SHARP RESTRICTION ESTIMATES FOR SEVERAL DEGENERATE HIGHER CO-DIMENSIONAL QUADRATIC SURFACES

ZHENBIN CAO, CHANGXING MIAO AND YIXUAN PANG

Abstract. Fourier restriction conjecture is an important problem in harmonic analysis. Guo-Oh in [17] studied the restriction estimates for quadratic surfaces of co-dimension 2 in $\mathbb{R}^5$. For one special surface $(\xi_1, \xi_2, \xi_3, \xi_1^2, \xi_2^2 + \xi_1\xi_3)$, they applied a nested induction argument to build essentially sharp restriction estimate. In this paper, we simplify their method, and extend it to a variant of the broad-narrow analysis. As applications, we will prove essentially sharp restriction estimates for some kinds of degenerate higher co-dimensional quadratic surfaces.

1. Introduction

Let $d, n \geq 1$, and $\mathbf{Q}(\xi) = (Q_1(\xi), ..., Q_n(\xi))$ be an n-tuple of real quadratic forms in d variables. The graph of such a tuple, $S_{\mathbf{Q}} = \{(\xi, \mathbf{Q}(\xi)) \in [0,1]^d \times \mathbb{R}^n\}$, is a d-dimensional submanifold of $\mathbb{R}^{d+n}$. Define the extension operator associated to $S_{\mathbf{Q}}$ by

$$E^{\mathbf{Q}} f(x) := \int_{[0,1]^d} e^{2\pi i x \cdot (\xi, \mathbf{Q}(\xi))} f(\xi) d\xi, \qquad x \in \mathbb{R}^{d+n}.$$

We mainly focus on the following Fourier restriction problem: Find optimal ranges of p and q such that

$$\|E^{\mathbf{Q}} f\|_{L^q(\mathbb{R}^{d+n})} \leq C_{p,q,\mathbf{Q},d,n} \|f\|_{L^p(\mathbb{R}^d)}, \tag{1.1}$$

for every measurable function f.

Originating from a deep observation of Stein [31] in 1967, the Fourier restriction conjecture has been widely studied for hypersurfaces ($n = 1$), especially the paraboloid. Although the $d = 1$ case was solved by Fefferman [14] and Zygmund [38] half a century ago, the $d \geq 2$ case is still far from been fully understood. In the special case when $p = 2$, the sharp estimate up to the endpoint for the sphere was first proved by Tomas [35], and the endpoint result was later established by Stein [32] through complex interpolation. The Stein-Tomas framework is very influential and turns out to work for any hypersurface with nonvanishing Gaussian curvature, including the paraboloid. However, things become much trickier for general p, and

Date: April 16, 2024.

2010 *Mathematics Subject Classification.* 42B20, 42B37.

Key words and phrases. quadratic surface, decoupling, broad-narrow analysis.

people have to search for other methods. After early major progress made by Bourgain [5], several influential methods have been developed, such as the bilinear method [33, 34], the multilinear method/broad-narrow analysis [9], and the polynomial method [21, 22]. In recent years, these methods are combined with more delicate techniques from incidence geometry and real algebraic geometry [24, 25, 36], and the current world records are set up by Wang-Wu [37] for $d = 2$ and Guo-Wang-Zhang [20] for $d \geq 3$.

In contrast to the paraboloid, people know much less about restriction estimates for higher co-dimensional quadratic surfaces ($n > 1$). And now we briefly summarize the history.

For a long time, the state-of-the-art methods are hardly beyond the Stein-Tomas framework and interpolation of analytic families of operators, so the $p = 2$ case is the main focus of many papers. When $n = 2$, Christ [11] and Mockenhaupt [28] introduced a geometric notion of nondegeneracy (Definition 2.3) which is equivalent to optimal $L^2 \to L^q$ estimates. However, when $n \geq 3$, things become much more complicated and the results are quite sparse, see [4] for some partial progress.

Another line of attack is based on the Fefferman-Zgymund framework and mapping properties of multilinear fractional integral operators. Such a method breaks through the L^2 barrier, and can sometimes yield full sharp ranges of $L^p \to L^q$ estimates (i.e., even including the critical line). For example, when $d = n = 2$, Christ [12] fully solved the restriction problem for non-degenerate quadratic surfaces, and when $n = \frac{d(d+1)}{2}$, Bak-Lee [3] and Oberlin [30] fully solved the restriction problem for extremal quadratic surfaces. However, such a method seems to require certain unnatural relations between d and n, which prevents us from obtaining sharp results in general cases, see [29] for a discussion on the non-degenerate surface $(\xi_1, \xi_2, \xi_3, \xi_1^2 + \xi_2^2, \xi_2^2 + \xi_3^2)$ using similar techniques.

In recent years, the developments when $n = 1$ mentioned before have greatly facilitated the research when $n > 1$. For example, Bak-Lee-Lee [2] applied the bilinear method to obtain some restriction estimates for higher co-dimensional surfaces with nonvanishing rotational curvature, such as the complex paraboloid $(\xi_1, \xi_2, \xi_3, \xi_4, \xi_1^2 - \xi_2^2 + \xi_3^2 - \xi_4^2, \xi_1\xi_2 + \xi_3\xi_4)$, and Lee-Lee [26] applied the broad-narrow analysis to study holomorphic complex hypersurfaces. However, the assumptions in these two works are too strong due to technical reasons. Noteworthy progress was made by Guo-Oh [17], who systematically investigated restriction estimates for all quadratic surfaces of co-dimension 2 in $\mathbb{R}^5$, such as $(\xi_1, \xi_2, \xi_3, \xi_1^2, \xi_2^2 + \xi_1\xi_3)$. Later on, Guo-Oh-Zhang-Zorin-Kranich [19] established sharp decoupling inequalities for all higher co-dimensional quadratic surfaces by combining a transversality condition originated from [7] with a scale-dependent Brascamp-Lieb inequality due to Maldague [27], and the decoupling constant is characterized by "the minimal number of variables" (Definition 2.2). As an application, they obtained a unified restriction estimate for quadratic surfaces of any dimensions and co-dimensions through the broad-narrow analysis. However, the full power of the broad-narrow analysis has not been exploited until Gan-Guth-Oh [15] further developed the framework in [19] by adopting a more delicate notion of transversality, which enables them to capture lower dimensional contribution. But in this case, the narrow set may lie in a small neighborhood of an algebraic variety with very bad singularities. To overcome this obstacle, Gan-Guth-Oh devised a covering lemma for varieties by using Tarski's projection theorem from real algebraic geometry.

In this paper, we consider the restriction estimates for some kinds of degenerate higher co-dimensional quadratic surfaces. Our main result is as follows:

Theorem 1.1. *Let $d, n \geq 2$ and $\mathbf{Q} = (Q_1, ..., Q_n)$ be an n-tuple of real quadratic forms defined on $\mathbb{R}^d$.*
(1) ($d = n$, *monomial*) *Suppose that*

$$\mathbf{Q} = (\xi_{\lambda_1}\xi_1, \xi_{\lambda_2}\xi_2, ..., \xi_{\lambda_n}\xi_n).$$

Here each λ_j is an integer and $1 \leq \lambda_j \leq n$. Denote the number of j in the set $\{\lambda_1, \lambda_2, ..., \lambda_n\}$ by w_j. If $\lambda_j \leq j$ for each $1 \leq j \leq n$, then the estimate (1.1) *holds for*

$$q > \max_j w_j + 3, \quad \frac{1}{p} + \frac{\max_j w_j + 2}{q} < 1. \tag{1.2}$$

(2) ($d = n$, *polynomial*) *Suppose that*

$$\mathbf{Q} = (\xi_1^2, \xi_1\xi_2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1} + P_{w_1+1}(\xi), ..., \xi_\lambda\xi_n + P_n(\xi)).$$

Denote the total number of variables on which $P_{w_1+1}(\xi), ..., P_n(\xi)$ depend except the variables ξ_1 and ξ_λ by θ. Let $w_\lambda = n - w_1$. We assume that one of the following conditions holds:

- (2a) *Each P_j is a quadratic form that is independent of the variables $\xi_1, \xi_j, ..., \xi_n$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta$.*
- (2b) *Each P_j is a quadratic form without mixed terms that is independent of the variables $\xi_j, ..., \xi_n$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta$.*
- (2c) *P_j is a quadratic form without mixed terms that is independent of the variables $\xi_j, ..., \xi_n$ for some $w_1 + 1 \leq j \leq n$ and $P_{j'} = 0$ for all $j' \neq j$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta/2$.*

Then the estimate (1.1) *holds for*

$$q > w_1 + 3, \quad \frac{1}{p} + \frac{w_1 + 2}{q} < 1. \tag{1.3}$$

(3) ($d = n + k$, *monomial*) *Suppose that*

$$\mathbf{Q} = (\xi_{\lambda_{1+k}}\xi_{1+k}, \xi_{\lambda_{2+k}}\xi_{2+k}, ..., \xi_{\lambda_{n+k}}\xi_{n+k}),$$

with $k \geq 1$ and $1 \leq \lambda_j \leq n+k$ for each j. Denote the number of j in the set $\{\lambda_{1+k}, \lambda_{2+k}, ..., \lambda_{n+k}\}$ by w_j. If $\lambda_j \leq j$ for each $1+k \leq j \leq n+k$, and $w_j \geq 1$ for each $1 \leq j \leq k$, then the estimate (1.1) *holds for*

$$q > \max\Big\{ \max_{j=1,...,k} w_j + 2, \max_{j=1+k,...,n+k} w_j + 3 \Big\} =: q_1, \quad \frac{1}{p} + \frac{q_1 - 1}{q} < 1. \quad (1.4)$$

(4) *($d = n+k$, monomial) Suppose that*

$$\mathbf{Q} = (\xi_1^2, ..., \xi_\eta^2, \xi_{\lambda_{\eta+1+k}}\xi_{\eta+1+k}, \xi_{\lambda_{\eta+2+k}}\xi_{\eta+2+k}, ..., \xi_{\lambda_{n+k}}\xi_{n+k}),$$

with $k \geq 1$, $1 \leq \eta < n$, and $1 \leq \lambda_j \leq n+k$ for each j. Denote the number of j in the set $\{\lambda_{\eta+1+k}, \lambda_{\eta+2+k}, ..., \lambda_{n+k}\}$ by w_j. If $\lambda_j \leq j$ for each $\eta+1+k \leq j \leq n+k$, and $w_j \geq 1$ for each $\eta+1 \leq j \leq \eta+k$, then the estimate (1.1) *holds for*

$$q > \max\Big\{ \max_{j=1,...,\eta} w_j + 4, \max_{j=\eta+1,...,\eta+k} w_j + 2, \max_{j=\eta+1+k,...,n+k} w_j + 3 \Big\} =: q_2, \quad (1.5)$$

$$\frac{1}{p} + \frac{q_2 - 1}{q} < 1. \quad (1.6)$$

(5) *($d = n+k$, polynomial) Suppose that*

$$\mathbf{Q} = (\xi_1^2, \xi_1\xi_2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1} + P_{w_1+1}(\xi), ..., \xi_\lambda\xi_{n-1} + P_{n-1}(\xi), \xi_\lambda\xi_n + P_n(\xi) + \xi_{n+1}^2 + ... + \xi_{n+k}^2).$$

Denote the total number of variables on which $P_{w_1+1}(\xi), ..., P_n(\xi)$ depend except the variables ξ_1 and ξ_λ by θ. Let $w_\lambda = n - w_1 > 0$. We assume that one of the following conditions holds:

- (5a) *Each P_j is a quadratic form that is independent of the variables $\xi_1, \xi_j, ..., \xi_{n+k}$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta$.*
- (5b) *Each P_j is a quadratic form without mixed terms that is independent of the variables $\xi_j, ..., \xi_{n+k}$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta$.*
- (5c) *P_j is a quadratic form without mixed terms that is independent of the variables $\xi_j, ..., \xi_{n+k}$ for some $w_1 + 1 \leq j \leq n$ and $P_{j'} = 0$ for all $j' \neq j$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta/2$.*
- (5d) *$P_j = 0$ for every $w_1 + 1 \leq j \leq n$; $1 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda$.*

Then the estimate (1.1) *holds for*

$$q > w_1 + 3, \quad \frac{1}{p} + \frac{w_1 + 2}{q} < 1. \quad (1.7)$$

Moreover, the ranges (1.2) – (1.7) *are all sharp up to endpoints.*

Remark 1.1. *We restrict ourselves to $d \geq n$ so that we can apply the bilinear restriction estimates independent of the transversality parameter (Theorem 3.1). The quadratic surfaces in Theorem 1.1 are all degenerate—except the case (3)—due to $\mathfrak{d}_{d,1}(\boldsymbol{Q}) = 1$, see* [18, Lemma 2.2].

The idea of our proof of Theorem 1.1 initially comes from Guo-Oh's work [17], which studied the Fourier restriction estimates for quadratic surfaces of co-dimension 2 in $\mathbb{R}^5$. In particular, for one special degenerate case, $(\xi_1, \xi_2, \xi_3, \xi_1^2, \xi_2^2 + \xi_1\xi_3)$, they developed a *nested* induction technique to reach the sharp restriction estimate. More precisely, they decomposed the frequency domain into rectangular boxes with dimensions $K^{-1} \times K^{-1/2} \times 1$ and carried out the broad-narrow analysis. Such scale is adapted to this case, which means that the related surface $S_{\mathbf{Q}}$ remains unchanged if we scale it to $\chi_{[0,1]^3}$. After this, the narrow part may be restricted to a rectangular box with dimensions $K^{-1} \times 1 \times 1$ or $1 \times K^{-1/2} \times 1$. In order to apply induction on scales to the narrow part, one needs first to reduce it to the adapted scale by using decoupling. Unfortunately, the loss of decoupling keeps the optimal range out of reach. Instead, Guo-Oh chose to consider a more general restriction estimate in which the Fourier support of f lies in a rectangular box with dimensions $L^{-1} \times 1 \times 1$ (see (2.11), and the argument for the other case is similar) for every $K \le L \le R^{1/2}$. This seemingly harder estimate is actually more manageable due to an inductive structure. On the one hand, they employed several techniques to validate the base case $L = R^{1/2}$. On the other hand, they further combined the broad-narrow analysis with backward induction on L to finally prove the sharp estimate. It is worth mentioning that a key ingredient in the above argument is certain bilinear restriction estimates with favourable dependence on transversality parameters.

