*)_{z_S^{(1)}, z_S^{(2)}} \cdots (\mathcal{K}\mathcal{K}^*)_{z_S^{(K+2)}, z_S^{(1)}}. \quad (5.8)$$

Combining (5.7) with (5.8) yields

$$\begin{aligned} & \mathrm{tr}((\mathcal{K}\mathcal{K}^*)^{K+2}) \\ &= \sum_{\substack{z_S^{(1)}, \dots, z_S^{(K+2)} \\ \in [R]^S}} \sum_{\substack{z_T^{(1)}, \dots, z_T^{(K+2)} \\ \in [R]^T}} e \left(\sum_{i \in S, j \in T} \Theta_{ij} \left[\sum_{h=1}^{K+1} (z_i^{(h)} - z_i^{(h+1)}) z_j^{(h)} + (z_i^{(K+2)} - z_i^{(1)}) z_j^{(K+2)} \right] \right). \end{aligned}$$

Now integrate each Θ_{ij} over $\mathbb{T}$ and apply the orthogonality relation. The equations imposed by orthogonality are precisely the entries of the matrix equation

$$\sum_{h=1}^{K+1} (z_S^{(h)} - z_S^{(h+1)}) (z_T^{(h)})^\top + (z_S^{(K+2)} - z_S^{(1)}) (z_T^{(K+2)})^\top = 0. \quad (5.9)$$

Consequently, the integral of the trace is the number of choices $z_S^{(h)} \in [R]^S$ and $z_T^{(h)} \in [R]^T$, $1 \leq h \leq K+2$, satisfying this matrix equation.

We next put this constraint into the form required by Lemma B.1. The differences on the S -side telescope:

$$(z_S^{(1)} - z_S^{(2)}) + \cdots + (z_S^{(K+2)} - z_S^{(1)}) = 0.$$

We may therefore subtract $z_T^{(K+2)}$ from every T -variable in (5.9): the change in its left-hand side is the outer product of the zero vector above with $(z_T^{(K+2)})^\top$. The final term then vanishes, and the matrix equation becomes

$$\sum_{h=1}^{K+1} (z_S^{(h)} - z_S^{(h+1)}) (z_T^{(h)} - z_T^{(K+2)})^\top = 0.$$

For $1 \leq h \leq K+1$, set

$$U_h := z_S^{(h)} - z_S^{(h+1)} \in \mathbb{Z}^S, \quad V_h := z_T^{(h)} - z_T^{(K+2)} \in \mathbb{Z}^T,$$

and place these vectors in the columns of

$$U = (U_1 \cdots U_{K+1}) \in \mathbb{Z}^{|S| \times (K+1)}, \quad V = (V_1 \cdots V_{K+1}) \in \mathbb{Z}^{|T| \times (K+1)}.$$

Since $UV^\top = \sum_{h=1}^{K+1} U_h V_h^\top$, equation (5.9) is precisely $UV^\top = 0$.

Once U and V are fixed, the variables

$$z_S^{(1)}, \dots, z_S^{(K+2)} \quad \text{and} \quad z_T^{(1)}, \dots, z_T^{(K+2)}$$

are determined by the two base points $z_S^{(1)} \in [R]^S$ and $z_T^{(K+2)} \in [R]^T$. Indeed,

$$z_S^{(h+1)} = z_S^{(h)} - U_h, \quad z_T^{(h)} = z_T^{(K+2)} + V_h \quad (1 \leq h \leq K+1).$$

Thus each pair U, V arises from at most $R^{|S|} R^{|T|} = R^K$ choices of these variables. Some choices of the two base points may reconstruct variables outside $[R]^S$ or $[R]^T$, which only decreases this number.

All entries of U and V lie in $[-R, R]$. Moreover, balancedness gives $2 \max(|S|, |T|) \leq K + 1$, while $|S| + |T| = K < K + 1$. Thus Lemma B.1 applies with $n = K + 1$ and shows that the number of possible pairs U, V is

$$\ll_K R^{(K+1)(|S|+|T|)-2|S||T|} = R^{K(K+1)-2|S||T|}.$$

Combining this with the R^K choices of base points gives

$$\int_{\mathbb{T}^{|S||T|}} \text{tr}((\mathcal{K}\mathcal{K}^*)^{K+2}) d\Theta \ll_K R^{K(K+2)-2|S||T|}. \quad (5.10)$$

Finally, (5.6) yields

$$m_R^{S,T}(\Theta)^{2K+4} = R^{-K(K+2)} \|\mathcal{K}\|_{2 \rightarrow 2}^{2K+4} \leq R^{-K(K+2)} \text{tr}((\mathcal{K}\mathcal{K}^*)^{K+2}).$$

Integrating this inequality and applying (5.10) gives

$$\int_{\mathbb{T}^{|S||T|}} m_R^{S,T}(\Theta)^{2K+4} d\Theta \ll_K R^{-2|S||T|},$$

which is (2.10). □

5.5 Balanced cuts and fractional Hölder

Recall that a cut $S \mid T$ is *balanced* if

$$S \sqcup T = \{0, \dots, K-1\}, \quad ||S| - |T|| \leq 1.$$

Let $\mathcal{B}_K$ denote the set of all balanced cuts, with $S \mid T$ and $T \mid S$ regarded as the same cut. Give every cut in $\mathcal{B}_K$ the common weight $w_K := L/(\lfloor K^2/4 \rfloor |\mathcal{B}_K|)$.

For a cut $S \mid T$, regard the edges crossing $S \mid T$ as a subset of $\mathcal{E}$. A *fractional cover* of $\mathcal{E}$ by such sets assigns a nonnegative weight to each cut so that the total weight of the cuts crossed by any fixed edge is at least one. The following lemma shows that, for the weights above, this total is exactly one for every edge.

Lemma 5.4. *For every edge $e \in \mathcal{E}$,*

$$w_K \# \{S \mid T \in \mathcal{B}_K : e \text{ crosses } S \mid T\} = 1. \quad (5.11)$$

Proof. Permuting the vertices preserves $\mathcal{B}_K$ and acts transitively on the edges of $\mathcal{E}$. Hence every edge crosses the same number of balanced cuts.

Count pairs consisting of a balanced cut and one of its crossing edges. Choosing the cut first gives $|\mathcal{B}_K| \lfloor K^2/4 \rfloor$ pairs, since every balanced cut has $\lfloor K^2/4 \rfloor$ crossing edges. Choosing the edge first gives

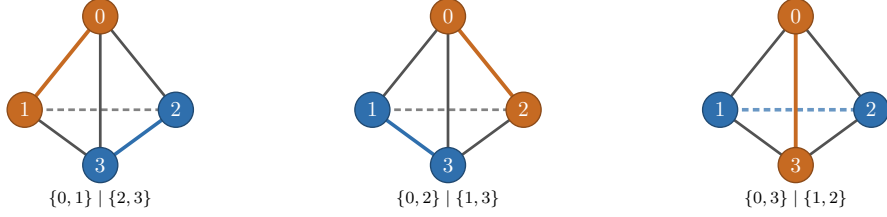
$$L \# \{S \mid T \in \mathcal{B}_K : e \text{ crosses } S \mid T\}$$

pairs, since there are L edges and the number of balanced cuts crossed by an edge is independent of the edge. Therefore

$$L \# \{S \mid T \in \mathcal{B}_K : e \text{ crosses } S \mid T\} = |\mathcal{B}_K| \lfloor K^2/4 \rfloor.$$

Multiplying by w_K/L proves (5.11). □

For $K = 4$, the cover can be read directly from the three diagrams below. The orange and blue vertices lie on the two sides of each cut. Every edge is internal in exactly one panel and crosses the cut in the other two. Since $w_4 = 1/2$, every edge therefore has total crossing weight one.



The fractional cover identity (5.11) is exactly the hypothesis in the following special case of Finner's fractional form of Hölder's inequality [4].

Lemma 5.5 (Fractional Hölder inequality). *Let $\mathcal{E}$ be a finite set. For each $e \in \mathcal{E}$, let (X_e, μ_e) be a σ -finite measure space, and let $\mu := \bigotimes_{e \in \mathcal{E}} \mu_e$ be the product measure. Let J be a positive integer. For each $1 \leq j \leq J$, choose a nonempty subset $\mathcal{E}_j \subseteq \mathcal{E}$, a weight $c_j > 0$, and a nonnegative integrable function $f_j : \prod_{e \in \mathcal{E}_j} X_e \rightarrow [0, \infty)$. Denote the product measure on the domain of f_j by $\mu_j := \bigotimes_{e \in \mathcal{E}_j} \mu_e$. Suppose that for every $e \in \mathcal{E}$, we have*

$$\sum_{\substack{1 \leq j \leq J \\ e \in \mathcal{E}_j}} c_j = 1.$$

Then, writing $x = (x_e)_{e \in \mathcal{E}}$,

$$\int \prod_{j=1}^J f_j((x_e)_{e \in \mathcal{E}_j})^{c_j} d\mu(x) \leq \prod_{j=1}^J \left(\int f_j d\mu_j \right)^{c_j}.$$

This is [4, Theorem 2.1], applied with exponents $1/c_j$ to the functions $f_j^{c_j}$.

6 The minor arcs

We begin the section by decomposing $\mathbb{T}^L$ into major and minor arcs. For $2 \leq Q \leq R$, define the one-dimensional major arcs by

$$\mathfrak{M}_Q^{(1)} := \bigcup_{q \leq Q} \bigcup_{\substack{a \in \mathbb{Z}/q\mathbb{Z} \\ (a, q) = 1}} \left\{ \xi \in \mathbb{T} : \|\xi - a/q\|_{\mathbb{T}} \leq \frac{Q}{qR^2} \right\}. \quad (6.1)$$

For $r \geq 1$, put

$$\mathfrak{M}_Q^{(r)} := (\mathfrak{M}_Q^{(1)})^r, \quad \mathfrak{m}_Q^{(r)} := \mathbb{T}^r \setminus \mathfrak{M}_Q^{(r)}.$$

Since the scalar coefficients arising from the cut operators are $-2\beta_e$, define

$$\mathfrak{M}_Q := \{\beta \in \mathbb{T}^L : -2\beta \in \mathfrak{M}_Q^{(L)}\}, \quad \mathfrak{m}_Q := \mathbb{T}^L \setminus \mathfrak{M}_Q.$$

Thus

$$\mathfrak{m}_Q = \bigcup_{e \in \mathcal{E}} \{\beta \in \mathbb{T}^L : -2\beta_e \in \mathfrak{m}_Q^{(1)}\}. \quad (6.2)$$

The goal of the section is to prove the following minor arc estimate.

Proposition 6.1 (minor arc estimate). *Assume $2 \leq Q \leq R$ and $d \geq 4K + 4$. There is $\gamma = \gamma(K, d) > 0$ such that, for all functions $F_0, \dots, F_{K-1} : \mathbb{Z}^d \rightarrow \mathbb{C}$ supported on $[N]^d$,*

$$|\Lambda_N^{\mathfrak{m}Q}(F_0, \dots, F_{K-1})| \ll_{\sigma, \Delta, d} Q^{-\gamma} R^{Kd/2-2L} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \quad (6.3)$$

6.1 Applying the cut reduction

Since each F_i is supported on $[N]^d$, the sum $\mathcal{S}_\beta$ may be taken over $([N]^d)^K$. For a cut $S \mid T$, take $X_i = [N]^d$ with counting measure and set $f_i = F_i$. The crossing part of the phase is

$$\mathcal{Q}_\beta^{S,T}(z_S, z_T) = -2 \sum_{i \in S, j \in T} \beta_{ij} z_i \cdot z_j,$$

so the cut operator in Lemma 5.1 is exactly $\mathcal{K}_{N,d}^{S,T}(-2\beta_{S,T})$. That lemma therefore gives

$$|\mathcal{S}_\beta| \leq \|\mathcal{K}_{N,d}^{S,T}(-2\beta_{S,T})\|_{2 \rightarrow 2} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}.$$

Applying Lemma 5.2 now gives the required cut-operator estimate directly.

Proposition 6.2 (Cut-operator estimate). *For every $\beta \in \mathbb{T}^L$, cut $S \mid T$, and functions $F_0, \dots, F_{K-1} : \mathbb{Z}^d \rightarrow \mathbb{C}$ supported on $[N]^d$,*

$$|\mathcal{S}_\beta| \leq N^{Kd/2} m_N^{S,T}(-2\beta_{S,T})^d \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \quad (6.4)$$

6.2 The large-sieve estimate

We now prove the scalar Weyl estimate stated in Lemma 2.3 by applying the analytic large sieve.

Proof of Lemma 2.3. For the proof, write $P := R$ and $\theta := \xi$. Dirichlet's approximation theorem gives coprime integers a, q such that

$$1 \leq q \leq \frac{P^2}{Q}, \quad \|\theta - a/q\|_{\mathbb{T}} \leq \frac{Q}{qP^2}.$$

If $q \leq Q$, this contradicts the hypothesis of the lemma. Therefore $Q < q \leq P^2/Q$. The upper bound on q also gives $\|\theta - a/q\|_{\mathbb{T}} \leq Q/(qP^2) \leq 1/q^2$.

If $q < 8$, then $Q < 8$, and the conclusion follows immediately from $m_P(\theta)^2 \leq 1 \leq 8Q^{-1}$. We may therefore suppose that $q \geq 8$.

