ANALYSIS & PDE

Volume 11 No. 6 2018

SANGHYUK LEE

SQUARE FUNCTION ESTIMATES FOR THE BOCHNER–RIESZ MEANS

msp

ANALYSIS AND PDE
Vol. 11, No. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1535

SQUARE FUNCTION ESTIMATES FOR THE BOCHNER–RIESZ MEANS

SANGHYUK LEE

We consider the square-function (known as Stein's square function) estimate associated with the Bochner–Riesz means. The previously known range of the sharp estimate is improved. Our results are based on vector-valued extensions of Bennett, Carbery and Tao's multilinear (adjoint) restriction estimate and an adaptation of an induction argument due to Bourgain and Guth. Unlike the previous work by Bourgain and Guth on L^p boundedness of the Bochner–Riesz means in which oscillatory operators associated to the kernel were studied, we take more direct approach by working on the Fourier transform side. This enables us to obtain the correct order of smoothing, which is essential for obtaining the sharp estimates for the square functions.

1. Introduction

We consider the Bochner–Riesz mean of order α, which is defined by

$$\widehat{\mathcal{R}_t^\alpha f}(\xi) = \left(1 - \frac{|\xi|^2}{t^2}\right)_+^\alpha \hat{f}(\xi), \quad t > 0,\ \xi \in \mathbb{R}^d,\ d \geq 2.$$

Let $1 \leq p \leq \infty$. The Bochner–Riesz conjecture is that the estimate

$$\|\mathcal{R}_t^\alpha f\|_p \leq C\|f\|_p \tag{1}$$

holds (except for $p = 2$) if and only if

$$\alpha > \alpha(p) = \max\left(d\left|\frac{1}{2} - \frac{1}{p}\right| - \frac{1}{2}, 0\right). \tag{2}$$

The Bochner–Riesz mean, which is a kind of summability method, has been studied in order to understand convergence properties of Fourier series and integrals. In fact, for $1 \leq p < \infty$, L^p boundedness of $\mathcal{R}_t^\alpha$ implies $\mathcal{R}_t^\alpha f \to f$ in L^p as $t \to \infty$. The necessary condition (2) has been known for a long time [Fefferman 1971; Stein 1993, p. 389].

When $d = 2$, the conjecture was verified by Carleson and Sjölin [1972]; also see [Fefferman 1971]. In higher dimensions $d \geq 3$, the problem is still open and partial results are known. The conjecture was shown to be true for $\max(p, p') \geq 2(d+1)/(d-1)$ by an argument due to Stein [Fefferman 1970], also see [Stein 1993, Chapter 9], and the sharp $L^2 \to L^{\frac{2(d+1)}{d-1}}$ restriction estimate (the Stein–Tomas theorem) for the sphere [Tomas 1975; Stein 1986]. It was Bourgain [1991a; 1991b] who first made progress beyond this result when $d = 3$. Since then, subsequent progress has been tied to that of the restriction problem. Bilinear or multilinear generalizations under transversality assumptions have turned out to be the most effective and

MSC2010: primary 42B15; secondary 35B65.
Keywords: square function, Bochner–Riesz means.

fruitful tools. These results have propelled progress in this area and there is a large body of literature on restriction estimates and related problems. See [Tao et al. 1998; Tao and Vargas 2000a; Wolff 2001; Tao 2003; Lee and Vargas 2008; 2010; Lee 2004; 2006a; 2006b] for bilinear restriction estimates and related results, [Bennett et al. 2006; 2015; Bourgain and Guth 2011; Lee and Vargas 2012; Bourgain 2013; Temur 2014; Bourgain and Demeter 2015; Bennett 2014; Bejenaru 2016; 2017; Ramos 2016] for multilinear restriction estimates and their applications, and [Guth 2016a; 2016b; Shayya 2017; Du et al. 2017; Zhang 2015; Ou and Wang 2017] for the most recent developments related to the polynomial partitioning method.

Concerning improved L^p boundedness of the Bochner–Riesz means in higher dimensions, the sharp L^p bounds for the Bochner–Riesz operator on the range $\max(p, p') \geq 2(d+2)/d$ were established by the author [Lee 2004], making use of the sharp bilinear restriction estimate due to Tao [2003]. When $d \geq 5$, further progress was recently made by Bourgain and Guth [2011]. They improved the range of the sharp (linear) estimates for the oscillatory integral operators of Carleson–Sjölin-type with phases that additionally satisfy an elliptic condition (see [Stein 1986; Bourgain 1991c; Lee 2006a] for earlier results) by using the multilinear estimates for oscillatory integral operators due to Bennett, Carbery and Tao [Bennett et al. 2006] and a factorization theorem. Also see [Carleson and Sjölin 1972; Hörmander 1973; Stein 1986; 1993, Chapter 11] for the relationship between the Bochner–Riesz problem and the oscillatory integral operators of Carleson–Sjölin-type.

The following is currently the best known result for the sharp L^p boundedness of the Bochner–Riesz operator.

Theorem 1.1 [Carleson and Sjölin 1972; Lee 2004; Bourgain and Guth 2011]. *Let* $d \geq 2,\ p \in [1, \infty]$, *and* $p_\circ$ *be defined by*

$$p_\circ = p_\circ(d) = 2 + \frac{12}{4d-3-k} \quad \textit{if } d \equiv k \pmod 3,\ k = -1, 0, 1. \tag{3}$$

If $\max(p, p') \geq p_\circ$, *then* (1) *holds for* $\alpha > \alpha(p)$.[†]

There are also results concerning the endpoint estimates at the critical exponent $\alpha = \alpha(p)$; for example, see [Christ 1987; 1988; Seeger 1996; Tao 1996]. It was shown by Tao [1998] that the sharp L^p bounds of $\mathcal{R}_t^\alpha$ for $1 < p < p_\circ < 2d/(d-1)$ imply the weak-type bounds of $\mathcal{R}_t^{\alpha(p)}$ for $1 < p < p_\circ$. We refer interested readers to [Lee et al. 2014] for variants and related problems.

Square function estimate. We now consider the square function $\mathcal{G}^\alpha f$ which is defined by

$$\mathcal{G}^\alpha f(x) = \Big(\int_0^\infty \Big|\frac{\partial}{\partial t}\mathcal{R}_t^\alpha f(x)\Big|^2 t\,dt\Big)^{\frac{1}{2}}.$$

It was introduced by Stein [1958] to study the almost-everywhere summability of Fourier series. Due to the derivative in t, the square function behaves as if it is a multiplier of order $\alpha - 1$ and the derivative $\partial/\partial t$ makes an L^p estimate possible by mitigating bad behavior near the origin. In this paper we are concerned

[†]For the sharp bound for $\max(p, p') \geq p_*$, we have the following relationships: bilinear, $p_* = 2 + 4/d$; multilinear, $p_* = 3 + 3/d + O(d^{-2})$; conjecture, $p_* = 2 + 2/d + O(d^{-2})$.

with the estimate

$$\|\mathcal{G}^\alpha f\|_p \le C\|f\|_p. \tag{4}$$

The L^p estimate for the square function has various consequences and applications. First of all, it is related to smoothing estimates for solutions to dispersive equations associated to radial symbols such as wave and Schrödinger operators. See [Lee et al. 2012; 2013] for the details; see also Remark 3.3. The sharp square-function estimate implies the sharp maximal bounds for Bochner–Riesz means, which will be discussed below in connection to pointwise convergence. It also gives L^p and maximal L^p boundedness of general radial Fourier multipliers, especially the sharp L^p boundedness result of Hörmander–Mikhlin-type; see Corollary 1.3 below and [Carbery et al. 1984; Carbery 1985; Lee et al. 2012].

For $1 < p \le 2$, the inequality (4) is well understood. In this range of p, we know $\mathcal{G}^\alpha$ is bounded on L^p if and only if $\alpha > d\big(\frac{1}{p} - \frac{1}{2}\big) + \frac{1}{2}$; see [Sunouchi 1967; Lee et al. 2014]. Sufficiency can be shown by using vector-valued Calderón–Zygmund theory. In contrast with the case $1 < p \le 2$, if $p > 2$, due to a smoothing effect resulting from averaging in time, the problem has more interesting features and may be considered as a vector-valued extension of the Bochner–Riesz conjecture in that its sharp L^p bound also implies that of the Bochner–Riesz operator. The condition $\alpha > \max\big\{\frac{1}{2}, d\big(\frac{1}{2} - \frac{1}{p}\big)\big\}$ is known to be necessary for (4), see, for example [Lee et al. 2014], and it is natural to conjecture that this is also sufficient for $p > 2$. This conjecture in two dimensions was proven by Carbery [1983], and in higher dimensions $d \ge 3$, sharp estimates for $p > 2(d+1)/(d-1)$ were obtained by Christ [1985] and Seeger [1986] and it was later improved to the range of $p \ge 2(d+2)/d$ by the author, Rogers, and Seeger [Lee et al. 2012]. There are also endpoint estimates at the critical exponent $\alpha = d/2 - d/p$ and weaker $L^{p,2} \to L^p$ endpoint estimates were obtained in [Lee et al. 2014] for $2(d+1)/(d-1) < p < \infty$.

There are two notable approaches for the study of the Bochner–Riesz problem. One, which may be called the *spatial-side approach*, is to prove the sharp estimates for the oscillatory integral operators of Carleson–Sjölin-type [Carleson and Sjölin 1972; Hörmander 1973; Stein 1986]. These operators are natural variable coefficient generalizations of the adjoint restriction operators [Bourgain 1991c; Lee 2006a; Wisewell 2005] for hypersurfaces with nonvanishing Gaussian curvature such as spheres, paraboloids, and hyperboloids. The other, which we may call the *frequency-side approach*, is more related to the Fourier transform side, based on a suitable decomposition in the frequency side and orthogonality between the decomposed pieces [Fefferman 1973; Carbery 1983; Christ 1985; 1987; Seeger 1996; Tao 1998; Lee 2004]. As has been demonstrated in related works, the latter approach makes it possible to carry out finer analysis and to obtain refined results such as the sharp maximal bounds, square-function estimates, and various endpoint estimates.

The recently improved bound for the Bochner–Riesz operator in [Bourgain and Guth 2011] was obtained from the sharp estimate for the oscillatory integral operators of Carleson–Sjölin-type with an additional elliptic assumption. However, this approach doesn't seem appropriate for the study of the square function. In particular, there is an obvious difficulty when one tries to make use of the disjointness of the singularity of the Fourier transform of $\mathcal{R}_t^\alpha f$ which occurs as t varies; for example, see (76). This is where the extra smoothing of order $\frac{1}{2}$ for the square-function estimate comes in, which is most important for the sharp estimates for $\mathcal{G}^\alpha f$ [Carbery 1983; Christ 1985; Lee 2004; Lee et al. 2012]. This kind of smoothing can be seen clearly in the Fourier transforms of Bochner–Riesz means but is not easy to exploit in the oscillatory kernel side. As is already known [Bourgain 1991c; Wisewell 2005; Lee 2006a; Bourgain and Guth 2011], the behavior of the oscillatory integral operators of Carleson–Sjölin-type are more subtle and generally considered to be difficult to analyze when compared to their constant-coefficient counterparts, the adjoint restriction operators. So, we take the frequency-side approach, in which we directly handle the associated multiplier by working in the frequency space rather than dealing with the oscillatory integral operator given by the kernel of the Bochner–Riesz operator.

In this paper, we obtain the sharp square-function estimates which are new when $d \ge 9$.

Theorem 1.2. *Let us define* $p_s = p_s(d)$ *by*

$$p_s = 2 + \frac{12}{4d-6-k}, \quad d \equiv k \pmod 3),\ k = 0, 1, 2. \tag{5}$$

Then, if $p \ge \min(p_s, 2(d+2)/d)$ *and* $\alpha > d/2 - d/p$, *the estimate* (4) *holds.*

The range here does not match that of Theorem 1.2. This results from an additional time average which increases the number of decomposed frequency pieces. (See Section 3F.)

Maximal estimate and pointwise convergence. A straightforward consequence of the estimate (4) is the maximal estimate

$$\Big\| \sup_{t>0} |\mathcal{R}_t^\alpha f| \Big\|_p \le C \|f\|_p \tag{6}$$

for $\alpha > \alpha(p)$, which follows from Sobolev imbedding and (4). Hence, Theorem 1.2 yields the sharp maximal bounds for $p \ge p_s(d)$. When $p \ge 2$, it has been conjectured that (6) holds as long as (2) is satisfied. The sharp L^2 bound goes back to Stein [1958]. The conjecture in $\mathbb{R}^2$ and the sharp bounds for $p > 2(d+1)/(d-1)$, $d \ge 3$, were verified by square-function estimates [Christ 1985; Seeger 1986]. The bounds were later improved to the range $p > 2(d+2)/d$ by the author [Lee 2004] using an $L^p \to L^p(L_t^4)$ estimate. The inequality (6) has been studied in connection with almost-everywhere convergence of Bochner–Riesz means. However, the problem of showing $\mathcal{R}_t^\alpha f \to f$ a.e. for $f \in L^p$, $p > 2$, $\alpha > \alpha(p)$, was settled by Carbery, Rubio de Francia and Vega [Carbery et al. 1988]. Their result relies on weighted L^2 estimates. There are also results on pointwise convergence at the critical $\alpha = \alpha(p)$. See [Lee and Seeger 2015; Annoni 2017]. When $1 < p < 2$, by Stein's maximal theorem almost-everywhere convergence of $\mathcal{R}_t^\alpha f \to f$ for $f \in L^p$ is equivalent to the $L^p \to L^{p,\infty}$ estimate for the maximal operator and it was shown by Tao [1998] that the stronger condition $\alpha \ge (2d-1)/(2p) - d/2$ is necessary for (6). Except for $d = 2$ [Tao 2002], little is known beyond the classical result which follows from interpolation between L^2 ($\alpha > 0$) and L^1 ($\alpha > (d-1)/2$) estimates.

Radial multiplier. Let m be a function defined on $\mathbb{R}_+$. Combining the inequality due to Carbery, Gasper and Trebels [Carbery et al. 1984] and Theorem 1.2, we obtain the following L^p boundedness result of Hörmander–Mikhlin-type, which is sharp in that the regularity assumption cannot be improved. A similar result for the maximal function $f \to \sup_{t>0} |\mathcal{F}^{-1}(m(t|\cdot|)\hat{f})|$ is also possible thanks to the inequality due to Carbery [1985].

Corollary 1.3. *Let $d \geq 2$, and φ be a nontrivial smooth function with compact support contained in $(0,\infty)$. If $\min(p_s, 2(d+2)/d) \leq \max(p, p') < \infty$ and $\alpha > d\left|\frac{1}{p}-\frac{1}{2}\right|$, then*

$$\left\|\mathcal{F}^{-1}[m(|\cdot|)\hat{f}]\right\|_p \lesssim \sup_{t>0} \|\varphi m(t\,\cdot)\|_{L^2_\alpha(\mathbb{R})}\|f\|_p.$$

About the paper. In Section 2, by working in the frequency side we provide an alternative proof of Theorem 1.1. Although, this doesn't give an improvement over the current range, we include this because it has some new consequences, clarifies several issues which were not clearly presented in [Bourgain and Guth 2011], and provides preparation for Section 3, in which we work in a vector-valued setting. The proof in that paper is sketchy and doesn't look readily accessible. Also the heuristic that a function with Fourier support in a ball of radius σ behaves as if it is constant on balls of radius $\frac{1}{\sigma}$ is now widely accepted and has important role in the induction argument but it doesn't seem justified at high level of rigor. We provide a rigorous argument by making use of Fourier series (see Lemmas 2.13 and 3.14). Another problem of the induction argument is that the primary object (the associated surfaces or phase functions) changes in the course of induction. However, these issues are not properly addressed in literature. We handle this matter by introducing a stronger induction assumption (see Remark 2.4) and carefully handling the stability of various estimates. We also use a different type of multilinear decomposition which is more systematic, easier and more efficient for dealing with multiplier operators (see Section 2E, especially the discussion at the beginning of Section 2E).

Section 3 is very much built on the frequency-side analysis in Section 2, as it may be regarded a vector-valued extension of Section 2. Consequently, the structure of Section 3 is similar to that of Section 2 and some of the arguments commonly work in both sections. In such cases we try to minimize repetition while keeping readability as much as possible. We first obtain vector-valued extensions of multilinear estimates (Propositions 3.6 and 3.10) which serve as basic estimates for the sharp-square function estimate. Then, to derive the linear estimate (Theorem 1.2) we adapt the frequency side approach in Section 2 to the vector-valued setting and prove our main theorem.

Finally, the oscillatory integral approach has its own limits for proving Bochner–Riesz conjecture. As is now well known [Bourgain 1991c; Wisewell 2005; Lee 2006a; Bourgain and Guth 2011], the sharp L^p–L^q estimates for the oscillatory operators of Carleson–Sjölin-type fail for $q < q_\circ,\ q_\circ > 2d/(d-1)$, even under the elliptic condition on the phase [Wisewell 2005; Lee 2006a; Bourgain and Guth 2011]. The Fourier-transform-side approach may help further development in a different direction and thanks to its flexibility may have applications to related problems.

Notation. The following is a list of notation we frequently use in the rest of the paper:

- C, c are constants which depend only on d and may differ at each occurrence.
- For $A, B \geq 0$, we say $A \lesssim B$ if there is a constant C such that $A \leq CB$.
- $I = [-1, 1]$ and $I^d = [-1, 1]^d \subset \mathbb{R}^d$.
- $\tau_h f(x) = f(x-h)$ and $\tau_i f$ denotes $\tau_{h_i} f$ for some $h_i \in \mathbb{R}^d,\ i = 1, \ldots, m$.

- We denote by $\mathfrak{q}(a,\ell) \subset \mathbb{R}^d$ the closed cube centered at a with side length 2ℓ, namely, $a + \ell I^d$. If $\mathfrak{q} = \mathfrak{q}(a,\ell)$, denote a, the center of $\mathfrak{q}$, by $\boldsymbol{c}(\mathfrak{q})$.
- For $r > 0$ and a given cube or rectangle Q, we denote by rQ the cube or rectangle which is the r-times dilation of Q from the center of Q.
- Let $\rho \in \mathcal{S}(\mathbb{R}^d)$ be a function with Fourier support in $\mathfrak{q}(0,1)$ and $\rho \geq 1$ on $\mathfrak{q}(0,1)$. And we also set $\rho_{B(z,r)}(x) := \rho((\cdot - z)/r)$.
- For a given set $A \subset \mathbb{R}^d$, we define the set $A + O(\delta)$ by
$$A + O(\delta) := \{x \in \mathbb{R}^d : \operatorname{dist}(x, A) < C\delta\}.$$
- For a given dyadic cube $\mathfrak{q}$ and function f, we define $f_{\mathfrak{q}}$ by $\hat{f}_{\mathfrak{q}} = \chi_{\mathfrak{q}} \hat{f}$.
- Besides $\hat{}$ and $\check{}$, we also denote by $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ the Fourier transform and the inverse Fourier transform, respectively.
- For a smooth function G on I^k, we set $\|G\|_{C^N(I^k)} := \max_{|\alpha| \leq N} \max_{x \in I^k} |\partial^\alpha G(x)|$.

2. Estimates for multiplier operators

In this section we consider the multiplier operators of Bochner–Riesz-type which are associated with elliptic-type surfaces. They are natural generalizations of the Bochner–Riesz operator $\mathcal{R}_1^\alpha$. We prove the sharp L^p boundedness of these of operators and this provides an alternative proof of Theorem 1.1. Basically we adapt the induction argument in [Bourgain and Guth 2011]. However, compared to the (adjoint) restriction counterpart, the induction argument becomes less obvious when we consider it for the Fourier multiplier operator. However, exploiting sharpness of bounds for the frequency-localized operator T_δ, see (9)–(10), we manage to carry out a similar argument. See Section 2F.

From now on we write
$$\xi = (\zeta, \tau) \in \mathbb{R}^{d-1} \times \mathbb{R}.$$
Let ψ be a smooth function defined on I^d and $\chi_\circ$ be a smooth function supported in a small neighborhood of the origin. We consider the multiplier operator $T^\alpha = T^\alpha(\psi)$ which is defined by
$$\mathcal{F}(T^\alpha f)(\xi) = (\tau - \psi(\zeta))_+^\alpha \chi_\circ(\xi) \hat{f}(\xi).$$
By a finite decomposition, rotation and translation and by discarding the harmless smooth multiplier, it is easy to see that the L^p boundedness of $\mathcal{R}_1^\alpha$ is equivalent to that of T^α, which is given by $\psi(\zeta) = 1 - (1 - |\zeta|^2)^{\frac{1}{2}}$. A natural generalization of the Bochner–Riesz problem is as follows: If $\det H\psi \neq 0$ on the support of $\chi_\circ$ (here, $H\psi$ is the Hessian matrix of ψ), we may conjecture that, for $1 \leq p \leq \infty$, $p \neq 2$,
$$\|T^\alpha f\|_p \leq C\|f\|_p \tag{7}$$
if and only if $\alpha > \alpha(p)$. From explicit computation of the kernel of T^α it is easy to see that the condition $\alpha > \alpha(p)$ is necessary for (7). However, in this paper we only work with specific choices of ψ.

2A. ***Elliptic function.*** Let us set

$$\psi_\circ(\zeta) = \tfrac{1}{2}|\zeta|^2.$$

For $0 < \varepsilon_\circ \ll \frac{1}{2}$ and an integer $N \geq 100d$ we denote by $\mathfrak{G}(\varepsilon_\circ, N)$ the collection of smooth functions which is given by

$$\mathfrak{G}(\varepsilon_\circ, N) = \{\psi : \|\psi - \psi_\circ\|_{C^N(I^{d-1})} \leq \varepsilon_\circ\}.$$

If $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $a \in \frac{1}{2}I^{d-1}$, then $H\psi(a)$ has eigenvalues $\lambda_1, \dots, \lambda_{d-1}$ close to 1 and we may write $H\psi(a) = P^{-1}DP$ for an orthogonal matrix P, while D is a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_{d-1}$. We denote by $\sqrt{H\psi(a)}$ the matrix $P^{-1}D'P$, where D' is the diagonal matrix with diagonal entries $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{d-1}}$. So, $(\sqrt{H\psi(a)})^2 = H\psi(a)$.