We find that Guo-Oh's approach can be rewritten through an iteration of the broad-narrow analysis, which allows us to reduce the original term $D_p(1, 1, 1; R)$ (see Definition 2.1) to several terms: one bilinear term $BD_p(\mu_1, \mu_2, 1; R)$, two linear terms $D_p(L, 1, 1; R)$ and $D_p(1, M, 1; R)$, and two additional terms $D_p(L, L^{1/2}, 1; R)$ and $D_p(M^2, M, 1; R)$. For the additional terms, their scales are adapted to the surface, for which Guo-Oh was able to apply rescaling and backward induction on the parameters L and M. One important observation made by us is that these additional terms can be further reduced to the preceding linear terms through finitely many steps of the broad-narrow iteration. Apart from simplifying Guo-Oh's proof, there are two advantages of this alternative argument: The first is that it relies little on the specific expression of $\mathbf{Q}$ as we do not resort to rescaling and induction on scales, and so can be generalized to study some abstract quadratic surfaces; the second is that it eventually turns the narrow part into two terms $D_p(R^{1/2}, 1, 1; R)$ and $D_p(1, R^{1/2}, 1; R)$, which can be regarded as "genuine" lower dimensional cases compared with $D_p(K, 1, 1; R)$ and $D_p(1, K^{1/2}, 1; R)$ in the classical broad-narrow analysis. These "less"-narrow terms can be estimated via more techniques, which offers many conveniences. Furthermore, we will extend Guo-Oh's bilinear restriction estimates with favourable dependence on transversality parameters to more general situations, by which we further develop our method into a complete version. As applications, we will prove sharp restriction estimates for some kinds of degenerate quadratic surfaces (Theorem 1.1).

Before ending this section, we point out that our method can be used to study more degenerate quadratic surfaces than those listed in Theorem 1.1, although we choose to only include several typical examples in this theorem. However, our method may not be well-suited for all degenerate quadratic surfaces. We will discuss these issues in detail in Section 5.

Outline of the paper. In Section 2, we first sketch Guo-Oh's proof on the restriction estimate for the surface $(\xi_1,\xi_2,\xi_3,\xi_1^2,\xi_2^2+\xi_1\xi_3)$, and then simplify their proof by iterating the broad-narrow analysis without rescaling and induction on scales. As an application, we give an alternative proof of Guo-Oh's result for the non-degenerate case in [17]. In Section 3, by establishing a general bilinear restriction estimate independent of the transversality parameter, we further develop our method into a complete version which can be used to deal with more degenerate cases. We will compare it with the classical broad-narrow analysis at the end of this section. In Section 4, we prove Theorem 1.1. In Section 5, we offer additional remarks on our method and Theorem 1.1 through several examples.

Notation. We will use $\#X$ to denote the cardinality of a finite set X, and use $|X|$ to denote the Lebesgue measure of a measurable set X. We will use B_R^m to represent a ball with radius R in $\mathbb{R}^m$, and we abbreviate B_R^m to B_R if m is clear from the context. We will write $A \lesssim_\epsilon B$ to mean that there exists a constant C depending on ϵ such that $A \le CB$. Moreover, $A \sim B$ means $A \lesssim B$ and $A \gtrsim B$, and $A \lessapprox B$ means $A \lesssim \log R \cdot B$, where R is some parameter. Let $e(b) := e^{2\pi i b}$ for each $b \in \mathbb{R}$. We will use $\operatorname{supp} f$ to denote the support of a function f, and use χ_E to denote the characteristic function of a set E.

2. Review and simplification of Guo-Oh's proof

For $\mathbf{Q} = (\xi_1^2, \xi_2^2+\xi_1\xi_3)$, Guo-Oh proved that (1.1) holds for $q > 4$ and $p > q/(q-3)$. By interpolation with the trivial $L^1 \to L^\infty$ bound, it suffices to show

$$\|E^{\mathbf{Q}} f\|_{L^p(\mathbb{R}^5)} \le C_p \|f\|_{L^p(\mathbb{R}^3)}, \tag{2.1}$$

for every $p > 4$. By testing (2.1) on $\chi_{[0,R^{-1}]\times[0,R^{-1/2}]\times[0,1]}$, we can easily see that this range is sharp up to endpoint, see [17, Section 3]. Also, the surface $S_{\mathbf{Q}}$ remains unchanged if we scale this example to $\chi_{[0,1]^3}$. This inspires Guo-Oh to consider frequency decomposition adapted to such a scale.

Definition 2.1. *Let $R, \mu_1, ..., \mu_d \ge 1$. Let $\mathbf{Q}(\xi) = (Q_1(\xi), ..., Q_n(\xi))$ be an n-tuple of real quadratic forms in d variables.*

Define $D_p(\mu_1, ..., \mu_d; R)$ to be the smallest constant such that

$$\|E^{\mathbf{Q}} f\|_{L^p(B_R^{d+n})} \le D_p(\mu_1, \dots, \mu_d; R) \|f\|_{L^p(\mathbb{R}^d)}, \tag{2.2}$$

where f satisfies $\operatorname{supp} f \subset [a_1, a_1 + \mu_1^{-1}] \times \cdots \times [a_d, a_d + \mu_d^{-1}] \subset [0,1]^d$. *When $\mu_1 = \cdots = \mu_d = 1$, we abbreviate $D_p(R) = D_p(1, ..., 1; R)$.*

Define $BD_p(\mu_1, ..., \mu_d; R)$ to be the smallest constant such that

$$\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^p(B_R^{d+n})} \le BD_p(\mu_1, ..., \mu_d; R) \|f_1\|_{L^p(\mathbb{R}^d)}^{\frac{1}{2}} \|f_2\|_{L^p(\mathbb{R}^d)}^{\frac{1}{2}}, \tag{2.3}$$

where f_1 satisfies $\operatorname{supp} f_1 \subset [a_1, a_1 + \mu_1^{-1}] \times \cdots \times [a_d, a_d + \mu_d^{-1}] \subset [0,1]^d$ *and f_2 satisfies* $\operatorname{supp} f_2 \subset [b_1, b_1 + \mu_1^{-1}] \times \cdots \times [b_d, b_d + \mu_d^{-1}] \subset [0,1]^d$ *with $|a_j - b_j| \ge 10\mu_j^{-1}$ if $\mu_j > 1$ $(1 \le j \le d)$.* If $\mu_1 = 1$, we say f_1 and f_2 are both supported on $[0,1]$ with respect to the first coordinate; definitions in other cases ($\mu_j = 1$ for $j = 2, ..., d$) are similar.

Note that

$$D_p(\mu_1, ..., \mu_d; R) \le D_p(\mu_1', ..., \mu_d'; R) \tag{2.4}$$

whenever $\mu_j \ge \mu_j'$ for any $1 \le j \le d$. And our aim is to prove

$$D_p(R) \le C_p, \tag{2.5}$$

for every $p > 4$.

Guo-Oh showed several estimates, which laid the foundation for induction on scales:

$$D_p(R^{\frac{1}{2}}, 1, 1; R) \lesssim R^{\frac{2}{p} - \frac{1}{2}}, \qquad p > 4; \tag{2.6}$$

$$D_p(1, R^{\frac{1}{2}}, 1; R) \lesssim R^{\frac{1}{p} - \frac{1}{3}}, \qquad p > 4; \tag{2.7}$$

$$BD_p(\mu_1, \mu_2, 1; R) \lesssim 1, \qquad p \ge 4, \qquad \forall\ \mu_1, \mu_2 > 1. \tag{2.8}$$

Let K be a large dyadic integer satisfying $K \sim \log R$. Divide $[0,1]^3$ into rectangular boxes τ of the form $[a_1, a_1 + K^{-1}] \times [a_2, a_2 + K^{-1/2}] \times [0,1]$, and write $f = \sum_\tau f_\tau$, where $f_\tau = f\chi_\tau$. For each $x \in B_R$, we define its *significant set* as

$$\mathcal{S}(x) := \Big\{ \tau : |E^{\mathbf{Q}} f_\tau(x)| \ge \frac{1}{100\#\{\tau\}} |E^{\mathbf{Q}} f(x)| \Big\}.$$

We say x is *broad* if there exist $\tau_1, \tau_2 \in \mathcal{S}(x)$ such that for any $\xi = (\xi_1, \xi_2, \xi_3) \in \tau_1$, $\eta = (\eta_1, \eta_2, \eta_3) \in \tau_2$,

$$|\xi_1 - \eta_1| \ge 10K^{-1}, \quad |\xi_2 - \eta_2| \ge 10K^{-\frac{1}{2}}. \tag{2.9}$$

Otherwise, we say x is *narrow*. Via the classical broad-narrow argument, we get

$$|E^{\mathbf{Q}} f(x)| \lesssim K^{\frac{3}{2}} \sup_{\tau_1, \tau_2 \text{ satisfy } (2.9)} |E^{\mathbf{Q}} f_{\tau_1}(x) E^{\mathbf{Q}} f_{\tau_2}(x)|^{\frac{1}{2}} + \Big| \sum_{\tau \in \mathcal{S}(x)} E^{\mathbf{Q}} f_\tau(x) \Big|,$$

where each $\tau \in \mathcal{S}(x)$ lies in one rectangular box with dimensions $\sim K^{-1} \times 1 \times 1$ or $1 \times K^{-1/2} \times 1$. We integrate over B_R on both sides to obtain

$$\begin{aligned}\|E^{\mathbf{Q}} f(x)\|_{L^p(B_R)} \lesssim & K^{\frac{3}{2}} \sum_{\tau_1,\tau_2 \text{ satisfy } (2.9)} \Big\| |E^{\mathbf{Q}} f_{\tau_1}(x) E^{\mathbf{Q}} f_{\tau_2}(x)|^{\frac{1}{2}} \Big\|_{L^p(B_R)} \\ & + \Big\| \sum_{\tau \in \mathcal{S}(x)} E^{\mathbf{Q}} f_\tau(x) \Big\|_{L^p(B_R)}.\end{aligned}$$

By Definition 2.1, this in fact implies that

$$D_p(R) \lesssim K^{\frac{9}{2}} B D_p(K, K^{\frac{1}{2}}, 1; R) + D_p(K, 1, 1; R) + D_p(1, K^{\frac{1}{2}}, 1; R). \tag{2.10}$$

For the broad part, i.e., the first bilinear term in (2.10), (2.8) gives one good bound for $p \geq 4$. So it suffices to consider the narrow part, i.e., the last two linear terms in (2.10). We take $D_p(K, 1, 1; R)$ as an example. Suppose we plan to use rescaling and induction. To ensure that the surface remains unchanged after rescaling, we have to use decoupling to reduce $D_p(K, 1, 1; R)$ to $D_p(K, K^{1/2}, 1; R)$ in the first place. Unfortunately, the loss of decoupling makes our argument fail when p is close to 4.

However, Guo-Oh considered the following new proposition instead:

$$D_p(L, 1, 1; R) \leq C_p L^{-\epsilon}, \tag{2.11}$$

for every $K \leq L \leq R^{1/2}$. Note that (2.6) is just the base case when $L = R^{1/2}$. Guo-Oh applied a further backward induction on L to deal with (2.11). More precisely, they used the broad-narrow argument again to get

$$D_p(L, 1, 1; R) \lesssim K^{\frac{9}{2}} B D_p(KL, K^{\frac{1}{2}}, 1; R) + D_p(KL, 1, 1; R) + D_p(L, K^{\frac{1}{2}}, 1; R). \tag{2.12}$$

For the bilinear term in (2.12), they can still apply (2.8). For the term $D_p(KL, 1, 1; R)$ in (2.12), they can use induction on L. For the term $D_p(L, K^{\frac{1}{2}}, 1; R)$ in (2.12), they made use of the anisotropic rescaling

$$\xi_1 \to \frac{\xi_1}{K}, \quad \xi_2 \to \frac{\xi_2}{K^{1/2}}, \quad \xi_3 \to \xi_3, \tag{2.13}$$

and then got

$$D_p(L, K^{\frac{1}{2}}, 1; R) \lesssim K^{\frac{6}{p} - \frac{3}{2}} D_p\Big(\frac{L}{K}, 1, 1; R\Big).$$

Now if $L/K \geq K$, then they can use the induction hypothesis of (2.11). If $L/K \leq K$, then by using (2.4) and cutting B_R into smaller balls $B_{R/2}$, they had

$$D_p\Big(\frac{L}{K}, 1, 1; R\Big) \leq D_p(1, 1, 1; R) \lesssim D_p\Big(1, 1, 1; \frac{R}{2}\Big).$$

Finally, through an induction on R of (2.5), they proved (2.11) for $p > 4$, which of course implied (2.5) for $p > 4$.

Before simplifying their proof, we first rewrite it in another manner, while lacking of rigour. We start from (2.10). For the narrow part, we use the broad-narrow analysis again to get

$$D_p(K,1,1;R) \lesssim K^{\frac{9}{2}} BD_p(K^2,K^{\frac{1}{2}},1;R) + D_p(K^2,1,1;R) + D_p(K,K^{\frac{1}{2}},1;R),$$

and

$$D_p(1,K^{\frac{1}{2}},1;R) \lesssim K^{\frac{9}{2}} BD_p(K,K,1;R) + D_p(K,K^{\frac{1}{2}},1;R) + D_p(1,K,1;R).$$

Thus (2.10) becomes

$$\begin{aligned} D_p(R) \lesssim & K^{\frac{9}{2}} \sup_{\mu_1,\mu_2>1} BD_p(\mu_1,\mu_2,1;R) + D_p(K^2,1,1;R) \\ & + D_p(1,K,1;R) + D_p(K,K^{\frac{1}{2}},1;R). \end{aligned} \tag{2.14}$$

For the first bilinear term in (2.14), (2.8) offers a good bound. For the second and third terms in (2.14), in view of (2.6) and (2.7), we can use induction on the first and second parameters respectively. For the last term in (2.14), we can use rescaling as in (2.13) and induction on R.

Our idea is to reduce the original term $D_p(R)$ to only three terms (2.6)-(2.8). Note that only the last term $D_p(K,K^{\frac{1}{2}},1;R)$ in (2.14) relies on rescaling. But if we apply the broad-narrow analysis to it rather than rescaling, then we will obtain

$$\begin{aligned} D_p(K,K^{\frac{1}{2}},1;R) &\lesssim K^{\frac{9}{2}} BD_p(K^2,K,1;R) + D_p(K^2,K^{\frac{1}{2}},1;R) + D_p(K,K,1;R) \\ &\le K^{\frac{9}{2}} BD_p(K^2,K,1;R) + D_p(K^2,1,1;R) + D_p(1,K,1;R), \end{aligned} \tag{2.15}$$

where we used (2.4) in the second line. Note that the last two terms in (2.15) have occurred in (2.14). In other words, the term $D_p(K,K^{\frac{1}{2}},1;R)$ which seems to require both rescaling and induction can actually be reduced to terms that only need induction. If we continue iterating such steps, only three terms (2.6)-(2.8) will remain. This alternative argument does not rely heavily on the specific expression of $\mathbf{Q}$, since we do not make use of any rescaling. So we can just use uniform decomposition instead in the first place, which can be easily adapted to general cases.

