ON RESTRICTION ESTIMATES FOR DISCRETE QUADRATIC SURFACES

KEVIN HENRIOT, KEVIN HUGHES

This work is dedicated to the memory of Kevin Henriot with whom it was a pleasure to work.

Abstract. We obtain truncated restriction estimates of an unexpected form for discrete surfaces

$$S_N = \{\, (n_1, \ldots, n_d, R(n_1, \ldots, n_d))\,,\, n_i \in [-N, N] \cap \mathbb{Z}\,\},$$

where R is an indefinite quadratic form with integer matrix.

1. Introduction

We fix a non-degenerate quadratic form R in d variables with integer matrix. We are interested in restriction estimates for quadratic surfaces in $\mathbb{Z}^{d+1}$ of the form

$$S_N = \{\, (R(n_1, \ldots, n_d), n_1, \ldots, n_d)\,,\, n_i \in [-N, N] \cap \mathbb{Z}\,\}, \tag{1.1}$$

in the case where R is indefinite. This paper should be seen as a companion to [7], which concerned the case $R(n) = n^k$ of k-th powers and the case $R(\mathbf{n}) = n_1^d + \cdots + n_d^k$ of 'k-paraboloids' when $k \geqslant 3$; the methods employed here are similar but our results take a different shape.

In the case $d = 1$, $R(x) = x^2$ of the $2D$ parabola, Bourgain [2] resolved the natural restriction conjecture in the supercritical range, via discrete versions of the Tomas–Stein argument [13, Chapter 7] and the Hardy–Littlewood circle method. By powerful new methods of multilinear harmonic analysis, Bourgain and Demeter [4, Theorem 2.4] later established the natural restriction conjecture for arbitrary definite irrational paraboloids $R(\mathbf{x}) = \theta_1 x_1^2 + \cdots + \theta_d x_d^2$, $\theta_i > 0$, up to ε losses. In a subsequent work [3, Corollary 1.3], they also obtained the conjectured estimate for indefinite paraboloids. To state those results precisely, we set up some notation. The extension operator acting on a sequence $a : \mathbb{Z}^d \to \mathbb{C}$ supported on $[-N, N]^d$ is denoted by

$$F_a(\alpha, \boldsymbol{\theta}) = \sum_{\mathbf{n} \in \mathbb{Z}^d} a(\mathbf{n}) e(\alpha R(\mathbf{n}) + \boldsymbol{\theta} \cdot \mathbf{n}) \qquad (\alpha \in \mathbb{T}, \boldsymbol{\theta} \in \mathbb{T}^d). \tag{1.2}$$

Theorem 1.1 (Bourgain–Demeter [3], special case)**.** *Suppose that R is a non-degenerate indefinite quadratic form in d variables with integer matrix and signature (p, q), and let*

$s = \min(p, q)$. *We have*

$$\|F_a\|_p^p \lesssim_\varepsilon \begin{cases} N^{\frac{sp}{2}-s+\varepsilon}\|a\|_2^p & \text{for } 2 \leqslant p \leqslant \frac{2(d-s+2)}{d-s}, \\ N^{\frac{dp}{2}-(d+2)}\|a\|_2^p & \text{for } p > \frac{2(d-s+2)}{d-s}. \end{cases}$$

We take a moment to clarify that the Bourgain–Demeter restriction estimate is stated for an extension operator. This is because the restriction operator

$$\mathcal{R}f(n_0, \mathbf{n}) := \widehat{f}(n_0, \mathbf{n}) \mid_{S_N}, \tag{1.3}$$

defined for functions $f : \mathbb{T}^{d+1} \to \mathbb{C}$, is the adjoint operator to the extension operator F_a. (Here $(n_0, \mathbf{n})$ in $\mathbb{Z} \times \mathbb{Z}^d$ and $\widehat{f}$ denotes the Fourier transform on $\mathbb{T}^{d+1}$.) Historically the restriction operator was introduced by Stein, but the extension operator is often more natural to work with. By duality, estimates for one operator are equivalent to estimates for the other. As is standard in harmonic analysis we will refer to any estimates of the extension operator as restriction estimates. We refer the reader to [11, Chapter VII] or [13, Chapter 7] for more information on this duality and the classical Tomas–Stein method.

While Theorem 1.1 is stated only for diagonal forms in [3], a simple diagonalization argument allows one to reduce to this case. There is also an extra N^ε factor in the supercritical range in that reference, which can be removed by (a minor variant of) Bourgain's ε-removal estimate [2]; we refer to Appendix B for the details. Our result goes beyond this to give a range of L^p spaces, independent of the signature of R, for which the sharp bounds (bounds without the extra N^ε factor) are obtained. Let $\mathrm{d}m$ denote Lebesgue measure on any torus.

Theorem 1.2. *Let R be a non-degenerate quadratic form in $d \geqslant 1$ variables with integer matrix. There exists $C_R > 0$ such that*

$$\int_{|F_a| \geqslant C_R N^{d/4}\|a\|_2} |F_a|^p \mathrm{d}m \lesssim N^{\frac{dp}{2}-(d+2)}\|a\|_2^p \tag{1.4}$$

for $p > \frac{2(d+2)}{d}$.

Our main result adapts the proof of Bourgain's ε-removal lemma [2] to include any non-degenerate integral paraboloids. Moreover, we obtain an intriguing bound for the truncated integral when the paraboloid is indefinite. To explain why this in interesting in the indefinite case, we begin by contrasting with the positive definite case. When $R(\mathbf{n}) = n_1^2 + \cdots + n_d^2$, Theorem 1.2 is due to Bourgain in [2]. The trivial bound, which follows from the Cauchy–Schwartz inequality, is that $\|F_a\|_\infty \lesssim N^{d/2}\|a\|_2$. Furthermore in this case we know that for any fixed $p > \frac{2(d+2)}{d}$ there is some $C > 0$ and some small

$\delta > 0$ such that

$$\int_{|F_a|\leqslant CN^{d/4}\|a\|_2} |F_a|^p \mathrm{d}m \lesssim N^{\frac{dp}{2}-(d+2)-\delta}\|a\|_2^p. \tag{1.5}$$

Now instead suppose that R is a non-degenerate quadratic form in $d \geqslant 1$ variables with integer coefficients and signature $s \geqslant 1$. The exponent of N in Theorem 1.1 is sharp for exponents $p > \frac{2(d-s+2)}{d-s}$, as can be seen by taking $a \equiv 1$ and using the circle method to obtain an asymptotic. As explained in [3], the lower bound

$$\|F_a\|_p^p \gtrsim N^{\frac{sp}{2}-s}\|a\|_2^p \tag{1.6}$$

holds for a sequence supported on a subspace of dimension s. No such subspace, nor lower bound exists when R is definite and non-degenerate. More precisely, assume for simplicity that $R(\mathbf{x}) = \sum_{i=1}^s x_i^2 - \sum_{i=s+1}^d x_i^2$ with $s \leqslant d/2$, then $a(\mathbf{n}) = 1_{([-N,N]\cap\mathbb{Z})^{2s}}(\mathbf{n})\prod_{i=1}^s 1_{n_i=n_{s+i}}$ satisfies (1.6) uniformly as N tends to infinity. Once again, the trivial bound (for any sequence) is $\|F_a\|_\infty \lesssim N^{d/2}\|a\|_2$ by the Cauchy–Schwartz inequality. However, $|F_a| \leqslant N^{s/2}\|a\|_2 \leqslant N^{d/4}\|a\|_2$ is substantially smaller for our present choice of a. In fact it obtains 'square-root cancellation', by which we mean that we are comparing the bound $N^{d/4}$ for our present choice of a to the trivial bound $N^{d/2}$ for general sequences. Therefore, for our choice of a, we have

$$\int_{|F_a|>N^{d/4}\|a\|_2} |F_a|^p \mathrm{d}m = 0 \tag{1.7}$$

for all p. Since (1.6) is sharp and $\frac{sp}{2} - s > \frac{dp}{2} - (d+2)$ for $2 \leqslant p < \frac{2(d+2-s)}{d-s}$, (1.5) is impossible for non-degenerate indefinite quadratic forms.

The upper bound (1.4) in Theorem 1.2 is of order less than the order $N^{\frac{sp}{2}-s+O(\varepsilon)}\|a\|_2^p$ of the complete integral, as given by Theorem 1.1. This says that a weak form of (1.7) remains true for general sequences; in particular, we do not need to assume that the sequence a is supported on a special affine subvariety. We interpret this as an inverse result for non-degenerate indefinite forms saying that for sequences $a : \mathbb{Z}^d \to \mathbb{C}$ satisfying (1.6), the "mass" of the integral $\int |F_a|^p \mathrm{d}m$ is concentrated on a set where $|F_a|$ has square-root cancellation $|F_a| \leqslant N^{d/4}$ (in comparison with the trivial Cauchy–Schwartz bound $|F_a| \leqslant N^{d/2}\|a\|_2$) as in the above example. In the parlance of the circle method, the major arcs contribute a small amount and instead the mass is concentrated on the minor arcs. While this is consistent with our example, it contradicts the standard circle method paradigm that *the major arcs dominate*. One should compare this with higher degree k-paraboloids in [7] which is consistent with the standard circle method paradigm. In light of these observations, we pose the following question.