For $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, $a \in \frac{1}{2}I^{d-1}$, and $0 < \varepsilon \leq \frac{1}{2}$, we define

$$\psi_a^\varepsilon(\zeta) = \frac{1}{\varepsilon^2}\big(\psi\big(\varepsilon[\sqrt{H\psi(a)}]^{-1}\zeta + a\big) - \psi(a) - \varepsilon\nabla\psi(a)\cdot[\sqrt{H\psi(a)}]^{-1}\zeta\big). \tag{8}$$

Since $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, by Taylor's theorem it is easy to see that $\|\psi_a^\varepsilon - \psi_\circ\|_{C^N(I^{d-1})} \leq C\varepsilon$ for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.[†] Hence we get the following.

Lemma 2.1. *Let $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $a \in \frac{1}{2}I^{d-1}$. Then there is a constant $\kappa = \kappa(\varepsilon_\circ, N)$, independent of a, ψ, such that $\psi_a^\varepsilon \in \mathfrak{G}(\varepsilon_\circ, N)$ provided that $0 < \varepsilon \leq \kappa$.*

Remark 2.2. If ψ is smooth and $H\psi(a)$ has $d-1$ positive eigenvalues, after finite decomposition and affine transformations we may assume $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ for arbitrarily small $\varepsilon_\circ$ and large N. Indeed, for given $\varepsilon > 0$, decomposing the multiplier $(\tau - \psi(\zeta))_+^\alpha \chi_\circ(\xi)$ to multipliers supported in balls of small radius ε/C with some large C, one may assume that $\mathcal{F}f$ is supported in $B((a, \psi(a)), \varepsilon/C)$. Then, the change of variables (12) transforms $\psi \to \psi_a^\varepsilon$ and gives rise to a new multiplier operator $T^\alpha(\psi_a^\varepsilon)$ and, as can be easily seen by a simple change of variables, the operator norm $\|T^\alpha(\psi_a^\varepsilon)\|_{p\to p}$ remains same. (See the proof of Proposition 2.5.) By Lemma 2.1 we see $\psi_a^\varepsilon \in \mathfrak{G}(\varepsilon_\circ, N)$ if ε is small enough.

2B. ***Multiplier operator with localized frequency.*** Let ϕ be a smooth function supported in $2I$. For $\delta > 0$, $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and f with Fourier transform supported in $\frac{1}{2}I^d$ we define the (frequency-localized) multiplier operator $T_\delta = T_\delta(\psi)$ by

$$\widehat{T_\delta f}(\xi) = \phi\left(\frac{\tau - \psi(\zeta)}{\delta}\right)\hat{f}(\xi). \tag{9}$$

As is well known, the L^p bound for T_δ largely depends on the curvature of the surface $\tau = \psi(\zeta)$. By decomposing the multiplier dyadically away from the singularity $\tau = \psi(\zeta)$, in order to prove (7) for $p > 2d/(d-1)$ and $\alpha > \alpha(p)$, it is enough to show that, for any $\varepsilon > 0$,

$$\|T_\delta f\|_p \leq C\delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon}\|f\|_p \tag{10}$$

whenever $\hat{f}$ is supported in $\frac{1}{2}I^d$. The following recovers the sharp L^p bound up to the currently best known range in [Bourgain and Guth 2011].

[†]Indeed, since $|\partial^\alpha(\psi_a^\varepsilon - \psi_\circ)| \lesssim \varepsilon^{|\alpha|-2}$ for any multi-index α, we need only to show $|\partial^\alpha(\psi_a^\varepsilon - \psi_\circ)| \lesssim \varepsilon$ for $|\alpha| = 0, 1, 2$. This follows by Taylor's theorem since $N \geq 100d$.

Proposition 2.3. *Let $\varepsilon > 0$. If $p \ge p_\circ(d)$ and $\varepsilon_\circ$ is small enough, there is an $N = N(\varepsilon)$ such that* (10) *holds uniformly provided that $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and* $\operatorname{supp} \hat f \subset \frac{1}{2} I^d$.

It is possible to remove the loss of $\delta^{-\varepsilon}$ in (10) by the ε-removal argument in [Tao 1998, Section 4].

Induction quantity. To control the L^p norm of T_δ, for $0 < \delta$, we define $A(\delta) = A_p(\delta)$ by

$$A(\delta) = \sup\big\{\|T_\delta(\psi) f\|_{L^p} : \psi \in \mathfrak{G}(\varepsilon_\circ, N),\ \|f\|_p \le 1,\ \operatorname{supp} \hat f \subset \tfrac{1}{2} I^d\big\}.$$

Remark 2.4. Though the induction argument in [Bourgain and Guth 2011] heavily relies on the stability of the multilinear estimates, such issue doesn't seem properly addressed. In particular, after (multiscale) decomposition and rescaling, the associated phase functions (or surfaces) are no longer fixed-phase functions (or surfaces).† This requires the induction quantity defined over a class of phase functions or surfaces. This leads us to consider $A(\delta)$.

From the estimate for the kernel of T_δ (see Lemma 2.9), it is easy to see that $A(\delta) \le C$ uniformly in $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ if $\delta \ge 1$ and $A(\delta) \le C\delta^{-\frac{d-1}{2}}$ if $0 < \delta \le 1$, because the L^1-norm of the kernel is uniformly $O(\delta^{-\frac{d-1}{2}})$. To prove Proposition 2.3, we need to show $A(\delta) \lesssim \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon}$ for any $\varepsilon > 0$. However, due to the lack of monotonicity $A(\delta)$ is not suitable for closing the induction. So, we need to modify $A(\delta)$. For $\beta, \delta > 0$, we define

$$\mathcal{A}^\beta(\delta) = \mathcal{A}_p^\beta(\delta) := \sup_{\delta < s \le 1} s^{\frac{d-1}{2} - \frac{d}{p} + \beta} A_p(s).$$

Hence, Proposition 2.3 follows if we show $\mathcal{A}^\beta(\delta) \le C$ for any $\beta > 0$.

The following lemma makes precise the heuristic that the bound of T_δ improves if it acts on functions with Fourier transforms supported in a smaller set. However, this becomes less obvious for the multiplier operator when it is compared to the restriction (adjoint) operator; see [Bourgain and Guth 2011]. This type of improvement is basically due to the parabolic rescaling structure of the operator, and generally appears in $L^p - L^q$ estimates for p, q satisfying $(d+1)/q < (d-1)(1-1/p),\ p \le q$, which are not invariant under the parabolic rescaling. The following is important for the induction argument to work.

Proposition 2.5. *Let $0 < \delta \ll 1$, $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and $(a, \mu) \in \mathbb{R}^{d-1} \times \mathbb{R}$. Suppose that* $\operatorname{supp} \hat f \subset \mathfrak{q}((a, \mu), \varepsilon) \subset \frac{1}{2} I^d$, $0 < \varepsilon < \frac{1}{2}$ *and $\delta \le (10)^{-2}\varepsilon^2$. Then, there is a $\kappa = \kappa(\varepsilon_\circ, N)$ such that for $0 < \varepsilon \le \kappa$*

$$\|T_\delta f\|_p \le C A(\varepsilon^{-2}\delta)\|f\|_p \tag{11}$$

holds with C independent of ψ and ε.

Proof. Decomposing $\mathfrak{q}(a, \varepsilon)$ into as many as $O(d^d)$, we may assume $\hat f$ is supported in $\mathfrak{q}((a, \mu), \varepsilon/(10d))$. Since $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $\operatorname{supp} \hat f \subset \mathfrak{q}((a, \mu), \varepsilon/(10d))$, by Taylor's theorem we have $\phi((\tau - \psi(\zeta))/\delta)\hat f(\xi)$, which is supported in the set

$$\left\{(\xi, \tau) : |\tau - \psi(a) - \nabla\psi(a) \cdot (\zeta - a)| \le \frac{(1+\varepsilon_0)\varepsilon^2}{2 \times 10^2}\right\}.$$

† It is only true for the paraboloid.

Hence, we may write

$$\phi\Big(\frac{\tau-\psi(\zeta)}{\delta}\Big)\hat f(\xi)=\phi\Big(\frac{\tau-\psi(\zeta)}{\delta}\Big)\tilde\chi\Big(\frac{\tau-\psi(a)-\nabla\psi(a)\cdot(\zeta-a)}{\varepsilon^2}\Big)\hat f(\xi),$$

where $\tilde\chi$ is a smooth function supported in $\frac{1}{2}I$ such that $\tilde\chi=1$ on $\frac{1}{4}I$. Let us set $M=(\sqrt{H\psi(a)})^{-1}$ and make the change of variables in the frequency domain

$$(\zeta,\tau)\to L(\zeta,\tau)=(\varepsilon M\zeta+a,\varepsilon^2\tau+\psi(a)+\varepsilon\nabla\psi(a)\cdot M\zeta). \tag{12}$$

Then it follows that

$$\mathcal{F}(T_\delta(\psi)f)(L\xi)=\phi\Big(\frac{\tau-\psi_a^\varepsilon(\zeta)}{\varepsilon^{-2}\delta}\Big)\tilde\chi(\tau)\hat f(L\xi).$$

Since L is an invertible affine transformation, it is easy to see

$$\|\mathcal{F}^{-1}(\hat g(L\cdot))\|_p=\varepsilon^{(d+1)(\frac{1}{p}-1)}\|g\|_p$$

for any g. We also note that $\operatorname{supp}(\tilde\chi(\tau)\hat f(L\cdot))\subset\frac{1}{2}I^d$ and by Lemma 2.1 there exists a $\kappa>0$ such that $\psi_a^\varepsilon\in\mathfrak{G}(\varepsilon_\circ,N)$ if $0<\varepsilon\le\kappa$ whenever $\psi\in\mathfrak{G}(\varepsilon_\circ,N)$. So, by the definition of $A(\delta)$ it follows that, for $0<\varepsilon\le\kappa$,

$$\begin{aligned}\|T_\delta(\psi)f\|_p&=\varepsilon^{(d+1)(1-\frac{1}{p})}\Big\|\mathcal{F}^{-1}\Big(\phi\Big(\frac{\tau-\psi_a^\varepsilon(\zeta)}{\varepsilon^{-2}\delta}\Big)\tilde\chi(\tau)\hat f(L\xi)\Big)\Big\|_p\\&\le\varepsilon^{(d+1)(1-\frac{1}{p})}A(\varepsilon^{-2}\delta)\|\mathcal{F}^{-1}(\tilde\chi(\tau)\hat f(L\xi))\|_p\le CA(\varepsilon^{-2}\delta)\|f\|_p.\end{aligned}$$

For the last inequality we also use the trivial bound $\|\mathcal{F}^{-1}(\tilde\chi(\tau)\hat g)\|_p\le C\|g\|_p$ for any $1\le p\le\infty$. The inequality is valid for any $\psi\in\mathfrak{G}(\varepsilon_\circ,N)$. This gives the desired bound. □

We will need the following estimate which is easy to show by making use of Rubio de Francia's one-dimensional inequality [Rubio de Francia 1985].

Lemma 2.6. *Let* $\{\mathfrak{q}\}$ *be a collection of* (*distinct*) *dyadic cubes of the same side length* σ. *Let* $2\le p<\infty$. *Then*, *there is a constant* C, *independent of the collection* $\{\mathfrak{q}\}$, *such that*

$$\Big(\sum_{\mathfrak{q}}\|\mathcal{F}^{-1}(\hat f\chi_{\mathfrak{q}})\|_p^p\Big)^{\frac{1}{p}}\le C\|f\|_p.$$

2C. *Multilinear estimates.* In this subsection we consider various multilinear estimates which are basically consequences of multilinear restriction and Kakeya estimates in [Bennett et al. 2006].

For $\psi\in\mathfrak{G}(\varepsilon_\circ,N)$ let us set

$$\Gamma=\Gamma(\psi)=\big\{(\zeta,\psi(\zeta)):\zeta\in\tfrac{1}{2}I^{d-1}\big\}.$$

Let $2\le k\le d$, and let $U_1,U_2,\dots,U_k$ be compact subsets of I^{d-1}. For $i=1,\dots,k$, and $\lambda>0$, set

$$\Gamma_i=\{(\zeta,\psi(\zeta)):\zeta\in U_i\},\quad \Gamma_i(\lambda)=\Gamma_i+O(\lambda).$$

For $\xi=(\zeta,\psi(\zeta))\in\Gamma(\psi)$, let $\mathrm{N}(\xi)$ be the upward unit normal vector at $(\zeta,\psi(\zeta))$.

For $v_1, \dots, v_k \in \mathbb{R}^d$, denote by $\mathrm{Vol}(v_1, \dots, v_k)$ the k-dimensional volume of the parallelepiped given by $\{s_1 v_1 + \cdots + s_k v_k : s_i \in [0,1],\ 1 \le i \le k\}$. Transversality among the surfaces $\Gamma_1, \dots, \Gamma_k$ is important for the multilinear estimates. The degree of transversality is quantitatively stated as follows:

$$\mathrm{Vol}(\mathrm{N}(\xi_1), \mathrm{N}(\xi_2), \dots, \mathrm{N}(\xi_k)) \ge \sigma \tag{13}$$

for some $\sigma > 0$ whenever $\xi_i \in \Gamma_i,\ i = 1, \dots, k$. Since $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, we know $\nabla\psi$ is a diffeomorphism which is close to the identity map. The condition (13) may be replaced by a simpler one: $\mathrm{Vol}(\zeta_1, \zeta_2, \dots, \zeta_k|) \gtrsim \sigma$ whenever $\zeta_i \in U_i,\ i = 1, \dots, k$. The following is due to Bennett, Carbery and Tao [Bennett et al. 2006].

Theorem 2.7. *Let $0 < \delta \ll \sigma \ll 1$ and $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$. Suppose that $\Gamma_1, \dots, \Gamma_k$ are given as in the above and (13) is satisfied whenever $\xi_i \in \Gamma_i,\ i = 1, \dots, k$, and suppose that $\widehat{F_i} \subset \Gamma_i(\delta),\ i = 1, \dots, k$. Then, if $p \ge 2k/(k-1)$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon > 0$ there are constants $N = N(\varepsilon)$ such that, for $x \in \mathbb{R}^d$,*

$$\Big\| \prod_{i=1}^{k} F_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon} \prod_{i=1}^{k} \delta^{\frac{1}{2}} \|F_i\|_2$$

holds with C, C_ε, independent of ψ.

Besides the stability issue, this estimate is essentially the same as the multilinear restriction estimate in [Bennett et al. 2006]; see Theorem 1.16 of that paper, as well as Lemma 2.2, for the case $k = d$ and see Section 5 for the case of lower linearity $2 \le k < d$. Though we are considering only the surfaces which are the graphs of $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, the theorem remains true for surfaces even with vanishing curvature as long as the transversality condition is satisfied. Uniformity of the estimate follows from the fact that the multilinear Kakeya and restriction estimates are stable under perturbation of the associated surfaces. The estimate is conjectured to be true without $\delta^{-\varepsilon}$ loss (this is equivalent with the endpoint k-linear restriction estimate) but it remains open when $k \ge 3$ even though the corresponding endpoint case for the multilinear Kakeya estimate was obtained by Guth [2010].

Remark 2.8. The proof of Theorem 2.7 is based on the multilinear Kakeya estimate and an induction-on-scale argument, which involves iteration of the induction assumption to reduce the exponent of δ^{-1}. Such an improvement of exponent is possible at the expense of extra loss of bounds in terms of σ^{-c}. By following the argument in [Bennett et al. 2006], one can easily see that one may take $C_\varepsilon \lesssim C \log \frac{1}{\varepsilon}$; see the paragraph below (20). Hence, the bound becomes less efficient when σ gets as small as δ^c for some $c > 0$. In $\mathbb{R}^3$ the sharp bound depending on σ was recently obtained by Ramos [2016]. However, the argument of Bourgain and Guth avoids such problems by keeping the Fourier supports of functions largely separated while being decomposed. In contrast with the conventional approach in which functions are usually decomposed into finer frequency pieces, this was achieved by decomposing the input functions into those of relatively large frequency supports.

Lemma 2.9. *Let $\varphi \in C_c^\infty(2I)$ and $\eta \in C_c^\infty(I^d)$, where $\frac{1}{2} \le \eta \le 2$. Let $0 < \delta \ll \sigma \le 1$. Set*

$$K_\delta = \mathcal{F}^{-1}\bigg(\varphi\Big(\frac{\eta(\xi)(\tau - \psi(\zeta))}{C\delta}\Big)\tilde{\chi}(\xi)\bigg),$$

and $\mathfrak{K}_M(x) = (1+\delta|x|)^{-M}$. *Suppose* $\tilde\chi$ *is supported in a cube of side length* $C\sigma$ *and* $|\partial_\xi^\alpha \tilde\chi| \lesssim \sigma^{-|\alpha|}$ *for any* α. *Then, for any* M, *there is an* $N = N(M)$ *such that*

$$|K_\delta(x)| \le C\delta\sigma^{d-1}\mathfrak{K}_M(x) \tag{14}$$

with C *depending only on* $\|\psi\|_{C^N(I^{d-1})}$.

Proof. Changing variables $\tau \to \delta\tau + \psi(\zeta)$, we write

$$K_\delta(x) = (2\pi)^{-d}\delta \int e^{i\delta\tau x_d} \int e^{i(x'\cdot\zeta + x_d\psi(\zeta))}\tilde\varphi(\xi)\,d\zeta\,d\tau,$$

where

$$\tilde\varphi(\xi) = \varphi\biggl(\frac{\eta(\zeta,\delta\tau+\psi(\zeta))\tau}{C}\biggr)\tilde\chi(\zeta,\delta\tau+\psi(\zeta)).$$

We note that $|\partial_\zeta^\alpha\tilde\varphi| \lesssim \sigma^{-|\alpha|}(\|\psi\|_{C^{|\alpha|}} + \|\eta\|_{C^{|\alpha|}})$. Then, if $|x'|/100 \ge |x_d|$, by integration by parts it follows that

$$\biggl|\int e^{i(x'\cdot\zeta + x_d\psi(\zeta))}\tilde\phi(\xi)\,d\zeta\biggr| \le C\sigma^{d-1}\bigl(\|\psi\|_{C^M(I^{d-1})} + \|\eta\|_{C^M(I^d)}\bigr)(1+\sigma|x'|)^{-M}.$$

Note that $\tilde\phi(\xi) = 0$ if $|\tau| \ge 5C$ since $\frac12 \le \eta \le 1$. This gives the desired inequality (14) by taking integration in τ since $\delta \ll \sigma$. On the other hand, if $|x'|/100 < |x_d|$, we integrate in τ first. Since $|\partial_\tau^l\tilde\varphi| \lesssim (\|\psi\|_{C^l} + \|\eta\|_{C^l})$, by integration by parts again we have

$$\biggl|\int e^{i\delta\tau x_d}\tilde\phi(\xi)\,d\tau\biggr| \le C\bigl(\|\psi\|_{C^M(I^{d-1})} + \|\eta\|_{C^M(I^d)}\bigr)(1+|\delta x_d|)^{-M}.$$

This and taking integration in ζ yield (14). □

From Theorem 2.7 and Lemma 2.9 we can obtain the sharp multilinear L^p estimate for T_δ under the transversality condition without localizing the multilinear operator on a ball of radius $\frac{1}{\delta}$. In fact, since $T_\delta f = K_\delta * f$ and the kernel K_δ (from Lemma 2.9) is rapidly decaying outside of $B(0, C/\delta)$, one may handle f as if it were supported in a ball B of radius $\delta^{-1-\varepsilon}$. This type of localization and Hölder's inequality make it possible to lift the L^2 estimate to that of L^p, $p \ge 2$, with sharp bound. Such an idea of deducing L^p estimates from L^2 ones goes back to Stein [1993, pp. 442–443], see also [Fefferman 1970; 1973], and in [Lee 2004; Lee et al. 2012] a similar idea was used to make use of the L^2 bilinear restriction estimate. The same argument also works with the multilinear estimates with a little modification. We make it precise in what follows.

Proposition 2.10. *Let* $0<\delta\ll\sigma\ll\tilde\sigma\ll 1$ *and* $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, *and let* $Q_1,\dots,Q_k \in \frac12 I^d$ *be dyadic cubes of side length* $\tilde\sigma$. *Suppose that* (13) *is satisfied whenever* $\xi_i \in \Gamma\cap Q_i$, $i=1,\dots,k$, *and* $\operatorname{supp}\hat f_i \subset Q_i$, $i=1,\dots,k$. *Then, if* $p \ge 2k/(k-1)$ *and* $\varepsilon_\circ$ *is small enough, for* $\varepsilon>0$ *there is an* $N = N(\varepsilon)$ *such that*

$$\biggl\|\prod_{i=1}^k T_\delta f_i\biggr\|_{L^{p/k}(\mathbb{R}^d)} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon}\prod_{i=1}^k \delta^{\frac{d}{p}-\frac{d-1}{2}}\|f_i\|_p \tag{15}$$

holds with C, C_ε *independent of* ψ.

Proof. Set $\widetilde{Q}_i = \{\xi : \mathrm{dist}(\xi, Q_i) \le \tilde{c}\sigma\}$, and let $\tilde{\chi}_i$ be a smooth function supported in $\widetilde{Q}_i$ which satisfies $\tilde{\chi}_i = 1$ on Q_i and $|\partial_\xi^\alpha \tilde{\chi}_i| \lesssim \sigma^{-|\alpha|}$. Let us define K_i by

$$\mathcal{F}(K_i)(\xi) = \phi\Big(\frac{\tau - \psi(\zeta)}{\delta}\Big)\tilde{\chi}_i(\xi).$$

Since $\hat{f}_i$ is supported in Q_i, we have $T_\delta f_i = K_i * f_i$.