Firstly, by the ϵ-removal argument in [17, Section 4], we can reduce our aim (2.5) to: For every $\epsilon > 0$, we have

$$D_p(R) \le C_p R^{\epsilon}, \qquad p \ge 4,$$

holds for every $R \ge 1$. Divide $[0,1]^3$ into cubes τ of K^{-1}-scale, and write $f = \sum_\tau f_\tau$, where $f_\tau = f\chi_\tau$. By a standard broad-narrow analysis as before, we have

$$D_p(R) \lesssim K^9 BD_p(K,K,1;R) + D_p(K,1,1;R) + D_p(1,K,1;R). \tag{2.16}$$

Now if we apply the broad-narrow analysis again to the last two terms in (2.16), then

$$D_p(K,1,1;R) \lesssim K^9 BD_p(K^2,K,1;R) + D_p(K^2,1,1;R) + D_p(K,K,1;R),$$

$$D_p(1,K,1;R) \lesssim K^9 BD_p(K,K^2,1;R) + D_p(1,K^2,1;R) + D_p(K,K,1;R).$$

Note that there is a common term $D_p(K,K,1;R)$ above, and we further apply the broad-narrow analysis to it as follows:

$$\begin{aligned} D_p(K,K,1;R) &\lesssim K^9 BD_p(K^2,K^2,1;R) + D_p(K^2,K,1;R) + D_p(K,K^2,1;R) \\ &\leq K^9 BD_p(K^2,K^2,1;R) + D_p(K^2,1,1;R) + D_p(1,K^2,1;R), \end{aligned}$$

where we used (2.4) in the second line. Putting all these estimates together, we can conclude that

$$D_p(R) \lesssim K^9 \sup_{\mu_1,\mu_2>1} BD_p(\mu_1,\mu_2,1;R) + D_p(K^2,1,1;R) + D_p(1,K^2,1;R).$$

Now we can iterate this process to obtain

$$D_p(R) \lessapprox K^9 \sup_{\mu_1,\mu_2>1} BD_p(\mu_1,\mu_2,1;R) + D_p(R^{\frac{1}{2}},1,1;R) + D_p(1,R^{\frac{1}{2}},1;R). \tag{2.17}$$

Note that all the three terms in (2.17) can be covered by the estimates (2.6)-(2.8), so we have completed our proof.

As an application of the idea, in the remainder of this section, we reprove: For every $\epsilon > 0$, we have

$$\|E^{\mathbf{Q}} f\|_{L^p(B_R^5)} \lesssim R^{\epsilon} \|f\|_{L^p(\mathbb{R}^3)}, \qquad p \geq 4, \tag{2.18}$$

for every $R \geq 1$, where Q satisfies *the* (CM) *condition* (defined below). This result has been obtained through the multilinear method and the k-linear method in [17, 19, 15].

Definition 2.2 ([19])**.** *Given a tuple* $\mathbf{Q}(\xi) = (Q_1(\xi), ..., Q_n(\xi))$ *of real quadratic forms with* $\xi \in \mathbb{R}^d$*, we define*

$$\mathrm{NV}(\mathbf{Q}) := \#\{1 \leq d' \leq d : \partial_{\xi_{d'}} Q_{n'} \not\equiv 0 \text{ for some } 1 \leq n' \leq n\}.$$

For $0 \leq d' \leq d$ *and* $0 \leq n' \leq n$*, we define*

$$\mathfrak{d}_{d',n'}(\mathbf{Q}) := \inf_{\substack{M \in \mathbb{R}^{d \times d} \\ rank(M) = d'}} \inf_{\substack{M' \in \mathbb{R}^{n \times n'} \\ rank(M') = n'}} \mathrm{NV}((\mathbf{Q} \circ M) \cdot M'), \tag{2.19}$$

where $\mathbf{Q} \circ M$ *is the composition of* $\mathbf{Q}$ *and* M*. We say that* $\mathbf{Q}$ *and* $\mathbf{Q}'$ *are equivalent and write* $\mathbf{Q} \equiv \mathbf{Q}'$ *if there exist two invertible real matrices* $M_1 \in \mathbb{R}^{d \times d}$ *and* $M_2 \in \mathbb{R}^{n \times n}$ *such that*

$$\mathbf{Q}'(\xi) = \mathbf{Q}(M_1 \cdot \xi) \cdot M_2, \qquad \forall\, \xi \in \mathbb{R}^d.$$

Definition 2.3 ([11, 12, 28]). *Given a tuple $\mathbf{Q}(\xi) = (Q_1(\xi), ..., Q_n(\xi))$ of quadratic forms with $\xi \in \mathbb{R}^d$, we say that $\mathbf{Q}$ satisfies the* (CM) *condition if*

$$\int_{\mathbb{S}^{n-1}} |\det(y_1 Q_1 + ... + y_n Q_n)|^{-\gamma} d\sigma(y) < \infty, \qquad \forall\ 0 < \gamma < \frac{n}{d}, \tag{2.20}$$

where each Q_j, $j = 1, ..., n$, denotes the matrix associated with the quadratic form $Q_j(\xi)$, and $d\sigma$ is the surface measure on $\mathbb{S}^{n-1}$.

How to establish a relationship between these two conditions is an interesting question. When $d = 3$ and $n = 2$, Guo-Oh essentially proved that if $\mathbf{Q}$ satisfies $\mathfrak{d}_{3,2}(\mathbf{Q}) = 3$, then

$$\mathbf{Q} \text{ satisfies the (CM) condition} \iff \mathfrak{d}_{3,1}(\mathbf{Q}) = 2,\ \mathfrak{d}_{2,2}(\mathbf{Q}) = 2. \tag{2.21}$$

Readers can also see [10, Theorem 1.1] for reference. Now we start to show (2.18) by our method illustrated before. Let $\mathbf{Q} = (Q_1, Q_2)$ be a quadratic form satisfying the (CM) condition and $\mathfrak{d}_{3,2}(\mathbf{Q}) = 3$. Then (2.21) implies that $\mathfrak{d}_{3,1}(\mathbf{Q}) = 2$. We pick $M \in \mathbb{R}^{3\times 3}$ and $M' \in \mathbb{R}^{2\times 1}$ such that the equality in (2.19) is achieved with $d' = 3, n' = 1$. By linear transformations, we can assume that $M = I_{3\times 3}$ and $M' = (1, 0)^T$. Then $\mathfrak{d}_{3,1}(\mathbf{Q}) = 2$ implies that Q_1 depends on 2 variables. We use linear transformations again to diagonalize Q_1, then

$$(Q_1(\xi), Q_2(\xi)) \equiv (\xi_1^2 \pm \xi_2^2, b_{11}\xi_1^2 + b_{22}\xi_2^2 + b_{33}\xi_3^2 + 2b_{12}\xi_1\xi_2 + 2b_{13}\xi_1\xi_3 + 2b_{23}\xi_2\xi_3).$$

It suffices to consider the case $Q_1(\xi) = \xi_1^2 + \xi_2^2$, since the argument of the other case is similar. Moreover, $\mathfrak{d}_{3,2}(\mathbf{Q}) = 3$ implies that b_{13}, b_{23} and b_{33} are not simultaneously zero. For the same reason, we assume $b_{33} \neq 0$. By a change of variables in ξ_3, we obtain

$$(Q_1(\xi), Q_2(\xi)) \equiv (\xi_1^2 + \xi_2^2, b_{11}\xi_1^2 + b_{22}\xi_2^2 + 2b_{12}\xi_1\xi_2 + \xi_3^2). \tag{2.22}$$

We now prove a bilinear restriction estimate with favourable dependence on transversality parameters. Although the proof is similar to Guo-Oh's, for completeness and later convenience, we still sketch the argument.

Lemma 2.4. *Let $\mathbf{Q}$ be defined by* (2.22). *Suppose that f_i, $i = 1, 2$, are functions satisfying* $\operatorname{supp} f_i \subset [0,1] \times [b_i, b_i + \mu_2^{-1}] \times [c_i, c_i + \mu_3^{-1}] \subset [0,1]^3$ *with*

$$|b_1 - b_2| \geq 10\mu_2^{-1},\ |c_1 - c_2| \geq 10\mu_3^{-1}.$$

Then for every $p \geq 4$, we have

$$\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^p(\mathbb{R}^5)} \lesssim_p (\mu_2\mu_3)^{\frac{4}{p}-1} \|f_1\|_{L^p}^{\frac{1}{2}} \|f_2\|_{L^p}^{\frac{1}{2}}. \tag{2.23}$$

Proof. Note that for $p = \infty$, by Hölder's inequality, we have the trivial estimate

$$\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^\infty(\mathbb{R}^5)} \leq \|f_1\|_{L^1}^{\frac{1}{2}} \|f_2\|_{L^1}^{\frac{1}{2}} \leq \mu_2^{-1}\mu_3^{-1} \|f_1\|_{L^\infty}^{\frac{1}{2}} \|f_2\|_{L^\infty}^{\frac{1}{2}}.$$

Thus, by interpolation, we only need to show the case $p=4$. Since

$$\Big\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \Big\|_{L^4(\mathbb{R}^5)}^4 = \int |E^{\mathbf{Q}} f_1(x) E^{\mathbf{Q}} f_2(x)|^2 dx, \tag{2.24}$$

we can write

$$E^{\mathbf{Q}} f_1(x) E^{\mathbf{Q}} f_2(x) = \int\int f_1(\xi) f_2(\xi') \cdot \\ e\Big[x'(\xi+\xi') + x_4(Q_1(\xi)+Q_1(\xi')) + x_5(Q_2(\xi)+Q_2(\xi'))\Big] d\xi d\xi', \tag{2.25}$$

with $x' = (x_1,x_2,x_3)$, $\xi=(\xi_1,\xi_2,\xi_3)$ and $\xi'=(\xi_1',\xi_2',\xi_3')$. Apply the change of variables

$$\begin{cases} \eta = \xi+\xi'; \\ \eta_4 = Q_1(\xi)+Q_1(\xi'); \\ \eta_5 = Q_2(\xi)+Q_2(\xi'); \\ \eta_6 = \xi_1, \end{cases} \tag{2.26}$$

where $\eta=(\eta_1,\eta_2,\eta_3)$. Direct computation shows that the Jacobian J of the change of variables (2.26) is $4|\xi_2-\xi_2'||\xi_3-\xi_3'|$. Then (2.24) becomes:

$$\begin{aligned} \Big\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \Big\|_{L^4(\mathbb{R}^5)}^4 &= \int \Big| \int\int g_1(\eta) g_2(\eta) e(x\cdot\eta') J^{-1} d\eta' d\eta_6 \Big|^2 dx \\ &\le \int\int |g_1|^2 |g_2|^2 J^{-2} \\ &\le \int\int |f_1|^2 |f_2|^2 J^{-1} \\ &\lesssim \mu_2\mu_3 \|f_1\|_{L^2}^2 \|f_2\|_{L^2}^2 \\ &\le \|f_1\|_{L^4}^2 \|f_2\|_{L^4}^2, \end{aligned}$$

where g_1 and g_2 are appropriate functions, and $\eta' = (\eta,\eta_4,\eta_5)$. Here in the second line we first use Hölder's inequality in the variable η_6 and then apply the Plancherel theorem in the variables η', in the third line we change variables back, and in the last line we use Hölder's inequality to exploit the support condition of f_1 and f_2. This finishes our proof.

□

Inspired by Lemma 2.4, we consider the following type of broad-narrow analysis. Let $D_p(R)$ be the smallest constant such that (2.18) holds, and let K be a large dyadic integer satisfying $K \sim \log R$. Divide $[0,1]^3$ into cubes τ of K^{-1}-scale, and write $f=\sum_\tau f_\tau$, where $f_\tau = f\chi_\tau$. For each $x\in B_R$, we define its *significant set* as

$$\mathcal{S}(x) := \Big\{ \tau : |E^{\mathbf{Q}} f_\tau(x)| \ge \frac{1}{100\#\{\tau\}} |E^{\mathbf{Q}} f(x)| \Big\}.$$

We say x is *broad* if there exist $\tau_1, \tau_2 \in \mathcal{S}(x)$ such that for any $\xi = (\xi_1, \xi_2, \xi_3) \in \tau_1$, $\eta = (\eta_1, \eta_2, \eta_3) \in \tau_2$

$$|\xi_2 - \eta_2| \geq 10K^{-1}, \quad |\xi_3 - \eta_3| \geq 10K^{-1}. \tag{2.27}$$

Otherwise, we say x is *narrow*. Via the classical broad-narrow argument, we get

$$\begin{aligned}\|E^{\mathbf{Q}} f(x)\|_{L^p(B_R)} \lesssim & K^3 \sum_{\tau_1,\tau_2 \text{ satisfy } (2.27)} \Big\| |E^{\mathbf{Q}} f_{\tau_1}(x) E^{\mathbf{Q}} f_{\tau_2}(x)|^{\frac12} \Big\|_{L^p(B_R)} \\ & + \Big\| \sum_{\tau \in \mathcal{S}(x)} E^{\mathbf{Q}} f_\tau(x) \Big\|_{L^p(B_R)},\end{aligned}$$

where each $\tau \in \mathcal{S}(x)$ lies in one rectangular box with dimensions $\sim 1 \times K^{-1} \times 1$ or $1 \times 1 \times K^{-1}$. This implies that

$$D_p(R) \lesssim K^9 BD_p(1, K, K; R) + D_p(1, K, 1; R) + D_p(1, 1, K; R). \tag{2.28}$$

Iterating this formula as in (2.17), we obtain

$$D_p(R) \lessapprox K^9 \sup_{\mu_2, \mu_3 > 1} BD_p(1, \mu_2, \mu_3; R) + D_p(1, R^{\frac12}, 1; R) + D_p(1, 1, R^{\frac12}; R). \tag{2.29}$$

Note Lemma 2.4 gives a good bound on the bilinear term in (2.29). So it suffices to consider the last two terms in (2.29). Take $D_p(1, R^{\frac12}, 1; R)$ as an example. Since $\mathbf{Q}$ satisfies the (CM) condition, we have the following sharp Stein-Tomas type inequality (see [12, 28]):

$$\|E^{\mathbf{Q}} f\|_{L^{\frac{14}{3}}(B_R)} \lesssim \|f\|_{L^2}.$$

On the other hand, we have the trivial L^2 estimate:

$$\|E^{\mathbf{Q}} f\|_{L^2(B_R)} \lesssim R\|f\|_{L^2}.$$

By interpolation, we get

$$\|E^{\mathbf{Q}} f\|_{L^p(B_R)} \lesssim R^{\frac{7}{2p} - \frac34} \|f\|_{L^2},$$

for $2 \leq p \leq 14/3$. Thus for any function f with $\operatorname{supp} f \subset [0,1] \times [b, b + R^{-1/2}] \times [0,1] \subset [0,1]^3$, we can apply Hölder's inequality to obtain

$$\|E^{\mathbf{Q}} f\|_{L^p(B_R)} \lesssim R^{\frac{7}{2p} - \frac34} R^{-\frac12(\frac12 - \frac1p)} \|f\|_{L^p} = R^{\frac4p - 1} \|f\|_{L^p}.$$

This just tells us that

$$D_p(1, R^{\frac12}, 1; R) \lesssim R^{\frac4p - 1} \lesssim 1,$$

for $4 \leq p \leq 14/3$. Therefore, we complete the proof of (2.18).