Question 1.3. *Fix a non-degenerate, indefinite integral quadratic form R. Let V denote the affine subvariety $\{R(\mathbf{x}) = 0\}$. If a sequence $a : \mathbb{Z}^d \to \mathbb{C}$ satisfies* (1.6) *(uniformly in N), then is a necessarily, essentially supported on V?*

In proving Theorem 1.2, we do not have a simple diagonalization argument at our disposal to estimate the truncated integral (1.4). To overcome this obstacle, we adapt the discrete Tomas–Stein approach of Bourgain [2] for the paraboloid $(x_1, \dots, x_d, x_1^2 + \dots + x_d^2)$ to use multidimensional exponential sum estimates, whereas in the diagonal case the relevant exponential sum (4.3) splits into one-dimensional quadratic Weyl sums. The Tomas–Stein method is a powerful method that allows one to effectively leverage bounds for the unweighted exponential sums F_1 against the weighted exponential sums F_a. A simple, limited version of this method in the discrete setting is Lemma 3.2 of our companion paper [7]. This method is successful since optimal bounds on quadratic exponential sums are known classically, and we do not encounter certain difficulties described in [7] for surfaces of high degree.

We note finally that the related problem of obtaining ε-free estimates in the full supercritical range for indefinite irrational quadratic forms R is still open, although there is partial progress in this direction by Godet and Tzvetkov [6] and Wang [12]. In the definite case, the question has been settled recently by Killip and Vişan [9].

Acknowledgements. The first author would like to thank Akos Magyar for stimulating discussions on exponential sums in many variables. The second author thanks Hiro Oh, Yuzhao Wang and Trevor Wooley for stimulating discussions on restriction theory.

2. Notation

For functions $f : \mathbb{T}^d \to \mathbb{C}$ and $g : \mathbb{Z}^d \to \mathbb{C}$ we define the Fourier transforms of f and g by $\widehat{f}(\mathbf{k}) = \int_{\mathbb{T}^d} f(\boldsymbol{\alpha}) e(-\boldsymbol{\alpha} \cdot \mathbf{k}) \mathrm{d}\boldsymbol{\alpha}$ and $\widehat{g}(\boldsymbol{\alpha}) = \sum_{\mathbf{n} \in \mathbb{Z}^d} g(\mathbf{n}) e(\boldsymbol{\alpha} \cdot \mathbf{n})$. For a function $h : \mathbb{R}^d \to \mathbb{C}$ we define the Fourier transform by $\widehat{h}(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} f(\mathbf{x}) e(-\boldsymbol{\xi} \cdot \mathbf{x}) \mathrm{d}\mathbf{x}$. Given a function $f : \mathbb{R}^d \to \mathbb{R}$ and two subsets A, B of $\mathbb{R}^d$, we write $A \prec f \prec B$ when $0 \leqslant f \leqslant 1$ everywhere, $f = 1$ on A and $f = 0$ outside B. For two complex valued functions f and g, $f \lesssim g$ means that $|f| \leqslant C|g|$ for some constant uniform in the inputs for f and g; furthermore, $f \sim g$ means $f \lesssim g$ and $g \lesssim f$.

When $\mathcal{P}$ is a certain property, we let $1_{\mathcal{P}}$ denote the boolean equal to 1 when $\mathcal{P}$ holds and 0 otherwise, and when E is a set we define the indicator function of E by $1_E(x) = 1_{x \in E}$. When $p \in [1, +\infty]$ is an exponent, we systematically denote by $p' \in [1, +\infty]$ its dual exponent satisfying $\frac{1}{p} + \frac{1}{p'} = 1$. We let $\mathrm{d}m$ denote the Lebesgue measure on $\mathbb{R}^d$, or on $\mathbb{T}^d$ identified with any fundamental domain of the form $[\theta, 1+\theta)^d$. For $q \geqslant 2$ we occasionally use $\mathbb{Z}_q$ as a shorthand for the group $\mathbb{Z}/q\mathbb{Z}$.

Throughout the article, we use the letter ε generically to denote a constant which can be taken arbitrarily small, and whose value may change in each occurence.

3. Arc mollifiers

This section is a specialization of [7, Section 6] to the quadratic case $k = 2$, and we include it for completeness. Its aim is to describe a technical tool due to Bourgain [2, Section 3] and used in the proof of Theorem 1.2, which consists essentially in a partition of unity adapted to major arcs.

We fix an integer $N \geqslant 2$, to be thought of as large. We fix a smooth bump function κ with $[-1,1] \prec \kappa \prec [-2,2]$. Let $\widetilde{N} = 2^{\lfloor \log_2 N \rfloor}$, and for every integer $0 \leqslant s \leqslant \lfloor \log_2 N \rfloor$ define

$$\phi^{(s)} := \begin{cases} \kappa(2^s N\,\cdot\,) - \kappa(2^{s+1} N\,\cdot\,) & \text{if } 1 \leqslant 2^s < \widetilde{N}, \\ \kappa(2^s N\,\cdot\,) & \text{if } 2^s = \widetilde{N}. \end{cases} \tag{3.1}$$

Note that we have $\mathrm{Supp}(\phi^{(s)}) \subset \frac{1}{2^s N} I_s$, where $I_s = \pm[\frac{1}{2}, 2]$ if $1 \leqslant 2^s < \widetilde{N}$, and $I_s = [-2,2]$ if $2^s = \widetilde{N}$. Furthermore, for every dyadic integer $1 \leqslant Q \leqslant N$, we have

$$\sum_{Q \leqslant 2^s \leqslant N} \phi^{(s)} = \kappa(QN\,\cdot\,). \tag{3.2}$$

Throughout the paper, Q will denote a dyadic integer; that is, $Q \in \{2^j : j \in \mathbb{Z}, j \geqslant 0\}$.

We let $N_1 = c_1 N$, for a small constant $c_1 \in (0,1]$. It is easy to check that the intervals $\frac{a}{q} + [-\frac{2}{QN}, \frac{2}{QN}]$, $1 \leqslant a \leqslant q$, $q \sim Q$, $1 \leqslant Q \leqslant N_1$ are all disjoint. For a dyadic integer Q and an integer $0 \leqslant s \leqslant \log_2 N$, we define

$$\Phi_{Q,s} = \sum_{\substack{(a,q)=1 \\ q \sim Q}} \tau_{-a/q} \phi^{(s)}, \tag{3.3}$$

where $\tau_{-a/q}\phi^{(s)}(\alpha) := \phi^{(s)}(\alpha - a/q)$ is translation by a/q, so that

$$\mathrm{Supp}(\Phi_{Q,s}) \subset \bigsqcup_{\substack{(a,q)=1 \\ q \sim Q}} \left(\frac{a}{q} + \frac{I_s}{2^s N} \right). \tag{3.4}$$

We also define the functions

$$\lambda = \sum_{Q \leqslant N_1} \sum_{Q \leqslant 2^s \leqslant N} \Phi_{Q,s}, \qquad \rho = 1 - \lambda \tag{3.5}$$

where we emphaisze that any sum over Q is over dyadic integers.

Proposition 3.1. *We have* $0 \leqslant \lambda, \rho \leqslant 1$ *and*

$$\lambda = 1,\ \rho = 0 \quad on \quad \bigsqcup_{Q \leqslant N_1} \bigsqcup_{\substack{(a,q)=1 \\ q \sim Q}} \left(\frac{a}{q} + \left[-\frac{1}{QN}, \frac{1}{QN} \right] \right).$$

Proof. By (3.2), we can rewrite λ as

$$\lambda = \sum_{Q \leqslant N_1} \sum_{\substack{(a,q)=1 \\ q \sim Q}} \tau_{-a/q} \bigg(\sum_{Q \leqslant 2^s \leqslant N} \phi^{(s)} \bigg) = \sum_{Q \leqslant N_1} \sum_{\substack{(a,q)=1 \\ q \sim Q}} \tau_{-a/q} \kappa(QN\,\cdot\,).$$

The proposition follows from the localization properties of κ. □

At this stage we define the fundamental domain $\mathcal{U} = (\frac{1}{2N_1}, 1 + \frac{1}{2N_1}]$, and we note that when N is large, then for every $1 \leqslant a \leqslant q \leqslant Q \leqslant N_1$, the intervals $\frac{a}{q} + [-\frac{2}{QN}, \frac{2}{QN}]$ are contained in the interior of $\mathcal{U}$. Therefore for $1 \leqslant Q \leqslant 2^s \leqslant N$, the functions $\phi^{(s)}$, $\Phi_{Q,s}$ and λ are supported on the interior of $\mathcal{U}$, and they may be viewed as smooth functions over the torus $\mathbb{T}$, by 1-periodization from the interval $\mathcal{U}$. We will view $\Phi_{Q,s}$ alternatively as a smooth function on the torus $\mathbb{T}$ or on the real line, but note that for an integer n, $\widehat{\Phi_{Q,s}}(n)$ has the same definition under both points of view.