Partition $[P]$ into $j \ll 1 + P/q$ intervals $I_1, \dots, I_j$, each of diameter at most $q/8$. Thus, for every $1 \leq r \leq j$,

$$|u - u'| \leq \frac{q}{8} \quad (u, u' \in I_r).$$

Fix one such interval I_r . Suppose $u, u' \in I_r$ are distinct. Since $(a, q) = 1$ and $|u - u'| \leq q/8 < q$, the integer $a(u - u')$ is not divisible by q , so $\|a(u - u')/q\|_{\mathbb{T}} \geq 1/q$. Moreover,

$$|u - u'| \|\theta - a/q\|_{\mathbb{T}} \leq \frac{q}{8} \cdot \frac{1}{q^2} = \frac{1}{8q}.$$

It follows from the reverse triangle inequality that

$$\begin{aligned}
\|\theta u - \theta u'\|_{\mathbb{T}} &= \|\theta(u - u')\|_{\mathbb{T}} \\
&\geq \left\| \frac{a(u - u')}{q} \right\|_{\mathbb{T}} - \left\| (u - u') \left(\theta - \frac{a}{q} \right) \right\|_{\mathbb{T}} \\
&\geq \left\| \frac{a(u - u')}{q} \right\|_{\mathbb{T}} - |u - u'| \left\| \theta - \frac{a}{q} \right\|_{\mathbb{T}} \\
&\geq \frac{1}{q} - \frac{1}{8q} = \frac{7}{8q}.
\end{aligned}$$

Thus the points $\{\theta u : u \in I_r\}$ are pairwise separated in $\mathbb{T}$ by at least $7/(8q)$.

Applying the analytic large sieve [18, Theorem 1, equation (1.4)] to the trigonometric polynomial $\sum_{s \in [P]} h(s)e(sx)$ at the points $\{\theta u : u \in I_r\}$ and using their separation yields

$$\sum_{u \in I_r} \left| \sum_{s \in [P]} e(\theta u s) h(s) \right|^2 \leq \left(P + \frac{8q}{7} \right) \|h\|_{\ell^2([P])}^2.$$

Summing this inequality over $r \in [j]$ and using the fact that the intervals I_r partition $[P]$, we obtain

$$\sum_{u \in [P]} \left| \sum_{s \in [P]} e(\theta u s) h(s) \right|^2 \ll \left(1 + \frac{P}{q} \right) (P + q) \|h\|_{\ell^2([P])}^2.$$

Taking the supremum over $h \in \ell^2([P])$ with $\|h\|_{\ell^2([P])} \leq 1$ yields

$$\|\mathcal{K}_P(\theta)\|_{2 \rightarrow 2}^2 \ll \left(1 + \frac{P}{q} \right) (P + q).$$

Since $m_P(\theta) = P^{-1} \|\mathcal{K}_P(\theta)\|_{2 \rightarrow 2}$, we conclude that

$$m_P(\theta)^2 \ll P^{-2} \left(1 + \frac{P}{q} \right) (P + q) \ll q^{-1} + P^{-1} + qP^{-2} \ll Q^{-1}.$$

Here we used

$$q > Q, \quad Q \leq P, \quad q \leq \frac{P^2}{Q}.$$

□

6.3 Proof of the minor arc estimate

We now prove the minor arc estimate by combining fractional Hölder, the large-sieve estimate, and the high moment estimate.

Proof of Proposition 6.1. Put $\tau := d \lfloor K^2/4 \rfloor / (2L)$. Since $2\tau w_K = d/|\mathcal{B}_K|$, taking the geometric mean of Proposition 6.2 over all balanced cuts gives

$$|\mathcal{S}_\beta| \leq N^{Kd/2} \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \quad (6.5)$$

The dimension hypothesis gives

$$\tau \geq \frac{(4K+4)(K^2-1)}{8L} = \frac{(K+1)^2}{K} > K+2.$$

Set $\gamma := \tau - (K + 2) > 0$. In view of (4.12), (4.13), and (6.5), it remains to prove

$$\int_{\mathfrak{m}_Q} \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K} d\beta \ll_{K,d} Q^{-\gamma} N^{-2L}. \quad (6.6)$$

For each $e \in \mathcal{E}$, put

$$I_e := \int_{\mathbb{T}^L} \mathbf{1}_{\mathfrak{m}_Q^{(1)}}(-2\beta_e) \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K} d\beta.$$

By (6.2) and nonnegativity,

$$\int_{\mathfrak{m}_Q} \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K} d\beta \leq \sum_{e \in \mathcal{E}} I_e.$$

It therefore suffices to prove

$$I_e \ll_{K,d} Q^{-\gamma} N^{-2L} \quad (6.7)$$

for every $e \in \mathcal{E}$.

Fix $e \in \mathcal{E}$. In applying Lemma 5.5, take $J = |\mathcal{B}_K|$ and label the indices $1, \dots, J$ by the balanced cuts $S | T \in \mathcal{B}_K$. Thus, for the index j corresponding to $S | T$, the set $\mathcal{E}_j$ in that lemma is the set of edges crossing $S | T$, and the weight c_j is w_K . We write $f_{S,T}$ for the corresponding function f_j .

For each balanced cut $S | T$, define

$$f_{S,T}(\beta_{S,T}) := \begin{cases} \mathbf{1}_{\mathfrak{m}_Q^{(1)}}(-2\beta_e) m_N^{S,T}(-2\beta_{S,T})^{2\tau}, & e \text{ crosses } S | T, \\ m_N^{S,T}(-2\beta_{S,T})^{2\tau}, & e \text{ does not cross } S | T. \end{cases}$$

When e crosses $S | T$, the coordinate β_e is one of the coordinates in $\beta_{S,T}$, so the first function is indeed defined on the product of the coordinate spaces indexed by $\mathcal{E}_j$.

For every $e' \in \mathcal{E}$, Lemma 5.4 gives

$$\sum_{\substack{S|T \in \mathcal{B}_K \\ e' \text{ crosses } S|T}} w_K = 1,$$

which is exactly the hypothesis of Lemma 5.5. Moreover,

$$\begin{aligned} \prod_{S|T \in \mathcal{B}_K} f_{S,T}(\beta_{S,T})^{w_K} &= \left(\prod_{\substack{S|T \in \mathcal{B}_K \\ e \text{ crosses } S|T}} \mathbf{1}_{\mathfrak{m}_Q^{(1)}}(-2\beta_e)^{w_K} \right) \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K} \\ &= \mathbf{1}_{\mathfrak{m}_Q^{(1)}}(-2\beta_e) \prod_{S|T \in \mathcal{B}_K} m_N^{S,T}(-2\beta_{S,T})^{2\tau w_K}, \end{aligned}$$

where the last equality follows from (5.11). Hence

$$I_e = \int_{\mathbb{T}^L} \prod_{S|T \in \mathcal{B}_K} f_{S,T}(\beta_{S,T})^{w_K} d\beta.$$

Lemma 5.5 therefore gives

$$\begin{aligned}
I_e &\leq \prod_{\substack{S|T \in \mathcal{B}_K \\ e \text{ crosses } S|T}} \left(\int_{\mathbb{T}^{|S||T|}} \mathbf{1}_{\mathfrak{m}_Q^{(1)}}(-2\beta_e) m_N^{S,T}(-2\beta_{S,T})^{2\tau} d\beta_{S,T} \right)^{w_K} \\
&\times \prod_{\substack{S|T \in \mathcal{B}_K \\ e \text{ does not cross } S|T}} \left(\int_{\mathbb{T}^{|S||T|}} m_N^{S,T}(-2\beta_{S,T})^{2\tau} d\beta_{S,T} \right)^{w_K}. \tag{6.8}
\end{aligned}$$

We now estimate the integrals on the right. Fix a balanced cut $S | T$. The coordinatewise map

$$(\beta_{ij})_{i \in S, j \in T} \mapsto (-2\beta_{ij})_{i \in S, j \in T}$$

preserves normalized Haar measure on $\mathbb{T}^{|S||T|}$. Hence Lemma 2.1 gives

$$\int_{\mathbb{T}^{|S||T|}} m_N^{S,T}(-2\beta_{S,T})^{2K+4} d\beta_{S,T} \ll_K N^{-2|S||T|}. \tag{6.9}$$

Since $0 \leq m_N^{S,T}(-2\beta_{S,T}) \leq 1$ and $2\tau > 2K + 4$, (6.9) gives

$$\int_{\mathbb{T}^{|S||T|}} m_N^{S,T}(-2\beta_{S,T})^{2\tau} d\beta_{S,T} \ll_K N^{-2|S||T|}. \tag{6.10}$$

Suppose that e crosses $S | T$. Whenever $-2\beta_e \in \mathfrak{m}_Q^{(1)}$, the definition of $\mathfrak{m}_Q^{(1)}$ and the inequality $N \geq R$ show that $-2\beta_e$ satisfies the hypothesis of Lemma 2.3 with $P = N$; indeed, $Q/(qN^2) \leq Q/(qR^2)$. Therefore Lemma 5.3 gives

$$m_N^{S,T}(-2\beta_{S,T}) \leq m_N(-2\beta_e) \ll Q^{-1/2}.$$

Since $2\tau = 2K + 4 + 2\gamma$, we have, on this set,

$$m_N^{S,T}(-2\beta_{S,T})^{2\tau} \ll_{K,d} Q^{-\gamma} m_N^{S,T}(-2\beta_{S,T})^{2K+4}.$$

Using (6.9), we obtain

$$\int_{\mathbb{T}^{|S||T|}} \mathbf{1}_{\mathfrak{m}_Q^{(1)}}(-2\beta_e) m_N^{S,T}(-2\beta_{S,T})^{2\tau} d\beta_{S,T} \ll_{K,d} Q^{-\gamma} N^{-2|S||T|}. \tag{6.11}$$

For the cuts crossed by e , (6.11) contributes both a factor $Q^{-\gamma w_K}$ and a factor $N^{-2w_K|S||T|}$ to the fractional Hölder estimate (6.8). For the remaining cuts, (6.10) contributes the factor $N^{-2w_K|S||T|}$. Hence

$$I_e \ll_{K,d} Q^{-\gamma w_K \#\{S|T \in \mathcal{B}_K : e \text{ crosses } S|T\}} N^{-2w_K \sum_{S|T \in \mathcal{B}_K} |S||T|}.$$

The fractional-cover identity (5.11) gives

$$w_K \#\{S | T \in \mathcal{B}_K : e \text{ crosses } S | T\} = 1.$$

Moreover, every balanced cut has $|S||T| = \lfloor K^2/4 \rfloor$, so the definition of w_K gives

$$w_K \sum_{S|T \in \mathcal{B}_K} |S||T| = w_K |\mathcal{B}_K| \lfloor K^2/4 \rfloor = L. \tag{6.12}$$

Putting these facts together proves (6.7), and hence (6.6). Combining that estimate with (4.12), (4.13), and (6.5) gives

$$|\Lambda_N^{\mathfrak{m}_Q}(F_0, \dots, F_{K-1})| \ll_{\Delta,d} Q^{-\gamma} N^{Kd/2-2L} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}.$$

Finally, $N \geq 2/\sigma$ implies $R = \lfloor \sigma N \rfloor \geq \sigma N/2$, so the second bound in (6.3) follows as well. $\square$

7 The major arcs

Throughout this section assume

$$R \geq 4Q^{L(L+3)+3}. \quad (7.1)$$

For $1 \leq q \leq 2Q^L$, put

$$N_q := \left\lfloor \frac{R}{qQ^{L(L+2)+3}} \right\rfloor. \quad (7.2)$$

Since $q \leq 2Q^L$, condition (7.1) gives $N_q \geq 1$. Define the probability measure μ_q on $\mathbb{Z}^d$ by

$$\mu_q(x) := \frac{1}{N_q^d} \mathbf{1}_{q\{0, \dots, N_q-1\}^d}(x).$$

Thus, for every finitely supported $F : \mathbb{Z}^d \rightarrow \mathbb{C}$,

$$(\mu_q * F)(y) = \frac{1}{N_q^d} \sum_{h \in \{0, \dots, N_q-1\}^d} F(y - qh).$$

The goal of the section is to prove the following major arc estimate.

Proposition 7.1 (major arc estimate). *Assume $2 \leq Q \leq R$, $d \geq 4K + 4$, and (7.1). There is $\eta = \eta(K, d) > 1$ such that, for all functions $F_0, \dots, F_{K-1} : \mathbb{Z}^d \rightarrow \mathbb{C}$ supported on $[N]^d$,*

$$\begin{aligned} |\Lambda_N^{\mathfrak{M}_Q}(F_0, \dots, F_{K-1})| &\ll_{\Delta, d, \sigma} R^{(K-1)d-2L} \sum_{q \leq 2Q^L} q^{-1-\eta} \prod_{i=0}^{K-1} \|\mu_q * F_i\|_{\ell^K(\mathbb{Z}^d)} \\ &\quad + Q^{-2} R^{Kd/2-2L} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \end{aligned} \quad (7.3)$$

7.1 Reduction to the major arc form

For $q \geq 1$, $a \in (\mathbb{Z}/q\mathbb{Z})^L$, $\theta \in \mathbb{T}^L$, and finitely supported functions $G_i : \mathbb{Z}^d \times (\mathbb{Z}/q\mathbb{Z})^d \rightarrow \mathbb{C}$, define the major arc form

$$\mathcal{M}_{q,a,\theta}(G_0, \dots, G_{K-1}) := \mathbb{E}_{r \bmod q} \sum_{u \in (\mathbb{Z}^d)^K} e_q(\mathcal{Q}_a(r)) e(\mathcal{Q}_\theta(u)) \prod_{i=0}^{K-1} G_i(u_i, r_i). \quad (7.4)$$

Here $r = (r_0, \dots, r_{K-1})$ ranges over $((\mathbb{Z}/q\mathbb{Z})^d)^K$, and $u = (u_0, \dots, u_{K-1})$ ranges over $(\mathbb{Z}^d)^K$. This is a form of the general type controlled by the cut-operator estimates, with each vertex variable given by the pair (u_i, r_i) . Its kernel has a complete-sum factor in the residue variables and a discrete factor in the lattice variables. The form need not factor, since each G_i may depend on both u_i and r_i .