Let $\{\mathcal{B}\}$ be the collection of boundedly overlapping balls of radius δ^{-1} which cover $\mathbb{R}^d$. For $\varepsilon > 0$ we denote by $\widetilde{\mathcal{B}}$ the balls $B(a, \delta^{-1-\varepsilon})$ if $\mathcal{B} = B(a, \delta^{-1})$. By decomposing $f_i = \chi_{\widetilde{\mathcal{B}}} f_i + \chi_{\widetilde{\mathcal{B}}^c} f_i$, we bound the (p/k)-th power of the left-hand side of (15) by

$$\sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} |T_\delta f_i(x)|^{\frac{p}{k}}\,dx = \sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} |K_i * f_i(x)|^{\frac{p}{k}}\,dx \lesssim I + II,$$

where

$$I = \sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} |K_i * (\chi_{\widetilde{\mathcal{B}}} f_i)(x)|^{\frac{p}{k}}, \quad II = \sum_{\mathcal{B}} \Big(\sum_{\substack{g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i \\ \text{for some } i}} \int_{\mathcal{B}} \prod_{i=1}^{k} |K_\delta * g_i(x)|^{\frac{p}{k}}\,dx \Big).$$

The second sum in II is summation over all possible choices of g_i with $g_i = \chi_{\widetilde{\mathcal{B}}} f_i$ or $\chi_{\widetilde{\mathcal{B}}^c} f_i$, and $g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i$ for some i. So, in $\prod_{i=1}^k K_\delta * g_i(x)$ there is at least one g_i which satisfies $g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i$.

Since $\mathcal{F}(K_i * (\chi_{\widetilde{\mathcal{B}}} f_i)) \subset \Gamma(\delta) \cap \widetilde{Q}_i$, taking a sufficiently small $\tilde{c} > 0$, from continuity it is easy to see that $F_1 = K_1 * (\chi_{\widetilde{\mathcal{B}}} f_1), \dots, F_k = K_k * (\chi_{\widetilde{\mathcal{B}}} f_k)$ satisfy the assumption of Theorem 2.7. So, by Theorem 2.7 and Plancherel's theorem we see

$$I \lesssim \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{\varepsilon} \sum_{\mathcal{B}} \prod_{i=1}^{k} \delta^{\frac{p}{2k}} \|K_i * (\chi_{\widetilde{\mathcal{B}}} f_i)\|_2^{\frac{p}{k}} \le \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{\varepsilon} \sum_{\mathcal{B}} \prod_{i=1}^{k} \delta^{\frac{p}{2k}} \|\chi_{\widetilde{\mathcal{B}}} f_i\|_2^{\frac{p}{k}}$$

for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $\varepsilon_\circ$ small enough. Since $p > 2$, by applying Hölder's inequality twice we have

$$I \lesssim \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{\varepsilon} \prod_{i=1}^{k} \delta^{\frac{p}{k}(\frac{1}{2} + d(1+\varepsilon)(\frac{1}{p} - \frac{1}{2}))} \Big(\sum_{\mathcal{B}} \|\chi_{\widetilde{\mathcal{B}}} f_i\|_p^p \Big)^{\frac{1}{k}} \lesssim \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{c\varepsilon} \Big(\prod_{i=1}^{k} \delta^{\frac{d}{p} - \frac{d-1}{2}} \|f_i\|_p \Big)^{\frac{p}{k}}.$$

For II, we use Lemma 2.9. There is a constant $C = C(\|\psi\|_{C^N(I^{d-1})})$ such that $|K_i * (\chi_{\widetilde{\mathcal{B}}^c} f_i)(x)| \le C\delta\delta^{\varepsilon(M-d-1)} \mathfrak{K}_{d+1} * |f_i|(x)$ if $x \in B$, and $|K_i * g_i(x)| \le C\delta \mathfrak{K}_{d+1} * |f_i|(x)$. Thus, we get

$$II \lesssim \delta^{\frac{(k-1)p}{k}} \delta^{\varepsilon(N-d-1)\frac{p}{k}} \int \prod_{i=1}^{k} (\mathfrak{K}_{d+1} * |f_i|(x))^{\frac{p}{k}}\,dx \lesssim \delta^{c_2 N\varepsilon - c_1} \prod_{i=1}^{k} \|f_i\|_p^{\frac{p}{k}}$$

for some $c_1, c_2 > 0$ because $\|\mathfrak{K}_{d+1} * f\|_p \le C\delta^{-d}\|f\|_p$ for $1 \le p \le \infty$ by Young's convolution inequality. Combining the two estimates for I and II with N large enough, we see that for $\varepsilon > 0$ there is an N such that

$$\Big\| \prod_{i=1}^{k} T_\delta f_i \Big\|_{L^{p/k}(\mathbb{R}^d)} \le C\sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{c\varepsilon} \prod_{i=1}^{k} \delta^{\frac{d}{p} - \frac{d-1}{2}} \|f_i\|_p$$

for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $\varepsilon_\circ$ small enough. Therefore, choosing $\varepsilon = \varepsilon/c$, we get the desired bound (15). $\square$

In what follows we show that if the normal vectors of the surfaces are confined in a $C\delta$-neighborhood of a k-plane in Proposition 2.11, then the associated multilinear restriction estimate has an improved bound. In particular, if one takes $p = 2k/(k-1)$, the bound in (17) is $\sim \delta^{-\varepsilon}\delta^{\frac{d}{2}}$, which is better than the corresponding bound $\sim \delta^{-\varepsilon}\delta^{\frac{k}{2}}$ in Proposition 2.10. However, it seems difficult to make use of such an improvement to get a better linear bound without using the square sum function (see Corollary 2.12 below).

Proposition 2.11. *Let $0 < \delta \ll \sigma \ll 1$, $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and Π be a k-plane containing the origin. Suppose that $\Gamma(\psi)$, $\Gamma_1, \dots, \Gamma_k$ are given as in the above and* (13) *is satisfied whenever $\xi_i \in \Gamma_i$, $i = 1, \dots, k$. Suppose that*

$$\operatorname{supp} \widehat{F_i} \subset \Gamma_i(\delta) \cap \mathrm{N}^{-1}(\Pi + O(\delta)), \quad i = 1, \dots, k. \tag{16}$$

Then, if $2 \le p \le 2k/(k-1)$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\| \prod_{i=1}^k F_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon}\delta^{dk(\frac{1}{2}-\frac{1}{p})} \prod_{i=1}^k \|F_i\|_2 \tag{17}$$

holds with C, C_ε, independent of ψ.

If $p/k \ge 1$, the inequality could be shown by using Hölder's inequality and the k-linear multilinear restriction estimate in [Bennett et al. 2006]. However, this is not true in general and we prove Proposition 2.11 by making use of the induction-on-scale argument and the multilinear Kakeya estimate. The following is a consequence of Proposition 2.11.

Corollary 2.12. *Suppose that the same assumptions in Proposition 2.11 hold. Let $\{\mathfrak{q}\}$, $\mathfrak{q} \subset \frac{1}{2}I^d$, be the collection of dyadic cubes of side length ℓ, $2^{-2}\delta < \ell \le 2^{-1}\delta$. Then, if $2 \le p \le 2k/(k-1)$, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that, for $x \in \mathbb{R}^d$,*

$$\Big\| \prod_{i=1}^k F_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon} \prod_{i=1}^k \Big\| \Big(\sum_{\mathfrak{q}} |F_{i\mathfrak{q}}|^2 \Big)^{\frac{1}{2}} \rho_{B(x,\delta^{-1})} \Big\|_p \tag{18}$$

holds with C, C_ε, independent of $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.

This may be compared with a discrete formulation of the multilinear inequality in [Bourgain and Guth 2011, (1.1), p. 1250]. The inequality (18) can be easily deduced from Proposition 2.11 by the standard argument using Plancherel's theorem and orthogonality; see the proof of Corollary 3.11. So, we omit the proof.

Proof of Proposition 2.11. For $p = 2$ the estimate (17) follows from Hölder's inequality and Plancherel's theorem. Hence, in view of interpolation, it is enough to show (17) for $p = 2k/(k-1)$.

We prove (17) by adapting the proof of the multilinear restriction estimate in [Bennett et al. 2006]. By translation we may assume $x = 0$. We make the following assumption that, for $0 < \delta \ll \sigma$ and some $\alpha > 0$,

$$\Big\| \prod_{i=1}^k F_i \Big\|_{L^{2/(k-1)}(B(0,\delta^{-1}))} \lesssim \delta^{-\alpha}\delta^{\frac{d}{2}} \prod_{i=1}^k \|F_i\|_2 \tag{19}$$

holds uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ whenever (16) holds and (13) is satisfied for $\xi_i \in \Gamma_i$, $i = 1, \dots, k$. It is clearly true with a large $\alpha > 0$, as can be seen by making use of Lemma 2.9. We show (19) implies that, for $\varepsilon > 0$, there is an N such that

$$\Big\| \prod_{i=1}^{k} F_i \Big\|_{L^{2/(k-1)}(B(0,\delta^{-1}))} \lesssim C_\varepsilon \sigma^{-\kappa} \delta^{-\frac{\alpha}{2} - c\varepsilon} \delta^{\frac{d}{2}} \prod_{i=1}^{k} \| F_i \|_2 \tag{20}$$

holds uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$. In what follows we set $R = \delta^{-1}$.

Iteration of the implication from (19) to (20) allows us to suppress α as small as $\sim \varepsilon$. In fact, since the implication remains valid as long as $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, by fixing an ε and iterating the implication (19) to (20) l times, we have the bound

$$C_\varepsilon^l \sigma^{-\kappa l} R^{2^{-l}\alpha + c\varepsilon(1 + 2^{-1}\varepsilon + \dots + 2^{-l+1})} \le C_\varepsilon^l \sigma^{-\kappa l} R^{2^{-l}\alpha + 2c\varepsilon}.$$

Choosing l such that $2^{-l}\alpha \sim \varepsilon$ gives the bound $\widetilde{C}_\varepsilon \sigma^{Ck \log \frac{\alpha}{\varepsilon}} R^{C\varepsilon}$. Hence, taking $\varepsilon = \varepsilon/C$, we get the desired bound.

Let $\{\boldsymbol{q}\}$ be the collection of dyadic cubes (hence essentially disjoint) of side length ℓ, $\ell < R^{-\frac{1}{2}} \le 2\ell$, such that $\mathbb{R}^d = \bigcup \boldsymbol{q}$. Since the Fourier transform of $\rho_{B(z,\sqrt{R})} F_i$ is supported in $\Gamma(\delta^{\frac{1}{2}}) \cap \mathrm{N}^{-1}(\Pi + O(\delta^{\frac{1}{2}}))$, by the assumption it follows that

$$\begin{aligned} \Big\| \prod_{i=1}^{k} F_i \Big\|_{L^{2/(k-1)}(B(z,R^{1/2}))} &\lesssim \Big\| \prod_{i=1}^{k} \rho_{B(z,\sqrt{R})} F_i \Big\|_{L^{2/(k-1)}(B(z,R^{1/2}))} \\ &\lesssim \delta^{-\frac{\alpha}{2}} \delta^{\frac{d}{4}} \prod_{i=1}^{k} \| \rho_{B(z,\sqrt{R})} F_i \|_2 \lesssim \delta^{-\frac{\alpha}{2}} \delta^{\frac{d}{4}} \prod_{i=1}^{k} \Big\| \rho_{B(z,\sqrt{R})} \Big(\sum_{\boldsymbol{q}} |F_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_2. \end{aligned}$$

Here $F_{i\boldsymbol{q}}$ is given by $\mathcal{F}(F_{i\boldsymbol{q}}) = \widehat{F}_i \chi_{\boldsymbol{q}}$. Since the supports of $\mathcal{F}(\rho_{B(z,\sqrt{R})} F_{i\boldsymbol{q}})$ are boundedly overlapping, the last inequality follows from Plancherel's theorem. By the rapid decay of ρ we have, for a large $M > 0$,

$$\Big\| \prod_{i=1}^{k} F_i \Big\|_{L^{2/(k-1)}(B(z,\sqrt{R}))} \lesssim \delta^{-\frac{\alpha}{2}} \delta^{\frac{d}{4}} \prod_{i=1}^{k} \Big\| \chi_{B(z,R^{1/2+\varepsilon})} \Big(\sum_{\boldsymbol{q}} |F_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_2 + \delta^M \prod_{i=1}^{k} \| F_i \|_2. \tag{21}$$

For a given $\xi \in \mathrm{N}^{-1}(\Pi)$, let $\{v_1, \dots, v_{k-1}\}$ be an orthonormal basis for the tangent space $T_\xi(\mathrm{N}^{-1}(\Pi))$ at ξ, $v_k = \mathrm{N}(\xi)$, and let $v_{k+1}, \dots, v_d$ form an orthonormal basis for $(\mathrm{span}\{v_1, \dots, v_{k-1}, v_k\})^\perp$. (So, the vectors $v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_d$ depend on $\xi \in \mathrm{N}^{-1}(\Pi)$.) Then, we define $\boldsymbol{p}(\xi)$ and $\boldsymbol{P}(\xi)$ by

$$\begin{aligned} \boldsymbol{p}(\xi) &= \xi + \big\{ x : |x \cdot v_j| \le C_1 \sqrt{\delta},\ j = 1, \dots, k-1, \text{ and } |x \cdot v_j| \le C_1 \delta,\ j = k+1, \dots, d \big\}, \\ \boldsymbol{P}(\xi) &= \big\{ x : |x \cdot v_j| \le C \sqrt{\delta},\ j = 1, \dots, k-1, \text{ and } |x \cdot v_j| \le C,\ j = k+1, \dots, d \big\}. \end{aligned}$$

Since $\mathrm{N}^{-1}(\Pi)$ is smooth, $\mathrm{N}^{-1}(\Pi) + O(\delta)$ can be covered by a collection of boundedly overlapping $\{\boldsymbol{p}(\xi_\alpha)\}$, $\xi_\alpha \in \mathrm{N}^{-1}(\Pi)$ (here, we are seeing $\mathrm{N}^{-1}(\Pi)$ as a subset of $\mathbb{R}^d$), such that for any $\boldsymbol{q}$ there exists ξ_α satisfying

$$\operatorname{supp} \widehat{F}_i \cap \boldsymbol{q} \subset \tfrac{1}{2} \boldsymbol{p}(\xi_\alpha) \tag{22}$$

with a sufficiently large $C_1 > 0$.

For $(i,\boldsymbol{q})$ satisfying $\operatorname{supp}\widehat{F_i}\cap\boldsymbol{q}\neq\varnothing$ let us denote by $\xi_{i,\boldsymbol{q}}$ the ξ_α which satisfies (22) (if there are more than one, we simply choose one of them). We also denote by $L(i,\boldsymbol{q})$ the bijective affine map from $\frac{1}{2}\boldsymbol{p}(\xi_{i,\boldsymbol{q}})$ to $\mathfrak{q}(0,1)$. Then we define $\widetilde{F}_{i\boldsymbol{q}}$ by

$$\mathcal{F}(\widetilde{F}_{i\boldsymbol{q}})(\xi)=\frac{1}{\rho(L(i,\boldsymbol{q})\xi)}\widehat{F}_{i\boldsymbol{q}}(\xi).$$

We also set $\boldsymbol{P}_{i,\boldsymbol{q}}=\boldsymbol{P}(\xi_{i,\boldsymbol{q}})$ and $\boldsymbol{K}_{i,\boldsymbol{q}}=\mathcal{F}^{-1}(\rho(L(i,\boldsymbol{q})\,\cdot))$. By $R\boldsymbol{P}_{i,\boldsymbol{q}}$ we denote the rectangle which is the R-times dilation of $\boldsymbol{P}_{i,\boldsymbol{q}}$ from the center of $\boldsymbol{P}_{i,\boldsymbol{q}}$. Also denote by $\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}$ the set $R^{1+\varepsilon}\boldsymbol{P}_{i,\boldsymbol{q}}$ which is the $R^{1+\varepsilon}$-times dilation of $\boldsymbol{P}_{i,\boldsymbol{q}}$ from its center. Since $\boldsymbol{K}_{i,\boldsymbol{q}}*\widetilde{F}_{i\boldsymbol{q}}=F_{i\boldsymbol{q}}$ and $|\boldsymbol{K}_{i,\boldsymbol{q}}|\lesssim\chi_{R\boldsymbol{P}_{i,\boldsymbol{q}}}/|R\boldsymbol{P}_{i,\boldsymbol{q}}|$, we have, for $y\in B(x,2R^{\frac{1}{2}+\varepsilon})$ and some $c>0$,

$$|F_{i\boldsymbol{q}}(y)|^2=|\boldsymbol{K}_{i,\boldsymbol{q}}|*|\widetilde{F}_{i\boldsymbol{q}}|^2(y)\lesssim\frac{\chi_{R\boldsymbol{P}_{i,\boldsymbol{q}}}}{|R\boldsymbol{P}_{i,\boldsymbol{q}}|}*|\widetilde{F}_{i\boldsymbol{q}}|^2(y)\lesssim R^{c\varepsilon}\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|}*|\widetilde{F}_{i\boldsymbol{q}}|^2(x).$$

The last inequality is trivial since $|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|\sim R^{c\varepsilon}|R\boldsymbol{P}_{i,\boldsymbol{q}}|$ for some $c>0$. Hence, for $x,y\in B(z,R^{\frac{1}{2}+\varepsilon})$ we have

$$\sum_{\boldsymbol{q}}|F_{i\boldsymbol{q}}|^2(y)\lesssim R^{c\varepsilon}\sum_{\boldsymbol{q}}\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|}*|\widetilde{F}_{i\boldsymbol{q}}|^2(x). \tag{23}$$

Integrating in y over $B(z,R^{\frac{1}{2}+\varepsilon})$ for each $1\le i\le k$, we see that, for $x\in B(z,R^{\frac{1}{2}+\varepsilon})$,

$$\prod_{i=1}^k\Big\|\chi_{B(z,R^{1/2+\varepsilon})}\Big(\sum_{\boldsymbol{q}}|F_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_2\lesssim R^{c\varepsilon}R^{\frac{dk}{4}}\prod_{i=1}^k\Big(\sum_{\boldsymbol{q}}\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|}*|\widetilde{F}_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}(x). \tag{24}$$

Now, integration in x over $B(z,R^{\frac{1}{2}+\varepsilon})$ yields

$$\prod_{i=1}^k\Big\|\chi_{B(z,R^{1/2+\varepsilon})}\Big(\sum_{\boldsymbol{q}}|F_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_2\lesssim R^{c\varepsilon}R^{\frac{d}{4}}\Big\|\prod_{i=1}^k\Big(\sum_{\boldsymbol{q}}\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|}*|\widetilde{F}_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_{L^{2/(k-1)}(B(z,R^{1/2+\varepsilon}))}. \tag{25}$$

Combining this with (21) we have, for any large $M>0$,

$$\begin{aligned}\Big\|\prod_{i=1}^k F_i\Big\|_{L^{2/(k-1)}(B(z,\sqrt{R}))}&\\ \lesssim\delta^{-\frac{\alpha}{2}-c\varepsilon}\Big\|&\prod_{i=1}^k\Big(\sum_{\boldsymbol{q}}\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|}*|\widetilde{F}_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_{L^{2/(k-1)}(B(z,R^{1/2+\varepsilon}))}+\delta^M\prod_{i=1}^k\|F_i\|_2.\end{aligned} \tag{26}$$

We now cover $B(0,R)$ with boundedly overlapping balls $B(z,\sqrt{R})$ and use the above inequality for each of them. Then we get

$$\Big\|\prod_{i=1}^k F_i\Big\|_{L^{2/(k-1)}(B(0,R))}\lesssim\delta^{-\frac{\alpha}{2}-c\varepsilon}\Big\|\prod_{i=1}^k\Big(\sum_{\boldsymbol{q}}\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|}*|\widetilde{F}_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_{L^{2/(k-1)}(B(0,2R))}+\delta^{M-C}\prod_{i=1}^k\|F_i\|_2.$$

Here we have an increased c because of the overlapping of the balls $B(z, R^{\frac{1}{2}+\varepsilon})$ in the right-hand side. Since $\sum_{\boldsymbol{q}} \|\widetilde{F}_{i\boldsymbol{q}}\|_2^2 \sim \|F_i\|_2^2$, for (20) it is sufficient to show

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2\Big)\Big\|_{L^{1/(k-1)}(B(0,2R))} \lesssim \sigma^{-\kappa}\delta^{\frac{d}{2}-c\varepsilon}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}}\|\widetilde{F}_{i\boldsymbol{q}}\|_2^2\Big).$$

By rescaling this is equivalent to

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}} \frac{\chi_{\boldsymbol{P}_{i,\boldsymbol{q}}}}{|\boldsymbol{P}_{i,\boldsymbol{q}}|} * f_{i,\boldsymbol{q}}\Big)\Big\|_{L^{1/(k-1)}(B(0,2))} \lesssim \sigma^{-\kappa}R^{c\varepsilon}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}}\|f_{i,\boldsymbol{q}}\|_1\Big). \tag{27}$$

Let $\mathcal{I}_i = \{\boldsymbol{q} : \operatorname{supp}\widehat{F}_i \cap \boldsymbol{q} \neq \varnothing\}$, $I_i \subset \mathcal{I}_i$ and $\mathcal{T}_{i,\boldsymbol{q}}$ be a finite subset of $\mathbb{R}^d$. Allowing the loss of $(\log R)^C$ in bound, by a standard reduction with pigeon-holing it suffices to show

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}\in I_i}\sum_{\tau\in\mathcal{T}_{i,\boldsymbol{q}}} \chi_{\boldsymbol{P}_{i,\boldsymbol{q}}+\tau}\Big)\Big\|_{L^{1/(k-1)}(B(0,2))} \lesssim \sigma^{-\frac{\kappa}{2}}R^{c\varepsilon}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}\in I_i}\sum_{\tau\in\mathcal{T}_{i,\boldsymbol{q}}}|\boldsymbol{P}_{i,\boldsymbol{q}}+\tau|\Big). \tag{28}$$

We write $x = (u, v) \in \Pi \times \Pi^{\perp}(= \mathbb{R}^d)$. Then the left-hand side is clearly bounded by

$$\sup_{v\in\Pi^{\perp}}\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}\in I_i}\sum_{\tau\in\mathcal{T}_{i,\boldsymbol{q}}} \chi_{\boldsymbol{P}_{i,\boldsymbol{q}}+\tau}(\,\cdot\,, v)\Big)\Big\|_{L^{1/(k-1)}(\widetilde{B}(0,2))},$$

where $\widetilde{B}(0, \rho) \subset \mathbb{R}^k$ is the ball of radius ρ which is centered at the origin.