For $n \in \mathbb{Z}$ and an integer $Q \geqslant 1$ we define a truncated divisor function

$$d(n, Q) = \sum_{\substack{1 \leqslant d \leqslant Q: \\ d | n}} 1.$$

The following useful lemma is due to Bourgain [2].

Lemma 3.2. *Let* δ_x *be the Dirac function at* x*. Then*

$$\widehat{\sum_{\substack{(a,q)=1 \\ q \sim Q}} \delta_{a/q}}(n) \lesssim Q \cdot d(n, 2Q) \qquad (n \in \mathbb{Z}).$$

Proposition 3.3. *We have*

$$\int_{\mathbb{T}} \Phi_{Q,s} \mathrm{d}m \lesssim \frac{Q^2}{2^s N}, \tag{3.6}$$

$$\widehat{\Phi_{Q,s}}(n) \lesssim \frac{Q}{2^s N} d(n, 2Q) \qquad (n \in \mathbb{Z}) \tag{3.7}$$

Proof. Let $\gamma^{(s)} = \kappa - \kappa(2\,\cdot\,)$ for $0 \leqslant s < \lfloor \log_2 N \rfloor$ and $\gamma^{(s)} = \kappa$ when $s = \lfloor \log_2 N \rfloor$. By (3.1) and (3.3), we can write

$$\Phi_{Q,s} = \sum_{\substack{(a,q)=1 \\ q \sim Q}} \tau_{-a/q} \gamma^{(s)}(2^s N\,\cdot\,) = \bigg(\sum_{\substack{(a,q)=1 \\ q \sim Q}} \delta_{a/q} \bigg) * \gamma^{(s)}(2^s N\,\cdot\,).$$

By Lemma 3.2, we deduce the pointwise bound

$$|\widehat{\Phi_{Q,s}}(n)| = \Big| \widehat{\sum_{\substack{(a,q)=1\\ q\sim Q}} \delta_{a/q}}(n) \cdot \frac{1}{2^s N} \widehat{\gamma^{(s)}}\Big(\frac{n}{2^s N}\Big) \Big| \lesssim \frac{Q}{2^s N} d(n, 2Q),$$

which is uniform in $n \in \mathbb{Z}$. When $n = 0$ the left-hand side is $\int \Phi_{Q,s} \mathrm{d}m$. □

Proposition 3.4. *For every $\varepsilon > 0$ and $A > 0$, we have*

$$\int_{\mathbb{T}} \rho \, \mathrm{d}m \asymp 1, \tag{3.8}$$

$$\widehat{\rho}(n) \lesssim_{\varepsilon,A} \frac{1}{N^{1-\varepsilon}} \quad \textit{for } 0 < |n| \leqslant AN^A. \tag{3.9}$$

Proof. From (3.5) and (3.6), it follows that

$$\begin{aligned} \int_{\mathbb{T}} \rho \, \mathrm{d}m &= 1 - O\Big(\sum_{Q \leqslant N_1} \sum_{Q \leqslant 2^s \leqslant N} \frac{Q^2}{2^s N} \Big) \\ &= 1 - O\Big(\frac{1}{N} \sum_{Q \leqslant N_1} Q \Big) \\ &= 1 - O\Big(\frac{N_1}{N} \Big). \end{aligned}$$

Since we have chosen $N_1 = c_0 N$ with c_1 small enough, we have $\int \rho \mathrm{d}m \asymp 1$ as desired. The bound on $\widehat{\rho}$ is derived from (3.7) in a similar fashion, using also the standard divisor bound $d(n, Q) \leqslant d(n) \lesssim_\varepsilon n^\varepsilon$. □

4. Restriction estimates

We fix a non-degenerate integer quadratic form R in d variables. In this section, we derive Theorem 1.2 from the introduction. When R is definite, the critical exponent is $p_d = \frac{2(d+2)}{d}$ instead of $p_{d,s} = \frac{2(d-s+2)}{d-s}$ when R is indefinite with signature $s \geqslant 1$. (The critical exponent is the L^p exponent p where the behavior in Theorem 1.1 of the restriction estimates changes as N tends to infinity.) The exponent p_d arises in our argument, even in the indefinite case, due to our use of d-dimensional exponential sum estimates which do not depend on the type of quadratic form. The larger critical exponent $p_{d,s} = \frac{2(d-s+2)}{d-s}$ of Theorem 1.1 accounts for the existence of a special linear subvariety contained in (1.1), but this does not influence our treatment of the truncated moment in Theorem 1.2.

We use a smooth weight function $\omega : \mathbb{R}^d \to [0,1]$ of the form

$$\omega = \eta\Big(\frac{\cdot}{N}\Big), \qquad \eta \text{ Schwartz function such that } [-1,1]^d \prec \eta \prec [-2,2]^d. \tag{4.1}$$

Given a sequence $a : \mathbb{Z}^d \to \mathbb{C}$ supported on $[-N, N]^d$ with $\|a\|_2 = 1$ and a weight function $\omega : \mathbb{Z}^d \to [0,1]$ of the form (4.1), we define

$$F_a(\alpha, \boldsymbol{\theta}) = \sum_{\mathbf{n} \in \mathbb{Z}^d} a(\mathbf{n}) e(\alpha R(\mathbf{n}) + \boldsymbol{\theta} \cdot \mathbf{n}) \qquad (\alpha \in \mathbb{T}, \boldsymbol{\theta} \in \mathbb{T}^d), \tag{4.2}$$

$$F(\alpha, \boldsymbol{\theta}) = \sum_{\mathbf{n} \in \mathbb{Z}^d} \omega(\mathbf{n}) e(\alpha R(\mathbf{n}) + \boldsymbol{\theta} \cdot \mathbf{n}) \qquad (\alpha \in \mathbb{T}, \boldsymbol{\theta} \in \mathbb{T}^d), \tag{4.3}$$

which are the extension operator of our surface S acting on the sequence a and the ω-smoothed Fourier transform of the counting measure on S, respectively.

We will quote the estimates of Section 3 extensively. Via the Tomas-Stein argument in Section 5, we will devote most of our attention to the complete exponential sum (4.3). The minor arc estimates of Appendix A yield the following in our context.

Proposition 4.1. *Uniformly in $\alpha \in \mathbb{T}$, $\boldsymbol{\theta} \in \mathbb{T}^d$, we have*

$$\rho(\alpha) \neq 0 \quad \Rightarrow \quad |F(\alpha, \boldsymbol{\theta})| \lesssim N^{d/2}.$$

Proof. We prove the contrapositive. If $|F(\alpha, \boldsymbol{\theta})| \geqslant C_1 N^{d/2}$ for a large enough $C_1 > 0$, then by Proposition A.1 there exist $a, q \in \mathbb{Z}$ such that $|\alpha - \frac{a}{q}| \leqslant \frac{c_1}{qN}$, $1 \leqslant q \leqslant c_1 N$ and $(a, q) = 1$. Consequently there exists a dyadic integer Q such that $q \sim Q \Rightarrow Q \leqslant c_1 N = N_1$ and $|\alpha - \frac{a}{q}| \leqslant \frac{1}{QN}$, so that $\rho(\alpha) = 0$ by Proposition 3.1. □

For each dyadic integer $Q \geqslant 1$ and each integer $s \geqslant 0$ such that $1 \leqslant Q \leqslant 2^s$, we define a piece of our original exponential sum by

$$F_{Q,s}(\alpha, \boldsymbol{\theta}) = F(\alpha, \boldsymbol{\theta}) \Big[\Phi_{Q,s}(\alpha) - \frac{\int_{\mathbb{T}} \Phi_{Q,s}}{\int_{\mathbb{T}} \rho} \rho(\alpha) \Big]. \tag{4.4}$$

Remark 4.2. *The decomposition induced by* (4.4) *was introduced by Bourgain in [2]. It may seem strange at first and in particular, it may be more natural to consider the pieces $F(\alpha, \boldsymbol{\theta}) \Phi_{Q,s}(\alpha)$ which corresponds to a piece of a major arc. While this piece has some nice localization properties, it is insufficient for our purposes. However, the mean zero property of $\Phi_{Q,s}(\alpha) - \frac{\int \Phi_{Q,s}}{\int \rho} \rho(\alpha)$ and the minor arc bound of Proposition 4.4 allow us to obtain a superior estimate in Proposition 4.3 compared to that of Proposition 6.3 of [7] which uses $F(\alpha, \boldsymbol{\theta}) \Phi_{Q,s}(\alpha)$ in place of* (4.4)*. The superior estimate of Proposition 4.3 then feeds into Proposition 4.8 to obtain a sharper range of L^p spaces than if we used $F(\alpha, \boldsymbol{\theta}) \Phi_{Q,s}(\alpha)$.*

We establish physical and Fourier bounds for the exponential sums $F_{Q,s}$ via the major and minor arc estimates of Appendix A. It turns out to be important to have square-root cancellation of the exponential sum F on the minor arcs. We also introduce a technical device to ensure that the Fourier transforms under consideration stay inside an $N^2 \times N \times \cdots \times N$ box, a fact that will prove useful later on. Specifically, we fix a trigonometric polynomial ψ_N on $\mathbb{T}^{d+1}$ such that, for a constant C_R large enough with respect to R,

$$[-C_R N^2, C_R N^2] \times [-2N, 2N]^d \prec \widehat{\psi}_N \prec [-2C_R N^2, 2C_R N^2] \times [-4N, 4N]^d,$$

which in particular implies that $\widehat{\psi_N}(0) := \int_{\mathbb{T}^{d+1}} \psi_N = 1$ since the origin is in the box $[-C_R N^2, C_R N^2] \times [-2N, 2N]^d$. When $H : \mathbb{T}^{d+1} \to \mathbb{C}$ is a bounded measurable function, we write $\dot{H} = H * \psi_N$ for brevity; note that $\|\dot{H}\|_p \leqslant \|H\|_p$ for any $p \geqslant 1$ by Young's inequality, and that $F = \dot{F}$ by Fourier inversion (since $\widehat{F}$ is supported on the surface (1.1)).