For every $(u, r) \in \mathbb{Z}^d \times (\mathbb{Z}/q\mathbb{Z})^d$, there is a unique $y \in \mathbb{Z}^d$ satisfying

$$y \equiv r \pmod{q}, \quad u - y \in \{0, \dots, q-1\}^d.$$

For an input $F_i : \mathbb{Z}^d \rightarrow \mathbb{C}$, define $G_{i,q} : \mathbb{Z}^d \times (\mathbb{Z}/q\mathbb{Z})^d \rightarrow \mathbb{C}$ by

$$G_{i,q}(u, r) := (\mu_q * F_i)(y), \quad (7.5)$$

where y is the unique lattice point just described. Thus r is the residue class modulo q of the lattice point at which $\mu_q * F_i$ is evaluated.

Lemma 7.2. Suppose $1 \leq q \leq 2Q^L$, $a \in (\mathbb{Z}/q\mathbb{Z})^L$, and $\|\theta\|_\infty \leq Q/R^2$. Then, for all functions $F_0, \dots, F_{K-1} : \mathbb{Z}^d \rightarrow \mathbb{C}$ supported on $[N]^d$,

$$\left| \mathcal{S}_{a/q+\theta}(F_0, \dots, F_{K-1}) - \mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}) \right| \ll_{K,d,\sigma} Q^{-L(L+2)-2} N^{Kd/2} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \quad (7.6)$$

Proof. Since every $|z_i - z_j|^2$ is an integer,

$$e(\mathcal{Q}_{a/q+\theta}(z)) = e_q(\mathcal{Q}_a(z))e(\mathcal{Q}_\theta(z)).$$

Consequently,

$$\mathcal{S}_{a/q+\theta}(F_0, \dots, F_{K-1}) = \sum_{z \in (\mathbb{Z}^d)^K} e_q(\mathcal{Q}_a(z))e(\mathcal{Q}_\theta(z)) \prod_{i=0}^{K-1} F_i(z_i). \quad (7.7)$$

We now rewrite the major arc form in the variables appearing in (7.7). Fix $r = (r_0, \dots, r_{K-1})$. For each $0 \leq i \leq K-1$, every $u_i \in \mathbb{Z}^d$ has a unique representation

$$u_i = y_i + s_i, \quad y_i \equiv r_i \pmod{q}, \quad s_i \in \{0, \dots, q-1\}^d.$$

It follows from (7.5) that

$$\begin{aligned} & \mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}) \\ &= \mathbb{E}_{r \bmod q} \sum_{\substack{y \in (\mathbb{Z}^d)^K \\ y_i \equiv r_i \pmod{q} \ (0 \leq i \leq K-1)}} e_q(\mathcal{Q}_a(r)) \prod_{i=0}^{K-1} (\mu_q * F_i)(y_i) \sum_{\substack{s_i \in \{0, \dots, q-1\}^d \\ 0 \leq i \leq K-1}} e(\mathcal{Q}_\theta(y + s)). \end{aligned}$$

Expanding the convolutions gives

$$\begin{aligned} & \mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}) \\ &= \frac{1}{N^{Kd}} \mathbb{E}_{r \bmod q} \sum_{\substack{h_i \in \{0, \dots, N_q-1\}^d \\ 0 \leq i \leq K-1}} \sum_{\substack{y \in (\mathbb{Z}^d)^K \\ y_i \equiv r_i \pmod{q} \ (0 \leq i \leq K-1)}} e_q(\mathcal{Q}_a(r)) \prod_{i=0}^{K-1} F_i(y_i - qh_i) \\ & \quad \times \sum_{\substack{s_i \in \{0, \dots, q-1\}^d \\ 0 \leq i \leq K-1}} e(\mathcal{Q}_\theta(y + s)). \end{aligned}$$

For fixed $h = (h_0, \dots, h_{K-1})$, make the change of variables

$$z_i := y_i - qh_i \quad (0 \leq i \leq K-1).$$

Then $z_i \equiv r_i \pmod{q}$, $y + s = z + qh + s$, and $e_q(\mathcal{Q}_a(r)) = e_q(\mathcal{Q}_a(z))$. For each $z \in (\mathbb{Z}^d)^K$, there is exactly one $r \bmod q$ satisfying $z_i \equiv r_i \pmod{q}$ for every i . Since the average over r is normalized, it contributes a factor q^{-Kd} . Therefore

$$\begin{aligned} & \mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}) \\ &= \frac{1}{(qN_q)^{Kd}} \sum_{z \in (\mathbb{Z}^d)^K} e_q(\mathcal{Q}_a(z)) \prod_{i=0}^{K-1} F_i(z_i) \sum_{\substack{h_i \in \{0, \dots, N_q-1\}^d \\ s_i \in \{0, \dots, q-1\}^d \\ 0 \leq i \leq K-1}} e(\mathcal{Q}_\theta(z + qh + s)). \end{aligned}$$

Finally, for each i , the map $(h_i, s_i) \mapsto t_i := qh_i + s_i$ is a bijection from

$$\{0, \dots, N_q - 1\}^d \times \{0, \dots, q - 1\}^d \quad \text{onto} \quad \{0, \dots, qN_q - 1\}^d.$$

Consequently,

$$\mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}) = \sum_{z \in (\mathbb{Z}^d)^K} e_q(\mathcal{Q}_a(z)) \prod_{i=0}^{K-1} F_i(z_i) \times \frac{1}{(qN_q)^{Kd}} \sum_{\substack{t_i \in \{0, \dots, qN_q - 1\}^d \\ 0 \leq i \leq K-1}} e(\mathcal{Q}_\theta(z + t)). \quad (7.8)$$

Comparing (7.7) and (7.8), the rational phase and the values of the F_i are identical. The only difference is that $e(\mathcal{Q}_\theta(z))$ is replaced by the average of $e(\mathcal{Q}_\theta(z + t))$ over the indicated choices of t .

For $0 \leq i \leq K - 1$,

$$\nabla_{x_i} \mathcal{Q}_\theta(x) = 2 \sum_{j: \{i,j\} \in \mathcal{E}} \theta_{ij}(x_i - x_j).$$

If $z_i \in [N]^d$ and $t_i \in \{0, \dots, qN_q - 1\}^d$ for every i , then every point x on the segment joining z to $z + t$ satisfies

$$\max_{i,j} |x_i - x_j| \ll_d N + qN_q \ll_d N.$$

Since $\|\theta\|_\infty \leq Q/R^2$ and $qN_q \leq RQ^{-L(L+2)-3}$, the mean value theorem gives

$$|e(\mathcal{Q}_\theta(z)) - e(\mathcal{Q}_\theta(z + t))| \ll_{K,d} \frac{QN}{R^2} qN_q \ll_{K,d,\sigma} Q^{-L(L+2)-2}.$$

Here we used $R = \lfloor \sigma N \rfloor$ and (7.1), which imply $N \ll_\sigma R$. Averaging this estimate in t and using (7.7)–(7.8) gives

$$\left| \mathcal{S}_{a/q+\theta}(F_0, \dots, F_{K-1}) - \mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}) \right| \ll_{K,d,\sigma} Q^{-L(L+2)-2} \prod_{i=0}^{K-1} \sum_{z_i \in [N]^d} |F_i(z_i)|. \quad (7.9)$$

Cauchy–Schwarz in each factor proves (7.6). $\square$

7.2 The complete sum

With normalized counting measure, define the complete cut operator

$$\begin{aligned} \mathcal{K}_{q,a,d}^{S,T} : L^2((\mathbb{Z}/q\mathbb{Z})^d)^T &\longrightarrow L^2((\mathbb{Z}/q\mathbb{Z})^d)^S, \\ (\mathcal{K}_{q,a,d}^{S,T} H)(r_S) &:= \mathbb{E}_{r_T \bmod q} e_q \left(-2 \sum_{i \in S, j \in T} a_{ij} r_i \cdot r_j \right) H(r_T). \end{aligned}$$

We now compute the $2 \rightarrow 2$ norm of this cut operator.

Lemma 7.3. *One has*

$$\|\mathcal{K}_{q,a,d}^{S,T}\|_{2 \rightarrow 2} = |\text{im}(2a_{S,T})|^{-d/2}. \quad (7.10)$$

Proof. By the same argument as in the proof of Lemma 5.2,

$$\|\mathcal{K}_{q,a,d}^{S,T}\|_{2 \rightarrow 2} = \|\mathcal{K}_{q,a,1}^{S,T}\|_{2 \rightarrow 2}^d. \quad (7.11)$$

It therefore remains to compute $\|\mathcal{K}_{q,a,1}^{S,T}\|_{2 \rightarrow 2}$.

Write $\mathcal{K} := \mathcal{K}_{q,a,1}^{S,T}$. For a function f on $(\mathbb{Z}/q\mathbb{Z})^S$, expanding $\mathcal{K}\mathcal{K}^*$ gives

$$\begin{aligned} (\mathcal{K}\mathcal{K}^* f)(r_S) &= \mathbb{E}_{r'_S \bmod q} f(r'_S) \mathbb{E}_{r_T \bmod q} e_q \left((r'_S - r_S)^\top 2a_{S,T} r_T \right) \\ &= q^{-|S|} \sum_{\substack{r'_S \bmod q \\ (2a_{S,T})^\top (r_S - r'_S) = 0}} f(r'_S), \end{aligned}$$

where the second equality follows from the orthogonality relation. Making the substitution $h = r_S - r'_S$ gives

$$(\mathcal{K}\mathcal{K}^* f)(r_S) = q^{-|S|} \sum_{h \in \ker((2a_{S,T})^\top)} f(r_S - h).$$

Hence, by Cauchy–Schwarz,

$$\begin{aligned} \|\mathcal{K}\mathcal{K}^* f\|_2^2 &= q^{-2|S|} \mathbb{E}_{r_S \bmod q} \left| \sum_{h \in \ker((2a_{S,T})^\top)} f(r_S - h) \right|^2 \\ &\leq q^{-2|S|} |\ker((2a_{S,T})^\top)| \sum_{h \in \ker((2a_{S,T})^\top)} \mathbb{E}_{r_S \bmod q} |f(r_S - h)|^2 \\ &= \left(\frac{|\ker((2a_{S,T})^\top)|}{q^{|S|}} \right)^2 \|f\|_2^2. \end{aligned}$$

This bound is attained by $f = \mathbf{1}$, since $\mathcal{K}\mathcal{K}^* \mathbf{1} = (|\ker((2a_{S,T})^\top)|/q^{|S|}) \mathbf{1}$. Thus

$$\|\mathcal{K}\mathcal{K}^*\|_{2 \rightarrow 2} = \frac{|\ker((2a_{S,T})^\top)|}{q^{|S|}}.$$

Finally, the orthogonality relation gives

$$\begin{aligned} |\ker((2a_{S,T})^\top)| &= \sum_{r_S \bmod q} \mathbb{E}_{r_T \bmod q} e_q \left(r_S^\top 2a_{S,T} r_T \right) \\ &= q^{|S|-|T|} |\ker(2a_{S,T})| \\ &= \frac{q^{|S|}}{|\text{im}(2a_{S,T})|}, \end{aligned}$$

where the last equality follows from

$$q^{|T|} = |\ker(2a_{S,T})| |\text{im}(2a_{S,T})|.$$

Consequently,

$$\|\mathcal{K}_{q,a,1}^{S,T}\|_{2 \rightarrow 2}^2 = \|\mathcal{K}\mathcal{K}^*\|_{2 \rightarrow 2} = |\text{im}(2a_{S,T})|^{-1}. \quad (7.12)$$

Combining (7.11) and (7.12) proves (7.10). $\square$

We need an unrestricted matrix moment and a stronger moment when one entry is a unit.

Lemma 7.4 (Matrix moments). *Let p be prime, and let m, s, t be positive integers. If $\tau > s + t - 1$, then, with normalized averaging over $\mathbf{M} \in \text{Mat}_{s \times t}(\mathbb{Z}/p^m\mathbb{Z})$, one has, for every fixed $1 \leq i \leq s$ and $1 \leq j \leq t$,*

$$\mathbb{E}_{\mathbf{M}} |\text{im}(2\mathbf{M})|^{-\tau} \ll_{s,t,\tau} p^{-mst}, \quad (7.13)$$

$$\mathbb{E}_{\mathbf{M}} \left[\mathbf{1}_{\mathbf{M}_{ij} \in (\mathbb{Z}/p^m\mathbb{Z})^\times} |\text{im}(2\mathbf{M})|^{-\tau} \right] \ll_{s,t,\tau} p^{-m[\tau + (s-1)(t-1)]}. \quad (7.14)$$

The implicit constants depend only on s, t , and τ .