3. A variant of the broad-narrow analysis

In this section, we aim to extend Guo-Oh's method. But before that, we shall first revisit the case $\mathbf{Q} = (\xi_1^2, \xi_2^2 + \xi_1\xi_3)$. In the previous section, we have discussed the broad-narrow analysis based on the transversality condition (2.9) in view of the example $\chi_{[0,R^{-1}]\times[0,R^{-1/2}]\times[0,1]}$, which yields the necessary condition $p > 4$. However, by some simple computations, we find that $\chi_{[0,R^{-1}]\times[0,1]\times[0,1]}$ can also give optimal necessary condition of the restriction estimate. A notable feature of $\chi_{[0,R^{-1}]\times[0,1]\times[0,1]}$ is that if we scale it to $\chi_{[0,1]^3}$, then the surface $S_{\mathbf{Q}}$ will change. This fact seems to indicate that Guo-Oh's frequency decomposition can not be well-suited. Nevertheless, recall that our simplified proof does not rely on rescaling, so it makes sense to ask: Can we further simplify Guo-Oh's proof by using the transversality condition induced by $\chi_{[0,R^{-1}]\times[0,1]\times[0,1]}$? To be more precise, if we use the broad-narrow analysis with a transversality condition in which only the first coordinate is separated, then we will get

$$D_p(R) \lesssim K^9 BD_p(K,1,1;R) + D_p(K,1,1;R). \tag{3.1}$$

Iterating this process as we did in (2.17) yields

$$D_p(R) \lessapprox K^9 \sup_{\mu_1>1} BD_p(\mu_1,1,1;R) + D_p(R^{\frac{1}{2}},1,1;R). \tag{3.2}$$

This argument seems simpler than that in the previous section. Unfortunately, it can not give optimal range $p > 4$. This is because the bound of the bilinear term in (3.2) depends on the transversality parameter if p is close to 4. In fact, we can prove

$$BD_p(\mu_1,1,1;R) \lesssim_{\mu_1} 1, \qquad p \geq 4,$$

and

$$\sup_{\mu_1>1} BD_p(\mu_1,1,1;R) \lesssim 1, \qquad p \geq 5.$$

On the other hand, we record (2.8) here for comparison:

$$\sup_{\mu_1,\mu_2>1} BD_p(\mu_1,\mu_2,1;R) \lesssim 1, \qquad p \geq 4.$$

Though both bilinear restriction estimates hold for $p \geq 4$ by the L^4 bi-orthogonality method, the latter which is independent of the transversality parameter may fail when p is close to 4. The main reason for such a distinction is that the forms of the Jacobian J of the change of variables are different when we estimate the bilinear terms. In other words, the Jacobian J characterizes bilinear restriction estimates. We will soon write out this characterization explicitly as a theorem.

Given a tuple $\mathbf{Q}(\xi) = (Q_1(\xi), ..., Q_n(\xi))$ of real quadratic forms with $\xi \in \mathbb{R}^d$ and $d \geq n$, we define

$$J(\xi; i_1, ..., i_{d-n}) := \begin{vmatrix} \nabla \mathbf{Q} \\ e_{i_1} \\ \vdots \\ e_{i_{d-n}} \end{vmatrix}, \tag{3.3}$$

where $\nabla \mathbf{Q} = (\partial_k Q_j)_{n \times d}$ is a matrix of size $n \times d$, and e_j denotes the j-th unit coordinate vectors in $\mathbb{R}^d$.

Theorem 3.1. *Let $\mathbf{Q}$ be an n-tuple of real quadratic forms defined on $\mathbb{R}^d$ with $d \geq n$. Suppose that there exist several parameters i_1,...,i_{d-n} such that*

$$|J(\xi; i_1, ..., i_{d-n})| \sim |\xi_1|^{w_1} |\xi_2|^{w_2} ... |\xi_t|^{w_t}, \tag{3.4}$$

where $t \leq d$, $w_j \in \mathbb{N} \backslash \{0\}$ for $1 \leq j \leq t$, and $\sum_{j=1}^{t} w_j = n$. Then we have

$$\sup_{\mu_1, ..., \mu_t > 1} BD_p(\mu_1, ..., \mu_t, 1, ..., 1; R) \lesssim 1, \tag{3.5}$$

holds for $p \geq \max_j w_j + 3$.

Proof. By Definition 2.1, we need to consider the bilinear restriction estimate

$$\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^p(\mathbb{R}^{d+n})} \lesssim \|f_1\|_{L^p}^{\frac{1}{2}} \|f_2\|_{L^p}^{\frac{1}{2}},$$

where f_1 and f_2 are supported on two separated rectangular boxes with dimensions $\mu_1^{-1} \times \cdots \times \mu_t^{-1} \times 1 \times \cdots \times 1$. When $p = \infty$, by Hölder's inequality, we trivially have

$$\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^\infty(\mathbb{R}^{d+n})} \leq \|f_1\|_{L^1}^{\frac{1}{2}} \|f_2\|_{L^1}^{\frac{1}{2}} \leq \mu_1^{-\frac{1}{2}} ... \mu_t^{-\frac{1}{2}} \|f_1\|_{L^2}^{\frac{1}{2}} \|f_2\|_{L^2}^{\frac{1}{2}}.$$

When $p = 4$, we write

$$E^{\mathbf{Q}} f_1(x) E^{\mathbf{Q}} f_2(x) = \int \int f_1(\xi) f_2(\xi') e\Big[x'(\xi + \xi') + x''(\mathbf{Q}(\xi) + \mathbf{Q}(\xi'))\Big] d\xi d\xi',$$

with $x' = (x_1, ..., x_d)$, $x'' = (x_{d+1}, ..., x_{d+n})$, $\xi = (\xi_1, .., \xi_d)$ and $\xi' = (\xi_1', ..., \xi_d')$. Apply the change of variables

$$\begin{cases} \eta' = \xi + \xi'; \\ \eta'' = \mathbf{Q}(\xi) + \mathbf{Q}(\xi'); \\ \eta_{d+n+1} = \xi_{d+i_1}; \\ \quad \vdots \\ \eta_{2d} = \xi_{d+i_{d-n}}, \end{cases} \tag{3.6}$$

where $\eta' = (\eta_1, ..., \eta_d)$ and $\eta'' = (\eta_{d+1}, ..., \eta_{d+n})$. Then we see that the Jacobian J of the change of variables is just $J(\xi' - \xi; i_1, ..., i_{d-n})$. Thus we can compute the L^4 norm as follows:

$$\begin{aligned}
\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^4(\mathbb{R}^{d+n})}^4 &= \int |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^2 \\
&= \int \left| \iint g_1(\eta) g_2(\eta) e(x\cdot\tilde{\eta}) J^{-1} d\tilde{\eta} d\overline{\eta} \right|^2 dx \\
&\le \iint |g_1|^2 |g_2|^2 J^{-2} \\
&\le \iint |f_1|^2 |f_2|^2 J^{-1} \\
&\lesssim \mu_1^{w_1} \cdots \mu_t^{w_t} \|f_1\|_{L^2}^2 \|f_2\|_{L^2}^2,
\end{aligned}$$

where g_1 and g_2 are appropriate functions, and $\tilde{\eta} = (\eta_1, ..., \eta_{d+n})$, $\overline{\eta} = (\eta_{d+n+1}, , , .\eta_{2d})$. Here in the third line we first use Hölder's inequality in the variables $\overline{\eta}$ and then apply the Plancherel theorem in the variables $\tilde{\eta}$, in the fourth line we change variables back, and in the last line we apply the assumption (3.4). It follows that

$$\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^4(\mathbb{R}^{d+n})} \lesssim \mu_1^{\frac{w_1}{4}} \cdots \mu_t^{\frac{w_t}{4}} \|f_1\|_{L^2}^{\frac{1}{2}} \|f_2\|_{L^2}^{\frac{1}{2}}.$$

Through Hölder's inequality, we can conclude that for $4 \le p \le \infty$,

$$\begin{aligned}
\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^p(\mathbb{R}^{d+n})} &\lesssim \mu_1^{\frac{w_1+2}{p}-\frac{1}{2}} \cdots \mu_t^{\frac{w_t+2}{p}-\frac{1}{2}} \|f_1\|_{L^2}^{\frac{1}{2}} \|f_2\|_{L^2}^{\frac{1}{2}} \\
&\lesssim \mu_1^{\frac{w_1+3}{p}-1} \cdots \mu_t^{\frac{w_t+3}{p}-1} \|f_1\|_{L^p}^{\frac{1}{2}} \|f_2\|_{L^p}^{\frac{1}{2}}.
\end{aligned}$$

As long as we take $p \ge \max_j w_j + 3$, by noting that $\mu_j > 1$ for all $1 \le j \le t$, we have

$$\left\| |E^{\mathbf{Q}} f_1 E^{\mathbf{Q}} f_2|^{\frac{1}{2}} \right\|_{L^p(\mathbb{R}^{d+n})} \lesssim \|f_1\|_{L^p}^{\frac{1}{2}} \|f_2\|_{L^p}^{\frac{1}{2}}$$

as desired. □

Note that the range of p in Theorem 3.1 is determined by $J(\xi; i_1, ..., i_{d-n})$, which relies on the choices of $i_1, ..., i_{d-n}$. In particular, for $\mathbf{Q} = (\xi_1^2, \xi_2^2 + \xi_1\xi_3)$, we have

$$|J(\xi; 2)| \sim |\xi_1|^2, \quad |J(\xi; 3)| \sim |\xi_1||\xi_2|.$$

For the former, it gives the range $p \ge 5$ in (3.5); for the latter, it gives the range $p \ge 4$ in (3.5). Therefore, to obtain optimal bilinear estimates independent of the transversality parameters, we need to locate suitable positions $i_1, ..., i_{d-n}$ such that the maximal power of the factors in $J(\xi; i_1, ..., i_{d-n})$ is minimized. We will expand on this point in the next section.

Now we sketch the key steps of our approach. Given a tuple $\mathbf{Q}(\xi) = (Q_1(\xi), ..., Q_n(\xi))$ of quadratic forms with $\xi \in \mathbb{R}^d$ and $d \ge n$, our aim is to prove: For every $\epsilon > 0$, we have

$$D_p(R) \le C_p R^{\epsilon}, \qquad \forall\ R \ge 1,$$

for a maximal range of p.

Firstly, we consider the bilinear estimate as in Theorem 3.1. By choosing appropriate positions $i_1, ..., i_{d-n}$ such that the maximal power of the factors in $J(\xi; i_1, ..., i_{d-n})$ is minimized, we can determine suitable transversality condition. Without loss of generality, assume that we have transversality when the first coordinate is separated. Then we can use the classical broad-narrow analysis to get

$$D_p(R) \lesssim K^{3d} BD_p(K, 1, ..., 1; R) + D_p(K, 1, ..., 1; R). \tag{3.7}$$

By iterating this formula, we obtain

$$D_p(R) \lessapprox K^{3d} \sup_{\mu_1 > 1} BD_p(\mu_1, 1, ..., 1; R) + D_p(R^{\frac{1}{2}}, 1, ..., 1; R). \tag{3.8}$$

Similar to (3.7) which is derived by the classical broad-narrow analysis, (3.8) also contains two terms. However, the first term in (3.8) includes more information than the broad part in (3.7), so we call it the *more-broad* part; while the second term in (3.8) includes less information than the narrow part in (3.7), so we call it the *less-narrow* part.

In applications, (3.7) and (3.8) each has its advantages and disadvantages. On the one hand, the broad part in (3.7) can be better controlled than the more-broad part in (3.8). On the other hand, for the narrow part in (3.7), due to $K \ll R$, rescaling and induction on scales seem to be the only way to utilize the lower dimensional information, while there are more freedom for the less-narrow part in (3.8) because we can apply many other tools to deal with it, such as the locally constant property and decoupling theory. Since adaptive decomposition is effective in the study of degenerate cases, considering (3.8) rather than (3.7) can offer great convenience. Nonetheless, we do not know beforehand which one is better for a specific situation. In fact, in the next section, we will use both types of the broad-narrow analysis to handle different cases in Theorem 1.1.

The idea of iteration has already appeared in many papers, such as [9, 19, 24, 36, 37]. In these works, the main advantage of iteration, compared with induction, is the potential for producing many different scales which can be combined to facilitate more delicate geometric analysis. However, the style and the effect of iteration in this paper are somewhat different, and now we take two examples to illustrate such difference.

The first example is decoupling for the parabola due to Bourgain-Demeter [6]. Via the classical broad-narrow analysis, they proved an estimate similar to (3.7) (in what follows we abuse $D_p(R)$ and $BD_p(K; R)$ to denote the optimal constants of decoupling and bilinear decoupling). By parabolic rescaling, the narrow part in (3.7) can be reduced to $D_p(R/K^2)$. Then they view R/K^2 as a new scale and repeat the broad-narrow analysis. Finally, they got

$$D_p(R) \lessapprox K^{3d} \sup_{1\le R'\le R} BD_p(K;R') + D_p(1) \lessapprox K^{3d} \sup_{1\le R'\le R} BD_p(K;R'). \quad (3.9)$$

In contrast to [6], our setting is incompatible with rescaling, so we have to repeat the broad-narrow analysis to $D_p(K,1,...,1;R)$ rather than just $D_p(1,...,1;R/K^2)$. This twist leads to two consequences: The first is that the transversality parameter becomes larger and larger in the bilinear term, which forces us into considering bilinear estimates independent of the transversality parameter; the second is that our iteration keeps going until we reach the scale $R^{1/2}$, instead of the scale 1 in (3.9).

The second example is Fourier restriction for the paraboloid in [36, 37], in which the polynomial partitioning is iterated at many smaller scales to derive estimates on B_R. These multiple scales are then effectively related through additional incidence geometry arguments, such as *brooms* and refined *hairbrushs*. Though we also use iteration, our aim is different from that in [36, 37]. We actually further divide the narrow part at K^{-1}-scale in (3.7) into broad/narrow parts at smaller scales and continue in this way until we successfully reduce the original broad part and narrow part in (3.7) to the more-broad part and less-narrow part in (3.8). Compared with the narrow part, the less-narrow part seems to be "genuine" lower dimensional contribution, since we can simply drop the term ξ_1^2 in $\mathbf{Q}$ by the locally constant property when we study $D_p(R^{\frac{1}{2}},1,...,1;R)$.

4. The proof of Theorem 1.1

In this section, we start to prove Theorem 1.1. By the trivial $L^1 \to L^\infty$ estimate and interpolation, it suffices to prove the local version of (1.1) when $p=q$, i.e.,

$$\|E^{\mathbf{Q}} f\|_{L^p(B_R^{d+n})} \lessapprox \|f\|_{L^p(\mathbb{R}^d)}.$$

Let $D_p(R)$ be the smallest constant for this estimate to hold. By the standard ϵ-removal argument, it suffices to show that for every $\epsilon > 0$, we have

$$D_p(R) \le C_p R^\epsilon, \quad (4.1)$$

for each $R \ge 1$. From now on, we abbreviate $D_p(\mu_1,...,\mu_d;R)$ to $D_p(\mu_1,...,\mu_d)$, and $BD_p(\mu_1,...,\mu_d;R)$ to $BD_p(\mu_1,...,\mu_d)$, as we will always fix one large scale R. Let K be a large dyadic integer satisfying $K \sim \log R$.