Proposition 4.3. *Uniformly for* $(m, \boldsymbol{\ell}) \in \mathbb{Z}^{d+1}$, *we have*

$$\|\dot{F}_{Q,s}\|_\infty \lesssim_\varepsilon \left(\frac{2^s N}{Q}\right)^{\frac{d}{2}} Q^\varepsilon,$$

$$\widehat{\dot{F}_{Q,s}}(m, \boldsymbol{\ell}) \lesssim_\varepsilon 1_{|m| \lesssim N^2,\ |\boldsymbol{\ell}| \lesssim N} \left(\frac{Q}{2^s N} d(m - R(\boldsymbol{\ell}), 2Q) + \frac{Q^2}{2^s N^{2-\varepsilon}}\right).$$

Proof. When $\Phi_{Q,s}(\alpha) \neq 0$, it follows from (3.4) that there exist $a, q \in \mathbb{Z}$ such that $q \sim Q$, $(a, q) = 1$ and $|\alpha - \frac{a}{q}| \asymp \frac{1}{2^s N}$ if $2^s < \widetilde{N}$, $|\alpha - \frac{a}{q}| \leqslant \frac{2}{2^s N}$ is $2^s = \widetilde{N}$. By Propositions A.2 and 4.1, and by (3.6) and (3.8), it follows that, uniformly in $\boldsymbol{\theta} \in \mathbb{R}^d$,

$$\begin{aligned} |F_{Q,s}(\alpha, \boldsymbol{\theta})| &\lesssim_\varepsilon Q^{-\frac{d}{2}+\varepsilon} (2^s N)^{\frac{d}{2}} + \frac{Q^2}{2^s N} N^{\frac{d}{2}} \\ &= \left(\frac{2^s}{Q}\right)^{\frac{d}{2}} Q^\varepsilon N^{\frac{d}{2}} + \frac{Q}{2^s} \cdot \frac{Q}{N} N^{\frac{d}{2}} \\ &\leqslant \left(\frac{2^s N}{Q}\right)^{\frac{d}{2}} (1 + Q^\varepsilon). \end{aligned}$$

We let $\Psi_{Q,s} = \Phi_{Q,s} - \left(\frac{\int_{\mathbb{T}} \Phi_{Q,s}}{\int_{\mathbb{T}} \rho}\right) \rho$ and note that $\int_{\mathbb{T}} \Psi_{Q,s} = 0$ for each Q, s. Next we observe that, for any $(m, \boldsymbol{\ell}) \in \mathbb{Z}^{d+1}$,

$$\begin{aligned} \widehat{F_{Q,s}}(m, \boldsymbol{\ell}) &= \int_{\mathbb{T}^{d+1}} \Psi_{Q,s}(\alpha) F(\alpha, \boldsymbol{\theta}) e(-\alpha m - \boldsymbol{\theta} \cdot \boldsymbol{\ell}) \mathrm{d}\alpha \mathrm{d}\boldsymbol{\theta} \\ &= \sum_{\mathbf{n} \in \mathbb{Z}^d} \omega(\mathbf{n}) \int_{\mathbb{T}^{d+1}} \Psi_{Q,s}(\alpha) e\big(\alpha(R(\mathbf{n}) - m) + \boldsymbol{\theta} \cdot (\mathbf{n} - \boldsymbol{\ell})\big) \mathrm{d}\alpha \mathrm{d}\boldsymbol{\theta} \\ &= \omega(\boldsymbol{\ell}) \widehat{\Psi}_{Q,s}(m - R(\boldsymbol{\ell})). \end{aligned}$$

Since $F = \dot{F}$, we have more localization than the above computation suggests, and the second bound of the proposition then follows from the identity

$$\widehat{\dot{F}_{Q,s}}(m, \boldsymbol{\ell}) = \psi_N(m, \boldsymbol{\ell}) \omega(\boldsymbol{\ell}) \widehat{\Psi}_{Q,s}(m - R(\boldsymbol{\ell})) 1_{m \neq R(\boldsymbol{\ell})},$$

and the estimates (3.6), (3.7), (3.8) and (3.9). □

We now define minor and major arc pieces of our exponential sum by

$$F_{\mathfrak{M}} = \sum_{Q \leqslant N_1} \sum_{Q \leqslant 2^s \leqslant N} F_{Q,s}, \qquad F_{\mathfrak{m}} = F - F_{\mathfrak{M}}. \tag{4.5}$$

We can readily derive a uniform bound on the minor arc piece $F_{\mathfrak{m}}$, as an immediate consequence of the definition (3.5) and Proposition 4.1.

Proposition 4.4. *We have* $\|F_{\mathfrak{m}}\|_\infty \lesssim N^{d/2}$.

The previous propositions also imply simple norm estimates for the operator of convolution with a major arc piece.

Proposition 4.5. *We have*

$$\|\dot{F}_{Q,s} * f\|_\infty \lesssim \Big(\frac{2^s N}{Q}\Big)^{\frac{d}{2}} Q^\varepsilon \|f\|_1, \tag{4.6}$$

$$\|\dot{F}_{Q,s} * f\|_2 \lesssim_\varepsilon \frac{Q}{2^s N^{1-\varepsilon}} \|f\|_2. \tag{4.7}$$

Proof. Note that for any bounded function $W : \mathbb{T}^{d+1} \to \mathbb{C}$, we have

$$\|W * f\|_\infty \leqslant \|W\|_\infty \|f\|_1, \qquad \|W * f\|_2 = \|\widehat{W}\widehat{f}\|_2 \leqslant \|\widehat{W}\|_\infty \|f\|_2.$$

It now suffices to apply these inequalities with $W = F_{Q,s}$ and insert the estimates of Proposition 4.3 (using also the bound $d(n, 2Q) \lesssim n^\varepsilon$). □

By interpolation, we can obtain an estimate for all moments.

Proposition 4.6. *Let* $p_0' = \frac{2(d+2)}{d}$. *For any* $p' \in [2, \infty)$, *we have*

$$\|\dot{F}_{Q,s} * f\|_{p'} \lesssim \Big(\frac{2^s N}{Q}\Big)^{(d+2)(\frac{1}{p_0'} - \frac{1}{p'})} N^\varepsilon \|f\|_p. \tag{4.8}$$

Proof. We interpolate between the estimates of Proposition 4.5 with $\theta \in (0, 1)$ given by

$$\frac{1}{p'} = \frac{1-\theta}{\infty} + \frac{\theta}{2}, \qquad \frac{1}{p} = \frac{1-\theta}{1} + \frac{\theta}{2}. \tag{4.9}$$

This yields

$$\begin{aligned} \|\dot{F}_{Q,s} * f\|_{p'} &\lesssim \Big(\frac{2^s N}{Q}\Big)^{(1-\theta)\frac{d}{2}} \cdot \Big(\frac{Q}{2^s N}\Big)^{\theta} \cdot N^\varepsilon \cdot \|f\|_p \\ &= \Big(\frac{2^s N}{Q}\Big)^{\frac{d}{2} - (1+\frac{d}{2})\theta} N^\varepsilon \|f\|_p. \end{aligned}$$

Since $\theta = \frac{2}{p'}$, we may rewrite the exponent of $\frac{2^s N}{Q}$ as $(d+2)(\frac{1}{p_0'} - \frac{1}{p'})$, which concludes the proof. □

At this stage we need a preparatory lemma on truncated divisor sums from [2].