Proof. Put

$$W_{s,t}(p^m) := \sum_{\mathbf{M} \in \text{Mat}_{s \times t}(\mathbb{Z}/p^m\mathbb{Z})} |\text{im } \mathbf{M}|^{-\tau},$$

with the conventions $W_{0,t}(p^m) = W_{s,0}(p^m) = W_{s,t}(1) = 1$.

The matrices divisible by p contribute $W_{s,t}(p^{m-1})$. Every other matrix has a unit entry. Choose such an entry and use it as a pivot. After permuting rows and columns, write

$$\mathbf{M} = \begin{pmatrix} u & \mathbf{b} \\ \mathbf{c} & \mathbf{D} \end{pmatrix},$$

where $\mathbf{b}$ is a row vector and $\mathbf{c}$ is a column vector. Invertible row and column operations reduce $\mathbf{M}$ to

$$\begin{pmatrix} u & 0 \\ 0 & \mathbf{M}' \end{pmatrix}, \quad \mathbf{M}' = \mathbf{D} - \mathbf{c}u^{-1}\mathbf{b}.$$

These operations preserve image cardinality, and hence $|\text{im } \mathbf{M}| = p^m |\text{im } \mathbf{M}'|$. There are $O_{s,t}(p^{m(s+t-1)})$ choices for the position of the pivot and for $u, \mathbf{b}$, and $\mathbf{c}$. For each such choice, the map $\mathbf{C} \mapsto \mathbf{M}' = \mathbf{C} - \mathbf{c}u^{-1}\mathbf{b}$ is a bijection of $\text{Mat}_{(s-1) \times (t-1)}(\mathbb{Z}/p^m\mathbb{Z})$, so $\mathbf{M}'$ may be summed freely. Therefore

$$W_{s,t}(p^m) \leq W_{s,t}(p^{m-1}) + c_{s,t} p^{-m(\tau-s-t+1)} W_{s-1,t-1}(p^m).$$

We now induct on $\min(s, t)$. The base case is $\min(s, t) = 0$, when $W_{s,t}(p^m) = 1$ by convention. Suppose that $s, t \geq 1$, and assume inductively that $W_{s-1,t-1}(p^r) \ll_{s,t,\tau} 1$ uniformly in p and r . Set $\delta := \tau - s - t + 1 > 0$. Iterating the preceding recursion from m down to 0 gives

$$\begin{aligned} W_{s,t}(p^m) &\leq W_{s,t}(p^{m-1}) + c_{s,t} p^{-m\delta} W_{s-1,t-1}(p^m) \\ &\leq W_{s,t}(p^{m-2}) + c_{s,t} p^{-(m-1)\delta} W_{s-1,t-1}(p^{m-1}) + c_{s,t} p^{-m\delta} W_{s-1,t-1}(p^m) \\ &\vdots \\ &\leq W_{s,t}(1) + c_{s,t} \sum_{r=1}^m p^{-r\delta} W_{s-1,t-1}(p^r) \\ &\ll 1 + \sum_{r=1}^m p^{-r\delta} \ll 1. \end{aligned}$$

Consequently,

$$\mathbb{E}_{\mathbf{M}} |\text{im } \mathbf{M}|^{-\tau} = p^{-mst} W_{s,t}(p^m) \ll p^{-mst}. \quad (7.15)$$

It remains to pass from $\mathbf{M}$ to $2\mathbf{M}$. If p is odd, multiplication by 2 is invertible modulo p^m , so $|\text{im}(2\mathbf{M})| = |\text{im } \mathbf{M}|$. The estimate (7.15) therefore yields (7.13).

Suppose now that $p = 2$ and $m \geq 2$, and let $\overline{\mathbf{M}}$ denote the reduction of $\mathbf{M}$ modulo 2^{m-1} . Then $|\text{im}(2\mathbf{M} \bmod 2^m)| = |\text{im } \overline{\mathbf{M}}|$. Every matrix modulo 2^{m-1} has exactly 2^{st} lifts modulo 2^m . Hence, under normalized averaging, we have

$$\mathbb{E}_{\mathbf{M} \bmod 2^m} |\text{im}(2\mathbf{M})|^{-\tau} = \mathbb{E}_{\overline{\mathbf{M}} \bmod 2^{m-1}} |\text{im } \overline{\mathbf{M}}|^{-\tau} \ll 2^{-(m-1)st} \ll 2^{-mst}.$$

The constant lost in the final inequality depends only on s, t , and τ . The case $m = 1$ is immediate.

Finally, (7.14) follows immediately from the proof of (7.13): use the prescribed unit entry M_{ij} as the pivot, and, when $p = 2$, note that reduction modulo 2^{m-1} preserves whether an entry is a unit. $\square$

For $a \in (\mathbb{Z}/q\mathbb{Z})^L$, define

$$\mathfrak{s}_q(a) := \prod_{S|T \in \mathcal{B}_K} |\text{im}(2a_{S,T})|^{-\tau w_K}. \quad (7.16)$$

We use the exponent $\tau := d\lfloor K^2/4 \rfloor / (2L)$ introduced in the proof of Proposition 6.1. Under the hypothesis $d \geq 4K + 4$, that proof shows that $\tau > K + 2$. Then put $\eta := (\tau - K)/2$. Since $\tau > K + 2$, we have $\eta > 1$.

Lemma 7.5. *If $d \geq 4K + 4$, then $\eta > 1$ and, for every $q \geq 1$,*

$$\sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a,q)=1}} \mathfrak{s}_q(a) \ll_{K,d} q^{-1-\eta}. \quad (7.17)$$

Proof. The case $q = 1$ is immediate. First let $q = p^m$. For each $e \in \mathcal{E}$, put

$$I_e := \mathbb{E}_{a \in (\mathbb{Z}/p^m\mathbb{Z})^L} \left[\mathbf{1}_{a_e \in (\mathbb{Z}/p^m\mathbb{Z})^\times} \mathfrak{s}_{p^m}(a) \right].$$

Fix $e \in \mathcal{E}$. In applying Lemma 5.5, take $J = |\mathcal{B}_K|$ and label the indices $1, \dots, J$ by the balanced cuts $S | T \in \mathcal{B}_K$. Thus, for the index j corresponding to $S | T$, the set $\mathcal{E}_j$ in that lemma is the set of edges crossing $S | T$, and the weight c_j is w_K . We write $f_{S,T}$ for the corresponding function f_j .

For each balanced cut $S | T$, define

$$f_{S,T}(a_{S,T}) := \begin{cases} \mathbf{1}_{a_e \in (\mathbb{Z}/p^m\mathbb{Z})^\times} |\text{im}(2a_{S,T})|^{-\tau}, & e \text{ crosses } S | T, \\ |\text{im}(2a_{S,T})|^{-\tau}, & e \text{ does not cross } S | T. \end{cases}$$

When e crosses $S | T$, the coordinate a_e is one of the coordinates in $a_{S,T}$, so the first function is indeed defined on the product of the coordinate spaces indexed by $\mathcal{E}_j$.

For every $e' \in \mathcal{E}$, Lemma 5.4 gives

$$\sum_{\substack{S|T \in \mathcal{B}_K \\ e' \text{ crosses } S|T}} w_K = 1,$$

which is exactly the hypothesis of Lemma 5.5. Moreover,

$$\prod_{S|T \in \mathcal{B}_K} f_{S,T}(a_{S,T})^{w_K} = \left(\prod_{\substack{S|T \in \mathcal{B}_K \\ e \text{ crosses } S|T}} \mathbf{1}_{a_e \in (\mathbb{Z}/p^m\mathbb{Z})^\times}^{w_K} \right) \prod_{S|T \in \mathcal{B}_K} |\text{im}(2a_{S,T})|^{-\tau w_K} = \mathbf{1}_{a_e \in (\mathbb{Z}/p^m\mathbb{Z})^\times} \mathfrak{s}_{p^m}(a),$$

where the last equality follows from (5.11). Hence

$$I_e = \mathbb{E}_{a \in (\mathbb{Z}/p^m\mathbb{Z})^L} \prod_{S|T \in \mathcal{B}_K} f_{S,T}(a_{S,T})^{w_K}.$$

Lemma 5.5 therefore gives

$$\begin{aligned} I_e &\leq \prod_{\substack{S|T \in \mathcal{B}_K \\ e \text{ crosses } S|T}} \left(\mathbb{E}_{a_{S,T}} \left[\mathbf{1}_{a_e \in (\mathbb{Z}/p^m\mathbb{Z})^\times} |\text{im}(2a_{S,T})|^{-\tau} \right] \right)^{w_K} \\ &\quad \times \prod_{\substack{S|T \in \mathcal{B}_K \\ e \text{ does not cross } S|T}} \left(\mathbb{E}_{a_{S,T}} |\text{im}(2a_{S,T})|^{-\tau} \right)^{w_K}, \end{aligned}$$

where each average in $a_{S,T}$ is normalized over $\text{Mat}_{|S| \times |T|}(\mathbb{Z}/p^m\mathbb{Z})$.

We now estimate the averages on the right. Fix a balanced cut $S \mid T$. If e crosses $S \mid T$, then (7.14) gives

$$\mathbb{E}_{a_{S,T}} \left[\mathbf{1}_{a_e \in (\mathbb{Z}/p^m\mathbb{Z})^\times} |\text{im}(2a_{S,T})|^{-\tau} \right] \ll_{K,d} p^{-m[\tau + (|S|-1)(|T|-1)]} = p^{-m(|S||T|+1+2\eta)},$$

where we used $|S| + |T| = K$ and $2\eta = \tau - K$. If e does not cross $S \mid T$, then (7.13) gives

$$\mathbb{E}_{a_{S,T}} |\text{im}(2a_{S,T})|^{-\tau} \ll_{K,d} p^{-m|S||T|}.$$

Consequently,

$$I_e \ll_{K,d} p^{-mw_K} \sum_{S \mid T \in \mathcal{B}_K} |S||T| p^{-m(1+2\eta)w_K \#\{S \mid T \in \mathcal{B}_K : e \text{ crosses } S \mid T\}}.$$

By (6.12) and (5.11), $I_e \ll_{K,d} p^{-m(L+1+2\eta)}$.

If $(a, p^m) = 1$, then a_e is a unit for at least one $e \in \mathcal{E}$. Therefore

$$\mathbf{1}_{(a, p^m)=1} \leq \sum_{e \in \mathcal{E}} \mathbf{1}_{a_e \in (\mathbb{Z}/p^m\mathbb{Z})^\times},$$

and hence

$$\sum_{\substack{a \in (\mathbb{Z}/p^m\mathbb{Z})^L \\ (a, p^m)=1}} \mathfrak{s}_{p^m}(a) = p^{mL} \mathbb{E}_{a \in (\mathbb{Z}/p^m\mathbb{Z})^L} \left[\mathbf{1}_{(a, p^m)=1} \mathfrak{s}_{p^m}(a) \right] \leq p^{mL} \sum_{e \in \mathcal{E}} I_e \ll_{K,d} p^{-m(1+2\eta)}.$$

Now write $q = \prod_{r=1}^u p_r^{m_r}$. Under the Chinese remainder theorem, $(\mathbb{Z}/q\mathbb{Z})^L \cong \prod_{r=1}^u (\mathbb{Z}/p_r^{m_r}\mathbb{Z})^L$. Write $a^{(r)}$ for the component of a modulo $p_r^{m_r}$. Then

$$(a, q) = 1 \iff (a^{(r)}, p_r^{m_r}) = 1 \text{ for every } 1 \leq r \leq u.$$

Moreover, for every balanced cut $S \mid T$, the map $2a_{S,T}$ corresponds under the Chinese remainder theorem to the product of the maps $2a_{S,T}^{(r)}$. Therefore

$$|\text{im}(2a_{S,T})| = \prod_{r=1}^u |\text{im}(2a_{S,T}^{(r)})|,$$

and consequently $\mathfrak{s}_q(a) = \prod_{r=1}^u \mathfrak{s}_{p_r^{m_r}}(a^{(r)})$. It follows that

$$\sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a, q)=1}} \mathfrak{s}_q(a) = \prod_{r=1}^u \left(\sum_{\substack{a^{(r)} \in (\mathbb{Z}/p_r^{m_r}\mathbb{Z})^L \\ (a^{(r)}, p_r^{m_r})=1}} \mathfrak{s}_{p_r^{m_r}}(a^{(r)}) \right) \ll_{K,d} C^{\omega(q)} \prod_{r=1}^u p_r^{-m_r(1+2\eta)} = C^{\omega(q)} q^{-1-2\eta}$$

for some $C = C(K, d)$. Since $C^{\omega(q)} \ll_{C, \eta} q^\eta$, this proves (7.17). $\square$

7.3 Integrating the major arc form

We equip $\mathbb{Z}^d \times (\mathbb{Z}/q\mathbb{Z})^d$ with counting measure in the first factor and normalized counting measure in the second. The estimates in this section give the following bound for the major arc form (7.4).