For $v \in \Pi^{\perp}$ let us set

$$(\boldsymbol{P}_{i,\boldsymbol{q}} + \tau)^v = \{u : (u, v) \in \boldsymbol{P}_{i,\boldsymbol{q}} + \tau\}.$$

Then $(\boldsymbol{P}_{i,\boldsymbol{q}} + \tau)^v$ is contained in a tube of length ~ 1 and width $CR^{-\frac{1}{2}}$, with axis parallel to $\mathrm{N}(\xi_{i,\boldsymbol{q}})$. This is because the longer sides of $\boldsymbol{P}_{i,\boldsymbol{q}}$, except the one parallel to $\mathrm{N}(\xi_{i,\boldsymbol{q}})$, are transversal to Π. More precisely, we can show that if $\varepsilon_\circ$ is sufficiently small and N is large enough, there a constant $c > 0$, independent of $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, such that, for $w \in \big(T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \operatorname{span}\{\mathrm{N}(\xi_{i,\boldsymbol{q}})\}\big)^{\perp}$,

$$\measuredangle(w, \Pi) \geq c > 0. \tag{29}$$

Since (13) is satisfied whenever $\xi_i \in \Gamma_i$, $i = 1, \dots, k$, we know $\mathrm{N}(\xi_{1,\boldsymbol{q}}), \dots, \mathrm{N}(\xi_{k,\boldsymbol{q}})$ which are, respectively, parallel to the axes of tubes $(\boldsymbol{P}_{1,\boldsymbol{q}} + \tau)^v, \dots, (\boldsymbol{P}_{k,\boldsymbol{q}} + \tau)^v$, satisfy $|\operatorname{Vol}(\mathrm{N}(\xi_{1,\boldsymbol{q}}), \dots, \mathrm{N}(\xi_{k,\boldsymbol{q}})) \gtrsim \sigma$. Also note that $|\boldsymbol{P}_{i,\boldsymbol{q}}^v + \tilde{\tau}| \sim |\boldsymbol{P}_{i,\boldsymbol{q}}|$. Hence, by the multilinear Kakeya estimate in $\mathbb{R}^k$ (Theorem 3.7) it follows that

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q},\tau} \chi_{\boldsymbol{P}_{i,\boldsymbol{q}}+\tau}(\,\cdot\,, v)\Big)\Big\|_{L^{1/(k-1)}(\widetilde{B}(0,2))} \lesssim \sigma^{-1}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q},\tau}|\boldsymbol{P}_{i,\boldsymbol{q}}+\tau|\Big).$$

This gives the desired inequality (28).

Now it remains to show (29). By continuity, taking sufficiently small $\varepsilon_\circ$, we only need to show (29) when $\psi = \psi_\circ$ since $\|\psi - \psi_\circ\|_{C^N(I^{d-1})} \leq \varepsilon_\circ$. Though it is easy to show and intuitively obvious, we include a proof for clarity. By rotation we may assume $\Pi \cap \{x_d = -1\} = \{(\boldsymbol{y}, \boldsymbol{a}, -1) : \boldsymbol{y} \in \mathbb{R}^{k-1}\}$ for some $\boldsymbol{a} \in \mathbb{R}^{d-k}$. Since Π contains the origin, Π can parametrized (except for $\Pi \cap \{x_d = 0\}$) as follows:

$$s(\boldsymbol{y}, \boldsymbol{a}, -1), \quad s \in \mathbb{R},\ \boldsymbol{y} \in \mathbb{R}^{k-1}. \tag{30}$$

We may assume $\Gamma_i(\delta) \cap (\mathrm{N}^{-1}(\Pi) + O(\delta)) \neq \varnothing$ because otherwise $F_i = 0$ and there is nothing to prove. Since $\mathrm{N}(\Gamma) \cap \Pi = \varnothing$ if $|\boldsymbol{a}|$ is large, we may assume that $|\boldsymbol{a}| \le C$ for some $C > 0$ and note that $\xi_{i,\boldsymbol{q}} \in \Gamma(\psi)$. Furthermore, it suffices to show that

$$\Pi \cap \big(T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \operatorname{span}\{\mathrm{N}(\xi_{i,\boldsymbol{q}})\}\big)^{\perp} = \{0\}, \tag{31}$$

which implies $\measuredangle(w, \Pi) > 0$ if $w \in \big(T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \operatorname{span}\{\mathrm{N}(\xi_{i,\boldsymbol{q}})\}\big)^{\perp}$. Then, by continuity and compactness (29) follows. We now verify (29) with $\psi = \psi_\circ$. By rotation we may assume $\boldsymbol{a} = (0, \dots, 0, a) =: (\boldsymbol{0}, a) \in \mathbb{R}^{d-k-1} \times \mathbb{R}$. Using the above parametrization of Π, we see that

$$\Pi = \operatorname{span}\{e_1, \dots, e_{k-1}, (0, \dots, 0, \boldsymbol{0}, a, -1)\}.$$

The normal vector at $(x', |x'|^2/2) \in \mathbb{R}^{d-1} \times \mathbb{R}$ is parallel to $(x', -1)$. Hence, if $(x', |x'|^2/2) \in \mathrm{N}^{-1}(\Pi)$, that is, $(x', -1) \in \Pi$, then x' takes the form $x' = (\boldsymbol{y}, \boldsymbol{a})$ because of (30). Hence, it follows that $\mathrm{N}^{-1}(\Pi) = \big\{\big(\boldsymbol{y}, \boldsymbol{0}, a, \tfrac{1}{2}(|\boldsymbol{y}|^2 + |a|^2)\big)\big\}$. Then, if $\xi_{i,\boldsymbol{q}} = \big(\boldsymbol{y}, \boldsymbol{0}, a, \tfrac{1}{2}(|\boldsymbol{y}|^2 + a^2)\big)$, we have $T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi))$ is spanned by $\boldsymbol{y}_1 = (1, 0, \dots, \boldsymbol{0}, 0, y_1)$, $\boldsymbol{y}_2 = (0, 1, \dots, 0, \boldsymbol{0}, 0, y_2)$, $\dots$, $\boldsymbol{y}_{k-1} = (0, 0, \dots, 1, \boldsymbol{0}, 0, y_{k-1})$. For (31) it is sufficient to show that $\mathfrak{P} := \Pi \cap (\operatorname{span}\{(\boldsymbol{y}, \boldsymbol{0}, a, -1), \boldsymbol{y}_1, \dots, \boldsymbol{y}_{k-1}\})^{\perp} = \{0\}$. Let $w \in \mathfrak{P}$. Then, since $w \in \operatorname{span}\{e_1, \dots, e_{k-1}, (0, \dots, 0, \boldsymbol{0}, a, -1)\}$, we may write $w = (c_1, \dots, c_{k-1}, \boldsymbol{0}, c_k a, -c_k)$. Also, $w \cdot \boldsymbol{y}_1 = \dots = w \cdot \boldsymbol{y}_{k-1} = w \cdot (\boldsymbol{y}, \boldsymbol{0}, a, -1) = 0$ gives $c_1 = \dots = c_k = 0$. So, $v = 0$ and, hence, we get (31). □

2D. *Scattered modulation sum of scale σ.* When the Fourier transform of a given function f is supported in a ball of radius σ, then f behaves as though it were constant on balls of radius σ^{-1}. This observation has important role in Bourgain and Guth's argument [2011] and is widely taken for granted without being made rigorous. There seems to be several ways which make this heuristic rigorous; see [Tao and Vargas 2000b; Tao 1999]. For this purpose we make use of the Fourier series expansion.

Fix $\sigma > 0$ and large positive constants $M = M(d) \ge 100d$ and C_M which are to be chosen to be large. For $l \in \sigma^{-1}\mathbb{Z}^d$ we set

$$A_l = A_l(\sigma) = C_M(1 + |\sigma l|)^{-M}, \quad \tau_l f(x) = f(x - l). \tag{32}$$

For $\sigma > 0$, we define $[F]_\sigma$, $|[F]|_\sigma$ (the scattered modulation sum of σ-scale) by

$$[F]_\sigma(x) = \sum_{l \in \sigma^{-1}\mathbb{Z}^d} A_l |\tau_l F(x)|, \quad |[F]|_\sigma(x) = \sum_{l_1, l_2 \in \sigma^{-1}\mathbb{Z}^d} A_{l_1} A_{l_2} |\tau_{l_1 + l_2} F(x)|. \tag{33}$$

Lemma 2.13. *Let $\xi_0, x_0 \in \mathbb{R}^d$. Suppose that F is a function with $\widehat{F}$ supported in $\mathfrak{q}(\xi_0, \sigma)$. Then, if $x \in \mathfrak{q}\big(x_0, \frac{1}{\sigma}\big)$,*

$$|F(x)| \le [F]_\sigma(x_0) \le |[F]|_\sigma(x).$$

It should be noted that the inequality holds regardless of ξ_0, x_0, and σ.

Proof. Let a be a smooth function supported in $[-\pi,\pi]^d$ and $a(x)=1$ if $|x_i|\le 1,\ i=1,\dots,d$. Set

$$A(x,\xi)=a(x)a(\xi)e^{ix\cdot\xi}.$$

Since $|\partial_\xi^\alpha A|\le C_\alpha$ for any multi-indices α, by expanding into Fourier series in ξ we have

$$a(x)a(\xi)e^{ix\cdot\xi}=\sum_{l\in\mathbb{Z}^d}a_l(x)e^{-i\xi\cdot l},\quad x,\xi\in[-\pi,\pi]^d, \tag{34}$$

while a_l satisfies $|a_l(x)|\le C_M(1+|l|)^{-M}$ for any large $M>0$. On the other hand, from the inversion formula we have

$$F(x)=(2\pi)^{-d}\int e^{i(x-x_0)\cdot\xi_0}e^{i(x-x_0)\cdot(\xi-\xi_0)}e^{ix_0\cdot\xi}\widehat{F}(\xi)\,d\xi.$$

Hence, since $x\in\mathfrak{q}\big(x_0,\frac{1}{\sigma}\big)$, inserting the harmless bump function a, we may write

$$F(x)=(2\pi)^{-d}e^{i(x-x_0)\cdot\xi_0}\int A\bigg(\sigma(x-x_0),\frac{\xi-\xi_0}{\sigma}\bigg)e^{ix_0\cdot\xi}\widehat{F}(\xi)\,d\xi.$$

Using (34) we have

$$F(x)=(2\pi)^{-d}e^{i(x-x_0)\cdot\xi_0}\sum_{l\in\mathbb{Z}^d}a_l(\sigma(x-x_0))\int e^{-i\frac{(\xi-\xi_0)}{\sigma}\cdot l}e^{ix_0\cdot\xi}\widehat{F}(\xi)\,d\xi.$$

Then it follows that

$$|F(x)|\le\sum_{l\in\sigma^{-1}\mathbb{Z}^d}A_l|\tau_lF(x_0)|\le\sum_{l_1,l_2\in\sigma^{-1}\mathbb{Z}^d}A_{l_1}A_{l_2}|\tau_{(l_1+l_2)}F(x)|. \tag{35}$$

The second inequality follows by applying the first one to each τ_lF with the roles of x,x_0 interchanged. □

2E. *Multiscale decomposition.* We now attempt to bound part of $T_\delta f$ with a sum of products which satisfy the transversality assumption, while the remaining parts are given by a sum of functions which have relatively small Fourier supports. The first is rather directly estimated by making use of the multilinear estimates and the latter is to be handled by Proposition 2.5, the induction assumption and Lemma 2.6.

In what follows, we basically adapt the idea in [Bourgain and Guth 2011]. However, concerning the decomposition in that paper, reappearance of many small-scale functions in large-scale decomposition becomes problematic when one attempts to sum up resulting estimates. For the adjoint restriction estimates this can be overcome by using L^∞-functions, as was done in [Bourgain and Guth 2011]. But such an argument doesn't work for the multiplier operators and leads to a loss in its bound. To get over this, unlike the decomposition in [Bourgain and Guth 2011] where one starts to decompose with d-linear products and proceeds by reducing the degree multilinearity based on dichotomy, we decompose the multiplier operator by increasing the degree of multilinearity in order to avoid small-scale functions appearing inside of large-scale ones. This has a couple of advantages. First, this allows us to keep the function relatively intact in the course of decomposition so that we can easily add up decomposed pieces to obtain the sharp L^p bound. Secondly, the decomposition makes it possible to obtain directly obtain the L^p-L^p estimate. Hence we don't need to rely on the factorization theorem to deduce L^p-L^p from $L^\infty-L^p$. (The same is also true for the adjoint restriction operators.) Hence, we can obtain the sharp L^p bounds for multiplier operators of Bochner–Riesz-type, which lack symmetry.

2E1. *Spatial and frequency dyadic cubes.* Let $0 < \varepsilon_\circ \ll 1,\ 1 \ll N,\ \psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and T_δ be given by (9). Let $\kappa = \kappa(\varepsilon_\circ, N)$ be the number given in Proposition 2.5 so that (11) holds whenever $0 < \varepsilon \le \kappa$ and $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$. Let m be an integer such that $2 \le m \le d-1$, and $\sigma_1, \dots, \sigma_m$ be dyadic numbers such that

$$\delta \ll \sigma_m \ll \cdots \ll \sigma_1 \ll \min(\kappa, 1). \tag{36}$$

These numbers will be specified to terminate induction. We call σ_i the *i-th scale*.

Let us denote by $\{\mathfrak{q}^i\}$ the collection of the dyadic cubes $\mathfrak{q}^i$ of side length $2\sigma_i$ which are contained in I^d (so, $\mathfrak{q}^i$ denotes the member of $\{\mathfrak{q}^i\}$ and the cubes $\mathfrak{q}^i$ are essentially disjoint). Rather than introducing a new notation to denote each collection of $\mathfrak{q}^i$, we take the convention that $\{\mathfrak{q}^i\}$ denotes the collection of all dyadic cubes of side length $2\sigma_i$ contained in I^d if it is not specified otherwise. For each i-th scale there is a unique collection so that there will be no ambiguity, and we also use $\mathfrak{q}^i$ as indices which run over the set $\{\mathfrak{q}^i\}$. Thus, we may write

$$\bigcup_{\mathfrak{q}^i} \mathfrak{q}^i = I^d. \tag{37}$$

For the rest of this section, we assume that

$$\operatorname{supp} \hat{f} \subset \tfrac{1}{2} I^d.$$

Since $f = \sum_{\mathfrak{q}^i} f_{\mathfrak{q}^i}$, for $i = 1, \dots, m$, we write

$$T_\delta f = \sum_{\mathfrak{q}^i} T_\delta f_{\mathfrak{q}^i}. \tag{38}$$

Clearly, we may assume that $\mathfrak{q}^i$ is contained in a $C\sigma_i$-neighborhood of the surface $\Gamma(\psi)$ because $T_\delta f_{\mathfrak{q}^i} = 0$ otherwise. So, in what follows, $\mathfrak{q}^i, \mathfrak{q}^i_1, \dots, \mathfrak{q}^i_{i+1}$ and $\mathfrak{q}^i_*$ denote the elements of $\{\mathfrak{q}^i\}$.

For convenience we extend in a trivial way the map N defined on $\Gamma(\psi)$ to the cube I^d by setting, for $\xi = (\zeta, \tau) \in I^d$,

$$\boldsymbol{n}(\zeta, \tau) = \mathrm{N}(\zeta, \psi(\zeta)).$$

This extension is not necessarily needed in what follows because we only consider a small neighborhood of $\Gamma(\psi)$. However, this allows us to define a normal vector for any point in I^d and makes exposition simpler. This definition of $\boldsymbol{n}$ agrees with the one given in the next section.

Definition 2.14. Let k be an integer such that $1 \le k \le m$ and fix a constant $c > 0$. Let $\mathfrak{q}^k_1, \dots, \mathfrak{q}^k_{k+1} \in \{\mathfrak{q}^k\}$ (k-th scale cubes). We say $\mathfrak{q}^k_1, \mathfrak{q}^k_2, \dots, \mathfrak{q}^k_{k+1}$ *are $(\sigma_1, \sigma_2, \dots, \sigma_k)$-transversal* if

$$\mathrm{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \dots, \boldsymbol{n}(\xi_{k+1})) \ge c\sigma_1\sigma_2 \cdots \sigma_k, \tag{39}$$

whenever $\xi_i \in \mathfrak{q}^k_i,\ i = 1, \dots, k+1$. And we simply denote this by "$\mathfrak{q}^k_1, \mathfrak{q}^k_2, \dots, \mathfrak{q}^k_{k+1}$ trans" and say $\mathfrak{q}^k_1, \mathfrak{q}^k_2, \dots, \mathfrak{q}^k_{k+1}$ are transversal, omitting dependence on $\sigma_1, \sigma_2, \dots, \sigma_k$.

Let us set

$$M_i = \frac{1}{\sigma_i}, \quad i = 1, \dots, m.$$

We denote by $\{\mathfrak{Q}^i\}$ the collection of the dyadic cubes of side length $2M_i$, which covers $\mathbb{R}^d$ (so, $\mathfrak{Q}^i$ again denotes a member of the sets $\{\mathfrak{Q}^i\}$). We write†

$$\bigcup_{\mathfrak{Q}^i} \mathfrak{Q}^i = \mathbb{R}^d. \tag{40}$$

Since the Fourier support of $T_\delta f_{\mathfrak{q}^i}$ is contained $\mathfrak{q}^i$, it may be thought of as a constant on $\mathfrak{Q}^i$ by invoking Lemma 2.13 with $\sigma = \sigma_i$. Since the scale σ_i is clear from the side length of the cube $\mathfrak{q}^i$, we simply set

$$[T_\delta f_{\mathfrak{q}^i}] := [T_\delta f_{\mathfrak{q}^i}]_{\sigma_i}, \quad \|[T_\delta f_{\mathfrak{q}^i}]\| := \|[T_\delta f_{\mathfrak{q}^i}]\|_{\sigma_i}.$$

2E2. *σ_1-scale decomposition.* Bilinear decomposition is rather elementary. Fix $x \in \mathbb{R}^d$. From (38),

$$|T_\delta f(x)| \le \sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)|.$$

We denote by $\mathfrak{q}^1_* = \mathfrak{q}^1_*(x)$ a cube $\mathfrak{q}^1 \in \{\mathfrak{q}^1\}$ such that $|T_\delta f_{\mathfrak{q}^1_*}(x)| = \max_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1_*}(x)|$. (There may be many such cubes but $\mathfrak{q}^1_*$ denotes just one of them.) Then we consider the following two cases separately:

$$\sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| \le 100^d |T_\delta f_{\mathfrak{q}^1_*}(x)|, \quad \sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| > 100^d |T_\delta f_{\mathfrak{q}^1_*}(x)|.$$

For the second case

$$\sum_{\operatorname{dist}(\mathfrak{q}^1,\mathfrak{q}^1_*)<10\sigma_1} |T_\delta f_{\mathfrak{q}^1}(x)| < 50^d |T_\delta f_{\mathfrak{q}^1_*}(x)| \le 2^{-d} \sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)|.$$

Hence there is $\mathfrak{q}^1_1 \in \{\mathfrak{q}^1\}$ such that $\operatorname{dist}(\mathfrak{q}^1_1, \mathfrak{q}^1_*) \ge 10\sigma_1$ and

$$\sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| \lesssim \sigma_1^{-(d-1)} |T_\delta f_{\mathfrak{q}^1_1}(x)| \le \sigma_1^{-(d-1)} |T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_*}(x)|^{\frac{1}{2}}.$$

From these two cases we get

$$\sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| \lesssim \max_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| + C\sigma_1^{-\frac{d-1}{2}} \max_{\operatorname{dist}(\mathfrak{q}^1_1,\mathfrak{q}^1_2)\gtrsim\sigma_1} |T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_2}(x)|^{\frac{1}{2}}. \tag{41}$$

Using the imbedding $\ell^p \subset \ell^\infty$, Proposition 2.5 and Lemma 2.6 give

$$\|\max_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}\|_p \le \Big(\sum_{\mathfrak{q}^1} \|T_\delta f_{\mathfrak{q}^1}\|_p^p\Big)^{\frac{1}{p}} \le \Big(\sum_{\mathfrak{q}^1} A(\sigma_1^{-2}\delta)^p \|f_{\mathfrak{q}^1}\|_p^p\Big)^{\frac{1}{p}} \lesssim A(\sigma_1^{-2}\delta)\|f\|_p. \tag{42}$$

Hence, combining this with (41), we have

$$\|T_\delta f\|_p \lesssim A(\sigma_1^{-2}\delta)\|f\|_p + \sigma_1^{-C} \max_{\operatorname{dist}(\mathfrak{q}^1_1,\mathfrak{q}^1_2)\gtrsim\sigma_1} \|T_\delta f_{\mathfrak{q}^1_1} T_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}}^{\frac{1}{2}}. \tag{43}$$

We now proceed to decompose the bilinear expression appearing in the left-hand side.