Lemma 4.7. *Let $D, Q, X \geqslant 1$ and $B \in \mathbb{N}$. When $Q \leqslant 2X^{1/B}$, we have*

$$\#\{|n| \leqslant X \,:\, d(n,Q) \geqslant D\} \lesssim_{\varepsilon,B} D^{-B} Q^{\varepsilon} X.$$

Proof. We show that

$$\sum_{|\ell| \leqslant X} d(\ell, Q)^B \lesssim_{\varepsilon,B} Q^{\varepsilon} X,$$

from which the result follows by Markov's inequality. In the sum above, the term $\ell = 0$ contributes at most Q^B, and by [1, Eq. (4.31)] the other terms contribute at most $C_{\varepsilon,B} Q^{\varepsilon} X$. The conclusion follows from our assumption on Q. □

We can now derive a more precise convolution bound using the previous lemma.

Proposition 4.8. *Let $B, D > 2$. Uniformly for $Q \leqslant N^{2/B}$ and $Q \leqslant 2^s \leqslant N$, we have*

$$\|\dot{F}_{Q,s} * f\|_2 \lesssim_{\varepsilon,B} \frac{DQ}{2^s N}\|f\|_2 + \frac{D^{-\frac{B}{2}} Q^{1+\varepsilon}}{2^s N} N^{\frac{d+2}{2}} \|f\|_1. \tag{4.10}$$

Proof. By Parseval's identity and the bounds of Proposition 4.3, we deduce that

$$\begin{aligned}
\|\dot{F}_{Q,s} * f\|_2 &= \Bigg[\sum_{\substack{|m| \lesssim N^2 \\ |\boldsymbol{\ell}| \lesssim N}} |\widehat{\dot{F}}_{Q,s}(m,\boldsymbol{\ell})|^2 |\widehat{f}(m,\boldsymbol{\ell})|^2 \Bigg]^{1/2} \\
&\lesssim \frac{Q}{2^s N} \Bigg[\sum_{\substack{|m| \lesssim N^2 \\ |\boldsymbol{\ell}| \lesssim N}} d(m - R(\boldsymbol{\ell}), 2Q)^2 |\widehat{f}(m,\boldsymbol{\ell})|^2 \Bigg]^{1/2} + \frac{Q^2}{2^s N^{2-\varepsilon}} \|\widehat{f}\|_2
\end{aligned}$$

We write $n = m - R(\boldsymbol{\ell})$, so that assuming $Q \leqslant N^{2/B}$ and invoking Lemma 4.7, we obtain

$$\begin{aligned}
\|\dot{F}_{Q,s} * f\|_2 &\lesssim_{\varepsilon,B} \frac{Q}{2^s N} \Big[D^2 \|\widehat{f}\|_2^2 + \|\widehat{f}\|_\infty^2 N^d \times \#\{|n| \lesssim N^2 \,:\, d(n,2Q) > D\} \Big]^{1/2} + \frac{Q^2}{2^s N^{2-\varepsilon}} \|f\|_2 \\
&\lesssim \frac{Q}{2^s N} \Big(D^2 \|f\|_2^2 + D^{-B} Q^{\varepsilon} N^{d+2} \|f\|_1^2 \Big)^{1/2} + \frac{Q}{2^s N} \cdot \frac{Q}{2^s N^{1-\varepsilon}} \|f\|_2.
\end{aligned}$$

Since $B > 2$, we have that $Q \leqslant N^{1-\epsilon}$ for some $\epsilon > 0$ and the last term may be absorbed into the first. Finally we obtain

$$\|\dot{F}_{Q,s} * f\|_2 \lesssim \frac{Q}{2^s N} \big(D\|f\|_2 + Q^{\varepsilon} D^{-\frac{B}{2}} N^{\frac{d+2}{2}} \|f\|_1 \big).$$

□

This new estimate can again be interpolated with the $L^1 \to L^\infty$ one, to obtain the following bound.

Proposition 4.9. *Let $B, D > 2$. Let $p_0' = \frac{2(d+2)}{d}$ and $p' \in [2,\infty)$. Uniformly for $Q \leqslant N^{2/B}$ and $Q \leqslant 2^s \leqslant N$, we have*

$$\|\dot{F}_{Q,s} * f\|_{p'} \lesssim_{\varepsilon,B} D^{\frac{2}{p'}} \Big(\frac{2^s N}{Q}\Big)^{(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} Q^{\varepsilon} \|f\|_p + D^{-\frac{B}{p'}} \Big(\frac{2^s N}{Q}\Big)^{(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} N^{\frac{d+2}{p'}} Q^{\varepsilon} \|f\|_1.$$

Proof. Let $\theta \in (0,1]$ and $p' \geqslant 2$ be such that (4.9) holds. By convexity and (4.6) and (4.10), we have

$$\begin{aligned}\|\dot{F}_{Q,s} * f\|_{p'} &\leqslant \|\dot{F}_{Q,s} * f\|_{\infty}^{1-\theta} \|\dot{F}_{Q,s} * f\|_2^{\theta}. \\ &\lesssim_{\varepsilon,B} Q^{\varepsilon}\Big(\frac{2^s N}{Q}\Big)^{(1-\theta)\frac{d}{2}} \cdot D^{\theta}\Big(\frac{Q}{2^s N}\Big)^{\theta} \cdot \|f\|_1^{1-\theta}\|f\|_2^{\theta} \\ &\quad + Q^{\varepsilon}\Big(\frac{2^s N}{Q}\Big)^{(1-\theta)\frac{d}{2}} \cdot D^{-\theta\frac{B}{2}}\Big(\frac{Q}{2^s N}\Big)^{\theta} (N^{\frac{d+2}{2}})^{\theta} \cdot \|f\|_1\end{aligned}$$

Since $|f|$ takes values in $\{0,1\}$, we may rewrite this as

$$\|\dot{F}_{Q,s} * f\|_{p'} \lesssim_{\varepsilon,B} D^{\theta}\Big(\frac{2^s N}{Q}\Big)^{\frac{d}{2}-(1+\frac{d}{2})\theta} Q^{\varepsilon}\|f\|_p + D^{-\theta\frac{B}{2}}\Big(\frac{2^s N}{Q}\Big)^{\frac{d}{2}-(1+\frac{d}{2})\theta} N^{\frac{d+2}{p'}} Q^{\varepsilon}\|f\|_1.$$

The proof is finished upon recalling that $\theta = \frac{2}{p'}$ by (4.9). ∎

We introduce a parameter $1 \leqslant Q_1 \leqslant N_1$ and write $F_{\mathfrak{M}} = F_1 + F_2$ with

$$F_1 = \sum_{Q \leqslant Q_1} \sum_{Q \leqslant 2^s \leqslant N} F_{Q,s}, \qquad F_2 = \sum_{Q_1 < Q < N_1} \sum_{Q \leqslant 2^s \leqslant N} F_{Q,s}. \tag{4.11}$$

Proposition 4.10. *Suppose that $p' > p_0' = \frac{2(d+2)}{d}$. Let $T \geqslant 1$ and suppose that $Q_1 \leqslant N^{2/B}$. Then*

$$\begin{aligned}\|\dot{F}_1 * f\|_{p'} &\lesssim T^2 N^{d-\frac{2(d+2)}{p'}}\|f\|_{p'} + T^{-B} N^{d-\frac{d+2}{p'}}\|f\|_1, \\ \|\dot{F}_2 * f\|_{p'} &\lesssim Q_1^{-(\frac{d}{2}-\frac{d+2}{p'})} N^{d-\frac{2(d+2)}{p'}}\|f\|_p.\end{aligned}$$

Proof. By the triangle inequality and Proposition 4.9 with $T = D^{1/p'}$, it follows that

$$\begin{aligned}\|\dot{F}_1 * f\|_{p'} &\lesssim \sum_{Q \leqslant Q_1} Q^{\varepsilon-(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} \sum_{2^s \leqslant N} (2^s)^{(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} N^{(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} \\ &\qquad \cdot \big(T^2\|f\|_p + T^{-B} N^{\frac{d+2}{p'}}\|f\|_1\big). \\ &\lesssim T^2 N^{2(d+2)(\frac{1}{p_0'}-\frac{1}{p'})}\|f\|_p + T^{-B} N^{2(d+2)(\frac{1}{p_0'}-\frac{1}{p'})-\frac{d+2}{p'}}\|f\|_1.\end{aligned}$$

It is easy to rewrite the exponents of N in the desired form.

Turning our attention to F_2, we deduce from the triangle inequality and (4.8) that

$$\|\dot{F}_2 * f\|_{p'} \lesssim \sum_{Q>Q_1} Q^{-(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} \sum_{2^s \leqslant N} (2^s)^{(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} \cdot N^{(d+2)(\frac{1}{p_0'}-\frac{1}{p'})} \cdot N^{\varepsilon}\|f\|_p$$
$$\lesssim N^{\varepsilon} Q_1^{-(\frac{d}{2}-\frac{d+2}{p'})} N^{d-\frac{2(d+2)}{p'}} \|f\|_p.$$

□

5. Proof of Theorem 1.2

In this section we prove our theorem using the restriction estimates from Section 4 and Bourgain's [1,2] discrete version of the Tomas–Stein argument [13, Chapter 7] from Euclidean harmonic analysis. We introduce a parameter $\lambda > 0$ and define

$$E_\lambda = \{|F_a| \geqslant \lambda\}, \qquad f = 1_{E_\lambda} \frac{F_a}{|F_a|}. \tag{5.1}$$

Note that, by Cauchy-Schwarz in (4.2), we always have $|F_a| \leqslant CN^{d/2}$, and thus we assume that the parameter λ lies in $(0, CN^{d/2}]$. Our theorem will quickly follow once we establish the following sharp level set bound.