Proposition 7.6. *Assume $d \geq 4K + 4$, and let $q \geq 1$. Suppose that*

$$G_i \in L^2(\mathbb{Z}^d \times (\mathbb{Z}/q\mathbb{Z})^d) \quad (0 \leq i \leq K-1).$$

For each i , suppose that

$$\text{supp } G_i \subseteq [P]^d \times (\mathbb{Z}/q\mathbb{Z})^d.$$

Then

$$\sum_{\substack{a \bmod q \\ (a,q)=1}} \int_{\mathbb{T}^L} |\mathcal{M}_{q,a,\theta}(G_0, \dots, G_{K-1})| d\theta \ll_{K,d} q^{-1-\eta} P^{Kd/2-2L} \prod_{i=0}^{K-1} \|G_i\|_2. \quad (7.18)$$

Proof. Fix a balanced cut $S \mid T$. For each i , take $X_i = [P]^d \times (\mathbb{Z}/q\mathbb{Z})^d$. Use the three phase pieces

$$\begin{aligned} \mathcal{Q}^S(u_S, r_S) &:= \mathcal{Q}_\theta^S(u_S) + \frac{\mathcal{Q}_a^S(r_S)}{q}, \\ \mathcal{Q}^{S,T}(u_S, r_S, u_T, r_T) &:= \mathcal{Q}_\theta^{S,T}(u_S, u_T) + \frac{\mathcal{Q}_a^{S,T}(r_S, r_T)}{q}, \\ \mathcal{Q}^T(u_T, r_T) &:= \mathcal{Q}_\theta^T(u_T) + \frac{\mathcal{Q}_a^T(r_T)}{q}. \end{aligned}$$

The cut operator in Lemma 5.1 is therefore

$$\begin{aligned} \mathcal{K}^{S,T} : L^2([P]^d \times (\mathbb{Z}/q\mathbb{Z})^d)^T &\longrightarrow L^2([P]^d \times (\mathbb{Z}/q\mathbb{Z})^d)^S, \\ (\mathcal{K}^{S,T} H)(u_S, r_S) &:= \mathbb{E}_{r_T \bmod q} \sum_{u_T \in ([P]^d)^T} e_q\left(\mathcal{Q}_a^{S,T}(r_S, r_T)\right) e\left(\mathcal{Q}_\theta^{S,T}(u_S, u_T)\right) H(u_T, r_T). \end{aligned}$$

Thus

$$|\mathcal{M}_{q,a,\theta}(G_0, \dots, G_{K-1})| \leq \|\mathcal{K}^{S,T}\|_{2 \rightarrow 2} \prod_{i=0}^{K-1} \|G_i\|_2.$$

We now identify the norm of $\mathcal{K}^{S,T}$. The product structure of its kernel gives

$$\mathcal{K}^{S,T} = \mathcal{K}_{P,d}^{S,T}(-2\theta_{S,T}) \otimes \mathcal{K}_{q,a,d}^{S,T}. \quad (7.19)$$

Hence (7.19), multiplicativity of the operator norm under tensor products, and Lemmas 7.3 and 5.2 give

$$\begin{aligned} \|\mathcal{K}^{S,T}\|_{2 \rightarrow 2} &= \|\mathcal{K}_{P,d}^{S,T}(-2\theta_{S,T})\|_{2 \rightarrow 2} \|\mathcal{K}_{q,a,d}^{S,T}\|_{2 \rightarrow 2} \\ &= P^{Kd/2} m_P^{S,T}(-2\theta_{S,T})^d |\text{im}(2a_{S,T})|^{-d/2}. \end{aligned}$$

Since $d/|\mathcal{B}_K| = 2\tau w_K$, taking the geometric mean of these bounds over the balanced cuts gives

$$|\mathcal{M}_{q,a,\theta}(G_0, \dots, G_{K-1})| \leq P^{Kd/2} \prod_{S \mid T \in \mathcal{B}_K} \left[|\text{im}(2a_{S,T})|^{-1/2} m_P^{S,T}(-2\theta_{S,T}) \right]^{2\tau w_K} \prod_{i=0}^{K-1} \|G_i\|_2.$$

By the definition (7.16),

$$|\mathcal{M}_{q,a,\theta}(G_0, \dots, G_{K-1})| \leq P^{Kd/2} \mathfrak{s}_q(a) \prod_{S|T \in \mathcal{B}_K} m_P^{S,T}(-2\theta_{S,T})^{2\tau w_K} \prod_{i=0}^{K-1} \|G_i\|_2.$$

It remains to integrate the product of the $m_P^{S,T}$ -terms. In applying Lemma 5.5, take $J = |\mathcal{B}_K|$ and label the indices $1, \dots, J$ by the balanced cuts $S | T \in \mathcal{B}_K$. Thus, for the index j corresponding to $S | T$, the set $\mathcal{E}_j$ in that lemma is the set of edges crossing $S | T$, and the weight c_j is w_K . We write $f_{S,T}$ for the corresponding function f_j and set

$$f_{S,T}(\theta_{S,T}) := m_P^{S,T}(-2\theta_{S,T})^{2\tau}.$$

For every $e \in \mathcal{E}$, Lemma 5.4 gives

$$\sum_{\substack{S|T \in \mathcal{B}_K \\ e \text{ crosses } S|T}} w_K = 1,$$

which is exactly the hypothesis of Lemma 5.5. Moreover,

$$\prod_{S|T \in \mathcal{B}_K} f_{S,T}(\theta_{S,T})^{w_K} = \prod_{S|T \in \mathcal{B}_K} m_P^{S,T}(-2\theta_{S,T})^{2\tau w_K}.$$

Lemma 5.5 therefore gives

$$\int_{\mathbb{T}^L} \prod_{S|T \in \mathcal{B}_K} m_P^{S,T}(-2\theta_{S,T})^{2\tau w_K} d\theta \leq \prod_{S|T \in \mathcal{B}_K} \left(\int_{\mathbb{T}^{|S||T|}} m_P^{S,T}(-2\theta_{S,T})^{2\tau} d\theta_{S,T} \right)^{w_K}.$$

Multiplication by -2 preserves normalized Haar measure on $\mathbb{T}^{|S||T|}$. Since $0 \leq m_P^{S,T} \leq 1$ and $2\tau > 2K + 4$, Lemma 2.1 gives

$$\int_{\mathbb{T}^{|S||T|}} m_P^{S,T}(-2\theta_{S,T})^{2\tau} d\theta_{S,T} \ll_K P^{-2|S||T|}.$$

Using (6.12), we therefore obtain

$$\int_{\mathbb{T}^L} \prod_{S|T \in \mathcal{B}_K} m_P^{S,T}(-2\theta_{S,T})^{2\tau w_K} d\theta \ll_K P^{-2L}.$$

Summing over a using Lemma 7.5 and applying this estimate gives

$$\sum_{\substack{a \bmod q \\ (a,q)=1}} \int_{\mathbb{T}^L} |\mathcal{M}_{q,a,\theta}(G_0, \dots, G_{K-1})| d\theta \ll_{K,d} q^{-1-\eta} P^{Kd/2-2L} \prod_{i=0}^{K-1} \|G_i\|_2.$$

This is (7.18). □

7.4 Proof of the major arc estimate

We can now prove the major arc estimate stated at the beginning of the section. It will simplify the argument to cover $\mathfrak{M}_Q$ by boxes centered at rational points a/q , with a single denominator q shared by all coordinates. For $1 \leq q \leq 2Q^L$ and $a \in (\mathbb{Z}/q\mathbb{Z})^L$ with $(a, q) = 1$, put

$$\mathfrak{B}_Q(a/q) := \prod_{e \in \mathcal{E}} \left\{ \beta_e \in \mathbb{T} : \|\beta_e - a_e/q\|_{\mathbb{T}} \leq \frac{Q}{R^2} \right\}.$$

Lemma 7.7. *Assume $2 \leq Q \leq R$. Then*

$$\mathfrak{M}_Q \subseteq \bigcup_{1 \leq q \leq 2Q^L} \bigcup_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a,q)=1}} \mathfrak{B}_Q(a/q). \quad (7.20)$$

Proof. Fix $\beta \in \mathfrak{M}_Q$. For every $e \in \mathcal{E}$, choose $1 \leq q_e \leq Q$ and $b_e \in \mathbb{Z}/q_e\mathbb{Z}$, with $(b_e, q_e) = 1$, such that

$$\|-2\beta_e - b_e/q_e\|_{\mathbb{T}} \leq \frac{Q}{q_e R^2}.$$

The two preimages of b_e/q_e under $x \mapsto -2x$ are $-b_e/(2q_e)$ and $(q_e - b_e)/(2q_e)$. Choose the one nearer to β_e and denote it by ρ_e . Then $-2\rho_e = b_e/q_e \in \mathbb{T}$ and

$$\|\beta_e - \rho_e\|_{\mathbb{T}} = \frac{1}{2} \|-2\beta_e - b_e/q_e\|_{\mathbb{T}} \leq \frac{Q}{2q_e R^2}.$$

Notice that the ρ_e divides $2q_e$.

Put $\rho := (\rho_e)_{e \in \mathcal{E}}$, and let q be the least common denominator of its coordinates. Then

$$q \mid \text{lcm}_{e \in \mathcal{E}}(2q_e) = 2 \text{lcm}_{e \in \mathcal{E}} q_e,$$

and hence

$$q \leq 2 \prod_{e \in \mathcal{E}} q_e \leq 2Q^L.$$

Write $\rho = a/q$ with $a \in (\mathbb{Z}/q\mathbb{Z})^L$. Since q is the least common denominator, $(a, q) = 1$. Finally, for every $e \in \mathcal{E}$,

$$\|\beta_e - a_e/q\|_{\mathbb{T}} = \|\beta_e - \rho_e\|_{\mathbb{T}} \leq \frac{Q}{2q_e R^2} \leq \frac{Q}{R^2}.$$

Therefore $\beta \in \mathfrak{B}_Q(a/q)$, proving (7.20). $\square$

Proof of Proposition 7.1. By definition,

$$\Lambda_N^{\mathfrak{M}_Q}(F_0, \dots, F_{K-1}) = \int_{\mathfrak{M}_Q} \widehat{w}_R(\mathcal{Q}_\beta(v)) \mathcal{S}_\beta(F_0, \dots, F_{K-1}) d\beta.$$

Lemma 7.7, the triangle inequality, and (4.13) therefore give

$$|\Lambda_N^{\mathfrak{M}_Q}(F_0, \dots, F_{K-1})| \ll_{\Delta} \sum_{1 \leq q \leq 2Q^L} \sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a,q)=1}} \int_{[-Q/R^2, Q/R^2]^L} |\mathcal{S}_{a/q+\theta}(F_0, \dots, F_{K-1})| d\theta. \quad (7.21)$$

For every q , a , and θ occurring in (7.21), put

$$E_{q,a,\theta} := \mathcal{S}_{a/q+\theta}(F_0, \dots, F_{K-1}) - \mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}).$$

Lemma 7.2 gives

$$|E_{q,a,\theta}| \ll_{K,d,\sigma} Q^{-L(L+2)-2} N^{Kd/2} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \quad (7.22)$$

Put

$$P_q := N + qN_q + q. \quad (7.23)$$

Then

$$\text{supp } G_{i,q} \subseteq [P_q]^d \times (\mathbb{Z}/q\mathbb{Z})^d.$$

Indeed, if $G_{i,q}(u, r) \neq 0$, then (7.5) and the definition of μ_q give $z \in [N]^d$, $h \in \{0, \dots, N_q - 1\}^d$, and $s \in \{0, \dots, q - 1\}^d$ such that

$$F_i(z) \neq 0, \quad u = z + qh + s.$$

Proposition 7.6 therefore gives

$$\begin{aligned} \sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a,q)=1}} \int_{[-Q/R^2, Q/R^2]^L} |\mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q})| d\theta &\leq \sum_{\substack{a \bmod q \\ (a,q)=1}} \int_{\mathbb{T}^L} |\mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q})| d\theta \\ &\ll_{K,d} q^{-1-\eta} P_q^{Kd/2-2L} \prod_{i=0}^{K-1} \|G_{i,q}\|_2. \end{aligned} \quad (7.24)$$

Observe that

$$\|G_{i,q}\|_2 \leq P_q^{d(1/2-1/K)} \|G_{i,q}\|_K. \quad (7.25)$$

For fixed $u \in \mathbb{Z}^d$, let $y = y(u, r)$ be the point in (7.5). The map $r \mapsto s := u - y(u, r)$ is a bijection from $(\mathbb{Z}/q\mathbb{Z})^d$ onto $\{0, \dots, q - 1\}^d$. Therefore

$$\begin{aligned} \|G_{i,q}\|_K^K &= \sum_{u \in \mathbb{Z}^d} \mathbb{E}_{r \bmod q} |G_{i,q}(u, r)|^K \\ &= \frac{1}{q^d} \sum_{s \in \{0, \dots, q-1\}^d} \sum_{u \in \mathbb{Z}^d} |(\mu_q * F_i)(u - s)|^K \\ &= \|\mu_q * F_i\|_{\ell^K(\mathbb{Z}^d)}^K. \end{aligned} \quad (7.26)$$

Moreover,

$$qN_q \leq R, \quad q \leq 2Q^L \leq R, \quad N \leq 2\sigma^{-1}R.$$

Here the first inequality follows from (7.2), the second from (7.1), and the third from $R = \lfloor \sigma N \rfloor$ and $N \geq 2/\sigma$. Thus $P_q \ll_\sigma R$. Equations (7.25) and (7.26) give

$$P_q^{Kd/2-2L} \prod_{i=0}^{K-1} \|G_{i,q}\|_2 \leq P_q^{(K-1)d-2L} \prod_{i=0}^{K-1} \|\mu_q * F_i\|_{\ell^K(\mathbb{Z}^d)}.$$