† Here we take the same convention for $\{\mathfrak{Q}^i\}$ as we do for $\{\mathfrak{q}^i\}$.

In the following section we explain how one can achieve trilinear decomposition out of (43) before we inductively obtain the full k-linear decomposition which we need for the proof of Theorem 1.1. Once one gets familiar with it, extension to a higher degree of multilinearity becomes more or less obvious.

2E3. *σ_2-scale decomposition.* Suppose that we are given two cubes $\mathfrak{q}_1^1$ and $\mathfrak{q}_2^1$ of first scale such that $\operatorname{dist}(\mathfrak{q}_1^1, \mathfrak{q}_2^1) \gtrsim \sigma_1$. For $i = 1, 2$, we denote by $\{\mathfrak{q}_i^2\}$ the collection of dyadic cubes $\mathfrak{q}_i^2$ of side length σ_2 contained in $\mathfrak{q}_i^1$ such that

$$\mathfrak{q}_i^1 = \bigcup_{\mathfrak{q}_i^2} \mathfrak{q}_i^2, \quad i = 1, 2. \tag{44}$$

We also denote by $\{\mathfrak{q}^2\}$ the set $\{\mathfrak{q}_1^2\} \cup \{\mathfrak{q}_2^2\}$. Then it follows that

$$T_\delta f_{\mathfrak{q}_i^1} = \sum_{\mathfrak{q}_i^2} T_\delta f_{\mathfrak{q}_i^2}, \quad i = 1, 2. \tag{45}$$

We may also assume that $\mathfrak{q}_1^2$, $\mathfrak{q}_2^2$ are contained in the $C\sigma_2$-neighborhood of $\Gamma(\psi)$ because $T_\delta f_{\mathfrak{q}_1^2}$, $T_\delta f_{\mathfrak{q}_2^2}$ are zero otherwise.

Decomposition from this stage is no longer as simple as in the σ_1-scale case. We need to use spatial localization in order to compare the values of the decomposed pieces. This makes it possible to bound large parts of the operator with transversal products.

Let us fix a cube $\mathfrak{Q}^2$ and x_0 be the center of $\mathfrak{Q}^2$. Let $\mathfrak{q}_{1*}^2 \in \{\mathfrak{q}_1^2\}$, $\mathfrak{q}_{2*}^2 \in \{\mathfrak{q}_2^2\}$ be the cubes such that

$$[T_\delta f_{\mathfrak{q}_{1*}^2}](x_0) = \max_{\mathfrak{q}_1^2} [T_\delta f_{\mathfrak{q}_1^2}](x_0), \quad [T_\delta f_{\mathfrak{q}_{2*}^2}](x_0) = \max_{\mathfrak{q}_2^2} [T_\delta f_{\mathfrak{q}_2^2}](x_0).$$

Let us define $\Lambda_i^2 \subset \{\mathfrak{q}_i^2\}$, $i = 1, 2$, by

$$\Lambda_i^2 = \bigl\{\mathfrak{q}_i^2 : [T_\delta f_{\mathfrak{q}_i^2}](x_0) \ge \sigma_2^{2d} \max([T_\delta f_{\mathfrak{q}_{1*}^2}](x_0), [T_\delta f_{\mathfrak{q}_{2*}^2}](x_0))\bigr\}.$$

Using (45), we split the summation to get

$$T_\delta f_{\mathfrak{q}_1^1} T_\delta f_{\mathfrak{q}_2^1} = \sum_{(\mathfrak{q}_1^2, \mathfrak{q}_2^2) \in \Lambda_1 \times \Lambda_2} T_\delta f_{\mathfrak{q}_1^2} T_\delta f_{\mathfrak{q}_2^2} + \sum_{(\mathfrak{q}_1^2, \mathfrak{q}_2^2) \notin \Lambda_1 \times \Lambda_2} T_\delta f_{\mathfrak{q}_1^2} T_\delta f_{\mathfrak{q}_2^2}. \tag{46}$$

Since there are at most $O(\sigma_2^{-2(d-1)})$ pairs $(\mathfrak{q}_1^2, \mathfrak{q}_2^2)$, the second sum in the right-hand side is bounded by

$$\sum_{(\mathfrak{q}_1^2, \mathfrak{q}_2^2) \notin \Lambda_1 \times \Lambda_2} |T_\delta f_{\mathfrak{q}_1^2}(x)| |T_\delta f_{\mathfrak{q}_2^2}(x)| \le \sigma_2^d \max_{\mathfrak{q}^2} ([T_\delta f_{\mathfrak{q}^2}](x_0))^2. \tag{47}$$

For a cube $\mathfrak{q}$ we denote by $\boldsymbol{c}(\mathfrak{q})$ the center of $\mathfrak{q}$. Let $\Pi = \Pi(\mathfrak{q}_{1*}^2, \mathfrak{q}_{2*}^2)$ be the 2-plane which is spanned by $\boldsymbol{n}_1 = \boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{1*}^2))$, $\boldsymbol{n}_2 = \boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{2*}^2))$, and define

$$\mathfrak{N} = \mathfrak{N}(\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1) = \{\mathfrak{q}^2 \in \Lambda_1^2 \cup \Lambda_2^2 : \operatorname{dist}(\boldsymbol{n}(\mathfrak{q}^2), \Pi) \le C\sigma_2\}. \tag{48}$$

Clearly, $\operatorname{Vol}(\boldsymbol{n}_1, \boldsymbol{n}_2) \gtrsim \sigma_1$ and $\operatorname{dist}(\boldsymbol{n}(\mathfrak{q}^2), \Pi) \gtrsim \sigma_2$ if $\mathfrak{q}^2 \notin \mathfrak{N}$. Since $\sigma_1 \gg \sigma_2$, if $\mathfrak{q}^2 \notin \mathfrak{N}$, then $\operatorname{Vol}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}(\xi)) \gtrsim \sigma_1\sigma_2$ for $\xi \in \mathfrak{q}^2$. Also, $\boldsymbol{n}(\mathfrak{q}_{i*}^2) \subset \boldsymbol{n}_i + O(\sigma_2)$, $i = 1, 2$. So, it follows that

$$\operatorname{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \boldsymbol{n}(\xi_3)) \gtrsim \sigma_1\sigma_2 \tag{49}$$

if $\xi_1 \in \mathfrak{q}^2_{1*}$, $\xi_2 \in \mathfrak{q}^2_{2*}$, and $\xi_3 \in \mathfrak{q}^2 \notin \mathfrak{N}$. That is, $\mathfrak{q}^2_{1*}$, $\mathfrak{q}^2_{2*}$, $\mathfrak{q}^2$ are transversal. Hence, we split $\sum_{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2} T_\delta f_{\mathfrak{q}^2_1} T_\delta f_{\mathfrak{q}^2_2}$ into

$$\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1,\mathfrak{q}^2_2\in\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x) + \sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1 \text{ or } \mathfrak{q}^2_2\notin\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x). \tag{50}$$

Each term appearing in the second sum can be bounded by a product of three operators which satisfy the transversality condition. Indeed, suppose that $(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2$ and $\mathfrak{q}^2_2\not\subset\mathfrak{N}$. The case $\mathfrak{q}^2_1\not\subset\mathfrak{N}$ can be handled similarly by symmetry. Since $[T_\delta f_{\mathfrak{q}^2_2}](x_0)\ge\sigma_2^{2d}[T_\delta f_{\mathfrak{q}^2_{1*}}](x_0)$, we have

$$\begin{aligned}[T_\delta f_{\mathfrak{q}^2_1}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0) &\le \big([T_\delta f_{\mathfrak{q}^2_{1*}}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0)\big)^{\frac23}\big([T_\delta f_{\mathfrak{q}^2_1}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0)\big)^{\frac13}\\ &\le \sigma_2^{-\frac{2d}{3}}\big([T_\delta f_{\mathfrak{q}^2_{1*}}](x_0)[T_\delta f_{\mathfrak{q}^2_{2*}}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0)\big)^{\frac23}.\end{aligned}$$

Hence, from this and (49) it follows that

$$\Bigg|\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1 \text{ or } \mathfrak{q}^2_2\notin\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x)\Bigg| \le \sigma_2^{-C}\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\Bigg(\prod_{i=1}^3[T_\delta f_{\mathfrak{q}^2_i}](x_0)\Bigg)^{\frac23}. \tag{51}$$

We combine (46), (47), (50) and (51) to get, for $x\in\mathfrak{Q}^2$,

$$\begin{aligned}&|T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_2}(x)|\\ &\quad\le \sigma_2^d\big(\max_{\mathfrak{q}^2}[T_\delta f_{\mathfrak{q}^2}](x_0)\big)^2 + \Bigg|\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1,\mathfrak{q}^2_2\in\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x)\Bigg| + \sigma_2^{-C}\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\Bigg(\prod_{i=1}^3[T_\delta f_{\mathfrak{q}^2_i}](x_0)\Bigg)^{\frac23}.\end{aligned}$$

Using Lemma 2.13 again, we have, for $x\in\mathfrak{Q}^2$,

$$\begin{aligned}|T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_2}(x)| &\le \sigma_2^d\big(\max_{\mathfrak{q}^2}|[T_\delta f_{\mathfrak{q}^2}]|(x)\big)^2\\ &\quad+\Bigg|\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1,\mathfrak{q}^2_2\in\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x)\Bigg| + \sigma_2^{-C}\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\Bigg(\prod_{i=1}^3|[T_\delta f_{\mathfrak{q}^2_i}]|(x)\Bigg)^{\frac23}.\end{aligned} \tag{52}$$

Taking $L^{\frac{p}{2}}$ on both sides of the inequality over each $\mathfrak{Q}^2$, summing along $\mathfrak{Q}^2$, and using Proposition 2.5 and Lemma 2.6, we get

$$\begin{aligned}\|T_\delta f_{\mathfrak{q}^1_1} T_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}} &\lesssim (A(\sigma_2^{-2}\delta))^2\|f\|_p^2 + \Bigg(\sum_{\mathfrak{Q}^2}\Bigg\|\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2\subset[\mathfrak{N}](\mathfrak{Q}^2,\mathfrak{q}^1_1,\mathfrak{q}^1_2)} T_\delta f_{\mathfrak{q}^2_1} T_\delta f_{\mathfrak{q}^2_2}\Bigg\|_{L^{p/2}(\mathfrak{Q}^2)}^{\frac{p}{2}}\Bigg)^{\frac2p}\\ &\quad+\sigma_2^{-C}\sup_{\tau_1,\tau_2,\tau_3}\max_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\|T_\delta(\tau_1 f_{\mathfrak{q}^2_1})T_\delta(\tau_2 f_{\mathfrak{q}^2_2})T_\delta(\tau_3 f_{\mathfrak{q}^2_3})\|_{\frac{p}{3}}^{\frac23},\end{aligned} \tag{53}$$

where $[\mathfrak{N}](\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)$ is a subset of $\mathfrak{N}(\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)$. Here, for simplicity we now denote $\tau_{l_i} f$ by $\tau_i f$ just to indicate translation by a vector. The precise value of l_i is not significant in the overall argument. To show (53), for the first term in the right-hand side of (52) we may repeat the same argument as in (42). In fact, by (33) and the rapid decay of A_l[†] combined with Hölder's inequality to summation along l, l', and using Proposition 2.5 and Lemma 2.6 we have

$$\Big\|\max_{\mathfrak{q}^2} |[T_\delta f_{\mathfrak{q}^2}]|\Big\|_p \lesssim \sup_{\tau_2} \Big\|\max_{\mathfrak{q}^2} |T_\delta(\tau_2 f_{\mathfrak{q}^2})|\Big\|_p \lesssim A(\sigma_1^{-2}\delta)\|f\|_p.$$

For the third term of the right-hand side of (52), thanks to (33) and the rapid decay of A_l, it is enough to note that there are as many as $O(\sigma_2^{-C})$ transversal $\mathfrak{q}_1^2, \mathfrak{q}_2^2, \mathfrak{q}_3^2$.

We combine (53) with (43) to get

$$\begin{aligned}\|T_\delta f\|_p \lesssim{}& A(\sigma_1^{-2}\delta)\|f\|_p + \sigma_1^{-C} A(\sigma_2^{-2}\delta)\|f\|_p \\ &+ \sigma_1^{-C} \sup_{\tau_1,\tau_2} \max_{\mathfrak{q}_1^1,\mathfrak{q}_2^1 \text{ trans}} \Bigg(\sum_{\mathfrak{Q}^2} \Bigg\| \sum_{\substack{\mathfrak{q}_1^2 \subset \mathfrak{q}_1^1,\ \mathfrak{q}_2^2 \subset \mathfrak{q}_2^1 \\ \mathfrak{q}_1^2, \mathfrak{q}_2^2 \subset [\mathfrak{N}](\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)}} T_\delta(\tau_1 f_{\mathfrak{q}_1^2}) T_\delta(\tau_2 f_{\mathfrak{q}_2^2}) \Bigg\|_{L^{p/2}(\mathfrak{Q}^2)}^{\frac{p}{2}} \Bigg)^{\frac{1}{p}} \\ &+ \sigma_1^{-C}\sigma_2^{-C} \sup_{\tau_1,\tau_2,\tau_3} \max_{\mathfrak{q}_1^2,\mathfrak{q}_2^2,\mathfrak{q}_3^2 \text{ trans}} \|T_\delta(\tau_1 f_{\mathfrak{q}_1^2}) T_\delta(\tau_2 f_{\mathfrak{q}_2^2}) T_\delta(\tau_3 f_{\mathfrak{q}_3^2})\|_{\frac{p}{3}}^{\frac{1}{3}}. \end{aligned} \tag{54}$$

Here $[\mathfrak{N}](\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)$ also depends on τ_1, τ_2. We keep decomposing the trilinear transversal part in order to achieve a higher level of multilinearity.

2E4. *From k-transversal to $(k+1)$-transversal.* Now we proceed inductively. Suppose that we are given dyadic cubes $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ of $(k-1)$-th scale which are transversal:

$$\mathrm{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \dots, \boldsymbol{n}(\xi_k)) \ge c\sigma_1\sigma_2\cdots\sigma_{k-1} \tag{55}$$

whenever $\xi_i \in \mathfrak{q}_i^{k-1}$, $i = 1, \dots, k$. As before, we denote by $\{\mathfrak{q}_i^k\}$ the collection of dyadic cubes of side length $2\sigma_k$ contained in $\mathfrak{q}_i^{k-1}$ such that

$$\bigcup_{\mathfrak{q}_i^k} \mathfrak{q}_i^k = \mathfrak{q}_i^{k-1}, \quad i = 1, \dots, k, \tag{56}$$

and we also denote by $\{\mathfrak{q}^k\}$ the set $\bigcup_{i=1}^k \{\mathfrak{q}_i^k\}$. Hence,

$$\prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^{k-1}} = \prod_{i=1}^k \Bigg(\sum_{\mathfrak{q}_i^k} T_\delta f_{\mathfrak{q}_i^k}\Bigg) = \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k} \prod_{i=1}^k (T_\delta f_{\mathfrak{q}_i^k}). \tag{57}$$

Fix a dyadic cube $\mathfrak{Q}^k$ of side length $2M_i$ and let x_0 be the center of $\mathfrak{Q}^k$. For $i = 1, \dots, k$, let $\mathfrak{q}_{i*}^k \in \{\mathfrak{q}_i^k\}$ be such that

$$[T_\delta f_{\mathfrak{q}_{i*}^k}](x_0) = \max_{\mathfrak{q}_i^k} [T_\delta f_{\mathfrak{q}_i^k}](x_0)$$

[†]Note that the sequence is independent of $\mathfrak{Q}^2$.

and we set, for $i = 1, \dots, k$,

$$\Lambda_i^k = \{\mathfrak{q}_i^k : [T_\delta f_{\mathfrak{q}_i^k}](x_0) \ge (\sigma_k)^{kd} \max_{i=1,\dots,k} [T_\delta f_{\mathfrak{q}_{i*}^k}](x_0)\}.$$

Then, it follows that

$$\sum_{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\notin\prod_{i=1}^k \Lambda_i^k} \prod_{i=1}^k [T_\delta f_{\mathfrak{q}_i^k}](x_0) \le \max [T_\delta f_{\mathfrak{q}^k}](x_0). \tag{58}$$

Let $\boldsymbol{n}_1, \dots, \boldsymbol{n}_k$ denote the normal vectors $\boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{1*}^k)), \dots, \boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{k*}^k))$, respectively, and let

$$\Pi^k = \Pi^k(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1})$$

be the k-plane spanned by $\boldsymbol{n}_1, \dots, \boldsymbol{n}_k$. Now, for a sufficiently large constant $C > 0$, we define

$$\mathfrak{N} = \mathfrak{N}(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}) = \{\mathfrak{q}^k : \mathrm{dist}(\boldsymbol{n}(\mathfrak{q}^k), \Pi^k) \le C\sigma_k\}. \tag{59}$$

By (55) it follows that if $\mathfrak{q}_i^k \notin \mathfrak{N}$, (39) holds whenever $\xi_1 \in \mathfrak{q}_{1*}^k, \dots, \xi_k \in \mathfrak{q}_{k*}^k$ and $\xi_{k+1} \in \mathfrak{q}_i^k$. Hence, $\mathfrak{q}_{1*}^k, \dots, \mathfrak{q}_{k*}^k, \mathfrak{q}_i^k$ are transversal.

We write

$$\sum_{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k} = \sum_{\substack{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in \mathfrak{N}}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k} + \sum_{\substack{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k \\ \mathfrak{q}_i^k \notin \mathfrak{N} \text{ for some } i}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k}. \tag{60}$$

Consider a k-tuple $(\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k)$ which appears in the second sum. There is a $\mathfrak{q}_i^k \notin \mathfrak{N}$. By the same manipulation as before, we get

$$\prod_{i=1}^k [T_\delta f_{\mathfrak{q}_i^k}](x_0) \le \sigma_k^{-\frac{dk^2}{k+1}} \prod_{i=1}^k ([T_\delta f_{\mathfrak{q}_{i*}^k}](x_0))^{\frac{k}{k+1}} ([T_\delta f_{\mathfrak{q}_i^k}](x_0))^{\frac{k}{k+1}}.$$

Since $\mathfrak{q}_{1*}^k, \dots, \mathfrak{q}_{k*}^k, \mathfrak{q}_i^k$ are transversal, by Lemma 2.13 we have, for $x \in \mathfrak{Q}^k$,

$$\Bigg| \sum_{\substack{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k \\ \mathfrak{q}_i^k \notin \mathfrak{N} \text{ for some } i}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k}(x) \Bigg| \lesssim \sigma_k^{-C} \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k \text{ trans}} \prod_{i=1}^{k+1} ([T_\delta f_{\mathfrak{q}_i^k}](x_0))^{\frac{k}{k+1}}. \tag{61}$$

Combining (58) and (61) with (57) and (60), and applying Lemma 2.13 yield, for $x \in \mathfrak{Q}^k$,

$$\begin{aligned} &\Bigg| \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^{k-1}}(x) \Bigg| \\ &\lesssim \big(\max_{\mathfrak{q}^k} |[T_\delta f_{\mathfrak{q}^k}]|(x)\big)^k + \sigma_k^{-C} \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k \text{ trans}} \prod_{i=1}^{k+1} \big(|[T_\delta f_{\mathfrak{q}_i^k}]|(x)\big)^{\frac{k}{k+1}} + \Bigg| \sum_{\substack{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in \\ [\mathfrak{N}](\mathfrak{Q}^k,\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1})}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k}(x) \Bigg|, \end{aligned}$$

where $[\mathfrak{N}](\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$ is a subset of $\mathfrak{N}(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$. After taking the (p/k)-th power of both sides of the inequality, we integrate on $\mathbb{R}^d$, and use Proposition 2.5 and Lemma 2.6 along with (33) to get

$$\Big\| \prod_{i=1}^{k} T_\delta f_{\mathfrak{q}_i^{k-1}}(x) \Big\|_{L^{p/k}}^{\frac{1}{k}} \lesssim A(\sigma_k^{-2}\delta)\|f\|_p + \sigma_k^{-C} \sup_{\tau_1,\dots,\tau_{k+1}} \max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k \text{ trans}} \Big\| \prod_{i=1}^{k+1} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/(k+1)}}^{\frac{1}{k+1}} + \Big(\sum_{\mathfrak{Q}^k} \Big\| \sum_{\substack{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in \\ [\mathfrak{N}](\mathfrak{Q}^k, \mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1})}} \prod_{i=1}^{k} T_\delta f_{\mathfrak{q}_i^k} \Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}} \Big)^{\frac{1}{p}}. \tag{62}$$

2E5. *Multiscale decomposition.* For $k = 2, \dots, d-1$, let us set

$$\mathfrak{M}^k f = \sup_{\tau_1,\dots,\tau_k} \max_{\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1} \text{ trans}} \Big(\sum_{\mathfrak{Q}^k} \Big\| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in [\mathfrak{N}](\mathfrak{Q}^k)}} \prod_{i=1}^{k} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}} \Big)^{\frac{1}{p}}.$$

Here $[\mathfrak{N}](\mathfrak{Q}^k)$ depends on $\tau_1, \dots, \tau_k$, and $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$, but $\boldsymbol{n}(\mathfrak{q}^k)$, $\mathfrak{q}^k \in [\mathfrak{N}](\mathfrak{Q}^k)$, is contained in a k-plan. Starting from (54) we iteratively apply (62) to the transversal products to get

$$\|T_\delta f\|_p \lesssim \sum_{k=1}^{m} \sigma_{k-1}^{-C} A(\sigma_k^{-2}\delta)\|f\|_p + \sum_{k=2}^{m} \sigma_{k-1}^{-C} \mathfrak{M}^k f + \sigma_l^{-C} \sup_{\tau_1,\dots,\tau_{m+1}} \max_{\mathfrak{q}_1^m,\dots\mathfrak{q}_{m+1}^m \text{ trans}} \Big\| \prod_{i=1}^{m+1} T_\delta \tau_i f_{\mathfrak{q}_i^m} \Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}}. \tag{63}$$

2F. ***Proof of Proposition 2.3.*** For given $\beta > 0$, we need to show that $\mathcal{A}^\beta(s) \le C$ for $0 < s \le 1$ if $p \ge p_\circ(d)$. Let $\varepsilon > 0$ be small enough such that $(100d)^{-1}\beta \ge \varepsilon$, and choose a small $\varepsilon_\circ > 0$ and $N = N(\varepsilon)$ large enough such that Proposition 2.10 and Corollary 2.12 hold uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.