For our purposes, we introduce several classical decoupling results.

Lemma 4.1 (ℓ^p decoupling for the paraboloid, [6])**.** *Let* $F:\mathbb{R}^m \to \mathbb{C}$ *be such that* $\mathrm{supp}\widehat{F} \subset N_{R^{-1}}(\mathbb{P}^{m-1})$. *Divide this neighborhood into slabs* θ *with* $m-1$ *long directions of length* $R^{-1/2}$ *and one short direction of length* R^{-1}. *Write* $F = \sum_\theta F_\theta$, *where* $\widehat{F}_\theta = \widehat{F}\chi_\theta$. *Then*

$$\|F\|_{L^p(\mathbb{R}^m)} \lesssim_\epsilon R^{\beta(p)+\epsilon} \Big(\sum_\theta \|F_\theta\|_{L^p(\mathbb{R}^m)}^p\Big)^{\frac{1}{p}}, \quad (4.2)$$

where $\beta(p) = \frac{m-1}{4} - \frac{m-1}{2p}$ *when* $2 \le p \le \frac{2(m+1)}{m-1}$ *and* $\beta(p) = \frac{m-1}{2} - \frac{m}{p}$ *when* $\frac{2(m+1)}{m-1} \le p \le \infty$.

Lemma 4.2 (Flat decoupling,[8, 23])**.** *Let R be a rectangular box in $\mathbb{R}^m$, and $R_1, ..., R_L$ be a partition of R into congruent boxes that are translates of each other. Write $F = \sum_j F_j$, where $\widehat{F}_j = \widehat{F}\chi_{B_j}$. Then for any $2 \le p, q \le \infty$, we have*

$$\|F\|_{L^p(\mathbb{R}^m)} \lesssim L^{1-\frac{1}{p}-\frac{1}{q}} \Big(\sum_j \|F_j\|^q_{L^p(\mathbb{R}^m)} \Big)^{\frac{1}{q}}. \tag{4.3}$$

(1) Firstly, we compute

$$|J(\xi)| \sim |\xi_1|^{w_1} |\xi_2|^{w_2} \cdots |\xi_n|^{w_n}.$$

Without loss of generality, we may assume that $w_j > 0$ for every $1 \le j \le n$, since other cases can be proved via similar arguments. We use the broad-narrow analysis and iteration to obtain

$$D_p(R) \lessapprox K^{3n} \sup_{\mu_1,...,\mu_n > 1} BD_p(\mu_1, ..., \mu_n) + \sum_{j=1}^n D_p(\underbrace{1, ..., 1}_{j-1}, R^{\frac{1}{2}}, 1..., 1). \tag{4.4}$$

For the first bilinear term in (4.4), we use Theorem 3.1 to conclude that

$$\sup_{\mu_1,...,\mu_n > 1} BD_p(\mu_1, ..., \mu_n) \lesssim 1, \qquad p \ge \max_j w_j + 3.$$

For the remaining linear terms in (4.4), we take the j-th term as an example. Since every Q_i in $\mathbf{Q}$ is a monomial, the related surface $S_{\mathbf{Q}}$ always stays unchanged after any rescaling. By the rescaling

$$\xi_j \to \frac{\xi_j}{R^{1/2}},$$

we obtain

$$D_p(\underbrace{1, ..., 1}_{j-1}, R^{\frac{1}{2}}, 1..., 1) \le R^{\frac{w_j+3}{2p} - \frac{1}{2}} D_p(R).$$

Covering B_R with balls of scale $R/2$, and using induction on R, one gets

$$D_p(\underbrace{1, ..., 1}_{j-1}, R^{\frac{1}{2}}, 1..., 1) \lesssim R^{\frac{w_j+3}{2p} - \frac{1}{2} + \epsilon}. \tag{4.5}$$

Therefore, we close the induction on this term for $p > w_j + 3$. Combining the estimates for all terms, we validate (4.1) for $p > \max_j w_j + 3$. And then the estimate (4.1) for $p = \max_j w_j + 3$ is a direct corollary of Hölder's inequality.

Remark 4.3. *We point out that the case* (1) (*also* (3) *and* (4) *as below*) *can also be proved by the classical broad-narrow analysis, since the methods we apply in the narrow case are rescaling and induction on scales. In fact, if we use the classical broad-narrow analysis, we have a better range* $p \ge 4$

for the broad part. Nevertheless, our method seems to be more suitable for such degenerate cases. We will explain this point in Section 5.

(2) We first assume that the case (2a) holds, i.e., each P_j is a quadratic form that is independent of the variables $\xi_1, \xi_j, ..., \xi_n$; $2 \le \lambda \le w_1 + 1$, $w_1 \ge w_\lambda + \theta$. Note that in this case

$$|J(\xi)| \sim |\xi_1|^{w_1} |\xi_\lambda|^{w_\lambda}.$$

We use the broad-narrow analysis and iteration to obtain

$$\begin{aligned} D_p(R) \lessapprox & K^{3n} \sup_{\mu_1, \mu_\lambda > 1} BD_p(\underbrace{\mu_1, 1, ..., 1}_{\lambda - 1}, \mu_\lambda, 1, ..., 1) + D_p(R^{\frac{1}{2}}, 1, ..., 1) \\ & + D_p(\underbrace{1, ..., 1}_{\lambda - 1}, R^{\frac{1}{2}}, 1, ..., 1). \end{aligned} \tag{4.6}$$

For the first bilinear term in (4.6), we apply Theorem 3.1 to conclude that

$$\sup_{\mu_1, \mu_\lambda > 1} BD_p(\underbrace{\mu_1, 1, ..., 1}_{\lambda - 1}, \mu_\lambda, 1, ..., 1) \lessapprox 1, \qquad p \ge w_1 + 3,$$

due to $w_1 \ge w_\lambda$. For the second term in (4.6), since $\lambda \ne 1$ and each P_j is independent of the variable ξ_1, we can use rescaling in the variable ξ_1 as in (4), and get

$$D_p(R^{\frac{1}{2}}, 1, ..., 1) \lessapprox R^{\frac{w_1 + 3}{2p} - \frac{1}{2} + \epsilon}. \tag{4.7}$$

For the last term in (4.6), recall that $P_{w_1+1}(\xi), ..., P_n(\xi)$ depend on θ many variables except the variables ξ_1 and ξ_λ. We apply flat decoupling (4.3) to these variables from 1-scale to $R^{1/2}$-scale. Here by "R^α-scale"($\alpha > 0$) we mean that the frequency support in these variables is within $R^{-\alpha}$-scale, and we will adopt this convention from now on. After this, by the locally constant property, the original form $\mathbf{Q}$ can be reduced to

$$\mathbf{Q}_1 = (\xi_1^2, \xi_1\xi_2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_n),$$

which has been investigated in the case (1). We use rescaling and the result of the case (1) to get

$$D_p(\underbrace{1, ..., 1}_{\lambda - 1}, R^{\frac{1}{2}}, 1, ..., 1) \lessapprox R^{\frac{\theta}{2}(1 - \frac{2}{p}) + \frac{w_\lambda + 3\theta + 3}{2p} - \frac{1 + \theta}{2} + \epsilon} = R^{\frac{w_\lambda + \theta + 3}{2p} - \frac{1}{2} + \epsilon} \lessapprox 1, \tag{4.8}$$

for $p > w_1 + 3$ due to $w_1 \ge w_\lambda + \theta$. And the estimate (4.1) for $p = w_1 + 3$ is a direct corollary of Hölder's inequality.

Suppose that we are in the case (2b), i.e., each P_j is a quadratic form without mixed terms that is independent of the variables $\xi_j, ..., \xi_n$; $2 \le \lambda \le w_1 + 1$, $w_1 \ge w_\lambda + \theta$. Compared with the case (2a), now each P_j can depend on the variable ξ_1. The existence of the variable ξ_1 can make the induction argument for $D_p(R^{\frac{1}{2}}, 1, ..., 1)$ fail. But we can easily remove the terms depending on the variable ξ_1 in P_j by some calculations. Since each

P_j is a quadratic form without mixed terms, if P_j depends on the variable ξ_1, there must exist one monomial $c\xi_1^2$ in P_j for some constant c. Adding a multiple of ξ_1^2 in Q_1 to Q_j, we can remove the term $c\xi_1^2$ in P_j. This helps us to reduce the case (2b) to the case (2a), and so we are done.

Suppose that we are in the case (2c), i.e., P_j is a quadratic form without mixed terms that is independent of the variables $\xi_j, ..., \xi_n$ for some $w_1 + 1 \le j \le n$ and $P_{j'} = 0$ for all $j' \ne j$; $2 \le \lambda \le w_1 + 1$, $w_1 \ge w_\lambda + \theta/2$. Without loss of generality, we assume that $j = n$ and

$$\mathbf{Q} = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}, \xi_\lambda\xi_n + \xi_\alpha^2 + ... + \xi_\gamma^2),$$

with $1 \le \alpha, ..., \gamma \le n-1$. Through the same argument as in the case (2b), we can further assume that $2 \le \alpha, ..., \gamma \le n-1$. We use the broad-narrow analysis and iteration to obtain

$$\begin{aligned} D_p(R) \lessapprox & K^{3n} \sup_{\mu_1, \mu_\lambda > 1} BD_p(\underbrace{\mu_1, 1, ..., 1}_{\lambda-1}, \mu_\lambda, 1, ..., 1) + D_p(R^{\frac{1}{2}}, 1, ..., 1) \\ & + D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1). \end{aligned} \tag{4.9}$$

The first two terms of (4.9) can be shown via the same discussion as in the case (2a). Now we focus on the last term of (4.9). In this case, we can without loss of generality assume that ξ_λ does not appear in $\xi_\alpha, ..., \xi_\gamma$, because otherwise we can safely remove ξ_λ^2 from $\xi_\alpha^2 + ... + \xi_\gamma^2$ by the locally constant property. We apply flat decoupling (4.3) to the variables $\xi_\alpha, ..., \xi_\gamma$ from 1-scale to $R^{1/4}$-scale, and then use the change of variables

$$\xi_\alpha \to \frac{\xi_\alpha}{R^{1/4}}, \quad ..., \quad \xi_\gamma \to \frac{\xi_\gamma}{R^{1/4}}, \quad \xi_\lambda \to \frac{\xi_\lambda}{R^{1/2}}.$$

Notice that such rescaling keeps the surface $S_{\mathbf{Q}}$ invariant. Covering B_R with balls of scale $R/2$, and using induction on R, one concludes

$$D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1..., 1) \lesssim R^{\frac{w_\lambda + \theta/2 + 3}{2p} - \frac{1}{2} + \epsilon} \lesssim 1, \tag{4.10}$$

for $p > w_1 + 3$ due to $w_1 \ge w_\lambda + \theta/2$. And the estimate (4.1) for $p = w_1 + 3$ is a direct corollary of Hölder's inequality.

(3) Since each Q_j in $\mathbf{Q}$ is a monomial, the narrow part can be easily proved via rescaling and induction on scales as in the case (1). The difference between the two cases (1) and (3) is that now we need to choose appropriate $i_1, ..., i_k$ such that the maximal power of the factors in $J(\xi; i_1, ..., i_k)$ is minimized. There will be two cases.

Suppose that q_1 reaches the maximum when $j = 1 + k$ (this argument also works when $j = 2 + k$, ..., $n + k$). Take $i_1 = 1, ..., i_k = k$, then

$$|J(\xi; 1, ..., k)| \sim |\xi_{\lambda_{1+k}}||\xi_{\lambda_{2+k}}|...|\xi_{\lambda_{n+k}}| = |\xi_1|^{w_1}|\xi_2|^{w_2}...|\xi_{n+k}|^{w_{n+k}}.$$

Using Theorem 3.1, we get (3.5) holds for $p \ge w_{1+k} + 3 = q_1$.

Suppose that q_1 reaches the maximum when $j = 1$ (this argument also works when $j = 2, ..., k$). If $\lambda_{1+k} = 1$, then by taking $i_1 = 2, i_2 = 3, ..., i_k = 1 + k$, we have

$$\begin{aligned}|J(\xi; 2, ..., 1+k)| &\sim |\xi_{1+k}||\xi_{\lambda_{2+k}}|...|\xi_{\lambda_{n+k}}|\\ &= |\xi_1|^{w_1-1}|\xi_2|^{w_2}...|\xi_{1+k}|^{w_{1+k}+1}...|\xi_{n+k}|^{w_{n+k}}.\end{aligned}$$

Using Theorem 3.1, we get (3.5) holds for $p \geq w_1 - 1 + 3 = w_1 + 2 = q_1$. If $\lambda_{1+k} \neq 1$, then we can find a minimal η such that $\lambda_{\eta+k} = 1$ as we assume $w_1 \geq 1$. Take $i_1 = 2, , ..., i_{k-1} = k, i_k = \eta + k$, then

$$\begin{aligned}|J(\xi; 2, ..., k, \eta+k)| &\sim |\xi_{\lambda_{1+k}}|...|\xi_{\lambda_{\eta-1+k}}||\xi_{\eta+k}||\xi_{\lambda_{\eta+1+k}}|...|\xi_{\lambda_{n+k}}|\\ &= |\xi_1|^{w_1-1}|\xi_2|^{w_2}...|\xi_{\eta+k}|^{w_{\eta+k}+1}...|\xi_{n+k}|^{w_{n+k}}.\end{aligned}$$

Using Theorem 3.1, we get (3.5) holds for $p \geq w_1 - 1 + 3 = w_1 + 2 = q_1$. In fact, when we argue that $p \geq w_1 - 1 + 3$, we need $w_1 \geq w_{\eta+k} + 2$, not just $w_1 \geq w_{\eta+k} + 1$. This minor gap can be easily filled: For the critical case $w_1 = w_{\eta+k} + 1$, q_1 actually also reaches the maximum when $j = \eta + k$, so it can be safely covered by the first case. This finishes the proof of the case (3).

The argument for the case (4) is essentially the same as that for the case (3), so we omit the details.

(5) Though the conditions for (2a)-(2c) and (5a)-(5c) are similar, the case (5) is more difficult due to the presence of $\xi_{n+1}^2 + ... + \xi_{n+k}^2$.