Consequently, summing (7.24) over $q \leq 2Q^L$ gives

$$\begin{aligned} \sum_{1 \leq q \leq 2Q^L} \sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a,q)=1}} \int_{[-Q/R^2, Q/R^2]^L} |\mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q})| d\theta \\ \ll_{K,d,\sigma} R^{(K-1)d-2L} \sum_{q \leq 2Q^L} q^{-1-\eta} \prod_{i=0}^{K-1} \|\mu_q * F_i\|_{\ell^K(\mathbb{Z}^d)}. \end{aligned} \quad (7.27)$$

For each q ,

$$\#\{a \in (\mathbb{Z}/q\mathbb{Z})^L : (a, q) = 1\} \leq q^L,$$

and the measure of $[-Q/R^2, Q/R^2]^L$ is $(2Q/R^2)^L$. Hence

$$\sum_{1 \leq q \leq 2Q^L} \sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a,q)=1}} \int_{[-Q/R^2, Q/R^2]^L} d\theta \leq \left(\frac{2Q}{R^2}\right)^L \sum_{q \leq 2Q^L} q^L \ll_L Q^{L(L+2)} R^{-2L}.$$

It follows from (7.22) that

$$\begin{aligned} \sum_{1 \leq q \leq 2Q^L} \sum_{\substack{\alpha \in (\mathbb{Z}/q\mathbb{Z})^L \\ (a,q)=1}} \int_{[-Q/R^2, Q/R^2]^L} |E_{q,a,\theta}| d\theta &\ll_{K,d,\sigma} Q^{-L(L+2)-2} N^{Kd/2} \cdot Q^{L(L+2)} R^{-2L} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)} \\ &\ll_{K,d,\sigma} Q^{-2} R^{Kd/2-2L} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)}. \end{aligned} \quad (7.28)$$

Since

$$\mathcal{S}_{a/q+\theta}(F_0, \dots, F_{K-1}) = \mathcal{M}_{q,a,\theta}(G_{0,q}, \dots, G_{K-1,q}) + E_{q,a,\theta},$$

the bounds (7.21), (7.27), and (7.28) prove (7.3). $\square$

The first term in (7.3) retains the averages $\mu_q * F_i$, which are used in the density-increment argument of the next section. After decreasing the exponent in Proposition 6.1 if necessary, assume from now on that $0 < \gamma \leq 1$; then the Q^{-2} error above is also $O(Q^{-\gamma})$.

8 The density increment and completion of the proof

The following proposition gives the dichotomy that will be iterated in the proof of Theorem 1.1.

Proposition 8.1 (Density-increment). *For the constant σ fixed in Section 4, there are constants $0 < \kappa < 1$ and $B_0 \geq 1$, depending only on Δ and d , with the following property. Let $A \subseteq [N]^d$ be nonempty and contain no nondegenerate similar copy of Δ , and set $\alpha := |A|/N^d$. Then one of the following alternatives holds:*

- (1) $N \ll_{\Delta,d} \alpha^{-B_0}$.
- (2) *There are an integer $N' \geq 1$ and a set $A' \subseteq [N']^d$, also containing no nondegenerate similar copy of Δ , such that $N' \gg_{\Delta,d} \alpha^{B_0} N$ and $|A'|/(N')^d > (1 + \kappa)\alpha$.*

Proof. Let $B_0 \geq 1$, $Q_0 \geq 2$, and $0 < \kappa < 1$ be constants to be chosen later depending only on Δ and d . Put $Q := Q_0 \alpha^{-B_0/(L(L+3)+3)}$. If $N < N_0$, then alternative (1) holds. Otherwise $R \geq \sigma N/2$, and if $R < 4Q^{L(L+3)+3}$, then

$$N \ll_{\Delta,d} Q^{L(L+3)+3} = Q_0^{L(L+3)+3} \alpha^{-B_0} \ll_{\Delta,d} \alpha^{-B_0},$$

so alternative (1) again holds. We may therefore assume (7.1). In particular, $2 \leq Q \leq R$, and the major- and minor arc estimates apply.

Let $\mathcal{B}$ and $s = |\mathcal{B}|$ be supplied by (4.9), and take $F_i = g$ for $i \in \mathcal{B}$ and $F_i = \mathbf{1}_{[N]^d}$ otherwise. Then

$$\alpha^s N^d R^{(K-1)d-2L} \ll_{\Delta,d} |\Lambda_N^{\mathfrak{M}Q}(F_0, \dots, F_{K-1})| + |\Lambda_N^{\mathfrak{m}Q}(F_0, \dots, F_{K-1})|.$$

The bounds in (4.7) give

$$\begin{aligned} \prod_{i=0}^{K-1} \|F_i\|_{\ell^2(\mathbb{Z}^d)} &= \|g\|_{\ell^2(\mathbb{Z}^d)}^s \|\mathbf{1}_{[N]^d}\|_{\ell^2(\mathbb{Z}^d)}^{K-s} \\ &\leq \alpha^{s/2} N^{Kd/2}. \end{aligned}$$

Since μ_q is a probability measure, Young's inequality gives

$$\|\mu_q * \mathbf{1}_{[N]^d}\|_{\ell^K(\mathbb{Z}^d)} \leq \|\mu_q\|_{\ell^1(\mathbb{Z}^d)} \|\mathbf{1}_{[N]^d}\|_{\ell^K(\mathbb{Z}^d)} = N^{d/K}.$$

Consequently,

$$\begin{aligned}
\prod_{i=0}^{K-1} \|\mu_q * F_i\|_{\ell^K(\mathbb{Z}^d)} &= \|\mu_q * g\|_{\ell^K(\mathbb{Z}^d)}^s \|\mu_q * \mathbf{1}_{[N]^d}\|_{\ell^K(\mathbb{Z}^d)}^{K-s} \\
&\leq N^{d(K-s)/K} \left(\sum_{x \in \mathbb{Z}^d} |\mu_q * g(x)|^K \right)^{s/K} \\
&= N^d \left(\frac{1}{N^d} \sum_{x \in \mathbb{Z}^d} |\mu_q * g(x)|^K \right)^{s/K},
\end{aligned}$$

where $d(K-s)/K + ds/K = d$.

Since $R \asymp_{\Delta} N$ and $Q^{-2} \leq Q^{-\gamma}$, Propositions 6.1 and 7.1 imply

$$\alpha^s \ll_{\Delta, d} \sum_{q \leq 2Q^L} q^{-1-\eta} \left(\frac{1}{N^d} \sum_{x \in \mathbb{Z}^d} |\mu_q * g(x)|^K \right)^{s/K} + Q^{-\gamma} \alpha^{s/2}. \quad (8.1)$$

Assuming B_0 is sufficiently large depending only on K and d , the exponent of α below is at least s , and hence

$$Q^{-\gamma} \alpha^{s/2} = Q_0^{-\gamma} \alpha^{\gamma B_0 / (L(L+3)+3) + s/2} \leq Q_0^{-\gamma} \alpha^s.$$

The contribution of this term to (8.1) is therefore $O_{\Delta, d}(Q_0^{-\gamma} \alpha^s)$. Assuming Q_0 is sufficiently large depending only on Δ and d , this contribution is at most $\alpha^s/2$ and may be moved to the left-hand side. Since $\sum_{q \geq 1} q^{-1-\eta} < \infty$, there is therefore some $q \leq 2Q^L$ such that

$$\sum_{x \in \mathbb{Z}^d} |\mu_q * g(x)|^K \gg_{\Delta, d} \alpha^K N^d. \quad (8.2)$$

Recall that $N_q = \lfloor R/(qQ^{L(L+2)+3}) \rfloor$. The function $\mu_q * g$ is supported on $[1, N + q(N_q - 1)]^d$, and

$$\begin{aligned}
|[1, N + q(N_q - 1)]^d \setminus [1 + q(N_q - 1), N]^d| &\ll_d q N_q N^{d-1} \\
&\ll_{\Delta, d} Q^{-L(L+2)-3} N^d \\
&\ll_{\Delta, d} Q_0^{-L(L+2)-3} \alpha^{B_0(L(L+2)+3)/(L(L+3)+3)} N^d.
\end{aligned}$$

Assuming B_0 is sufficiently large depending only on K , the exponent of α in the last line is at least K , so

$$|[1, N + q(N_q - 1)]^d \setminus [1 + q(N_q - 1), N]^d| \ll_{\Delta, d} Q_0^{-L(L+2)-3} \alpha^K N^d. \quad (8.3)$$

Since μ_q is a probability measure and $\sum_{x \in \mathbb{Z}^d} g(x) = 0$, we have $\sum_{x \in \mathbb{Z}^d} (\mu_q * g)(x) = 0$. Consequently,

$$\sum_{x \in \mathbb{Z}^d} (\mu_q * g)_+(x) = \sum_{x \in \mathbb{Z}^d} (\mu_q * g)_-(x) = \frac{1}{2} \sum_{x \in \mathbb{Z}^d} |(\mu_q * g)(x)|. \quad (8.4)$$

Since $g \geq -\alpha$ and $|g| \leq 1$, we also have

$$0 \leq (\mu_q * g)_- \leq \alpha, \quad |\mu_q * g| \leq 1.$$

Assuming Q_0 is sufficiently large depending only on Δ and d , (8.3) and (8.2) give

$$\sum_{x \in [1+q(N_q-1), N]^d} |(\mu_q * g)(x)|^K \gg_{\Delta, d} \alpha^K N^d. \quad (8.5)$$

On $[1 + q(N_q - 1), N]^d$ we have $\mu_q * g = \mu_q * \mathbf{1}_A - \alpha$. Since

$$|(\mu_q * g)(x)|^K = ((\mu_q * g)_+(x))^K + ((\mu_q * g)_-(x))^K,$$

(8.5) implies that at least one of the following two estimates holds:

$$\sum_{x \in [1+q(N_q-1), N]^d} ((\mu_q * g)_+(x))^K \gg_{\Delta, d} \alpha^K N^d, \quad (8.6)$$

$$\sum_{x \in [1+q(N_q-1), N]^d} ((\mu_q * g)_-(x))^K \gg_{\Delta, d} \alpha^K N^d. \quad (8.7)$$

If (8.6) holds, pigeonholing over at most N^d points gives some $y \in [1 + q(N_q - 1), N]^d$ such that $(\mu_q * g)(y) \gg_{\Delta, d} \alpha$. Suppose instead that (8.7) holds. Since $(\mu_q * g)_- \leq \alpha$,

$$\sum_{x \in [1+q(N_q-1), N]^d} (\mu_q * g)_-(x) \geq \alpha^{1-K} \sum_{x \in [1+q(N_q-1), N]^d} ((\mu_q * g)_-(x))^K \gg_{\Delta, d} \alpha N^d.$$

It follows from (8.4) that

$$\sum_{x \in \mathbb{Z}^d} (\mu_q * g)_+(x) = \sum_{x \in \mathbb{Z}^d} (\mu_q * g)_-(x) \gg_{\Delta, d} \alpha N^d. \quad (8.8)$$

Since $\mu_q * g$ is supported on $[1, N + q(N_q - 1)]^d$ and $(\mu_q * g)_+ \leq 1$, (8.3) gives

$$\sum_{x \in [1, N+q(N_q-1)]^d \setminus [1+q(N_q-1), N]^d} (\mu_q * g)_+(x) = O_{\Delta, d} \left(Q_0^{-L(L+2)-3} \alpha^K N^d \right).$$

Assuming Q_0 is sufficiently large depending only on Δ and d , this estimate, (8.8), and $\alpha^K \leq \alpha$ give

$$\sum_{x \in [1+q(N_q-1), N]^d} (\mu_q * g)_+(x) \gg_{\Delta, d} \alpha N^d.$$

Pigeonholing over at most N^d points again gives some $y \in [1 + q(N_q - 1), N]^d$ such that $(\mu_q * g)(y) \gg_{\Delta, d} \alpha$. Thus, assuming κ is sufficiently small depending only on Δ and d , there is some $y \in [1 + q(N_q - 1), N]^d$ such that

$$(\mu_q * \mathbf{1}_A)(y) = (\mu_q * g)(y) + \alpha > (1 + \kappa)\alpha.$$

Set $N' := N_q$. Since $q \leq 2Q^L$ and (7.1) holds,

$$\frac{R}{qQ^{L(L+2)+3}} \geq \frac{R}{2Q^{L(L+3)+3}} \geq 2.$$

It follows that

$$N' = \left\lfloor \frac{R}{qQ^{L(L+2)+3}} \right\rfloor \geq \frac{R}{2qQ^{L(L+2)+3}} \geq \frac{R}{4Q^{L(L+3)+3}} \gg_{\Delta, d} \alpha^{B_0} N.$$