Let $0 < s < \delta \le 1$, and let $\sigma_1, \dots, \sigma_m$ be dyadic numbers satisfying (36). Since $A(\delta) \le C$ for $\delta \gtrsim 1$ and $s \le \sigma_k^{-2}\delta$, we see

$$A(\sigma_k^{-2}\delta) \le A(\sigma_k^{-2}\delta)\chi_{(0,10^{-2}]}(\sigma_k^{-2}\delta) + C \le (\sigma_k^{-2}\delta)^{-\frac{d-1}{2}+\frac{d}{p}-\beta} \mathcal{A}^\beta(s) + C. \tag{64}$$

By Proposition 2.10 and Lemma 2.6 we have, for $p \ge 2(m+1)/m$,

$$\sup_{\tau_1,\dots,\tau_{m+1}} \max_{\mathfrak{q}_1^m,\dots,\mathfrak{q}_{m+1}^m \text{ trans}} \Big\| \prod_{i=1}^{m+1} T_\delta \tau_i f_{\mathfrak{q}_i^m} \Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}} \lesssim (\sigma_1 \cdots \sigma_m)^{-C_\varepsilon} \delta^{-\varepsilon} \delta^{\frac{d}{p}-\frac{d-1}{2}} \|f\|_p, \tag{65}$$

which uniformly holds for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.

We have two types of estimates for $\mathfrak{M}^k f$. Since $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are already transversal,

$$\Big| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \subset [\mathfrak{N}](\mathfrak{Q}^k)}} \prod_{i=1}^{k} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big| \le \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \text{ trans}} \Big| \prod_{i=1}^{k} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big|.$$

Here we slightly abuse the definition "trans" and "$\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k$ trans" means (55) holds if $\xi_i \in \mathfrak{q}_i^k$, $i = 1, \dots, k$. Since there are as many as $O(\sigma_k^{-C})$ k-tuples $(\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k)$ and the above inequality holds regardless of $\mathfrak{Q}^k$, we get

$$\mathfrak{M}^k f \lesssim \sigma_k^{-C} \sup_{\tau_1,\dots,\tau_k} \max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \text{ trans}} \Big\| \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{\frac{p}{k}}^{\frac{1}{k}}.$$

Since $\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k$ are transversal, by Proposition 2.10 (also see Remark 2.8) and Lemma 2.6, we get, for $p \ge 2k/(k-1)$,

$$\Big\| \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{\frac{p}{k}}^{\frac{1}{k}} \lesssim (\sigma_1 \cdots \sigma_{k-1})^{-C_\varepsilon} \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon} \prod_{i=1}^k \|\tau_i f_{\mathfrak{q}_i^k}\|_p^{\frac{1}{k}} \lesssim \sigma_k^{-C_\varepsilon} \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon} \|f\|_p.$$

Hence, for $p \ge 2k/(k-1)$, we have the uniform estimate for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$

$$\mathfrak{M}^k f \lesssim \sigma_k^{-C} \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon} \|f\|_p. \tag{66}$$

On the other hand, fixing $\tau_1, \dots, \tau_k$, $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ trans, and $\mathfrak{Q}^k$, we consider the integrals appearing in the definition of $\mathfrak{M}^k f$. Let us write $\mathfrak{Q}^k = \mathfrak{q}(z, 1/\sigma_k)$. Using Corollary 2.12, for $2 \le p \le 2k/(k-1)$, we have

$$\Big\| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in [\mathfrak{N}](\mathfrak{Q}^k)}} \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}(\mathfrak{Q}^k)} \lesssim \sigma_{k-1}^{-C_\varepsilon} \sigma_k^{-\varepsilon} \prod_{i=1}^k \Big\| \Big(\sum_{\mathfrak{q}_i^k \in [\mathfrak{N}](\mathfrak{Q}^k)} |T_\delta(\tau_i f_{\mathfrak{q}_i^k})|^2 \Big)^{\frac{1}{2}} \rho_{B(z, \frac{C}{\sigma_k})} \Big\|_p. \tag{67}$$

Since $[\mathfrak{N}](\mathfrak{Q}^k) \subset \mathfrak{N}(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$, it is clear that if $\mathfrak{q}_i^k \in [\mathfrak{N}](\mathfrak{Q}^k)$, then $\mathfrak{q}_i^k \subset \mathrm{N}^{-1}(\Pi) + O(\sigma_k)$ for a k-plane Π. Since $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal and $\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}$, $i = 1, \dots, k$, we know $\sum_{\mathfrak{q}_1^k \in \mathfrak{N}(\mathfrak{Q}^k)} T_\delta(\tau_1 f_{\mathfrak{q}_1^k}), \ \dots, \ \sum_{\mathfrak{q}_k^k \in \mathfrak{N}(\mathfrak{Q}^k)} T_\delta(\tau_k f_{\mathfrak{q}_k^k})$ satisfy the assumptions of Corollary 2.12 (Proposition 2.11) with $\delta = \sigma_k$ and $\sigma = \sigma_1 \cdots \sigma_{k-1}$. Hence, Corollary 2.12 gives (67).

Recalling that the $\mathfrak{q}_i^k$ are contained in a $C\sigma_k$-neighborhood of $\Gamma(\psi)$, we see that $\#\mathfrak{N}(\mathfrak{Q}^k)$ is $\lesssim \sigma_k^{1-k}$. So, by Hölder's inequality we have

$$\Big\| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \subset \mathfrak{N}(\mathfrak{Q}^k)}} \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{1}{k}} \lesssim \sigma_{k-1}^{-C} \sigma_k^{-\varepsilon - p(k-1)(\frac{1}{2} - \frac{1}{p})} \max_{1 \le i \le k} \Big\| \Big(\sum_{\mathfrak{q}^k} |T_\delta(\tau_i f_{\mathfrak{q}^k})|^p \Big)^{\frac{1}{p}} \rho_{B(z, \frac{C}{\sigma_k})} \Big\|_p.$$

Here we bound $\sigma_1, \dots, \sigma_{k-1}$ with σ_{k-1} using (36) and replace C_ε with a larger constant C, since ε is fixed. By using the rapid decay of ρ we sum the estimates along $\mathfrak{Q}^k$ to get

$$\mathfrak{M}^k f \lesssim \sigma_{k-1}^{-C} \sigma_k^{-\varepsilon - (k-1)(\frac{1}{2} - \frac{1}{p})} \sup_h \Big\| \Big(\sum_{\mathfrak{q}^k} |T_\delta(\tau_h f_{\mathfrak{q}^k})|^p \Big)^{\frac{1}{p}} \Big\|_p. \tag{68}$$

By Proposition 2.5, Lemma 2.6, and (64) we get, for $2 \le p \le 2k/(k-1)$,

$$\mathfrak{M}^k f \lesssim \big(\sigma_{k-1}^{-C} \sigma_k^{\beta + \frac{2d-k-1}{2} - \frac{2d-k+1}{p}} \delta^{-\frac{d-1}{2} + \frac{d}{p} - \beta} \mathcal{A}^\beta(s) + \sigma_k^{-C} \big) \|f\|_p.$$

Here we also use $(100d)^{-1}\beta \ge \varepsilon$. So, if $p \ge 2(2d-k+1)/(2d-k-1)$,

$$\mathfrak{M}^k f \lesssim \big(\sigma_{k-1}^{-C}\sigma_k^{\alpha}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\mathcal{A}^{\beta}(s)+\sigma_k^{-C}\big)\|f\|_p$$

for some $\alpha > 0$. Combining this with (66), we have for some $\alpha > 0$

$$\mathfrak{M}^k f \lesssim \big(\sigma_k^{-C}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\varepsilon}+\sigma_{k-1}^{-C}\sigma_k^{\alpha}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\mathcal{A}^{\beta}(s)+\sigma_k^{-C}\big)\|f\|_p$$

provided that

$$p \ge \min\bigg(\frac{2(2d-k+1)}{2d-k-1},\frac{2k}{k-1}\bigg).$$

Since $(100d)^{-1}\beta \ge \varepsilon$ and $p_\circ > 2d/(d-1)$, from (64) we note that $A(\sigma_k^{-2}\delta) \lesssim \sigma_k^{\alpha}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\mathcal{A}^{\beta}(s)$. Thus, by (63), the above inequality, (64), and (65) we obtain

$$\|T_\delta f\|_p \lesssim \sum_{k=1}^{m}(\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+\sigma_k^{-C})\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\|f\|_p+\sigma_m^{-C}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\|f\|_p \tag{69}$$

for some $\alpha > 0$ provided that

$$p \ge \min\bigg(\frac{2(2d-k+1)}{2d-k-1},\frac{2k}{k-1}\bigg),\quad k=2,\dots,m \ \text{ and } \ p \ge \frac{2(m+1)}{m}. \tag{70}$$

Since the estimates (65)–(68) hold uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, so does (69). Taking supremum along ψ and f, we have

$$A(\delta) \le \bigg(\sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+C\sigma_m^{-C}\bigg)\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}.$$

By multiplying $\delta^{\frac{d-1}{2}-\frac{d}{p}-\beta}$ to both sides, $\delta^{\frac{d-1}{2}-\frac{d}{p}+\beta}A(\delta) \le \sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+C\sigma_m^{-C}$. This is valid as long as $s < \delta \le 1$. Hence, taking supremum for $s < \delta \le 1$ yields

$$\mathcal{A}^{\beta}(s) \le \sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+C\sigma_m^{-C}$$

if (70) is satisfied. Therefore, choosing $\sigma_1 \ll \cdots \ll \sigma_m$, successively, we can make $\sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha} \le \frac{1}{2}$. This gives the desired $\mathcal{A}^{\beta}(s) \le C\sigma_m^{-C}$ provided that (70) holds.

Finally, we only need to check that the minimum of

$$\mathcal{P}(m) = \max\bigg(\frac{2(m+1)}{m},\max_{k=2,\dots,m}\min\bigg(\frac{2(2d-k+1)}{2d-k-1},\frac{2k}{k-1}\bigg)\bigg),\quad 2 \le m \le d-1$$

is $p_\circ(d)$ as can be done by routine computation. □

Remark 2.15. The minimum of $\mathcal{P}$ is achieved when m is near $2d/3$. So, it doesn't seem that the argument makes use of the full strength of the multilinear restriction estimates.

3. Square function estimates

In this section we prove Theorem 1.2. We firstly obtain multi(sub)linear square-function estimates which are vector-valued extensions of multilinear restriction estimates. Then, we modify the argument in Section 2F to obtain the sharp square-function estimate from these multilinear estimates. Although the basic strategy here is similar to the one in the previous section, due to the additional integration in t we need to handle a family of surfaces. This argument in this section is very much in parallel with that of the previous section.

3A. ***One-parameter family of elliptic functions.*** As before, for $0 < \varepsilon_\circ \ll \frac{1}{2}$ and an integer $N \geq 100d$, we denote by $\overline{\mathfrak{G}}(\varepsilon_\circ, N)$ the class of smooth functions defined on $I^{d-1} \times I$ which satisfy

$$\|\psi - \psi_\circ - t\|_{C^N(I^{d-1}\times I)} \leq \varepsilon_\circ. \tag{71}$$

This clearly implies that, for all $(x,t) \in I^{d-1} \times I$,

$$\partial_t \psi(x,t) \in [1-\varepsilon_\circ, 1+\varepsilon_\circ]. \tag{72}$$

For $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ and $z_0 = (\zeta_0, t_0) \in \frac{1}{2} I^d$, define

$$\psi^\varepsilon_{z_0}(\zeta,t) = \varepsilon^{-2}\bigg(\psi\bigg(\zeta_0 + \varepsilon \mathcal{H}^\psi_{z_0}\zeta, t_0 + \frac{\varepsilon^2 t}{\partial_t\psi(z_0)}\bigg) - \psi(z_0) - \varepsilon\nabla_\zeta\psi(z_0)\mathcal{H}^\psi_{z_0}\zeta\bigg),$$

where $\mathcal{H}^\psi_{z_0} = (\sqrt{H(\psi(\cdot\,,t_0))(\zeta_0)})^{-1}$. Then we have the following.

Lemma 3.1. *Let $z_0 \in \frac{1}{2} I^d$ and $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$. There is a $\kappa = \kappa(\varepsilon_\circ, N) > 0$, independent of ψ, ζ_0, t_0, such that $\psi^\varepsilon_{z_0}$ is contained in $\overline{\mathfrak{G}}(\varepsilon_\circ, N)$ if $0 < \varepsilon \leq \kappa$.*

Proof. It is sufficient to show that $|\partial^\alpha_\zeta \partial^\beta_t (\psi^\varepsilon_{z_0}(\zeta,t) - \psi_\circ(\zeta) - t)| \leq C\varepsilon$, with C independent of $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, if $|\alpha| + \beta \leq N$ and $(\zeta,t) \in I^d$.

Let $0 < \varepsilon \leq \frac{1}{4}$. If $(\zeta,t) \in I^d$ and $|\alpha| + 2\beta > 2$, trivially $|\partial^\alpha_\zeta \partial^\beta_t (\psi^\varepsilon_{z_0}(\zeta,t) - \psi_\circ(\zeta,t) - t)| \leq C\varepsilon$ because $z_0 = (\zeta_0, t_0) \in \frac{1}{2} I^d$. Thus, it is sufficient to consider the cases $\beta = 1,\ |\alpha| = 0$ and $\beta = 0,\ 0 \leq |\alpha| \leq 2$. The first case is easy to handle. Indeed, from Taylor's theorem and (72)

$$\partial_t(\psi^\varepsilon_{z_0}(\zeta,t) - \psi_\circ - t) = (\partial_t\psi(z_0))^{-1}\bigg(\partial_t\psi\bigg(\zeta_0 + \varepsilon\mathcal{H}^\psi_{z_0}\zeta, t_0 + \frac{\varepsilon^2 t}{\partial_t\psi(z_0)}\bigg) - \partial_t\psi(z_0)\bigg) = O(\varepsilon).$$

To handle the second case, we consider Taylor's expansion of ψ in t with integral remainder:

$$\psi(\zeta,t) = \psi(\zeta,t_0) + \partial_t\psi(\zeta,t_0)(t-t_0) + R_1(\zeta,t),$$

where

$$R_1(\zeta,t) = (t-t_0)^2\int_0^1 (1-s)\partial_t^2\psi(\zeta,(t-t_0)s+t_0)\,ds.$$

The change of variables $t \to t_0 + \varepsilon^2(\partial_t\psi(z_0))^{-1}t,\ \zeta \to \zeta_0 + \varepsilon\mathcal{H}^\psi_{z_0}\zeta$ gives

$$\begin{aligned}\psi\bigg(\zeta_0 + \varepsilon\mathcal{H}^\psi_{z_0}\zeta, t_0 + \frac{\varepsilon^2 t}{\partial_t\psi(z_0)}\bigg)&\\ = \varepsilon^2\psi(\cdot\,,t_0)^\varepsilon_{\zeta_0}(\zeta) + \psi(z_0) + \varepsilon\nabla_\zeta\psi(z_0)\mathcal{H}^\psi_{z_0}\zeta &+ \frac{\varepsilon^2\partial_t\psi(\zeta_0+\varepsilon\mathcal{H}^\psi_{z_0}\zeta,t_0)}{\partial_t\psi(z_0)}t + \widetilde{R}(\zeta,t),\end{aligned}$$

where $\psi(\cdot,t_0)^{\varepsilon}_{\zeta_0}$ is defined by (8) and $\widetilde{R}(\zeta,t)=R_1(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,t_0+\varepsilon^2(\partial_t\psi(z_0))^{-1}t)$. Hence, it follows that

$$\psi^{\varepsilon}_{z_0}-\psi_\circ-t=\psi(\cdot,t_0)^{\varepsilon}_{\zeta_0}(\zeta)-\psi_\circ+\frac{\partial_t\psi(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,t_0)-\partial_t\psi(z_0)}{\partial_t\psi(z_0)}t+\varepsilon^{-2}\widetilde{R}(\zeta,t).$$

Since $\psi(\cdot,t_0)-t_0\in\mathfrak{G}(\varepsilon_\circ,N)$ and $(\psi(\cdot,t_0)-t_0)^{\varepsilon}_{\zeta_0}=\psi(\cdot,t_0)^{\varepsilon}_{\zeta_0}$, we have $|\partial^{\alpha}_{\zeta}(\psi(\cdot,t_0)^{\varepsilon}_{\zeta_0}-\psi_\circ)|\le C\varepsilon$ on I^d for $|\alpha|=0,1,2$ (similarly to the proof of Lemma 2.1). By (72) and the mean value theorem we also have $(\partial_t\psi(z_0))^{-1}\,\partial^{\alpha}_{\zeta}(\partial_t\psi(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,t_0)-\partial_t\psi(z_0))t=O(\varepsilon)$ in $C^N(I^{d-1})$ for $|\alpha|=0,1,2$. Note that

$$\varepsilon^{-2}\widetilde{R}(\zeta,t)=\frac{\varepsilon^2t^2}{(\partial_t\psi(z_0))^2}\int_0^1(1-s)\partial_t^2\psi\big(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,\varepsilon^2(\partial_t\psi(z_0))^{-1}ts+t_0\big)\,ds.$$

Thus, again by (72) it is easy to see that $\partial^{\alpha}_{\zeta}(\varepsilon^{-2}\widetilde{R})=O(\varepsilon^{2+|\alpha|})$ for any α. Therefore, combining the all together we have $|\partial^{\alpha}_{\zeta}(\psi^{\varepsilon}_{z_0}(\cdot,t)-\psi_\circ-t)|\le C\varepsilon$ on I^{d-1} for $|\alpha|=0,1,2$. □

3B. ***Square function with localized frequency.*** Abusing the conventional notation we denote by $m(D)f$ the multiplier operator given by $\widehat{m(D)f}(\xi)=m(\xi)\hat{f}(\xi)$, and we also write $D=(D',D_d)$ where D', D_d correspond to the frequency variables ζ, τ, respectively.

In order to show (4), by the Littlewood–Paley decomposition, scaling, and further finite decompositions, it is sufficient to show

$$\bigg\|\bigg(\int_{1-\varepsilon^2}^{1+\varepsilon^2}\bigg|\frac{\partial}{\partial t}\mathcal{R}^{\alpha}_t f(x)\bigg|^2dt\bigg)^{\frac12}\bigg\|_p\le C\|f\|_p$$

for some small $\varepsilon>0$. And by decomposing $\hat{f}$, which may now be assumed to be supported in $S^{d-1}+O(\varepsilon^2)$, and rotation we may assume $\hat{f}$ is supported in $B(-e_d,c\varepsilon^2)$ with some $c>0$. Hence, by discarding the harmless smooth multiplier, the matter reduces to showing

$$\big\|\|(D_d+\sqrt{t^2-|D'|^2})^{\alpha-1}_+f\|_{L^2_t(1-\varepsilon^2,1+\varepsilon^2)}\big\|_p\le C\|f\|_p.$$

By changing variables in the frequency domain, $D_d\to D_d+1$, $(D',D_d)\to(\varepsilon D',\varepsilon^2D_d)$ and $t\to\varepsilon^2t+1$, this is equivalent to

$$\big\|\|(D_d-\psi_{br}(D',t)))^{\alpha-1}_+\chi_\circ(D)f\|_{L^2_t(I)}\big\|_p\le C\|f\|_p,\tag{73}$$

where $\psi_{br}(\zeta,t)=\varepsilon^{-2}(1-\sqrt{1+2\varepsilon^2t+\varepsilon^4t^2-\varepsilon^2|\zeta|^2})$ and $\chi_\circ$ is a smooth function supported in a small neighborhood of the origin. Clearly, ψ_{br} satisfies (71) with $\varepsilon_\circ=C\varepsilon^2$ for some $C>0$. Consequently, we are led to consider general $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ rather than the specific ψ_{br}.

Let us define the class $\mathcal{E}(N)$ of smooth functions by setting

$$\mathcal{E}(N)=\big\{\eta\in C^\infty(I^d\times I):\|\eta\|_{C^N(I^d\times I)}\le1,\ \tfrac12\le\eta\le1\big\}.$$

Let $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ and $\eta\in\mathcal{E}(N)$. For $0<\delta$ and f with $\hat{f}$ supported in $\frac12I^d$, we define $S_\delta=S_\delta(\psi,\eta)$ by

$$S_\delta f(x)=\bigg\|\phi\bigg(\frac{\eta(D,t)(D_d-\psi(D',t))}{\delta}\bigg)f\bigg\|_{L^2_t(I)}.\tag{74}$$

Compared to ψ, the role of η is less significant but this enables us to handle more general square functions (in particular, see Remark 3.3). By dyadic decomposition away from the singularity, the matter of showing (73) is reduced to obtaining the sharp bound

$$\|S_\delta f\|_p \le C\delta^{\frac{d}{p}-\frac{d-2}{2}-\varepsilon}\|f\|_p, \quad \varepsilon > 0, \tag{75}$$

when $\hat{f}$ is supported in a small neighborhood of the origin. This is currently verified for $p \ge 2(d+2)/d$ [Lee et al. 2012] by making use of the bilinear restriction estimate for the elliptic surfaces. The following is our main result concerning the estimate (75).