Proposition 5.1. *For $q > \frac{2(d+2)}{d}$, there exists $C = C(q) > 0$ such that*

$$|E_\lambda| \lesssim_q N^{\frac{dq}{2}-(d+2)} \lambda^{-q} \qquad \textit{for } \lambda \geqslant CN^{d/4}.$$

Proof. We view a and ω as functions of $(R(\mathbf{n}), \mathbf{n})$ for the sake of this argument, so that $F = \widehat{\omega 1_{S_{2N}}}$ and $F_a = \widehat{a 1_{S_N}}$, where S_N and S_{2N} are defined by (1.1). By Parseval, we have

$$\lambda |E_\lambda| \leqslant \langle f, F_a \rangle_{L^2(\mathbb{T}^{d+1})} = \langle f, \widehat{a 1_{S_N}} \rangle_{L^2(\mathbb{T}^{d+1})} = \langle \widehat{f}, a \rangle_{\ell^2(S_N)}.$$

By Cauchy-Schwarz and under the normalization $\|a\|_2 = 1$, it follows that

$$\lambda^2 |E_\lambda|^2 \leqslant \|f\|^2_{\ell^2(S_N)} \leqslant \langle f \cdot \omega 1_{S_{2N}}, f \rangle_{\ell^2(\mathbb{Z}^{d+1})}.$$

By another application of Parseval, we conclude that

$$\lambda^2 |E_\lambda|^2 \leqslant \langle f * F, f \rangle_{L^2(\mathbb{T}^{d+1})} \tag{5.2}$$

We will use this inequality to obtain bounds of the expected order on the level sets E_λ. By our earlier observation $F = \dot{F}$, inequality (5.2) becomes

$$\lambda^2 |E_\lambda|^2 \leqslant |\langle \dot{F} * f, f \rangle|,$$

and recalling the decompositions (4.5) and (4.11), we have

$$\begin{aligned}\lambda^2|E_\lambda|^2 &\leqslant |\langle \dot{F}_{\mathfrak{m}} * f, f\rangle| + |\langle \dot{F}_2 * f, f\rangle| + |\langle \dot{F}_1 * f, f\rangle| \\ &\leqslant \|F_{\mathfrak{m}}\|_\infty \|f\|_1^2 + \|\dot{F}_2 * f\|_{p'}\|f\|_p + \|\dot{F}_1 * f\|_{p'}\|f\|_p.\end{aligned}$$

Let $T \geqslant 1$ be a parameter to be determined later, and assume that we have chosen Q_1 so that $Q_1 \leqslant N^{2/B}$. Inserting the estimates of Propositions 4.4 and 4.10, this yields

$$\begin{aligned}\lambda^2|E_\lambda|^2 \lesssim N^{d/2}|E_\lambda|^2 &+ N^\varepsilon Q_1^{-(\frac{d}{2}-\frac{d+2}{p'})} N^{d-\frac{2(d+2)}{p'}}\|f\|_p^2 \\ &+ T^2 N^{d-\frac{2(d+2)}{p'}}\|f\|_p^2 + T^{-B}N^{d-\frac{d+2}{p'}}\|f\|_p\|f\|_1.\end{aligned}$$

Assume that $\lambda \geqslant CN^{d/4}$ for $C>0$ large enough and fix $Q_1 = N^{\varepsilon_1}$, where $\varepsilon_1 = 1/B$. For $p' > 2(d+2)/d$, and provided that ε is small enough, we have then

$$\lambda^2|E_\lambda|^2 \lesssim T^2 N^{d-\frac{2(d+2)}{p'}}|E_\lambda|^{2-\frac{2}{p'}} + T^{-B}N^{d-\frac{d+2}{p'}}|E_\lambda|^{2-\frac{1}{p'}}.$$

Writing $\lambda = \eta N^{d/2}$ with $\eta \in (0,1]$, we have therefore either

$$|E_\lambda|^{\frac{2}{p'}} \lesssim T^2 N^{-\frac{2(d+2)}{p'}}\eta^{-2} \quad \text{or} \quad |E_\lambda|^{\frac{1}{p'}} \lesssim T^{-B}N^{-\frac{d+2}{p'}}\eta^{-2}.$$

Write $D = T^{p'}$, so that in either case

$$|E_\lambda| \lesssim DN^{-(d+2)}\eta^{-p'} + D^{-B}N^{-(d+2)}\eta^{-2p'}.$$

Choose $D = \eta^{-\nu}$ for a parameter $\nu > 0$, so that

$$|E_\lambda| \lesssim N^{-(d+2)}\eta^{-p'-\nu}(1+\eta^{-p'+(B+1)\nu}).$$

Choosing $B \geqslant C/\nu$ with $C>0$ large enough, we deduce that $|E_\lambda| \lesssim N^{-(d+2)}\eta^{-(p'+\nu)}$. Since $q \coloneqq p'+\nu$ can be chosen arbitrarily close to $\frac{2(d+2)}{d}$, this finishes the proof, upon recalling that $\eta = \lambda N^{-d/2}$. □

Proof of Theorem 1.2. We may certainly assume that $\|a\|_2 = 1$ in proving this result. We apply Proposition 5.1 for a certain $\frac{2(d+2)}{d} < q < p$ to obtain

$$\begin{aligned}\int_{|F_a|\geqslant CN^{d/4}} |F_a|^p \mathrm{d}m &= p\int_{CN^{d/4}}^{N^{d/2}} \lambda^{p-1}|E_\lambda|\mathrm{d}\lambda \\ &\lesssim_p N^{\frac{dq}{2}-(d+2)}\int_1^{N^{d/2}} \lambda^{p-q-1}\mathrm{d}\lambda. \\ &\lesssim_p N^{\frac{dp}{2}-(d+2)}.\end{aligned}$$

□

Appendix A. Bounds on quadratic exponential sums

In this appendix we derive standard major and minor arc bounds on exponential sums associated to quadratic forms, which we could not locate precisely in the literature. We fix a nondegenerate quadratic form R in d variables with integer matrix, and we define

$$F_R(\alpha,\boldsymbol{\theta}) = \sum_{\mathbf{n}} \omega(\mathbf{n}) e(\alpha R(\mathbf{n}) + \boldsymbol{\theta}\cdot\mathbf{n}) \qquad (\alpha\in\mathbb{T}, \boldsymbol{\theta}\in\mathbb{T}^d).$$

Our first minor-arc-type bound is obtained by the standard Weyl differentiation process for forms of high dimension (see [10, Section 8.3.1.1] or Davenport [5, Chapter 13]).

Proposition A.1. *Let $d \geqslant 1$. For every $c_0 \in (0,1]$, there exists a constant $C > 0$ depending at most on c_0, d, R such that the following holds. If $|F_R(\alpha,\boldsymbol{\theta})| \geqslant CN^{d/2}$, there exist $a, q \in \mathbb{Z}$ such that $|\alpha - \frac{a}{q}| \leqslant \frac{c_0}{qN}$, $1 \leqslant q \leqslant c_0 N$ and $(a,q) = 1$.*

Proof. By definition, we have $R(\mathbf{x}) = \mathbf{x}^{\mathsf{T}} M \mathbf{x}$, where M is a symmetric, non-singular integer $d\times d$ matrix. For a vector $\mathbf{x}\in\mathbb{R}^d$, we write $|\mathbf{x}| = \max(|x_1|,\dots,|x_n|)$ and $\|x\| = \min_{\mathbf{n}\in\mathbb{Z}^d} |\mathbf{x}-\mathbf{n}|$. By squaring, we have

$$|F_R(\alpha,\boldsymbol{\theta})|^2 = \sum_{\mathbf{n},\mathbf{m}\in\mathbb{Z}^d} \omega(\mathbf{m})\omega(\mathbf{n}) e\big(\alpha(R(\mathbf{m}) - R(\mathbf{n})) + \boldsymbol{\theta}\cdot(\mathbf{m}-\mathbf{n})\big).$$

Letting $\mathbf{m} = \mathbf{n}+\mathbf{u}$ and $\Delta^{\times}_{\mathbf{u}}\omega(\mathbf{n}) = \omega(\mathbf{n})\omega(\mathbf{n}+\mathbf{u})$, we deduce that

$$\begin{aligned} |F_R(\alpha,\boldsymbol{\theta})|^2 &= \sum_{|\mathbf{u}|\leqslant 4N} e(\alpha R(\mathbf{u}) + \boldsymbol{\theta}\cdot\mathbf{u}) \sum_{\mathbf{n}\in\mathbb{Z}^d} \Delta^{\times}_{\mathbf{u}}\omega(\mathbf{n}) e(\mathbf{n}\cdot(2\alpha M\mathbf{u})) \\ &\leqslant \sum_{|\mathbf{u}|\leqslant 4N} |\widehat{\Delta^{\times}_{\mathbf{u}}\omega}(2\alpha M\mathbf{u})|. \end{aligned}$$