Suppose that we are in the case (5a), i.e., each P_j is a quadratic form that is independent of the variables $\xi_1, \xi_j, ..., \xi_{n+k}$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta$. Take $i_1 = n, ..., i_k = n + k - 1$, then

$$|J(\xi; n, n+1, ..., n+k-1)| \sim |\xi_1|^{w_1}|\xi_\lambda|^{w_\lambda - 1}|\xi_{n+k}|.$$

We use the broad-narrow analysis and iteration to obtain

$$\begin{aligned}D_p(R) \lessapprox &K^{3(n+k)} \sup_{\mu_1, \mu_\lambda, \mu_{n+k} > 1} BD_p(\underbrace{\mu_1, 1, ..., 1}_{\lambda-1}, \mu_\lambda, 1, ..., 1, \mu_{n+k})\\ &+ D_p(R^{\frac{1}{2}}, 1, ..., 1) + D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1) + D_p(1, ..., 1, R^{\frac{1}{2}}).\end{aligned} \tag{4.11}$$

The first two terms in (4.11) can be proved via the same argument as in the case (2a). For the last term in (4.11), by the locally constant property, the original $\mathbf{Q}$ can be reduced to

$$\begin{aligned}\mathbf{Q}_2 = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1} + P_{w_1+1}(\xi), ..., \xi_\lambda\xi_{n-1} + P_{n-1}(\xi),\\ \xi_\lambda\xi_n + P_n(\xi) + \xi_{n+1}^2 + ... + \xi_{n+k-1}^2).\end{aligned}$$

If $d - n = 1$ ($k = 1$), then $\mathbf{Q}_2$ is just in the case (2a), which has been proved. If $d - n \geq 2$ ($k \geq 2$), then $\mathbf{Q}_2$ is in fact the original $\mathbf{Q}$ with dimension $d - 1$ and co-dimension n. Thus we can use induction on $d - n$ to get the result of this case. Finally, we still need to consider the third term of (4.11). Without loss of generality, we can assume $\lambda \neq n$, otherwise we can proceed in a similar way as for the last term. Recall that $P_{w_1+1}(\xi), ..., P_n(\xi)$ depend on θ many variables except the variables ξ_1 and ξ_λ. We apply flat decoupling (4.3) to these variables as well as the variable ξ_n from 1-scale to $R^{1/2}$-scale. After this, by the locally constant property, the original $\mathbf{Q}$ is reduced to

$$\mathbf{Q}_3 = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}, \xi_{n+1}^2 + ... + \xi_{n+k}^2).$$

Note that $\mathbf{Q}_3$ has a tensor product structure. Write $\mathbf{Q}_3 = (\mathbf{Q}_3', \mathbf{Q}_3'')$ with

$$\mathbf{Q}_3' = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}), \quad \mathbf{Q}_3'' = \xi_{n+1}^2 + ... + \xi_{n+k}^2.$$

Then we can express the associated extension operator as

$$E^{\mathbf{Q}_3} f = E^{\mathbf{Q}_3'}(E^{\mathbf{Q}_3''} g),$$

where g denotes the inverse Fourier transform of f in the variable ξ_n. It follows that

$$\begin{aligned}
\|E^{\mathbf{Q}_3} f\|_{L^p(B_R)} &\lesssim R^{\frac{w_\lambda+3\theta+2}{2p} - \frac{1+\theta}{2} + \epsilon} \|E^{\mathbf{Q}_3''} g\|_{L^p} \\
&\lesssim R^{\frac{w_\lambda+3\theta+2}{2p} - \frac{1+\theta}{2} + \epsilon} \|g\|_{L^p} \\
&\lesssim R^{\frac{w_\lambda+3\theta+2}{2p} - \frac{1+\theta}{2} + \epsilon} \Big\| \|f\|_{L^{p'}_{\xi_n}} \Big\|_{L^p_{\xi_1, ..., \xi_{n-1}, \xi_{n+1}, ..., \xi_{n+k}}} \\
&\leq R^{\frac{w_\lambda+3\theta+2}{2p} - \frac{1+\theta}{2} - \frac{1}{2}(\frac{1}{p'} - \frac{1}{p}) + \epsilon} \|f\|_{L^p}.
\end{aligned}$$

Here in the first line we used rescaling and the restriction estimate of $\mathbf{Q}_3'$ for $p \geq w_1 + 3$, which has been proved in the case (1), in the second line we used the known restriction estimate for the paraboloid [9, Theorem 1], and in the third line we used the Hausdorff-Young inequality in the variable ξ_n. Therefore, we get

$$\begin{aligned}
D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1) &\lesssim R^{\frac{\theta+1}{2}(1-\frac{2}{p}) + \frac{w_\lambda+3\theta+2}{2p} - \frac{1+\theta}{2} - \frac{1}{2}(\frac{1}{p'} - \frac{1}{p}) + \epsilon} \\
&= R^{\frac{w_\lambda+\theta+2}{2p} - \frac{1}{2} + \epsilon} \lesssim 1, \qquad (4.12)
\end{aligned}$$

for $p > w_1 + 3$ due to $w_1 \geq w_\lambda + \theta$. And the estimate (4.1) for $p = w_1 + 3$ is a direct corollary of Hölder's inequality. This finishes the proof of the case (5a).

The argument for the case (5b) is essentially the same as that for the case (2b), so we omit the details.

Suppose that we are in the case (5c), i.e., P_j is a quadratic form without mixed terms that is independent of the variables $\xi_j, ..., \xi_{n+k}$ for some $w_1+1 \leq j \leq n$ and $P_{j'} = 0$ for all $j' \neq j$; $2 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda + \theta/2$. We divide it into two subcases: $j < n$ and $j = n$.

We first consider the subcase $j < n$. Without loss of generality, let $j = n-1$, then

$$\mathbf{Q} = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1} + \xi_\alpha^2 + ... + \xi_\gamma^2, \xi_\lambda\xi_n + \xi_{n+1}^2 + ... + \xi_{n+k}^2).$$

We use the broad-narrow analysis and iteration to obtain

$$\begin{aligned} D_p(R) \lessapprox & K^{3(n+k)} \sup_{\mu_1,\mu_\lambda,\mu_{n+k}>1} BD_p(\underbrace{\mu_1, 1, ..., 1}_{\lambda-1}, \mu_\lambda, 1, ..., 1, \mu_{n+k}) \\ & + D_p(R^{\frac{1}{2}}, 1, ..., 1) + D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1) + D_p(1, ..., 1, R^{\frac{1}{2}}). \end{aligned} \tag{4.13}$$

The first two terms in (4.13) can be controlled via the same argument as in the case (2a), while the third term in (4.13) can be controlled via the same argument as in the case (2c). The last term in (4.13) can be managed by an induction on $d-n$ which is the same as in the case (5a). So it remains to consider the subcase $j = n$, i.e.,

$$\mathbf{Q} = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}, \xi_\lambda\xi_n + \xi_\alpha^2 + ... + \xi_\gamma^2 + \xi_{n+1}^2 + ... + \xi_{n+k}^2).$$

We use the broad-narrow analysis and iteration to obtain

$$\begin{aligned} D_p(R) \lessapprox & K^{3(n+k)} \sup_{\mu_1,\mu_\lambda,\mu_{n+k}>1} BD_p(\underbrace{\mu_1, 1, ..., 1}_{\lambda-1}, \mu_\lambda, 1, ..., 1, \mu_{n+k}) \\ & + D_p(R^{\frac{1}{2}}, 1, ..., 1) + D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1) + D_p(1, ..., 1, R^{\frac{1}{2}}). \end{aligned} \tag{4.14}$$

The first two terms in (4.14) can be controlled via the same argument as in the case (5a). The last term in (4.14) can be controlled via an induction on $d-n$ as in the case (5a) with the constraint $w_1 \geq w_\lambda + \theta/2$. For the third term in (4.14), we can without loss of generality assume that ξ_λ does not appear in $\xi_\alpha, ..., \xi_\gamma$ as in the case (2c). In this case, we use flat decoupling (4.3) in the variables $\xi_\alpha, ..., \xi_\gamma$ and decoupling for the paraboloid (4.2) in the variables $\xi_{n+1}, ..., \xi_{n+k}$ from 1-scale to $R^{1/4}$-scale, and then apply the change of variables

$$\xi_\alpha \to \frac{\xi_\alpha}{R^{1/4}}, \quad ..., \quad \xi_\gamma \to \frac{\xi_\gamma}{R^{1/4}}, \quad \xi_\lambda \to \frac{\xi_\lambda}{R^{1/2}},$$
$$\xi_{n+1} \to \frac{\xi_{n+1}}{R^{1/4}}, \quad ..., \quad \xi_{n+k} \to \frac{\xi_{n+k}}{R^{1/4}}.$$

Notice that such rescaling keeps the surface $S_{\mathbf{Q}}$ invariant. Covering B_R with balls of scale $R/2$, and using induction on R, one concludes

$$\begin{aligned} D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1) &\lesssim R^{\frac{\theta}{4}(1-\frac{2}{p})+\frac{k}{4}-\frac{k+1}{2p}+\frac{2w_\lambda+3\theta+2k+4}{4p}-\frac{\theta+k+2}{4}+\epsilon} \\ &= R^{\frac{w_\lambda+\theta/2+1}{2p}-\frac{1}{2}+\epsilon} \lesssim 1, \end{aligned} \tag{4.15}$$

for $p > w_1 + 3$ due to $w_1 \geq w_\lambda + \theta/2$. And the estimate (4.1) for $p = w_1 + 3$ is a direct corollary of Hölder's inequality.

Suppose that we are in the case (5d), i.e., $P_j = 0$ for every $w_1 + 1 \leq j \leq n$; $1 \leq \lambda \leq w_1 + 1$, $w_1 \geq w_\lambda$. Then

$$\mathbf{Q} = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}, \xi_\lambda\xi_n + \xi_{n+1}^2 + ... + \xi_{n+k}^2).$$

We divide it into two subcases: $\lambda \neq 1$ and $\lambda = 1$.

We first assume that $\lambda \neq 1$. Take $i_1 = n, ..., i_k = n + k - 1$, then

$$|J(\xi; n, n+1, ..., n+k-1)| \sim |\xi_1|^{w_1}|\xi_\lambda|^{w_\lambda - 1}|\xi_{n+k}|.$$

We use the broad-narrow analysis and iteration to obtain

$$\begin{aligned} D_p(R) \lessapprox & K^{3(n+k)} \sup_{\mu_1, \mu_\lambda, \mu_{n+k} > 1} BD_p(\underbrace{\mu_1, 1, ..., 1}_{\lambda-1}, \mu_\lambda, 1, ..., 1, \mu_{n+k}) \\ & + D_p(R^{\frac{1}{2}}, 1, ..., 1) + D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1) + D_p(1, ..., 1, R^{\frac{1}{2}}). \end{aligned} \tag{4.16}$$

The first two terms in (4.16) can be controlled via the same argument as in the case (2a). For the last term in (4.16), by the locally constant property, the original $\mathbf{Q}$ can be reduced to

$$\mathbf{Q}_4 = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}, \xi_\lambda\xi_n + \xi_{n+1}^2 + ... + \xi_{n+k-1}^2).$$

If $d - n = 1$ ($k = 1$), $\mathbf{Q}_4$ is just included in the case (1), which has been proved. If $d - n \geq 2$ ($k \geq 2$), $\mathbf{Q}_4$ is in fact the original $\mathbf{Q}$ with dimension $d-1$ and co-dimension n. Then we can use induction on $d - n$ to get the result of this case. Finally, we still need to deal with the third term in (4.16). We use flat decoupling (4.3) in the variable ξ_n from 1-scale to $R^{1/2}$-scale, then

$$D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1) \lesssim R^{\frac{1}{2}(1-\frac{2}{p})} D_p(\underbrace{1, ..., 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, ..., 1, R^{\frac{1}{2}}).$$

By the locally constant property, $\mathbf{Q}_4$ is further reduced to

$$\mathbf{Q}_5 = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}, \xi_{n+1}^2 + ... + \xi_{n+k}^2).$$

Note that $\mathbf{Q}_5$ has a tensor product structure. Write $\mathbf{Q}_5 = (\mathbf{Q}_5', \mathbf{Q}_5'')$ with

$$\mathbf{Q}_5' = (\xi_1^2, ..., \xi_1\xi_{w_1}, \xi_\lambda\xi_{w_1+1}, ..., \xi_\lambda\xi_{n-1}), \quad \mathbf{Q}_5'' = \xi_{n+1}^2 + ... + \xi_{n+k}^2.$$

Then we express the associated extension operator as

$$E^{\mathbf{Q}_5} f = E^{\mathbf{Q}_5'}(E^{\mathbf{Q}_5''} g),$$

where g denotes the inverse Fourier transform of f in the variable ξ_n. It follows that

$$\begin{aligned}\|E^{\mathbf{Q}_5} f\|_{L^p(B_R)} &\lesssim \|E^{\mathbf{Q}_5''} g\|_{L^p} \\ &\lesssim \|g\|_{L^p} \\ &\leq \Big\| \|f\|_{L^{p'}_{\xi_n}} \Big\|_{L^p_{\xi_1,\dots,\xi_{n-1},\xi_{n+1},\dots,\xi_{n+k}}} \\ &\leq R^{-\frac{1}{2}(\frac{1}{p'}-\frac{1}{p})} \|f\|_{L^p}.\end{aligned}$$

Here in the first line we used the restriction estimate on $\mathbf{Q}_5'$ for $p \geq w_1 + 3$, which has been proved in the case (1), in the second line we used the known restriction estimate for the paraboloid [9, Theorem 1], and in the third line we used the Hausdorff-Young inequality in the variable ξ_n. Therefore, we get

$$D_p(\underbrace{1, \dots, 1}_{\lambda-1}, R^{\frac{1}{2}}, 1, \dots, 1) \lesssim R^{\frac{1}{2}(1-\frac{2}{p})} R^{-\frac{1}{2}(\frac{1}{p'}-\frac{1}{p})} = 1, \tag{4.17}$$

for $p \geq w_1 + 3$. This closes the proof for the subcase $\lambda \neq 1$.