Proposition 3.2. *Let $p_s = p_s(d)$ be given by* (5) *and* supp $\hat{f} \subset \frac{1}{2}I^d$. *If $p \ge \min(p_s(d), 2(d+2)/d)$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that* (75) *holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$.*

Proof of Theorem 1.2. By choosing a small $\varepsilon > 0$ in the above, we can make ψ_{br} be in $\overline{\mathfrak{G}}(\varepsilon_\circ, N)$ for any ε_0 and N. Hence, Proposition 3.2 gives (75) for any $\varepsilon > 0$ if $p \ge \min(p_s(d), 2(d+2)/d)$. Hence, dyadic decomposition of the multiplier operator in (73) and using (75) followed by summation along dyadic pieces gives (73) for $\alpha > d/2 - d/p$. This proves Theorem 1.2. □

Remark 3.3. As has been shown before, for the proof of Theorem 1.2 it suffices to consider an operator which is defined without η, but by allowing η in (74) we can handle the square-function estimates for the operator $f \to \phi((1-|D|/t)/\delta)f$, which is closely related to smoothing estimates for the solutions to the Schrödinger and wave equations; for example, see [Lee et al. 2012]. In fact, Proposition 3.2 implies, for $\varepsilon > 0$,

$$\left\| \left(\int_{\frac{1}{2}}^{2} \left| \phi\left(\frac{1-|D|/t}{\delta}\right) f \right|^2 dt \right)^{\frac{1}{2}} \right\|_p \le \delta^{\frac{d}{2}-\frac{d}{p}-\varepsilon} C \|f\|_p \tag{76}$$

if $p \ge p_s(d)$. Indeed, by finite decompositions, rotation and scaling, as before, it is sufficient to consider the time average over the interval $I_\varepsilon = (1-\varepsilon^2, 1+\varepsilon^2)$ and we may assume that $\hat{f}$ is supported in $B(-e_d, c\varepsilon^2)$. Writing

$$1 - |\xi|/t = t^{-2}(t+|\xi|)^{-1}(\tau - \sqrt{t^2 - |\zeta|^2})(\tau + \sqrt{t^2 - |\zeta|^2})$$

for $\xi \in B(-e_d, c\varepsilon^2)$, the same change of variables $D_d \to D_d + 1$, $(D', D_d) \to (\varepsilon D', \varepsilon^2 D_d)$ and $t \to \varepsilon^2 t + 1$ transforms $\phi((1-|\xi|/t)/\delta)$ to

$$\phi\left(\frac{\eta(\xi,t)(\tau - \psi_{br})}{\varepsilon^{-2}\delta/2}\right)$$

with a smooth η which satisfies $\eta \in (1 - c\varepsilon/2, 1 + c\varepsilon/2)$. Hence, we now apply Proposition 3.2 with sufficiently small ε to get (76).

As before, in order to control the L^p norm of S_δ we define $B(\delta) = B_p(\delta)$ by

$$B(\delta) \equiv \sup\{\|S_\delta(\psi,\eta)f\|_{L^p} : \psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N),\ \eta \in \mathcal{E}(N),\ \|f\|_p \le 1,\ \text{supp } \hat{f} \subset \tfrac{1}{2}I^d\}.$$

As before, using Lemma 2.9 it is easy to see that $B(\delta)\le C$ if $\delta\ge 1$, and $B(\delta)\le C\delta^{-c}$ for some $c>0$ otherwise (for example, see the paragraph below Proposition 3.6). We also define for $\beta>0$ and $\delta\in(0,1)$,

$$\mathcal{B}^{\beta}(\delta)=\mathcal{B}^{\beta}_{p}(\delta)\equiv\sup_{\delta<s\le 1} s^{\frac{d-2}{2}-\frac{d}{p}+\beta}B_p(s).$$

Thus, Theorem 1.2 follows if we show $\mathcal{B}^{\beta}(\delta)\le C$ for any $\beta>0$. As observed in the previous section, the bound for $S_\delta f$ improves if the Fourier transform of f is contained in a set of smaller diameter. The following plays a crucial role in the induction argument (see Section 3F).

Proposition 3.4. *Let* $0<\delta\ll 1$, $\psi\in\overline{\mathfrak{C}}(\varepsilon_\circ,N)$, *and* $\eta\in\mathcal{E}(N)$. *Suppose that* $\hat f$ *is supported in* $\mathfrak{q}(a,\varepsilon)$, $10\sqrt{\delta}\le\varepsilon\le\frac12$, *and* $a\in\frac12 I^d$. *Then, if* $\varepsilon_\circ>0$ *is small enough, there is a* $\kappa=\kappa(\varepsilon_\circ,N)$ *such that*

$$\|S_\delta(\psi,\eta)f\|_p\le C\varepsilon^{\frac1p+\frac12}B_p(\varepsilon^{-2}\delta)\|f\|_p \tag{77}$$

holds with C *independent of* ψ, *and* ε, *whenever* $\varepsilon\le\kappa$.

Proof. By breaking the support of $\hat f$ into a finite number of dyadic cubes, we may assume that $\hat f$ is supported in $\mathfrak{q}(a,\nu\varepsilon)$ for a small constant $\nu>0$ satisfying $\nu^2d^2\in[2^{-5},2^{-4})$. This only increases the bound by a constant multiple. Since $\hat f$ is supported in $\mathfrak{q}(a,\nu\varepsilon)$ and $a=(a',a_d)\in\frac12 I^d$, from (72) and the fact that $\frac12\le\eta\le 1$, it is clear that $\phi\big(\eta(D,t)(D_d-\psi(D',t))/\delta\big)f\not\equiv 0$ for t contained in an interval $[\alpha,\beta]$ of length $\lesssim\nu\varepsilon$ because $\phi\big(\eta(\xi,t)(\tau-\psi(\zeta,t))/\delta\big)$ is supported in an $O(\delta)$-neighborhood of $\tau=\psi(\zeta,t)$.

Let $\alpha=t_0<t_1<\cdots<t_l=\beta$, $l\le O(\varepsilon^{-1})$, be such that $t_{k+1}-t_k\le\nu^2\varepsilon^2$. Since $\delta\le 10^{-2}\varepsilon^2$, by (71) and (72) it follows that if $t\in[t_k,t_{k+1}]$, then $\phi\big(\eta(\xi,t)(\tau-\psi(\zeta,t))/\delta\big)\hat f(\xi)$ is supported in the parallelepiped

$$\mathcal{P}_k=\Big\{(\zeta,\tau):\max_{i=1,\dots,d-1}|\zeta_i-a_i'|<\nu\varepsilon,\ |\tau-\psi(a',t_k)-\nabla_\zeta\psi(a',t_k)(\zeta-a')|\le 2d^2\nu^2\varepsilon^2\Big\}.$$

This follows from Taylor's theorem since $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$. By (72) it is easy to see that $\{\mathcal{P}_k\}_{k=1}^{l}$ are overlapping boundedly. In fact, $\phi\big(\eta(\xi,t)(\tau-\psi(\zeta,t))/\delta\big)\hat f(\xi)$, $t\in[t_k,\ t_{k+1}]$, is supported in

$$\widetilde{\mathcal{P}}_k=\{\xi\in\mathfrak{q}(a,c\varepsilon):|\tau-\psi(\zeta,t_k))|\le C\varepsilon^2\},\quad k=0,\dots,l-1,$$

with $C\ge 3d^2\nu^2\varepsilon^2$ and the $\{\widetilde{\mathcal{P}}_k\}$ are boundedly overlapping because of (72), and by Taylor's expansion it is easy to see that $\mathcal{P}_k\subset\widetilde{\mathcal{P}}_k$ because the second remainder is uniformly $O(\varepsilon^2)$ for $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$.

Let φ be a smooth function supported in $2I^d$ and $\varphi=1$ on I^d. Let $L_{\mathcal{P}_k}$ be the affine map which bijectively maps $\mathcal{P}_k$ to I^d, and set $\varphi_{\mathcal{P}_k}=\varphi(L_{\mathcal{P}_k}\,\cdot)$ so that $\varphi_{\mathcal{P}_k}$ vanishes outside of $2\mathcal{P}_k$ and equals 1 on $\mathcal{P}_k$. Here $2\mathcal{P}_k$ denotes the parallelepiped which is given by dilating $\mathcal{P}_k$ twice from the center of $\mathcal{P}_k$. Then we have

$$(S_\delta f(x))^2=\sum_k\int_{I_k}\bigg|\phi\Big(\frac{\eta(D,t)(D_d-\psi(D',t))}{\delta}\Big)\varphi_{\mathcal{P}_k}(D)f(x)\bigg|^2dt.$$

Since $p\ge 2$, by Hölder's inequality it follows that

$$S_\delta f(x)\le C\varepsilon^{\frac1p-\frac12}\bigg(\sum_k\Big\|\phi\Big(\frac{\eta(D,t)(D_d-\psi(D',t))}{\delta}\Big)\varphi_{\mathcal{P}_k}(D)f(x)\Big\|^p_{L^2_t(I_k)}\bigg)^{\frac1p}.$$

Hence it is sufficient to show that

$$\bigg\| \bigg\| \phi\Big(\frac{\eta(D,t)(D_d-\psi(D',t))}{\delta}\Big)\varphi_{\mathcal{P}_k}(D)f\bigg\|_{L^2_t(I_k)}\bigg\|_p \le C\varepsilon B_p(\varepsilon^{-2}\delta)\|\varphi_{\mathcal{P}_k}(D)f\|_p \tag{78}$$

because $\big(\sum_k \|\varphi_{\mathcal{P}_k}(D)f\|_p^p\big)^{\frac{1}{p}} \le C\|f\|_p$ for $2 \le p \le \infty$. This follows by interpolation between the estimates for $p=2$ and $p=\infty$. The first is an easy consequence of Plancherel's theorem because the $\{2\mathcal{P}_k\}$ are boundedly overlapping and the latter is clear since $\mathcal{F}^{-1}(\phi_{\mathcal{P}_k}) \in L^1$ uniformly.

Now we make the change of variables

$$t \to \varepsilon^2(\partial_t\psi(a',t_k))^{-1}t + t_k, \quad \xi \to L(\xi) = (L'(\xi), L_d(\xi)),$$

where

$$L'(\xi) = \varepsilon\mathcal{H}^{\psi}_{(a',t_k)}\zeta + a', \quad L_d(\xi) = \varepsilon^2\tau + \psi(a',t_k) + \varepsilon\nabla_\zeta\psi(a',t_k)\mathcal{H}^{\psi}_{(a',t_k)}\zeta,$$

and

$$\varepsilon^2 x_d \to x_d, \quad \varepsilon\mathcal{H}^{\psi}_{(a',t_k)}(x' + x_d\nabla_\zeta\psi(a',t_k)) \to x'.$$

Then, (78) follows if we show

$$\bigg\| \bigg\| \phi\Big(\frac{\eta(L(D),t)(D_d-\psi^{\varepsilon}_{a',t_k}(D',t))}{\varepsilon^{-2}\delta}\Big)f\bigg\|_{L^r_t(0,2\nu^2)}\bigg\|_p \le CB_p(\varepsilon^{-2}\delta)\|f\|_p$$

when the support $\hat{f}$ is contained in $L^{-1}(2\mathcal{P}_k)$. Clearly, $\eta(L(\xi),t) \in \mathcal{E}(N)$ and $L^{-1}(2\mathcal{P}_k)$ is contained in the set $\{(\zeta,\tau) : |\zeta| \le 4\nu, |\tau| \le 8d^2\nu^2\} \subset \frac{1}{2}I^d$. From Lemma 3.1 there exists $\kappa > 0$ such that $\psi^{\varepsilon}_{a',t_k} \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ if $0 < \varepsilon \le \kappa$. Hence, using the definition of $B_p(\delta)$ we get the desired inequality for $\varepsilon \le \kappa$. □

3C. ***Multi(sub)linear square-function estimates.*** Let $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ and set

$$\Gamma^t = \Gamma^t(\psi) := \big\{(\zeta, \psi(\zeta,t)) : \zeta \in \tfrac{1}{2}I^d\big\}. \tag{79}$$

As before we denote by $\Gamma^t(\delta)$ the δ-neighborhood $\Gamma^t + O(\delta)$. Clearly, from (72) it follows that, for $\delta > 0$,

$$\Gamma^t(\delta) \cap \Gamma^s(\delta) = \varnothing \quad \text{if } |t-s| \ge C\delta \tag{80}$$

for some $C > 0$. We also denote by N^t the (upward) normal map from the surface Γ^t to $\mathbb{S}^{d-1}$.

Definition 3.5 (normal vector field $\boldsymbol{n} = \boldsymbol{n}(\psi)$). The map $(\zeta,t) \to (\zeta,\psi(\zeta,t))$ is clearly one-to-one and we may assume that the image of this map contains I^d by extending $\psi(\zeta,t)$ to a larger set $I^{d-1} \times CI$, while (71) is satisfied. Hence, for each $\xi = (\zeta,\tau) \in I^d$ there is a unique t such that $\xi = (\zeta,\psi(\zeta,t))$. Then we define $\boldsymbol{n}(\xi)$ to be the normal vector to Γ^t at ξ, which forms a vector field on I^d.

A natural attempt for the multilinear generalization of S_δ is to consider $\prod_{i=1}^k S_\delta f_i$ under a transversality condition between supp f_i. But, the induction-on-scale argument does not work well with this naive generalization and it doesn't seem easy to obtain the sharp multilinear square-function estimates directly. We get around this difficulty by considering a vector-valued extension in which we discard the exact structure of the operator S_δ. As is clearly seen in its proof, the estimate in Proposition 3.6 is not limited to the surfaces given by $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ but it holds for a more general class of surfaces as long as the transversality is satisfied.

Proposition 3.6. *Let $2 \le k \le d$ be an integer and $0 < \sigma \ll 1$, and let Γ^t be given by $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, and the functions G_i, $1 \le i \le k$, be defined on $\mathbb{R}^d \times I$. Suppose that, for each $t \in I$, $G_1(\cdot, t), \dots, G_k(\cdot, t)$ satisfy that, for $0 < \delta \ll \sigma$,*

$$\operatorname{supp} \widehat{G}_i(\cdot, t) \subset \Gamma^t(\delta), \quad t \in I, \tag{81}$$

and suppose that

$$\mathrm{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \dots, \boldsymbol{n}(\xi_k)) \gtrsim \sigma, \tag{82}$$

whenever $\xi_i \in \operatorname{supp} \widehat{G}_i(\cdot, t) + O(\delta)$ for some $t \in I$. Then, if $p \ge 2k/(k-1)$ and $\varepsilon_\circ > 0$ is small enough, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\| \prod_{i=1}^k \|G_i\|_{L_t^2(I)} \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon} \delta^{-\varepsilon} \prod_{i=1}^k (\delta^{\frac{1}{2}} \|G_i\|_{L^2_{x,t}}) \tag{83}$$

holds with C, C_ε independent of ψ.

Without being concerned about the optimal α for a while, we first observe that, for $p \ge 2$, there is an α such that

$$\big\| \|G_i\|_{L_t^2(I)} \big\|_{L^p(\mathbb{R}^d)} \le C\delta^{-\alpha} \|G_i\|_{L^2_{x,t}} \tag{84}$$

holds uniformly if $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ and N is large enough ($N \ge 100d$). (It is enough to keep $\|\psi\|_{C^N(I^d)}$ uniformly bounded.) To see this, let φ be a smooth function supported in $2I$ with $\varphi = 1$ on I, and we set $K_\delta^t = \mathcal{F}^{-1}\big(\varphi\big((\tau - \psi(\zeta, t))/C\delta\big)\tilde{\chi}(\xi)\big)$. Then, by Lemma 2.9 $|K_\delta^t(x)| \le C\delta\mathfrak{K}_M(x)$ for a large M with C depending only on $\|\psi\|_{C^N(I^d)}$. Since $\operatorname{supp} \mathcal{F}(G_i(\cdot, t)) \subset \Gamma^t(\delta)$, we have $G_i(\cdot, t) = K_\delta^t * G_i(\cdot, t)$. So, $|G_i(x, t)| \le C\delta\mathfrak{K}_M * |G_i(\cdot, t)|$, $t \in I$, and by Minkowski's inequality we get

$$\|G_i(x, t)\|_{L_t^2(I)} \le C\delta\mathfrak{K}_M * (\|G_i(\cdot, t)\|_{L_t^2(I)})(x). \tag{85}$$

Young's convolution inequality gives (84), namely with $\alpha = d - 1$, if taking sufficiently large M.

Proof of Proposition 3.6. Since

$$\mathcal{F}(G_i(\cdot, t)) = \varphi\Big(\frac{\tau - \psi(\zeta, t)}{C\delta}\Big)\tilde{\chi}(\xi)\mathcal{F}(G_i(\cdot, t)),$$

by Schwarz's inequality and Plancherel's theorem, $|G_i(x, t)| \lesssim \delta^{\frac{1}{2}} \|G_i(\cdot, t)\|_2$. So, this gives (83) for $p = \infty$. Thus, by interpolation it is sufficient to show (83) with $p = 2k/(k-1)$.

Let us set $R = \delta^{-1}$ and we may set $x = 0$. Following the same argument as in the proof of Proposition 2.11 we start with the assumption that, for $0 < \delta \ll \sigma$,

$$\Big\| \prod_{i=1}^k \|G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}(B(0,R))} \lesssim R^\alpha R^{-\frac{k}{2}} \prod_{i=1}^k \|G_i\|_{L^2_{x,t}} \tag{86}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ whenever (81) and (82) are satisfied. By (84) and Hölder's inequality, this is true for a large $\alpha > 0$. Hence, it is sufficient to show (86) implies that for $\varepsilon > 0$ there is an $N = N(\varepsilon)$

such that, for some $\kappa > 0$,

$$\Big\| \prod_{i=1}^{k} \|G_i\|_{L^2_t(I)} \Big\|_{L^{2/(k-1)}(B(0,R))} \lesssim C_\varepsilon \sigma^{-\kappa} R^{\frac{\alpha}{2}+c\varepsilon} R^{-\frac{k}{2}} \prod_{i=1}^{k} \|G_i\|_{L^2_{x,t}} \tag{87}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$. Then, iterating this implication from (86) to (87) gives the desired inequality; see the paragraph below (20).

Since $\hat{\rho}_{B(z,\sqrt{R})}$ is supported in a ball of radius $\sim R^{-\frac{1}{2}}$, the Fourier transform of $\rho_{B(z,\sqrt{R})} G_i(\cdot,t)$ is contained in $\Gamma^t + O(R^{-\frac{1}{2}})$ for each t and (82) holds with $\delta = R^{-\frac{1}{2}}$ since $\delta \ll \sigma$. Hence, by the assumption (86), it follows that

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_t(I)} \Big\|_{L^{2/(k-1)}(B(z,\sqrt{R}))} \le C R^{\frac{\alpha}{2}} R^{-\frac{k}{4}} \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_{x,t}}. \tag{88}$$

We now decompose $G_i(\cdot,t)$ into $\{G_{i,\boldsymbol{q}}(\cdot,t)\}$, which is defined by

$$\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot,t)) = \chi_{\boldsymbol{q}} \mathcal{F}(G_i(\cdot,t)). \tag{89}$$

Here $\{\boldsymbol{q}\}$ are the dyadic cubes of side length l, $R^{-\frac{1}{2}} < l \le 2R^{-\frac{1}{2}}$, which we already used in the proof of Proposition 2.11. We write

$$G_i(x,t) = \sum_{\boldsymbol{q}} G_{i,\boldsymbol{q}}(x,t).$$

In what follows we assume $G_{i,\boldsymbol{q}} \neq 0$. By (81) it follows that, for each t, the cubes $\{\boldsymbol{q}\}$ appearing in the sum are contained in $\Gamma^t(R^{-\frac{1}{2}})$ because $G_{i,\boldsymbol{q}}(\cdot,t) = 0$, otherwise. We also note from (72) that there is an interval $I_{i,\boldsymbol{q}}$ of length $CR^{-\frac{1}{2}}$ such that $G_{i,\boldsymbol{q}}(\cdot,t) = 0$ if $t \notin I_{i,\boldsymbol{q}}$. Hence we may multiply the characteristic function of $\chi_{I_{i,\boldsymbol{q}}}$ so that

$$G_{i,\boldsymbol{q}} = G_{i,\boldsymbol{q}}(\cdot,t)\chi_{I_{i,\boldsymbol{q}}}(t). \tag{90}$$

Since the Fourier supports of $\{\rho_{B(z,\sqrt{R})} G_{i,\boldsymbol{q}}(\cdot,t)\}$ are boundedly overlapping, by Plancherel's theorem it follows that

$$\prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_{x,t}} \le C \prod_{i=1}^{k} \Big\| \Big(\sum_{\boldsymbol{q}} |\rho_{B(z,\sqrt{R})} G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}}. \tag{91}$$

Combining this with (88) we have

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_t(I)} \Big\|_{L^{2/(k-1)}} \le C R^{\frac{\alpha}{2}} R^{-\frac{k}{4}} \prod_{i=1}^{k} \Big\| \Big(\sum_{\boldsymbol{q}} |\rho_{B(z,\sqrt{R})} G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}}.$$

Since $\rho_{B(z,\sqrt{R})}$ is rapidly decaying outside of $B(z,\sqrt{R})$, we have for any large $M > 0$

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_t(I)} \Big\|_{L^{2/(k-1)}} \lesssim R^{\frac{\alpha}{2}-\frac{k}{4}} \prod_{i=1}^{k} \Big\| \chi_{B(z,R^{1/2+\varepsilon})} \Big(\sum_{\boldsymbol{q}} |G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}} + R^{-M} \prod_{i=1}^{k} \|G_i\|_{L^2_{x,t}}. \tag{92}$$

We now partition the interval $I_{i,\boldsymbol{q}}$ further into intervals $I^l_{i,\boldsymbol{q}}=[t_l,t_{l+1}]$, $l=1,\dots,\ell_0$, of length $\sim R^{-1}$. Then the Fourier support of $G_{i,\boldsymbol{q}}(\cdot,t)$, $t\in I^l_{i,\boldsymbol{q}}=[t_l,t_{l+1}]$, is contained in an $O(R^{-1})$ neighborhood of Γ^{t_l}. Let $(\zeta_{\boldsymbol{q}},\tau_{\boldsymbol{q}})$ be the center of $\boldsymbol{q}$ and we define a set $\boldsymbol{r}^l_{i,\boldsymbol{q}}$ by

$$\boldsymbol{r}^l_{i,\boldsymbol{q}}=\big\{(\zeta,\tau):|\zeta-\zeta_{\boldsymbol{q}}|\le C\delta^{\frac12},\ |\tau-\psi(\zeta_{\boldsymbol{q}},t_l)-\nabla_\zeta\psi(\zeta_{\boldsymbol{q}},t_l)\cdot(\zeta-\zeta_{\boldsymbol{q}})|\le C\delta\big\}\tag{93}$$

with a constant $C>0$ large enough. It follows that the Fourier transform of $G_{i,\boldsymbol{q}}(\cdot,t)$, $t\in I^l_{i,\boldsymbol{q}}$, is supported in $\boldsymbol{r}^l_{i,\boldsymbol{q}}$. This is easy to see from the second-order Taylor approximation because $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$.