Since $\Delta^{\times}_{\mathbf{u}}\omega = \eta(\frac{\cdot}{N})\eta(\frac{\cdot+\mathbf{u}}{N})$ has support in $[-2N,2N]^d$ and satisfies $\|\partial^{\boldsymbol{\alpha}}\Delta^{\times}_{\mathbf{u}}\omega\|_\infty \lesssim N^{-|\boldsymbol{\alpha}|}$ for all $\boldsymbol{\alpha}\in(\mathbb{N}\cup\{0\})^d$, one can verify through an application of Poisson's formula that $|\widehat{\Delta^{\times}_{\mathbf{u}}\omega}(\boldsymbol{\xi})| \lesssim_A N^d(1+N\|\boldsymbol{\xi}\|)^{-A}$ uniformly for $\boldsymbol{\xi}\in\mathbb{R}^d$, for any $A>0$. Therefore

$$\begin{aligned} |F_R(\alpha,\boldsymbol{\theta})|^2 &\lesssim_d N^d \sum_{|\mathbf{u}|\leqslant 4N} \big(1+N\|2\alpha M\mathbf{u}\|\big)^{-(d+1)} \\ &\lesssim N^d \sum_{|\mathbf{r}|\leqslant \frac{N}{2}} (1+|\mathbf{r}|)^{-(d+1)} \cdot \#\big\{|\mathbf{u}|\leqslant 4N \,:\, 2\alpha M\mathbf{u} \in \tfrac{\mathbf{r}}{N} + [-\tfrac{1}{2N},\tfrac{1}{2N}] \bmod 1\big\}. \end{aligned}$$

If $\mathbf{u}$, $\mathbf{u}'$ belong to the set above, then $\|2\alpha M(\mathbf{u}-\mathbf{u}')\| \leqslant \frac{1}{N}$, and therefore

$$|F_R(\alpha,\boldsymbol{\theta})|^2 \lesssim N^d\cdot\#\{|\mathbf{u}|\leqslant 8N \,:\, \|2\alpha M\mathbf{u}\|\leqslant \tfrac{1}{N}\}.$$

Let $1\leqslant L\leqslant N$ be a new parameter. By a similar reasoning, if we partition the box $[-8N,8N]^d$ into subboxes of sidelength at most L, and if we partition the box $[-\frac{1}{N},\frac{1}{N}]^d$

into subboxes of sidelength at most $\frac{L}{N^2}$, we obtain

$$|F_R(\alpha,\boldsymbol{\theta})|^2 \lesssim_d L^{-2d}N^{3d}\cdot \#\{|\mathbf{u}|\leqslant L \,:\, \|2\alpha M\mathbf{u}\|\leqslant \tfrac{4L}{N^2}\}.$$

We choose $L = c_1 N$, where $c_1 \in (0,1]$ is to be determined later. If $|F_R(\alpha,\boldsymbol{\theta})| \geqslant CN^{d/2}$ for a large enough constant C (depending on d and c_1), then there exists $\mathbf{u}\neq 0$ such that $|\mathbf{u}|\leqslant c_1 N$ and $\|2\alpha M\mathbf{u}\| \leqslant \frac{c_1}{qN}$, and we let $q = 2|M\mathbf{u}|$. Since M is non-singular, we have $1 \leqslant q \lesssim_M c_1 N$ and $\|q\alpha\| \leqslant \frac{4c_1}{N}$, and therefore there exists $a\in\mathbb{Z}$ such that $|\alpha - a/q| \leqslant \frac{4c_1}{qN}$. This finishes the proof upon choosing c_1 small enough with respect to d, M and upon reducing a and q. □

On the major arcs, we use a standard majorant obtained through the Poisson formula, using the square-root level of cancellation in the Gaussian sum and oscillatory integral associated to non-degenerate quadratic forms.

Proposition A.2. *Let $d\geqslant 1$. Suppose that $\alpha\in\mathbb{R}$ is of the form $\alpha = \frac{a}{q}+\beta$ with $a,q\in\mathbb{Z}$, $\beta\in\mathbb{R}$ such that $\|\beta\| \lesssim \frac{1}{qN}$, $1\leqslant q\lesssim N$ and $(a,q)=1$. Then*

$$|F_R(\alpha,\boldsymbol{\theta})| \lesssim_\varepsilon q^{-d/2+\varepsilon}\min(N^d, |\beta|^{-d/2}).$$

Proof. We define a Gaussian sum and an oscillatory integral by

$$S(a,\mathbf{b};q) = \sum_{\mathbf{u}\in\mathbb{Z}_q^d} e_q(aR(\mathbf{u}) + \mathbf{b}\cdot\mathbf{u}), \qquad I(\beta,\boldsymbol{\gamma};N) = \int_{\mathbb{R}^d} \eta(\mathbf{x}) e(\beta N^2 R(\mathbf{x}) + N\boldsymbol{\gamma}\cdot\mathbf{x})\mathrm{d}\mathbf{x}.$$

We write $\alpha \equiv \frac{a}{q}+\beta \bmod 1$ and we sum over residue classes modulo q to obtain

$$F_R(\alpha,\boldsymbol{\theta}) = \sum_{\mathbf{u}\in\mathbb{Z}_q^d} e_q(aR(\mathbf{u})) \sum_{\substack{\mathbf{n}\in\mathbb{Z}^d:\\ \mathbf{n}\equiv\mathbf{u} \bmod q}} \omega(\mathbf{n}) e(\beta R(\mathbf{n}) + \boldsymbol{\theta}\cdot\mathbf{n}).$$

Writing $1_{\mathbf{n}\equiv\mathbf{u} \bmod q} = q^{-d}\sum_{\mathbf{b}\in\mathbb{Z}_q^d} e_q(\mathbf{b}\cdot(\mathbf{u}-\mathbf{n}))$, we arrive at

$$F_R(\alpha,\boldsymbol{\theta}) = \sum_{\mathbf{b}\in\mathbb{Z}_q^d} q^{-d} S(a,\mathbf{b};q) \sum_{\mathbf{n}\in\mathbb{Z}^d} \omega(\mathbf{n}) e(\beta R(\mathbf{n}) + (\boldsymbol{\theta} - \tfrac{\mathbf{b}}{q})\cdot\mathbf{n}).$$

By Poisson's formula and rescaling, it follows that

$$F_R(\alpha,\boldsymbol{\theta}) = \sum_{\mathbf{b}\in\mathbb{Z}_q^d} q^{-d} S(a,\mathbf{b};q) \sum_{\mathbf{m}\in\mathbb{Z}^d} N^d\cdot I(\beta, \boldsymbol{\theta} - \tfrac{\mathbf{b}}{q} - \mathbf{m}; N). \tag{A.1}$$

We write $I(\beta,\boldsymbol{\theta} - \frac{\mathbf{b}}{q} - \mathbf{m};N) = \int_{\mathbb{R}} \eta(\mathbf{x}) e(N\phi_{\mathbf{b},\mathbf{m}}(\mathbf{x}))\mathrm{d}\mathbf{x}$, where

$$\phi_{\mathbf{b},\mathbf{m}}(\mathbf{x}) = \beta N R(\mathbf{x}) + (\boldsymbol{\theta} - \tfrac{\mathbf{b}}{q} - \mathbf{m})\cdot\mathbf{x}.$$

On the support of η, we have $|x|\leqslant 2$ and therefore

$$\nabla\phi_{\mathbf{b},\mathbf{m}}(\mathbf{x}) = \theta - \tfrac{\mathbf{b}}{q} - \mathbf{m} + O(\tfrac{1}{q})$$

under our size condition on β. We fix a large enough constant $C > 0$.

For $|\mathbf{m}| \geqslant C$, we have $|\nabla\phi_{\mathbf{b},\mathbf{m}}| \asymp |\mathbf{m}|$ on $\operatorname{Supp}\eta$, and therefore by stationary phase [11, Chapter VII] we have $|\int_{\mathbb{R}} \eta e(N\phi_{\mathbf{b},\mathbf{m}})| \lesssim (N|\mathbf{m}|)^{-(d+1)}$. For $\|\boldsymbol{\theta} - \frac{\mathbf{b}}{q}\| \geqslant \frac{C}{q}$, we have $|\nabla\phi_{\mathbf{b},\mathbf{m}}| \asymp |\boldsymbol{\theta} - \frac{\mathbf{b}}{q} - \mathbf{m}| \gtrsim \|\boldsymbol{\theta} - \frac{\mathbf{b}}{q}\|$ on $\operatorname{Supp}\eta$ and $\|\frac{\phi_{\mathbf{b},\mathbf{m}}}{|\boldsymbol{\theta}-\frac{\mathbf{b}}{q}-\mathbf{m}|}\|_{C^2} \lesssim 1$, so that by stationary phase again we deduce that $|\int_{\mathbb{R}} \eta e(N\phi_{\mathbf{b},\mathbf{m}})| \lesssim (N\|\boldsymbol{\theta} - \frac{\mathbf{b}}{q}\|)^{-d}$. Finally, for $|\mathbf{m}| \leqslant C$ and $\|\boldsymbol{\theta} - \frac{\mathbf{b}}{q}\| \leqslant \frac{C}{q}$, we note that the phase is a non-degenerate quadratic form and therefore we have an oscillatory integral estimate [13, Section 6] of the form $|\int_{\mathbb{R}} \eta e(N\phi_{\mathbf{b},\mathbf{m}})| \lesssim (1 + |\beta|N^2)^{-d/2}$.