Now we assume that $\lambda = 1$, then

$$\mathbf{Q} = (\xi_1^2, \dots, \xi_1\xi_{n-1}, \xi_1\xi_n + \xi_{n+1}^2 + \dots + \xi_{n+k}^2).$$

Our goal is to show that the estimate (4.1) holds for $p \geq n + 2$. Take $i_1 = n, \dots, i_k = n + k - 1$, then

$$|J(\xi; n, n+1, \dots, n+k-1)| \sim |\xi_1|^{n-1} |\xi_{n+k}|.$$

We use the broad-narrow analysis and iteration to obtain

$$\begin{aligned}D_p(R) \lessapprox & K^{3(n+k)} \sup_{\mu_1, \mu_{n+k} > 1} BD_p(\mu_1, 1, \dots, 1, \mu_{n+k}) \\ & + D_p(R^{\frac{1}{2}}, 1, \dots, 1) + D_p(1, \dots, 1, R^{\frac{1}{2}}).\end{aligned} \tag{4.18}$$

The first bilinear term in (4.18) can be controlled via the same argument as in the case (2a). For the last term of (4.18), we can no longer use induction on $d - n$ as before, since the restriction estimate when $d - n = 0$ holds only for $p \geq n + 3$. We first consider the case $d - n = 1$. Then the original $\mathbf{Q}$ is reduced to $\mathbf{Q}_6 = (\xi_1^2, \dots, \xi_1\xi_n)$. Let g denote the inverse Fourier transform of f in the variable ξ_{n+1}. By the result of the case (1), we know that

$$\|E^{\mathbf{Q}_6} g\|_{L^{n+3}(B_R)} \lesssim R^{\epsilon} \|g\|_{L^{n+3}}.$$

On the other hand,

$$\|E^{\mathbf{Q}_6} g\|_{L^2(B_R)} \lesssim R^{\frac{n}{2}} \|g\|_{L^2}.$$

By interpolation, we have

$$\|E^{\mathbf{Q}_6} g\|_{L^p(B_R)} \lesssim R^{n \cdot \frac{n+3}{n+1}(\frac{1}{p} - \frac{1}{n+3}) + \epsilon} \|g\|_{L^p}, \qquad 2 \leq p \leq n+3.$$

By the restriction estimate of $\mathbf{Q}_6$ and the Hausdorff-Young inequality in the variable ξ_{n+1}, we get

$$\|E^{\mathbf{Q}}f\|_{L^p(B_R)} \lesssim R^{n\cdot\frac{n+3}{n+1}(\frac{1}{p}-\frac{1}{n+3})-\frac{1}{2}(\frac{1}{p'}-\frac{1}{p})+\epsilon}\|f\|_{L^p} = R^{\frac{n^2+4n+1}{(n+1)p}-\frac{3n+1}{2(n+1)}+\epsilon}\|f\|_{L^p},$$

which implies that

$$D_p(1,...,1,R^{\frac{1}{2}}) \lesssim R^{\frac{n^2+4n+1}{(n+1)p}-\frac{3n+1}{2(n+1)}+\epsilon}, \qquad d = n+1. \tag{4.19}$$

Note that the critical exponent $p = 2(n^2+4n+1)/(3n+1) \leq n+2$ for all $n \geq 1$. Having established the base case $d-n=1$, for general cases, we can now use induction on $d-n$ as before. In fact, via the same argument as when $d-n=1$, we can obtain the following better results:

$$D_p(1,...,1,R^{\frac{1}{2}}) \lesssim R^{\frac{n+3}{p}-\frac{3}{2}+\epsilon}, \qquad d \geq n+2. \tag{4.20}$$

Finally, we need to deal with the second term in (4.18). We will divide it into three subcases. Firstly, we assume that $n=2$ and $k=1$. By using flat decoupling (4.3) in the variable ξ_1 from $R^{1/2}$-scale to R-scale, we get

$$D_p(R^{\frac{1}{2}},1,1) \lesssim R^{\frac{1}{2}(1-\frac{2}{p})}D_p(R,1,1).$$

By the locally constant property, the original $\mathbf{Q}$ becomes $\mathbf{Q}_7 = (0,\xi_3^2)$. By the restriction estimate of the parabola and the Hausdorff-Young inequality, we have

$$\begin{aligned}\|E^{\mathbf{Q}}f\|_{L^p(B_R^5)} &\sim \|E^{\mathbf{Q}_7}f\|_{L^p(B_R^5)}\\ &\sim R^{\frac{1}{p}}\|E^{\mathbf{Q}_7}f\|_{L^p(B_R^4)}\\ &\lesssim R^{\frac{1}{p}}\Big\|\|f\|_{L^{p'}_{\xi_1}}\Big\|_{L^p_{\xi_2,\xi_3}}\\ &\lesssim R^{\frac{1}{p}-(\frac{1}{p'}-\frac{1}{p})}\|f\|_{L^p},\end{aligned}$$

and so

$$D_p(R^{\frac{1}{2}},1,1) \lesssim R^{\frac{1}{2}(1-\frac{2}{p})+\frac{1}{p}-(\frac{1}{p'}-\frac{1}{p})} = R^{\frac{2}{p}-\frac{1}{2}} \lesssim 1,$$

for $p \geq n+2 = 4$. Next, we assume that $n=3,4$ and $k=1$. We use decoupling for the parabola (4.2) in the variable ξ_{n+1} from 1-scale to $R^{1/2}$-scale, followed by (4.19), such that

$$\begin{aligned}D_p(R^{\frac{1}{2}},1,...,1) &\lesssim R^{\frac{1}{2}(\frac{1}{2}-\frac{1}{p})}D_p(R^{\frac{1}{2}},1,...,1,R^{\frac{1}{2}})\\ &\leq R^{\frac{1}{2}(\frac{1}{2}-\frac{1}{p})}D_p(1,...,1,R^{\frac{1}{2}})\\ &\lesssim R^{\frac{1}{2}(\frac{1}{2}-\frac{1}{p})+\frac{n^2+4n+1}{(n+1)p}-\frac{3n+1}{2(n+1)}+\epsilon}\\ &= R^{\frac{2n^2+7n+1}{2(n+1)p}-\frac{5n+1}{4(n+1)}+\epsilon}\end{aligned}$$

with the critical exponent $p = \frac{4n^2+14n+2}{5n+1} \leq n+2$ for $n \geq 3$. For all other n and k, we use decoupling for the paraboloid (4.2) in the variables

$\xi_{n+1}, ..., \xi_{n+k}$ from 1-scale to $R^{1/4}$-scale, followed by rescaling and induction on R, to obtain

$$\begin{aligned}D_p(R^{\frac{1}{2}}, 1, ..., 1) &\lesssim R^{\frac{k}{4}-\frac{k+1}{2p}} D_p(R^{\frac{1}{2}}, 1, ..., 1, R^{\frac{1}{4}}, ..., R^{\frac{1}{4}})\\ &\leq R^{\frac{k}{4}-\frac{k+1}{2p}+\frac{n+k+3}{2p}-\frac{k+2}{4}} D_p(R)\\ &\lesssim R^{\frac{n+2}{2p}-\frac{1}{2}+\epsilon}\end{aligned}$$

with the critical exponent $p = n+2$.

We have completed the proof of the estimates (1.2)-(1.7). Finally, we show that these estimates are all sharp up to the endpoints. This argument is essentially the same as that in [17, Section 3]. However, for later convenience we still sketch it here. We take the case (1) as an example, and the argument for other cases are similar. We first consider the characteristic function $\chi_{[0,R^{-t_1}]\times...\times[0,R^{-t_n}]}$ with $0 \leq t_1, ..., t_n \leq 1$. By the locally constant property, we have

$$\left|E^{\mathbf{Q}}(\chi_{[0,R^{-t_1}]\times...\times[0,R^{-t_n}]})(x)\right| \sim R^{-t_1-...-t_n}, \qquad x \in T, \tag{4.21}$$

where

$$\begin{aligned}T := \Big\{x : |x_1| \leq \frac{1}{100}R^{t_1}, ..., |x_n| \leq \frac{1}{100}R^{t_n},\\ |x_{n+1}| \leq \frac{1}{100}R^{t_{\lambda_1}+t_1}, ..., |x_{2n}| \leq \frac{1}{100}R^{t_{\lambda_n}+t_n}\Big\}.\end{aligned}$$

Divide $[0,1]^n$ to τ, where each $\tau = I_1 \times ... \times I_n$ is a rectangular box with $|I_1| = R^{-t_1}, ..., |I_n| = R^{-t_n}$. Let $\epsilon = (\epsilon_\tau)$ be a sequence of independent random variables where ϵ_τ takes values ± 1 with equal probability. Define

$$f = \sum_\tau \epsilon_\tau \chi_\tau,$$

and then (1.1) becomes

$$\Big\|E^{\mathbf{Q}}(\sum_\tau \epsilon_\tau \chi_\tau)\Big\|_{L^q} \lesssim 1.$$

Using Khintchine's inequality and Holder's inequality, we get

$$1 \gtrsim \mathbb{E}\Big\|E^{\mathbf{Q}}(\sum_\tau \epsilon_\tau\chi_\tau)\Big\|_{L^q}^q \sim \int \Big(\sum_\tau |E^{\mathbf{Q}}\chi_\tau|^2\Big)^{\frac{q}{2}} \geq \int \sum_\tau |E^{\mathbf{Q}}\chi_\tau|^q.$$

It follows from (4.21) that

$$1 \gtrsim R^{\sum_{j=1}^n (t_{\lambda_j}+3t_j)} R^{-(\sum_{j=1}^n t_j)q}.$$

And so

$$q \geq \frac{\sum_{j=1}^n t_{\lambda_j}}{\sum_{j=1}^n t_j} + 3 = \frac{\sum_{j=1}^n w_j t_j}{\sum_{j=1}^n t_j} + 3.$$

To obtain optimal necessary condition, we take maximum over all t_j, i.e.,

$$q \geq \max_{t_1,...,t_n \in \{0,1/2,1\}} \frac{\sum_{j=1}^n w_j t_j}{\sum_{j=1}^n t_j} + 3. \tag{4.22}$$

Here we restrict $t_1, ..., t_n$ to $\{0, 1/2, 1\}$, which is enough as $\mathbf{Q}$ is quadratic. Without loss of generality, we assume that $\max_j w_j = w_1$. If we take $t_1 = 1$ and $t_j = 0$ for $j \geq 2$, we immediately obtain $q \geq w_1 + 3$. For the second constraint of (1.2), we can take $f = \chi_{[0,R^{-1}]\times[0,1]\times...\times[0,1]}$ in (1.1) and do some elementary calculations.

5. Final remarks and more examples

In this final section, we give some additional remarks and examples regarding our method and Theorem 1.1.

Firstly, combining our method and Theorem 1.1, we can in fact obtain the essentially sharp restriction estimates for all quadratic forms with $d = n = 2$.

Theorem 5.1. *Let $\mathbf{Q} = (Q_1, Q_2)$ be an 2-tuple of real quadratic forms defined on $\mathbb{R}^2$. Suppose that $\mathfrak{d}_{2,2}(\mathbf{Q}) = 2$, then*

(1) *If $\mathbf{Q}$ satisfies $\mathfrak{d}_{2,1}(\mathbf{Q}) = 1$ and $\mathfrak{d}_{1,2}(\mathbf{Q}) = 0$, then $\mathbf{Q} \equiv (\xi_1^2, \xi_1\xi_2)$, and the estimate (1.1) holds for*

$$q > 5, \quad \frac{1}{p} + \frac{4}{q} < 1. \tag{5.1}$$

(2) *If $\mathbf{Q}$ satisfies $\mathfrak{d}_{2,1}(\mathbf{Q}) = 1$ and $\mathfrak{d}_{1,2}(\mathbf{Q}) = 1$, then $\mathbf{Q} \equiv (\xi_1^2, \xi_2^2)$, and the estimate (1.1) holds for*

$$q > 4, \quad \frac{1}{p} + \frac{3}{q} < 1. \tag{5.2}$$

(3) *If $\mathbf{Q}$ satisfies $\mathfrak{d}_{2,1}(\mathbf{Q}) = 2$, then $\mathbf{Q} \equiv (\xi_1\xi_2, \xi_1^2 - \xi_2^2)$, and the estimate (1.1) holds for*

$$q > 4, \quad \frac{1}{p} + \frac{3}{q} < 1. \tag{5.3}$$

Moreover, the estimates all above are sharp up to the endpoints.

Proof. Suppose that $\mathbf{Q}$ satisfies $\mathfrak{d}_{2,1}(\mathbf{Q}) = 1$. By Lemma 2.2 in [18], we have

$$\mathfrak{d}_{2,1}(\mathbf{Q}) = 1 \iff \mathbf{Q} \equiv (\xi_1^2, *). \tag{5.4}$$

If $\mathfrak{d}_{1,2}(\mathbf{Q}) = 0$, by (1.14) in [19], this quality must take minimal restricting on $\xi_1 = 0$. We can assume that

$$\mathbf{Q} \equiv (\xi_1^2, \xi_1 L(\xi)),$$

with a linear form L. Also note $\mathfrak{d}_{2,2}(\mathbf{Q}) = 2$, $L(\xi)$ must depend on the variable ξ_2. By a linear transformation, we get $\mathbf{Q} \equiv (\xi_1^2, \xi_1\xi_2)$. If $\mathfrak{d}_{1,2}(\mathbf{Q}) = 1$, we write

$$\mathbf{Q} \equiv (\xi_1^2, a\xi_2^2 + b\xi_1\xi_2),$$

for some a and b. We know $a \neq 0$, otherwise it just is the case (1). If $b = 0$, we have proved the case (2). If not, we add a multiple of ξ_1^2 to Q_2 such that the second term can form a perfect square. Using a linear transformation, we get $\mathbf{Q} \equiv (\xi_1^2, \xi_2^2)$.

Suppose that $\mathbf{Q}$ satisfies $\mathfrak{d}_{2,1}(\mathbf{Q}) = 2$. We pick $M \in \mathbb{R}^{2\times 2}$ and $M' \in \mathbb{R}^{2\times 1}$ such that the equality in (2.19) is achieved with $d' = 2, n' = 1$. By linear transformations, we can assume that $M = I_{2\times 2}$ and $M' = (1, 0)^T$. Then $\mathfrak{d}_{2,1}(\mathbf{Q}) = 2$ implies that Q_1 depends on 2 variables. We use linear transformations again to diagonalize Q_1, then

$$\mathbf{Q} \equiv (\xi_1^2 \pm \xi_2^2, *).$$

We first assume that $Q_1 = \xi_1^2 + \xi_2^2$. Then we can reduce it to

$$\mathbf{Q} \equiv (\xi_1^2 + \xi_2^2, \lambda\xi_1\xi_2 + \mu\xi_2^2).$$

If $\lambda = 0$, it can be further reduced to (ξ_1^2, ξ_2^2), which is just the case (2). If $\mu = 0$, by adding a multiple of $\xi_1\xi_2$ to Q_1 such that Q_1 forms a perfect square, which is contradiction with (5.4). Therefore, we can assume that $\lambda \neq 0$ and $\mu \neq 0$. In this case, we can add a multiple of Q_1 to Q_2 such that Q_2 forms a perfect square, which is contradiction with (5.4) again. Next we assume that $Q_1 = \xi_1^2 - \xi_2^2$. By a linear transformation, we can write

$$\mathbf{Q} \equiv (\xi_1\xi_2, a\xi_1^2 + b\xi_2^2).$$

We can assume $a \neq 0$ and $b \neq 0$, otherwise it can be further reduced to the case (1). So

$$\mathbf{Q} \equiv (\xi_1\xi_2, \xi_1^2 + b\xi_2^2).$$

If $b > 0$, we add a multiple of $\xi_1\xi_2$ to Q_2 such that Q_2 can form a perfect square, which is a contradiction with (5.4). Therefore $b < 0$, and $\mathbf{Q} \equiv (\xi_1\xi_2, \xi_1^2 - \xi_2^2)$.