Finally, put

$$A' := \left\{ n \in [N']^d : y - q(n - (1, \dots, 1)) \in A \right\}.$$

Because $y \in [1 + q(N_q - 1), N]^d$ and $N' = N_q$, the map $n \mapsto y - q(n - (1, \dots, 1))$ sends $[N']^d$ injectively into $[N]^d$. By the definition of μ_q ,

$$\frac{|A'|}{(N')^d} = (\mu_q * \mathbf{1}_A)(y) > (1 + \kappa)\alpha.$$

The map $n \mapsto y - q(n - (1, \dots, 1))$ is a composition of translations, a reflection, and a dilation, so A' remains Δ -free. This proves alternative (2). $\square$

Proof of Theorem 1.1. Fix the constants supplied by Proposition 8.1. If $A = \emptyset$, there is nothing to prove. Otherwise, set $A_0 = A$, $N_0 = N$, and, at every stage j , put $\alpha_j := |A_j|/N_j^d$. Apply Proposition 8.1 at each stage. If alternative (1) holds, stop. If alternative (2) holds, let $A_{j+1} \subseteq [N_{j+1}]^d$ be the set it supplies. Alternative (2) cannot hold indefinitely, since each iteration increases the density by a factor greater than $1 + \kappa$, whereas every density is at most 1. Let j_* be the first stage at which alternative (1) holds. Then

$$N_{j_*} \ll_{\Delta, d} \alpha_{j_*}^{-B_0}. \quad (8.9)$$

Iterating the density increase in alternative (2) gives $1 \geq \alpha_{j_*} \geq (1 + \kappa)^{j_*} \alpha_0$, and hence $j_* \ll_{\Delta, d} 1 + \log(1/\alpha_0)$. The bound $N_{j+1} \gg_{\Delta, d} \alpha_j^{B_0} N_j$ in alternative (2) gives, for $0 \leq j < j_*$,

$$\log N_j \leq \log N_{j+1} + B_0 \log(1/\alpha_j) + O_{\Delta, d}(1).$$

Summing over j and applying (8.9), we obtain

$$\log N_0 \leq B_0 \sum_{j=0}^{j_*} \log(1/\alpha_j) + O_{\Delta, d}(j_* + 1).$$

Since $\alpha_j \geq \alpha_0$ for every j , it follows that

$$\log N_0 \ll_{\Delta, d} (1 + \log(1/\alpha_0))^2.$$

Consequently, for suitable constants $C, c > 0$ depending only on Δ and d ,

$$\alpha_0 \leq C \exp(-c\sqrt{\log N_0}).$$

Since $\alpha_0 = |A|/N_0^d$, this proves the theorem. $\square$

A Application to a continuous variant of the problem

Proof of Corollary 1.3. Let $C_0, c_0 > 0$ be the constants in Theorem 1.1, and put $L_\alpha := 1 + \log(1/\alpha)$. Choose $D = D_{\Delta, d}$ sufficiently large and set

$$N := \left\lceil \exp(DL_\alpha^2) \right\rceil.$$

Then $C_0 \exp(-c_0\sqrt{\log N}) < \alpha$ and $N \leq 2 \exp(DL_\alpha^2)$.

For $u \in [0, 1/N)^d$, define

$$B_u := \left\{ m \in \{0, \dots, N-1\}^d : u + \frac{m}{N} \in E \right\}.$$

The translates of $[0, 1/N)^d$ by the vectors m/N partition $[0, 1)^d$, and therefore Fubini's theorem gives

$$\int_{[0, 1/N)^d} |B_u| du = |E|.$$

There is consequently some u such that $|B_u| \geq \alpha N^d$. By the choice of N and Theorem 1.1, the translate $B_u + (1, \dots, 1) \subseteq [N]^d$ contains a nondegenerate similar copy of Δ . Translating back, let $y_0, \dots, y_{K-1} \in B_u$ be its vertices, and let $\rho > 0$ be its similarity ratio.

For any fixed edge $\{i, j\}$ of Δ , the vector $y_i - y_j$ is a nonzero integer vector. Hence

$$\rho^2 = \frac{|y_i - y_j|^2}{|v_i - v_j|^2} \geq \frac{1}{|v_i - v_j|^2},$$

so $\rho \geq c_\Delta > 0$. Mapping the copy into E by $m \mapsto u + m/N$ gives a similar copy with ratio

$$\lambda = \frac{\rho}{N} \geq \exp\left(-C_{\Delta,d}(1 + \log(1/\alpha))^2\right),$$

after enlarging $C_{\Delta,d}$. Since the pairwise distances determine a nondegenerate simplex up to Euclidean isometry, this copy has the form $x + \lambda U \Delta$ for some $x \in \mathbb{R}^d$ and $U \in O(d)$. $\square$

B Counting solutions to the matrix equation

The trace expansion in Lemma 2.1 reduces the required high-moment estimate to counting bounded integer matrices U and V satisfying $UV^\top = 0$. The precise estimate needed in the main argument is the following.

Lemma B.1 (Integer matrix orthogonality). *Let s, t, n be positive integers with*

$$n \geq 2 \max(s, t) \quad \text{and} \quad n > s + t.$$

Then, for every positive integer R ,

$$\# \left\{ (U, V) : \begin{array}{l} U \in ([-R, R] \cap \mathbb{Z})^{s \times n}, \\ V \in ([-R, R] \cap \mathbb{Z})^{t \times n}, \\ UV^\top = 0 \end{array} \right\} \ll_{n,s,t} R^{n(s+t)-2st}.$$

In the application, $n = K + 1$, $s = |S|$, and $t = |T|$. Thus $n > s + t$ follows from $|S| + |T| = K$, while $n \geq 2 \max(s, t)$ follows from the balancedness of the cut $S \mid T$. We now develop the lattice terminology needed for the proof.

B.1 Lattice preliminaries

The definitions and elementary determinant facts below are standard; see [2, Chapter I]. We follow [12, §1] for the determinant convention and for the notation X_n , μ_n , $\text{Gr}(L, d)$, and rational flags.

A lattice in $\mathbb{R}^n$ is a subgroup of the form $\Gamma = \mathbb{Z}\gamma_1 + \cdots + \mathbb{Z}\gamma_d$, where $\gamma_1, \dots, \gamma_d \in \mathbb{R}^n$ are linearly independent. These vectors form a basis of Γ , and d is its rank. The real span of Γ is

$$\text{span}_{\mathbb{R}} \Gamma := \mathbb{R}\gamma_1 + \cdots + \mathbb{R}\gamma_d.$$

A sublattice of a lattice L is a subgroup $\Gamma \subseteq L$; every such subgroup is itself a lattice.

The Gram matrix of the basis $\gamma_1, \dots, \gamma_d$ is

$$G(\gamma_1, \dots, \gamma_d) := (\langle \gamma_i, \gamma_j \rangle)_{i,j=1}^d.$$

Following Kim, define the determinant of Γ by

$$\det \Gamma := \|\gamma_1 \wedge \cdots \wedge \gamma_d\| = (\det G(\gamma_1, \dots, \gamma_d))^{1/2}.$$

Any two bases of Γ differ by a matrix $A \in \text{GL}_d(\mathbb{Z})$. Under this change of basis, the Gram matrix G is replaced by $AGA^\top$, whose determinant equals $\det G$. Thus $\det \Gamma$ does not depend on the chosen basis.

If $\Gamma \subseteq \Lambda$ have the same rank, then the index $[\Lambda : \Gamma]$, namely the number of cosets of Γ in Λ , is finite, and

$$\det \Gamma = [\Lambda : \Gamma] \det \Lambda. \quad (\text{B.1})$$

Thus enlarging a lattice without changing its rank can only decrease its determinant. Moreover, if $\Gamma \subseteq \mathbb{Z}^n$, then the Gram matrix of any basis of Γ has integer entries and positive integer determinant, so

$$\det \Gamma \geq 1. \quad (\text{B.2})$$

A lattice in $\mathbb{R}^n$ is full-rank if it has rank n , and it is unimodular if it is full-rank and has determinant one. If $g \in \text{GL}_n(\mathbb{R})$, then $\mathbb{Z}^n g := \{zg : z \in \mathbb{Z}^n\}$ is a full-rank lattice of determinant $|\det g|$, where vectors are regarded as row vectors. In particular, $\mathbb{Z}^n g$ is unimodular when $g \in \text{SL}(n, \mathbb{R})$.

Following Kim, let $X_n := \text{SL}(n, \mathbb{Z}) \backslash \text{SL}(n, \mathbb{R})$. The coset of $g \in \text{SL}(n, \mathbb{R})$ corresponds to the unimodular row lattice $\mathbb{Z}^n g$. Left multiplication by an element of $\text{SL}(n, \mathbb{Z})$ merely changes the chosen lattice basis, and every unimodular lattice has such a basis matrix. Thus X_n is the space of unimodular lattices in $\mathbb{R}^n$. We equip it with the invariant probability measure μ_n induced by Haar measure on $\text{SL}(n, \mathbb{R})$.

A sublattice $\Gamma \subseteq L$ is primitive in L if $\Gamma = L \cap \text{span}_{\mathbb{R}} \Gamma$. For an arbitrary sublattice $\Gamma \subseteq L$, the lattice $L \cap \text{span}_{\mathbb{R}} \Gamma$ is called the saturation of Γ in L . It is primitive in L and has the same rank as Γ .

For $L \in X_n$ and $1 \leq d < n$, let $\text{Gr}(L, d)$ denote the set of primitive rank- d sublattices of L .

For a sublattice $\Gamma \subseteq \mathbb{Z}^n$, define its integral orthogonal complement by

$$\Gamma^{\perp_{\mathbb{Z}}} := \{x \in \mathbb{Z}^n : \langle x, \gamma \rangle = 0 \text{ for every } \gamma \in \Gamma\} = (\text{span}_{\mathbb{R}} \Gamma)^{\perp} \cap \mathbb{Z}^n.$$

If $\Gamma \subseteq \mathbb{Z}^n$ is primitive of rank d , standard lattice duality [21, §2, especially Lemma 1 and its corollary] gives

$$\Gamma^{\perp_{\mathbb{Z}}} \text{ is primitive of rank } n - d, \quad (\text{B.3})$$

$$\det(\Gamma^{\perp_{\mathbb{Z}}}) = \det \Gamma, \quad (\text{B.4})$$

and

$$(\Gamma^{\perp_{\mathbb{Z}}})^{\perp_{\mathbb{Z}}} = \Gamma. \quad (\text{B.5})$$

For a rank- d lattice Γ and $1 \leq j \leq d$, its j -th successive minimum $\lambda_j(\Gamma)$ is the least $\lambda > 0$ for which Γ contains j linearly independent vectors of Euclidean norm at most λ . Thus $\lambda_1(\Gamma) \leq \dots \leq \lambda_d(\Gamma)$. A standard consequence of the lattice-point estimate in [21, §5, Lemma 2] is

$$\#(\Gamma \cap [-R, R]^n) \ll_n \prod_{j=1}^d \left(1 + \frac{R}{\lambda_j(\Gamma)}\right), \quad (\text{B.6})$$

and Minkowski's second theorem [2, Chapter VIII, Theorem V] gives

$$\prod_{j=1}^d \lambda_j(\Gamma) \asymp_n \det \Gamma. \quad (\text{B.7})$$

If $\lambda_d(\Gamma) \ll_n R$, then $1 + R/\lambda_j(\Gamma) \ll_n R/\lambda_j(\Gamma)$ for every $1 \leq j \leq d$. Hence (B.6) and (B.7) give

$$\#(\Gamma \cap [-R, R]^n) \ll_n \frac{R^d}{\det \Gamma}. \quad (\text{B.8})$$

Finally, we fix Kim's notation for the flags appearing below. Let $V(i) := \pi^{i/2}/\Gamma(1+i/2)$ be the volume of the unit ball in $\mathbb{R}^i$, and adopt the convention $\zeta(1) = 1$. For $1 \leq d < n$, define

$$a(n, d) := \frac{1}{n} \binom{n}{d} \prod_{i=1}^d \frac{V(n-i+1)\zeta(i)}{V(i)\zeta(n-i+1)}.$$

For

$$d_0 = 0 < d_1 < \cdots < d_k < d_{k+1} = n, \quad \mathfrak{d} = (d_1, \dots, d_k),$$

Kim calls a chain

$$A_1 \subseteq \cdots \subseteq A_k \subseteq L, \quad A_i \in \text{Gr}(L, d_i),$$

a flag of type $\mathfrak{d}$, rational with respect to L , and defines

$$a(n, \mathfrak{d}) := a(n, d_1) \prod_{i=1}^{k-1} \frac{n - d_{i-1}}{d_{i+1} - d_{i-1}} a(n - d_i, d_{i+1} - d_i).$$

B.2 Kim's mean value formula for primitive lattice flags

We now record the only non-elementary input.

Theorem B.2 (Kim's mean value formula for rational flags [12, Theorem 5]). *For $H_1, \dots, H_k \geq 0$, the μ_n -average of the number of flags $A_1 \subseteq \cdots \subseteq A_k$ of type $\mathfrak{d}$, rational with respect to L , such that $\det A_i \leq H_i$ for $i = 1, \dots, k$, is equal to $a(n, \mathfrak{d}) \prod_{i=1}^k H_i^{d_{i+1} - d_{i-1}}$.*

We transfer the case $k = 2$ from a random unimodular lattice to the fixed lattice $\mathbb{Z}^n$.