For the Gaussian sum, we use the simple squaring-differencing bound $|S(a, \mathbf{b}; q)| \lesssim q^{d/2}$ for $(a, q) = 1$ (see e.g. [8] Lemma 20.12). Inserting these various estimates into (A.1) yields

$$\begin{aligned}
|F_R(\alpha, \boldsymbol{\theta})| \lesssim_\varepsilon\; & q^{-d/2} \sum_{\substack{\|\boldsymbol{\theta}-\frac{\mathbf{b}}{q}\| \leqslant \frac{C}{q} \\ |\mathbf{m}| \leqslant C}} N^d (1 + |\beta|N^2)^{-\frac{d}{2}} \\
& + q^{-d/2} \sum_{\substack{\|\boldsymbol{\theta}-\frac{\mathbf{b}}{q}\| \geqslant \frac{C}{q} \\ |\mathbf{m}| \leqslant C}} \|\boldsymbol{\theta} - \tfrac{\mathbf{b}}{q}\|^{-d} \;+\; q^{d/2} \sum_{|\mathbf{m}| \geqslant C} N^{-1} |\mathbf{m}|^{-(d+1)} \\
\lesssim\; & q^{-d/2+\varepsilon} N^d (1 + |\beta|N^2)^{-d/2} \;+\; q^{d/2+\varepsilon}.
\end{aligned}$$

The second term may be absorbed into the first since $|\beta| \lesssim \frac{1}{qN}$ and $1 \leqslant q \lesssim N$, and this concludes the proof. ☐

Appendix B. A diagonalization argument

In this section we present a simple argument, possibly well-known to experts, by which Theorem 1.1 follows from [3, Corollary 1.3].

Proof of Theorem 1.1.

Let Q be a non-singular quadratic form with integer coefficients. Fix a sequence $a : \mathbb{Z}^d \to \mathbb{C}$ supported on $[-N, N]^d$; by homogeneity we may assume $\|a\|_2 = 1$. We let

$$I = \|F_a\|_p^p = \int_{[-\frac{1}{2},\frac{1}{2}]^{d+1}} \Big| \sum_{\mathbf{n} \in \mathbb{Z}^d} a(\mathbf{n}) e\big(\alpha Q(\mathbf{n}) + \boldsymbol{\theta} \cdot \mathbf{n}\big) \Big|^p \mathrm{d}\alpha \mathrm{d}\boldsymbol{\theta}.$$

We pick a linear transformation T of $\mathbb{Q}^d$ such that $Q = D \circ T$, where D is a diagonal form with coefficients ± 1. Then by defining the lattice $\Lambda = T(\mathbb{Z}^d)$ and by a change of variables $\boldsymbol{\theta} = T^*(\boldsymbol{\xi})$, we have

$$I = \int_{[-\frac{1}{2},\frac{1}{2}]^{d+1}} \Big| \sum_{\mathbf{m}\in\Lambda} a(T^{-1}(\mathbf{m}))e\big(\alpha D(\mathbf{m}) + (T^{-1})^*(\boldsymbol{\theta})\cdot \mathbf{m}\big)\Big|^p \mathrm{d}\alpha \mathrm{d}\boldsymbol{\theta},$$

$$= |\det T| \int_E \Big| \sum_{\mathbf{m}\in\Lambda} a(T^{-1}(\mathbf{m}))e\big(\alpha D(\mathbf{m}) + \boldsymbol{\xi}\cdot \mathbf{m}\big)\Big|^p \mathrm{d}\alpha \mathrm{d}\boldsymbol{\xi},$$

where $E = [-\frac{1}{2},\frac{1}{2}] \times (T^*)^{-1}([-\frac{1}{2},\frac{1}{2}]^d)$. We have $\Lambda \subset q^{-1}\mathbb{Z}^d$ for some $q \in \mathbb{N}$ depending on Q, and by a change of variables $(\alpha,\boldsymbol{\xi}) \leftarrow (q^2\alpha, q\boldsymbol{\xi})$, we have

$$I = q^{d+2}|\det T| \int_F \Big| \sum_{\boldsymbol{\ell}\in q\Lambda} a(T^{-1}(\boldsymbol{\ell}/q))e\big(\alpha D(\boldsymbol{\ell}) + \boldsymbol{\xi}\cdot \boldsymbol{\ell}\big)\Big|^p \mathrm{d}\alpha \mathrm{d}\boldsymbol{\xi},$$

where $F = [-\frac{1}{2q^2},\frac{1}{2q^2}] \times (T^*)^{-1}([-\frac{1}{2q},\frac{1}{2q}]^d)$. Finally, we can cover F by finitely many translated copies of $[-\frac{1}{2},\frac{1}{2}]^{d+1}$, ostensibly by a power of q many translates, and since $q\Lambda \subset \mathbb{Z}^d \cap [-CN,CN]^d$, we may apply the usual restriction estimate for diagonal forms of Bourgain–Demeter [3, Corollary 1.3] to obtain the estimate

$$\|F_a\|_p^p \lesssim_\varepsilon \begin{cases} N^{\frac{sp}{2}-s+\varepsilon}\|a\|_2^p & \text{for } 2 \leqslant p \leqslant \frac{2(d-s+2)}{d-s}, \\ N^{\frac{dp}{2}-(d+2)+\varepsilon}\|a\|_2^p & \text{for } p > \frac{2(d-s+2)}{d-s}. \end{cases}$$

In applying the Bourgain–Demeter restriction estimate we are using that q and $|\det T|$ only depend on the quadratic form Q and do not depend on N. Therefore the implicit constants in the above estimate only depend on the form Q and not on N. The N^ε factor in the supercritical range can be removed via (a minor modification of) Bourgain's ε-removal lemma for the paraboloid $(x_1,\dots,x_d,x_1^2+\cdots+x_d^2)$. (Alternatively, one can use Theorem 1.2 to remove this factor, via Bourgain's ε-removal process [7, Lemma 3.1]). □

References

1. J. Bourgain, *On $\Lambda(p)$-subsets of squares*, Israel J. Math. **67** (1989), no. 3, 291–311.
2. ———, *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations. I. Schrödinger equations*, Geom. Funct. Anal. **3** (1993), no. 2, 107–156.
3. J. Bourgain and C. Demeter, *Decouplings for curves and hypersurfaces with nonzero Gaussian curvature*, to appear in J. Anal. Math.
4. ———, *The proof of the ℓ^2 decoupling conjecture*, Ann. of Math. (2) **182** (2015), no. 1, 351–389.
5. H. Davenport, *Analytic methods for Diophantine equations and Diophantine inequalities*, second ed., Cambridge University Press, Cambridge, 2005.
6. N. Godet and N. Tzvetkov, *Strichartz estimates for the periodic non-elliptic Schrödinger equation*, C. R. Math. Acad. Sci. Paris **350** (2012), no. 21-22, 955–958.
7. K. Henriot and K. Hughes, *Restriction estimates of ε-removal type for k-th powers and paraboloids*, to appear in Math. Ann.

8. H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society, Providence, RI, 2004.
9. R. Killip and M. Vişan, *Scale invariant Strichartz estimates on tori and applications*, Math. Res. Lett. **23** (2016), no. 2, 445–472. MR 3512894
10. A. Magyar, *On the distribution of solutions to Diophantine equations*, A panorama of discrepancy theory, Springer, Cham, 2014, pp. 487–538.
11. E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton University Press, Princeton, NJ, 1993.
12. Y. Wang, *Periodic cubic hyperbolic Schrödinger equation on* $\mathbb{T}^2$, J. Funct. Anal. **265** (2013), no. 3, 424–434.
13. T. H. Wolff, *Lectures on harmonic analysis*, vol. 29, American Mathematical Society, Providence, RI, 2003, Edited by Łaba and Carol Shubin.

Department of mathematics, University of British Columbia, Room 121, 1984 Mathematics Road, Vancouver BC V6T 1Z2, Canada

Email address: khenriot@math.ubc.ca

School of Mathematics, University of Bristol, Bristol, BS8 1TW, UK, and the Heilbronn Insitute for Mathematical Research, Bristol, UK

Email address: khughes.math@gmail.com