Now we show (5.1)-(5.3). For (5.1) and (5.2), we have proved them in the case (1) in Theorem 1.1. For (5.3), though $(\xi_1\xi_2, \xi_1^2 - \xi_2^2)$ is not included in Theorem 1.1, note

$$|J(\xi)| \sim \xi_1^2 + \xi_2^2 \geq \xi_1^2,$$

we can still apply Theorem 3.1 and the argument of Section 4. The sharpness of these estimates can be shown by testing on certain functions as in the final part of the previous section. □

Remark 5.1. *As a historical remark, Christ* [12] *showed that when $d = n = 2$, the* (CM) *condition (Definition 2.3) is equivalent to cases (2) and (3) in Theorem 5.1, and he was able to obtain endpoint results ($q > 4$, $\frac{1}{p} + \frac{3}{q} = 1$) via a different method.*

We point out that (5.1)-(5.3) can also be proved via the classical broad-narrow analysis as in [17, Proposition 8.2]. However, our method seems more suitable. We take the case (1) in Theorem 1.1 as an example. From the argument of the necessary condition, we see that the element determining the optimal necessary condition is $\max_j w_j$. This quality also dominates the range of p in Theorem 3.1. On the other hand, [19] and [15] both studied the restriction estimates for general higher co-dimensional surfaces via multilinear method and k-linear method, respectively. Their arguments are much more complex than ours. Despite all these complications, their approaches seem to be only more powerful for non-degenerate quadratic surfaces. In fact, for all cases in Theorem 5.1, the results derived by their methods are not sharp.

We say a few words on the last case in Theorem 5.1. $\mathbf{Q} = (\xi_1\xi_2, \xi_1^2 - \xi_2^2)$ is a very special example with the following property: $J(\xi) = 0$ implies $\xi = 0$. If we use the broad-narrow analysis to study it, the narrow part will just be 0-dimensional. By rescaling and induction straightforwardly, we easily obtain sharp restriction estimate on it. Will such phenomenon occur in higher dimensions? More precisely, given a tuple $\mathbf{Q}(\xi) = (Q_1(\xi), ..., Q_n(\xi))$ of quadratic forms with $\xi \in \mathbb{R}^d$, does there exist $\mathbf{Q}$ such that $\mathfrak{d}_{d,1}(\mathbf{Q}) = d$?

A deep algebraic result [1, Theorem 1] tells us that the maximal dimension of subspace of real invertible $d \times d$ matrices is given by the Hurwitz-Radon number $\rho(d)$, which is defined as follows: if $d = 2^{4a+b}c$ ($0 \le b \le 3$, c odd), then $\rho(d) = 8a + 2^b$. Therefore, if $\rho(d) < n$, then it's impossible to have $\mathfrak{d}_{d,1}(\mathbf{Q}) = d$. In particular, when d is odd, we must have $\rho(d) = 1$, so the only possible case is when $n = 1$, i.e., $\mathbf{Q}$ is a hypersurface with non-zero Gaussian curvature. Also, we can see that $\rho(d) \le d$, so it's impossible to have $\mathfrak{d}_{d,1}(\mathbf{Q}) = d$ when $d < n$.

In fact, a more detailed computation based on $\rho(d)$ shows that for any fixed ratio λ of co-dimension n and dimension d, there are at most finitely many pairs of (d, n) such that there exist some $\mathbf{Q}$ satisfying $\mathfrak{d}_{d,1}(\mathbf{Q}) = d$. (We have seen that when $\lambda > 1$ there is no such pair.) For example, when $\lambda = 1$(i.e., $d = n$), the only possible cases are $(d, n) = (1, 1), (2, 2), (4, 4), (8, 8)$. Here $(d, n) = (1, 1)$ is achieved by the real parabola $\mathbf{Q} \equiv (\xi_1^2)$ in $\mathbb{R}^2$, while $(d, n) = (2, 2)$ is achieved by the complex curve $\mathbf{Q} \equiv (\xi_1\xi_2, \xi_1^2 - \xi_2^2)$ in $\mathbb{C}^2$ as in case (3) of Theorem 5.1. However, we do not know if $\mathbf{Q}$ can be constructed when $(d, n) = (4, 4)$ or $(8, 8)$. Similarly, when $\lambda = \frac{1}{2}$(i.e., $d = 2n$), the only possible cases are $(d, n) = (2, 1), (4, 2), (8, 4), (16, 8)$. Here $(d, n) = (2, 1)$ is achieved by the paraboloid/hyperboloid $\mathbf{Q} = (\xi_1^2 \pm \xi_2^2)$ in $\mathbb{R}^3$, while $(d, n) = (4, 2)$ is achieved by the complex paraboloid $\mathbf{Q} \equiv (\xi_1^2 - \xi_2^2 + \xi_3^2 - \xi_4^2, \xi_1\xi_2 + \xi_3\xi_4)$ in $\mathbb{C}^3$. However, we do not know if $\mathbf{Q}$ can be constructed when $(d, n) = (8, 4)$ or $(16, 8)$.

Anyway, these facts in some sense indicate that $\mathfrak{d}_{d,1}(\mathbf{Q}) = d$ may not be the right way to describe "well-curvedness"(i.e., maximal nondegeneracy) in Fourier restriction theory. On the one hand, it will exclude a bunch of quadratic surfaces which are generally believed to be "well-curved", such as $\mathbf{Q} \equiv (\xi_i\xi_j)_{1 \le i \le j \le d}$ with $n = \frac{d(d+1)}{2}$ resolved in [30] (its well-curvedness can be tested using [16, Theorem 6]). On the other hand, the set of $\mathbf{Q}$ satisfying $\mathfrak{d}_{d,1}(\mathbf{Q}) = d$ can be very sparse, i.e., for "almost all" (d, n), $\mathfrak{d}_{d,1}(\mathbf{Q}) = d$ will never be satisfied.

In general, finding the right notion of well-curvedness is a subtle problem, especially in higher co-dimensional cases. One major breakthrough towards this direction is [16], which proposed a range of computable criteria for well-curvedness and identified the deep connections to geometric invariant theory. Interestingly, when $n = 2$, the criteria in [16] are shown to be equivalent to Definition 2.3, see [13, Theorem 2].

Though we mainly focus on Theorem 1.1 in this paper, our method can actually be used to study more quadratic surfaces than those in Theorem 1.1. For example, in the case (1) of Theorem 1.1, the constraint $\lambda_j \leq j$ for each $1 \leq j \leq n$ is convenient for us to directly apply Theorem 3.1. We can also consider other conditions: suppose that

$$\mathbf{Q} = (\xi_{\lambda_1}\xi_1, \xi_{\lambda_2}\xi_2, ..., \xi_{\lambda_n}\xi_n).$$

We assume that one of the following conditions holds:

- $\lambda_1 = 2$, $\lambda_j \leq j$ for each $2 \leq j \leq n$;
- $\lambda_1 = 3$, $2 \leq \lambda_j \leq j$ for each $2 \leq j \leq n-1$;
- $\lambda_1 = 3$, $\lambda_3 = 3$, $\lambda_j \leq j$ for each $j \neq 1, 3$.

Then the estimate (1.1) holds for

$$q > \max_j w_j + 3, \quad \frac{1}{p} + \frac{\max_j w_j + 2}{q} < 1. \tag{5.5}$$

Similarly, we can also consider other polynomial cases on (2) and (5) in Theorem 1.1.

From the above arguments, it's plausible that our method is more powerful in the study for degenerate higher co-dimensional surfaces. However, our method cannot cover all possible cases, even when the quadratic surface is monomial. For example, when

$$\mathbf{Q} = (\xi_1\xi_2, \xi_2\xi_3, \xi_3\xi_4, \xi_4\xi_1),$$

we can compute that $J(\xi) \equiv 0$, and then the condition in Theorem 3.1 is not met at all. What's more, even if Theorem 3.1 works, it may well be the case that the restriction estimate we obtain is not sharp. For the case

$$\mathbf{Q} = (\xi_1^2, ..., \xi_1\xi_{n-1}, \xi_1\xi_n + \xi_{n+1}^2),$$

we have proved sharp restriction estimate in the case (5d) of Theorem 1.1. However, for the case (2) of Theorem 1.1, we always require $\lambda \geq 2$. If $\lambda = 1$, a typical example is

$$\mathbf{Q} = (\xi_1^2, ..., \xi_1\xi_{n-1}, \xi_1\xi_n + \xi_2^2).$$

Through the same argument as in the final part of the previous section, we expect that (1.1) may hold for $p > n + 2$ if $p = q$. But the result in

Theorem 3.1 only covers $p > n+3$. Therefore, our method can not yield sharp bound in this case. If we consider the classical broad-narrow analysis, then the constraint from the broad part is $p > 4$, while the narrow part requires $p > n + 7/2$ due to the loss of decoupling.

Acknowledgements. This work was supported by the National Key R&D Program of China (No. 2022YFA1005700), NSFC Grant No. 12371095 and No. 12371100.

References

[1] J. F. Adams, P. D. Lax, and R. S. Phillips. On matrices whose real linear combinations are nonsingular. *Proc. Amer. Math. Soc*, 16(2):318–322, 1965.

[2] J.-G. Bak, J. Lee, and S. Lee. Bilinear restriction estimates for surfaces of codimension bigger than 1. *Anal. PDE*, 10(8):1961–1985, 2017.

[3] J.-G. Bak and S. Lee. Restriction of the Fourier transform to a quadratic surface in $\mathbb{R}^n$. *Math. Z.*, 247(2):409–422, 2004.

[4] A. D. Banner. Restriction of the fourier transform to quadratic submanifolds. *Ph.D. thesis, Princeton University*, 2002.

[5] J. Bourgain. Besicovitch type maximal operators and applications to fourier analysis. *Geom. Funct. Anal.*, 1(2):147–187, 1991.

[6] J. Bourgain and C. Demeter. The proof of the l^2 decoupling conjecture. *Ann. Math.*, 182(1):351–389, 2015.

[7] J. Bourgain and C. Demeter. Mean value estimates for weyl sums in two dimensions. *J. Lond. Math. Soc.*, 94(3):814–838, 2016.

[8] J. Bourgain and C. Demeter. Decouplings for curves and hypersurfaces with nonzero Gaussian curvature. *J. Anal. Math.*, 133:279–311, 2017.

[9] J. Bourgain and L. Guth. Bounds on oscillatory integral operators based on multilinear estimates. *Geom. Funct. Anal.*, 21(6):1239–1295, 2011.

[10] Z. Cao, C. Miao, and Z. Wang. Fourier decay of fractal measures on surfaces of co-dimension two in $\mathbb{R}^5$. *J. Funct. Anal.*, 286(8):110378, 2024.

[11] M. Christ. Restriction of the Fourier transform to submanifolds of low codimension. *Ph.D. thesis, University of Chicago*, 1982.

[12] M. Christ. On the restriction of the Fourier transform to curves: endpoint results and the degenerate case. *Trans. Amer. Math. Soc.*, 287(1):223–238, 1985.

[13] S. Dendrinos, A. Mustata, and M. Vitturi. A restricted 2-plane transform related to fourier restriction for surfaces of codimension 2. *arXiv:2209.15530*, 2022.

[14] C. Fefferman. Inequalities for strongly singular convolution operators. *Acta Math.*, 124(1):9–36, 1970.

[15] S. Gan, L. Guth, and C. Oh. Restriction estimates for quadratic manifolds of arbitrary codimensions. *arXiv.2308.06427v1*, 2023.

[16] P. T. Gressman. On the oberlin affine curvature condition. *Duke Math. J.*, 168(11):2075–2126, 2019.

[17] S. Guo and C. Oh. Fourier restriction estimates for surfaces of co-dimension two in $\mathbb{R}^5$. *J. Anal. Math.*, 148(2):471–499, 2022.

[18] S. Guo, C. Oh, J. Roos, P.-L. Yung, and P. Zorin-Kranich. Decoupling for two quadratic forms in three variables: a complete characterization. *Rev. Mat. Iberoam.*, 39(1):283–306, 2023.

[19] S. Guo, C. Oh, R. Zhang, and P. Zorin-Kranich. Decoupling inequalities for quadratic forms. *Duke Math. J.*, 172(2):387–445, 2023.

[20] S. Guo, H. Wang, and R. Zhang. A dichotomy for Hörmander-type oscillatory integral operators. *arXiv:2210.05851*, 2022.

[21] L. Guth. A restriction estimate using polynomial partitioning. *J. Amer. Math. Soc.*, 29(2):371–413, 2016.

[22] L. Guth. Restriction estimates using polynomial partitioning II. *Acta Math.*, 221(1):81–142, 2018.

[23] L. Guth, N. Solomon, and H. Wang. Incidence estimates for well spaced tubes. *Geom. Funct. Anal.*, 29(6):1844–1863, 2019.

[24] J. Hickman and M. Rogers. Improved Fourier restriction estimates in higher dimensions. *Camb. J. Math.*, 7(3):219–282, 2019.

[25] J. Hickman and J. Zahl. A note on Fourier restriction and nested Polynomial Wolff axioms. *J. Anal. Math.*, 2023.

[26] J. Lee and S. Lee. Restriction estimates to complex hypersurfaces. *J. Math. Anal. Appl.*, 506(2):125702, 2020.

[27] D. Maldague. Regularized Brascamp-Lieb inequalities and an application. *Quarterly J. Math.*, 73(1):311–331, 2022.

[28] G. Mockenhaupt. Bounds in Lebesgue spaces of oscillatory integral operators. *Habilitationsschrift, Universität Siegen*, 1996.

[29] D. M. Oberlin. Convolution and restriction estimates for a 3-surface in $\mathbb{R}^5$. *J. Fourier Anal. Appl.*, 10(4):377–382, 2004.

[30] D. M. Oberlin. A restriction theorem for a k-surface in $\mathbb{R}^n$. *Canad. Math. Bull.*, 48(2):260–266, 2005.

[31] E. M. Stein. Some problems in harmonic analysis. In *Harmonic analysis in Euclidean spaces, Proceedings of the Symposium in Pure Mathematics of the Amer. Math. Soc., Williams College, Mass, Proc. Sympos. Pure Math., XXXV Part I, 1979*, pages 3–20. Amer. Math. Soc., 1979.

[32] E. M. Stein. Oscillatory integrals in Fourier analysis (Beijing, 1984). In *Beijing lectures in harmonic analysis*, page 307–355. Ann. of Math. Stud., Princeton University Press, Princeton, NJ, 1986.

[33] T. Tao. A sharp bilinear restriction estimate for paraboloids. *Geom. Funct. Anal.*, 13(6):1359–1384, 2003.

[34] T. Tao, A. Vargas, and L. Vega. A bilinear approach to the restriction and Kakeya conjectures. *J. Amer. Math. Soc.*, 11(4):967–1000, 1998.

[35] P. A. Tomas. A restriction theorem for the Fourier transform. *Bull. Amer. Math. Soc.*, 81:477–478, 1975.

[36] H. Wang. A restriction estimate in $\mathbb{R}^3$ using brooms. *Duke Math. J.*, 171(8):1749–1822, 2022.

[37] H. Wang and S. Wu. An improved restriction in $\mathbb{R}^3$. *arXiv:2210.03878v2*, 2022.

[38] A. Zygmund. On Fourier coefficients and transforms of functions of two variables. *Studia Math.*, 50(2):189–201, 1974.

Institute of Mathematics, Henan Academy of Sciences, Zhengzhou 450046, China

Email address: `caozhenbin@hnas.ac.cn`

Institute of Applied Physics and Computational Mathematics, Beijing 100088, China

Email address: `miao_changxing@iapcm.ac.cn`

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104

Email address: pyixuan@sas.upenn.edu