Lemma B.3 (Two-height flag count). *Fix positive integers n, d_1, d_2 with $0 < d_1 < d_2 < n$. Uniformly for $H_1, H_2 \geq 1$,*

$$\# \left\{ (A_1, A_2) : \begin{array}{l} A_1 \in \text{Gr}(\mathbb{Z}^n, d_1), \\ A_2 \in \text{Gr}(\mathbb{Z}^n, d_2), \\ A_1 \subseteq A_2, \\ \det A_1 \leq H_1, \quad \det A_2 \leq H_2 \end{array} \right\} \ll_{n, d_1, d_2} H_1^{d_2} H_2^{n-d_1}.$$

Proof. Take $k = 2$ and $\mathfrak{d} = (d_1, d_2)$ in Theorem B.2. Since $d_0 = 0$ and $d_3 = n$, it gives

$$\begin{aligned} \int_{X_n} \# \left\{ (A_1, A_2) : \begin{array}{l} A_1 \in \text{Gr}(L, d_1), \\ A_2 \in \text{Gr}(L, d_2), \\ A_1 \subseteq A_2, \\ \det A_1 \leq H_1, \quad \det A_2 \leq H_2 \end{array} \right\} d\mu_n(L) &= a(n, \mathfrak{d}) H_1^{d_2 - d_0} H_2^{d_3 - d_1} \\ &= a(n, \mathfrak{d}) H_1^{d_2} H_2^{n-d_1}. \end{aligned} \quad (\text{B.9})$$

It remains to pass from this average to the particular lattice $\mathbb{Z}^n$. Let $\pi : \text{SL}(n, \mathbb{R}) \rightarrow X_n$ be the quotient map. Choose a relatively compact open neighborhood W of the identity on which π is injective, and put $\mathcal{U} = \pi(W)$. Then $\mu_n(\mathcal{U}) > 0$. For $A \subseteq \mathbb{Z}^n$ and $g \in \text{SL}(n, \mathbb{R})$, write

$Ag := \{ag : a \in A\}$. For $1 \leq j < n$, the relative compactness of W gives a constant $c_j < \infty$ such that

$$\det(Ag) \leq c_j \det A \quad (A \in \text{Gr}(\mathbb{Z}^n, j), g \in W).$$

Right multiplication by g maps $\text{Gr}(\mathbb{Z}^n, j)$ bijectively onto $\text{Gr}(\mathbb{Z}^n g, j)$ and preserves inclusions. Hence, for every $g \in W$,

$$\# \left\{ (A_1, A_2) : \begin{array}{l} A_i \in \text{Gr}(\mathbb{Z}^n, d_i) \quad (i = 1, 2), \\ A_1 \subseteq A_2, \\ \det A_1 \leq H_1, \quad \det A_2 \leq H_2 \end{array} \right\} \leq \# \left\{ (A_1, A_2) : \begin{array}{l} A_i \in \text{Gr}(\mathbb{Z}^n g, d_i) \quad (i = 1, 2), \\ A_1 \subseteq A_2, \\ \det A_1 \leq c_{d_1} H_1, \\ \det A_2 \leq c_{d_2} H_2 \end{array} \right\}.$$

Integrating this inequality over $\mathcal{U}$ and applying (B.9) with (H_1, H_2) replaced by $(c_{d_1} H_1, c_{d_2} H_2)$, we obtain

$$\mu_n(\mathcal{U}) \# \left\{ (A_1, A_2) : \begin{array}{l} A_i \in \text{Gr}(\mathbb{Z}^n, d_i) \quad (i = 1, 2), \\ A_1 \subseteq A_2, \\ \det A_1 \leq H_1, \quad \det A_2 \leq H_2 \end{array} \right\} \leq a(n, \mathfrak{d}) (c_{d_1} H_1)^{d_2} (c_{d_2} H_2)^{n-d_1}.$$

Since W is fixed in terms of n , this proves the result. $\square$

B.3 Proof of the matrix count

Proof of Lemma B.1. Decompose the count according to

$$d_1 = \text{rank } U, \quad d_2 = n - \text{rank } V.$$

If $d_1 = 0$, then $U = 0$, and there are $O(R^{nt})$ choices for V . This is at most the asserted bound because

$$n(s+t) - 2st - nt = s(n-2t) \geq 0.$$

The case $d_2 = n$, in which $V = 0$, is symmetric. We may therefore assume

$$1 \leq d_1 \leq s, \quad n-t \leq d_2 \leq n-1.$$

Since $d_1 + (n-d_2) \leq s+t < n$, we have $d_1 < d_2$.

The rows of U generate a sublattice of $\mathbb{Z}^n$. Its saturation is the intersection of $\mathbb{Z}^n$ with the real span of those rows. Define this saturated row lattice, and its analogue for V , by

$$\Gamma_U := \text{rowspan}_{\mathbb{R}}(U) \cap \mathbb{Z}^n, \quad \Gamma_V := \text{rowspan}_{\mathbb{R}}(V) \cap \mathbb{Z}^n.$$

Here $\text{rowspan}_{\mathbb{R}}$ denotes the real span of the rows. The lattices Γ_U and Γ_V are primitive of ranks d_1 and $n-d_2$, respectively. The equation $UV^T = 0$ says that their real row spaces are orthogonal. Consequently, if $\Lambda_V := \Gamma_V^{\perp_{\mathbb{Z}}}$, then $\Gamma_U \subseteq \Lambda_V$, and (B.3) shows that Λ_V is primitive of rank d_2 . Thus

$$\Gamma_U \subseteq \Lambda_V \subseteq \mathbb{Z}^n \tag{B.10}$$

is a flag of type (d_1, d_2) , rational with respect to $\mathbb{Z}^n$. Moreover, (B.4) gives $\det \Lambda_V = \det \Gamma_V$, while (B.5) gives $\Gamma_V = \Lambda_V^{\perp_{\mathbb{Z}}}$. Thus the flag in (B.10) determines the pair (Γ_U, Γ_V) .

We next count the matrices giving a fixed pair of row lattices. Since U has rank d_1 , it has d_1 linearly independent rows, each of Euclidean norm $O_n(R)$. Hence $\lambda_{d_1}(\Gamma_U) \ll_n R$. Moreover, the lattice generated by these d_1 rows has determinant $O_n(R^{d_1})$ by Hadamard's inequality. Since Γ_U

is its saturation, (B.1) shows that saturation can only decrease determinant. Together with (B.2), this gives $1 \leq \det \Gamma_U \ll_n R^{d_1}$. The same argument gives

$$\lambda_{n-d_2}(\Gamma_V) \ll_n R, \quad 1 \leq \det \Gamma_V \ll_n R^{n-d_2}.$$

It follows from (B.8) that, once Γ_U and Γ_V are fixed, the numbers of possible matrices are bounded by

$$\#\{U\} \ll_{n,s} \left(\frac{R^{d_1}}{\det \Gamma_U} \right)^s, \quad \#\{V\} \ll_{n,t} \left(\frac{R^{n-d_2}}{\det \Gamma_V} \right)^t. \quad (\text{B.11})$$

Let $N_{d_1,d_2}(R)$ denote the contribution from matrices with these fixed ranks. Choose integers $J_1, J_2 \geq 0$ such that all possible determinants are covered by the dyadic ranges

$$2^j \leq \det \Gamma_U < 2^{j+1} \quad (0 \leq j \leq J_1), \quad 2^\ell \leq \det \Gamma_V < 2^{\ell+1} \quad (0 \leq \ell \leq J_2),$$

with

$$2^{J_1} \asymp_n R^{d_1}, \quad 2^{J_2} \asymp_n R^{n-d_2}.$$

For fixed j and ℓ , the orthogonal-complement identity and Lemma B.3, applied with heights 2^{j+1} and $2^{\ell+1}$, show that the number of possible pairs (Γ_U, Γ_V) is

$$\ll_{n,d_1,d_2} 2^{jd_2} 2^{\ell(n-d_1)}.$$

For each such pair, (B.11) bounds the number of matrices by

$$\ll_{n,s,t} R^{d_1 s + (n-d_2)t} 2^{-js} 2^{-\ell t}.$$

Thus

$$\begin{aligned} N_{d_1,d_2}(R) &\ll_{n,s,t} R^{d_1 s + (n-d_2)t} \sum_{j=0}^{J_1} \sum_{\ell=0}^{J_2} 2^{j(d_2-s)} 2^{\ell(n-d_1-t)} \\ &= R^{d_1 s + (n-d_2)t} \left(\sum_{j=0}^{J_1} 2^{j(d_2-s)} \right) \left(\sum_{\ell=0}^{J_2} 2^{\ell(n-d_1-t)} \right) \\ &\ll_{n,s,t} R^{d_1 s + (n-d_2)t} 2^{J_1(d_2-s)} 2^{J_2(n-d_1-t)} \\ &\ll_{n,s,t} R^{d_1 s + (n-d_2)t} R^{d_1(d_2-s)} R^{(n-d_2)(n-d_1-t)} \\ &= R^{d_1 s + (n-d_2)t + d_1 d_2 - d_1 s + (n-d_2)(n-d_1) - (n-d_2)t} \\ &= R^{d_1 d_2 + (n-d_2)(n-d_1)} \\ &= R^{n^2 - n(d_1+d_2) + 2d_1 d_2}. \end{aligned}$$

It remains to maximize the exponent

$$E(d_1, d_2) := n^2 - n(d_1 + d_2) + 2d_1 d_2$$

over $1 \leq d_1 \leq s$ and $n-t \leq d_2 \leq n-1$. We have

$$E(d_1 + 1, d_2) - E(d_1, d_2) = 2d_2 - n \geq n - 2t \geq 0$$

and

$$E(d_1, d_2 - 1) - E(d_1, d_2) = n - 2d_1 \geq n - 2s \geq 0.$$

Thus $E(d_1, d_2)$ is maximized at $d_1 = s$ and $d_2 = n - t$, where

$$E(s, n - t) = n(s + t) - 2st.$$

Summing over the $O_n(1)$ choices for (d_1, d_2) completes the proof. $\square$

References

- [1] J. Bourgain, *A Szemerédi type theorem for sets of positive density in $\mathbb{R}^k$* , Israel J. Math. **54** (1986), no. 3, 307–316.
- [2] J. W. S. Cassels, *An introduction to the geometry of numbers*, Classics in Mathematics, Springer-Verlag, Berlin, 1997, Reprint of the corrected 1971 edition.
- [3] B. Cook, Á. Magyar, and T. Titichetrakun, *A multidimensional Szemerédi theorem in the primes via combinatorics*, Ann. Comb. **22** (2018), no. 4, 711–768.
- [4] H. Finner, *A generalization of Hölder’s inequality and some probability inequalities*, Ann. Probab. **20** (1992), no. 4, 1893–1901.
- [5] J. Fox and Y. Zhao, *A short proof of the multidimensional Szemerédi theorem in the primes*, Amer. J. Math. **137** (2015), no. 4, 1139–1145.
- [6] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), no. 1, 275–291.
- [7] A. V. Geramita and N. J. Pullman, *A theorem of Hurwitz and Radon and orthogonal projective modules*, Proc. Amer. Math. Soc. **42** (1974), no. 1, 51–56.
- [8] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), no. 3, 897–946.
- [9] L. Hochfilzer, *Systems of bihomogeneous forms of small bidegree*, Acta Arith. **213** (2024), no. 4, 325–368.
- [10] M. Jaber, Y. P. Liu, S. Lovett, A. Ostuni, and M. Sawhney, *Quasipolynomial bounds for the corners theorem*, 2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2025, pp. 2027–2036.
- [11] Z. Kelley and R. Meka, *Strong bounds for 3-progressions*, 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2023, pp. 933–973.
- [12] S. Kim, *Mean value formulas on sublattices and flags of the random lattice*, J. Number Theory **241** (2022), 330–351.
- [13] Y. Kitaoka, *Lectures on Siegel modular forms and representation by quadratic forms*, Tata Institute of Fundamental Research Lectures on Mathematics and Physics, vol. 77, Published for the Tata Institute of Fundamental Research by Springer-Verlag, Berlin, 1986.
- [14] J. Leng, A. Sah, and M. Sawhney, *Improved bounds for Szemerédi’s theorem*, 2024, arXiv:2402.17995.
- [15] ———, *Quasipolynomial bounds on the inverse theorem for the Gowers $U^{s+1}[N]$ -norm*, 2024, arXiv:2402.17994.
- [16] N. Lyall and Á. Magyar, *Weak hypergraph regularity and applications to geometric Ramsey theory*, Trans. Amer. Math. Soc. Ser. B **9** (2022), 160–207.
- [17] Á. Magyar, *k -point configurations in sets of positive density of $\mathbb{Z}^n$* , Duke Math. J. **146** (2009), no. 1, 1–34.

- [18] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika **20** (1973), no. 2, 119–134.
- [19] S. Prendiville, *Matrix progressions in multidimensional sets of integers*, Mathematika **61** (2015), no. 1, 14–48.
- [20] S. L. Rydin Myerson, *Quadratic forms and systems of forms in many variables*, Invent. Math. **213** (2018), no. 1, 205–235.
- [21] W. M. Schmidt, *Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height*, Duke Math. J. **35** (1968), no. 2, 327–339.
- [22] T. Tao, *A variant of the hypergraph removal lemma*, J. Combin. Theory Ser. A **113** (2006), no. 7, 1257–1280.
- [23] T. Tao and T. Ziegler, *A multi-dimensional Szemerédi theorem for the primes via a correspondence principle*, Israel J. Math. **207** (2015), no. 1, 203–228.