Also define $\mathfrak{m}^l_{i,\boldsymbol{q}}$ by

$$\mathfrak{m}^l_{i,\boldsymbol{q}}=\rho\bigg(\frac{\zeta-\zeta_{\boldsymbol{q}}}{C\sqrt{\delta}},\frac{\tau-\psi(\zeta_q,t_l)-\nabla_\zeta\psi(\zeta_q,t_l)\cdot(\zeta-\zeta_{\boldsymbol{q}})}{C\delta}\bigg)\tag{94}$$

with a suitable $C>0$ such that $\mathfrak{m}^l_{i,\boldsymbol{q}}$ is comparable to 1 on $\boldsymbol{r}^l_{i,\boldsymbol{q}}$. Now, we set

$$\mathcal{F}(G^l_{i,\boldsymbol{q}}(\cdot,t))=(\mathfrak{m}^l_{i,\boldsymbol{q}})^{-1}\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot,t))\chi_{I^l_{i,\boldsymbol{q}}}(t).\tag{95}$$

Denoting by $\boldsymbol{n}^l_{i,\boldsymbol{q}}$ the normal vector $\boldsymbol{n}(\zeta_{\boldsymbol{q}},\psi(\zeta_{\boldsymbol{q}},t_l))$, we also set with a large $C>0$

$$\boldsymbol{T}^l_{i,\boldsymbol{q}}=\big\{x:|x\cdot\boldsymbol{n}^l_{i,\boldsymbol{q}}|\le C,\ |x-(x\cdot\boldsymbol{n}^l_{i,\boldsymbol{q}})\boldsymbol{n}^l_{i,\boldsymbol{q}}|\le CR^{-\frac12}\big\}.$$

Let us set $K^l_{i,\boldsymbol{q}}=\mathcal{F}^{-1}(\mathfrak{m}^l_{i,\boldsymbol{q}})$ so that $G_{i,\boldsymbol{q}}(\cdot,t)=G^l_{i,\boldsymbol{q}}(\cdot,t)*K^l_{i,\boldsymbol{q}}$ if $t\in I^l_{i,\boldsymbol{q}}$. Since $\hat\rho$ is supported in $\mathfrak{q}(0,1)$,

$$|K^l_{i,\boldsymbol{q}}|\lesssim R^{-\frac{d+1}{2}}\chi_{R\boldsymbol{T}^l_{i,\boldsymbol{q}}}.$$

By (90) it follows that

$$\sum_{\boldsymbol{q}}\|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I)}=\sum_{\boldsymbol{q}}\|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I_{i,\boldsymbol{q}})}=\sum_{\boldsymbol{q},l}\|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I^l_{i,\boldsymbol{q}})}.$$

Thus, by (95) we have

$$\begin{aligned}\sum_{\boldsymbol{q}}\|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I)}&=\sum_{\boldsymbol{q},l}\|G^l_{i,\boldsymbol{q}}(\cdot,t)*K^l_{i,\boldsymbol{q}}\|^2_{L^2(I^l_{i,\boldsymbol{q}})}\\&\lesssim\sum_{\boldsymbol{q},l}\|G^l_{i,\boldsymbol{q}}(\cdot,t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})}*|K^l_{i,\boldsymbol{q}}|\\&\lesssim\sum_{\boldsymbol{q},l}\|G^l_{i,\boldsymbol{q}}(\cdot,t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})}*(R^{-\frac{d+1}{2}}\chi_{R\boldsymbol{T}^l_{i,\boldsymbol{q}}}).\end{aligned}\tag{96}$$

We denote by $\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}$ the tube $R^{1+\varepsilon}\boldsymbol{T}^l_{i,\boldsymbol{q}}$, which is an $R^{1+\varepsilon}$-times dilation of $\boldsymbol{T}^l_{i,\boldsymbol{q}}$ from its center. So, from (96) we have, for $x,y\in B(z,R^{\frac12+\varepsilon})$,

$$\sum_{\boldsymbol{q}}\|G_{i,\boldsymbol{q}}(y,\cdot)\|^2_{L^2_t(I)}\lesssim R^{c\varepsilon}\sum_{\boldsymbol{q},l}\|G^l_{i,\boldsymbol{q}}(\cdot,t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})}*\bigg(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\bigg)(x).$$

Once we have this equality we can repeat the argument from (23) to (26) which is in the proof of Proposition 2.11 and also using (92), we have

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L_t^2(I)}\Big\|_{L^{2/(k-1)}(B(0,R))} \lesssim R^{c\varepsilon+\frac{\alpha}{2}+\frac{d-k}{4}}\bigg\|\prod_{i=1}^{k}\bigg(\sum_{\boldsymbol{q},l}\|G_{i,\boldsymbol{q}}^{l}(\,\cdot\,,t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})} * \bigg(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\bigg)\bigg)^{\frac{1}{2}}\bigg\|_{L^{2/(k-1)}(B(0,2R))} + \mathcal{E},$$

where $\mathcal{E} = R^{-M}\prod_{i=1}^{k}\|G_i\|_{L^2_{x,t}}$ for any large $M > 0$. Hence, for (87) it suffices to show that

$$\bigg\|\prod_{i=1}^{k}\bigg(\sum_{\boldsymbol{q},l}\|G_{i,\boldsymbol{q}}^{l}(\,\cdot\,,t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})} * \bigg(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\bigg)\bigg)^{\frac{1}{2}}\bigg\|_{L^{2/(k-1)}(B(0,2R))} \lesssim \sigma^{-\kappa}R^{c\varepsilon}R^{-\frac{d+k}{4}}\prod_{i=1}^{k}\|G_i\|_{L^2_{x,t}}.$$

Since by (95) $\big\|\|G_{i,\boldsymbol{q}}^{l}\|_{L_t^2(I^l_{i,\boldsymbol{q}})}\big\|_2 \sim \big\|\|G_{i,\boldsymbol{q}}\|_{L_t^2(I^l_{i,\boldsymbol{q}})}\big\|_2$, making use of the disjointness of $I^l_{i,\boldsymbol{q}}$ and the supports of $\mathcal{F}(G_{i,\boldsymbol{q}}(\,\cdot\,,t))$, and by Plancherel's theorem,

$$\sum_{\boldsymbol{q},l}\big\|\|G_{i,\boldsymbol{q}}^{l}\|_{L_t^2(I^l_{i,\boldsymbol{q}})}\big\|_2^2 \sim \sum_{\boldsymbol{q}}\|G_{i,\boldsymbol{q}}\|^2_{L_t^2(I)} = \big\|\|G_i\|_{L_t^2(I)}\big\|_2^2.$$

Hence, the above inequality follows from

$$\bigg\|\prod_{i=1}^{k}\sum_{\boldsymbol{q},l} f^l_{i,\boldsymbol{q}} * \bigg(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\bigg)\bigg\|_{L^{1/(k-1)}(B(0,2R))} \le C\sigma^{-\kappa}R^{c\varepsilon}R^{-\frac{d+k}{2}}\prod_{i=1}^{k}\sum_{\boldsymbol{q},l}\|f^l_{i,\boldsymbol{q}}\|_1.$$

Let $\mathcal{I}_i = \{(\boldsymbol{q},l) : G^l_{i,\boldsymbol{q}} \neq 0\}$, $I_i \subset \mathcal{I}_i$ and $\mathcal{T}^l_{i,\boldsymbol{q}}$ be a finite subset of $\mathbb{R}^d$. By scaling and pigeonholing, losing $(\log R)^C$ in its bound, this reduces to

$$\bigg\|\prod_{i=1}^{k}\sum_{(\boldsymbol{q},l)\in\mathcal{I}_i}\sum_{\tau\in\mathcal{T}^l_{i,\boldsymbol{q}}}\chi_{\boldsymbol{T}^l_{i,\boldsymbol{q}}+\tau}\bigg\|_{L^{1/(k-1)}(B(0,2))} \le C\sigma^{-\kappa}R^{c\varepsilon}R^{\frac{d-k}{2}}\prod_{i=1}^{k}\sum_{(\boldsymbol{q},l)\in\mathcal{I}_i}\sum_{\tau\in\mathcal{T}^l_{i,\boldsymbol{q}}}|\boldsymbol{T}^l_{i,\boldsymbol{q}}+\tau|. \tag{97}$$

Here we note that if $G_{i,\boldsymbol{q}} \neq 0$, then $\boldsymbol{q} \in \operatorname{supp}\mathcal{F}(G_i(\,\cdot\,,t)) + O(\sqrt{\delta})$ for some t. So, by (82) we have $\mathrm{Vol}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k) \gtrsim \sigma$ whenever $\boldsymbol{n}_i \in \{\boldsymbol{n}^l_{i,\boldsymbol{q}} : G^l_{i,\boldsymbol{q}} \neq 0\}$, $i = 1,\dots,k$. Therefore, the estimate follows from the multilinear Kakeya estimate which is stated below in Theorem 3.7. □

Theorem 3.7 [Bennett et al. 2006; Guth 2010; Carbery and Valdimarsson 2013]. *Let $2 \le k \le d$, $1 \ll R$ and $\mathfrak{T}_i$, $i = 1,2,\dots,k$, be collections of tubes of width $R^{-\frac{1}{2}}$ (possibly with infinite length), with major axes parallel to the vectors in $\Theta_i \subset \mathbb{S}^{d-1}$. Suppose $\mathrm{Vol}(\theta_1,\theta_2,\dots,\theta_k) \ge \sigma$ holds whenever $\theta_i \in \Theta_i$, $i = 1,\dots,k$. Then there is a constant C such that, for any subset $\mathcal{T}_i \subset \mathfrak{T}_i$, $i = 1,\dots,k$,*

$$\bigg\|\prod_{i=1}^{k}\bigg(\sum_{T_i\in\mathcal{T}_i}\chi_{T_i}\bigg)\bigg\|_{L^{1/(k-1)}(B(0,1))} \le CR^{\frac{d-k}{2}}\sigma^{-1}\prod_{i=1}^{k}\bigg(\sum_{T_i\in\mathcal{T}_i}|T_i|\bigg).$$

This is a rescaled version of the estimate due to Guth [2010] (the case $d = k$) and Carbery and Valdimarsson [2013]; also see [Bennett et al. 2006]. However, we don't need the endpoint estimate for our purpose and the estimate in [Bennett et al. 2006] is actually enough because we allow a $\delta^{-\varepsilon}$ loss in our estimate.

Corollary 3.8. *Let $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$, and $0 < \delta \ll \sigma$. Suppose that* (82) *holds whenever $\xi_i \in \operatorname{supp} \hat{f}_i + O(\delta)$, $i = 1, 2, \dots, k$. Then, if $p \geq 2k/(k-1)$ and $\varepsilon_\circ$ is small enough, for $\varepsilon > 0$, there is an $N = N(\varepsilon)$ such that the following estimate holds with C, C_ε, independent of ψ and η:*

$$\Big\| \prod_{i=1}^{k} S_\delta(\psi, \eta) f_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \leq C\sigma^{-C_\varepsilon}\delta^{-\varepsilon} \prod_{i=1}^{k} (\delta \|f_i\|_2).$$

To show this we need only to replace G_i with $\phi\big(\eta(D,t)(D_d - \psi(D',t))/\delta\big) f_i$ and apply Proposition 3.6. The assumptions in Proposition 3.6 are satisfied with $G_1, \dots, G_k$. Thus, the estimate is straightforward because $\|\phi\big(\eta(D,t)(D_d - \psi(D',t))/\delta\big) f_i\|_{L^2_{x,t}} \lesssim \delta^{\frac{1}{2}} \|f\|_2$, which follows by Plancherel's theorem and taking t-integration first.

The following result is a consequence of Corollary 3.8 and localization argument in the proof of Proposition 2.10.

Proposition 3.9. *Let $0 < \delta \ll \sigma \ll \tilde{\sigma} \ll 1$ and $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$ and let $Q_1, \dots, Q_k \subset \frac{1}{2} I^d$ be dyadic cubes of side length $\tilde{\sigma}$. Suppose that* (82) *is satisfied whenever $\xi_i \in Q_i$, $i = 1, \dots, k$, and suppose that $\operatorname{supp} \hat{f}_i \subset Q_i$, $i = 1, \dots, k$. Then, if $p \geq 2k/(k-1)$ and $\varepsilon_\circ$ is small enough, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that*

$$\Big\| \prod_{i=1}^{k} S_\delta(\psi, \eta) f_i \Big\|_{\frac{p}{k}} \leq C\sigma^{-C_\varepsilon}\delta^{-\varepsilon} \prod_{i=1}^{k} (\delta^{\frac{d}{p} - \frac{d-2}{2}} \|f_i\|_p) \tag{98}$$

holds with C, C_ε, independent of ψ and η.

Proof. The proof is similar to that of Proposition 2.10. So, we shall be brief. Let φ, $\widetilde{Q}_i$, $\tilde{\chi}_i$, $\{\mathcal{B}\}$, and $\{\widetilde{\mathcal{B}}\}$ be the same as in the proof of Proposition 2.10. We set

$$K_i^t = \mathcal{F}^{-1}\bigg(\phi\bigg(\frac{\eta(\xi,t)(\tau - \psi(\zeta,t))}{\delta}\bigg)\tilde{\chi}_i(\xi)\bigg).$$

Then $S_\delta(\psi, \eta) f_i = \|K_i^t * f_i\|_{L^2_t(I)}$. The (p/k)-th power of the left-hand side of (98) is bounded by

$$\sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} \|K_i^t * f_i\|_{L^2_t(I)}^{\frac{p}{k}} \, dx \lesssim I + II,$$

where

$$I = \sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} \|K_i^t * (\chi_{\widetilde{\mathcal{B}}} f_i)\|_{L^2_t(I)}^{\frac{p}{k}} \, dx, \quad II = \sum_{\mathcal{B}} \bigg(\sum_{\substack{g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i \\ \text{for some } i}} \int_{\mathcal{B}} \prod_{i=1}^{k} \|K_i^t * g_i\|_{L^2_t(I)}^{\frac{p}{k}} \, dx \bigg).$$

As before, the second sum is taken over all choices with $g_i = \chi_{\widetilde{\mathcal{B}}} f_i$ or $\chi_{\widetilde{\mathcal{B}}^c} f_i$, and $g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i$ for some i. By choosing $c > 0$ small enough, we see that $\tilde{\chi}_1(D)(\chi_{\widetilde{\mathcal{B}}} f_1), \dots, \tilde{\chi}_k(D)(\chi_{\widetilde{\mathcal{B}}} f_k)$ satisfy the assumption of Corollary 3.8. Since $K_i^t * (\chi_{\widetilde{\mathcal{B}}} f_i)) = \phi\big(\eta(D,t)(D_d - \psi(D',t))/\delta\big)\tilde{\chi}_i(D)(\chi_{\widetilde{\mathcal{B}}} f_i)$, by Corollary 3.8

and Hölder's inequality

$$I \lesssim \sigma^{-C_\varepsilon}\Big(\frac{1}{\delta}\Big)^{\varepsilon}\sum_{\mathcal{B}}\prod_{i=1}^{k}\delta^{\frac{p}{k}}\|\chi_{\widetilde{\mathcal{B}}}f_i\|_2^{\frac{p}{k}} \lesssim \sigma^{-C_\varepsilon}\Big(\frac{1}{\delta}\Big)^{c\varepsilon}\Big(\prod_{i=1}^{k}\delta^{\frac{d}{p}-\frac{d-2}{2}}\|f_i\|_p\Big)^{\frac{p}{k}}.$$

To handle II we note from Lemma 2.9 that $|K_i^t(x)| \le C\delta\mathfrak{K}_M(x)$ with C depending only on $\|\psi\|_{C^N(I^{d-1})}$, $\|\eta\|_{C^N(I^d)}$. Thus,

$$\|K_i^t * (\chi_{\widetilde{\mathcal{B}}^c} f_i)(x)\|_{L_t^2} \le C\delta\delta^{\varepsilon(M-d-1)}\mathfrak{K}_{d+1} * |f_i|(x)$$

if $x \in B$, and $\|K_i * f_i(x)\|_{L_t^2(I)} \le C\delta\mathfrak{K}_{d+1} * |f_i|(x)$. The rest of proof is the same as before. We omit the details. □

3D. *Multilinear square-function estimate with confined direction sets.* From the point of view of Proposition 2.11 we may expect a better estimate thanks to the smallness of supports of the Fourier transforms of the input functions when they are confined in a small neighborhood of a k-dimensional submanifold. The following is a vector-valued generalization of Proposition 2.11.

Proposition 3.10. *Let k, $2 \le k \le d$, be an integer, $0 < \sigma \ll 1$ be fixed, and $\Pi \subset \mathbb{R}^d$ be a k-plane containing the origin. Let $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ and Γ^t be defined by* (79). *For $0 < \delta \ll \sigma$, suppose that the functions $G_1, \dots, G_k$ defined on $\mathbb{R}^d \times I$ satisfy* (81) *for $t \in I$ and* (82) *whenever $\xi_i \in \operatorname{supp}\mathcal{F}(G_i(\cdot,t)) + O(\delta)$, $i = 1, 2, \dots, k$, for some $t \in I$. Additionally we assume that, for all $t \in I$,*

$$\boldsymbol{n}(\operatorname{supp}\widehat{G}_1(\cdot,t)), \dots, \boldsymbol{n}(\operatorname{supp}\widehat{G}_k(\cdot,t)) \subset \mathbb{S}^{d-1} \cap (\Pi + O(\delta)). \tag{99}$$

Then, if $2 \le p \le 2k/(k-1)$ and ε_0 is sufficiently small, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L_t^2(I)}\Big\|_{L^{p/k}(B(x,\delta^{-1}))} \lesssim \sigma^{-C_\varepsilon}\delta^{dk(\frac{1}{2}-\frac{1}{p})-\varepsilon}\prod_{i=1}^{k}\|G_i\|_{L_{x,t}^2} \tag{100}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$.

The following is an easy consequence of (100).

Corollary 3.11. *Let $\{\mathfrak{q}\}$, $\mathfrak{q} \subset \frac{1}{2}I^d$, be the collection of dyadic cubes of side length l, $\delta < l \le 2\delta$. Define $G_{i,\mathfrak{q}}$ by $\mathcal{F}(G_{i,\mathfrak{q}}(\cdot,t)) = \chi_{\mathfrak{q}}\mathcal{F}(G_i(\cdot,t))$ and set $R = \frac{1}{\delta}$. Suppose that the same assumptions as in Proposition 3.10 are satisfied. Then, if $2 \le p \le 2k/(k-1)$ and $\varepsilon_\circ$ is small enough, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that*

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L_t^2(I)}\Big\|_{L^{p/k}(B(x,R))} \lesssim \sigma^{-C_\varepsilon}\delta^{-\varepsilon}\prod_{i=1}^{k}\Big\|\Big(\sum_{\mathfrak{q}}\|G_{i,\mathfrak{q}}\|_{L_t^2(I)}^2\Big)^{\frac{1}{2}}\rho_{B(x,R)}\Big\|_p \tag{101}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$.

Proof. Observe that

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L_t^2(I)}\Big\|_{L^{p/k}(B(x,R))} \le \Big\|\prod_{i=1}^{k}\Big\|\rho\Big(\frac{\cdot-x}{R}\Big)G_i\Big\|_{L_t^2(I)}\Big\|_{L^{p/k}}.$$

Then, the functions $\rho((\cdot-x)/R)G_i$, $i=1,\dots,k$, satisfy the assumption in Proposition 3.10 because $\operatorname{supp}\mathcal{F}\big(\rho((\cdot-x)/R)G_i(\cdot,t)\big)=\operatorname{supp}\widehat{G}(\cdot,t)+O(R^{-1})$. So, from Proposition 3.10 we get

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L^2_t(I)}\Big\|_{L^{p/k}(B(x,R))}\lesssim\sigma^{-C_\varepsilon}R^{\varepsilon}\prod_{i=1}^{k}R^{-d(\frac12-\frac1p)}\Big\|\Big\|\rho\Big(\frac{\cdot-x}{R}\Big)G_i\Big\|_{L^2}\Big\|_{L^2_t(I)}.$$

Since $G_i=\sum_{\mathfrak{q}}G_{i,\mathfrak{q}}$ and the supports of $\big\{\mathcal{F}\big(\rho((\cdot-x)/R)G_{i,\mathfrak{q}}(\cdot,t)\big)\big\}_{\mathfrak{q}}$ are boundedly overlapping, by Plancherel's theorem it follows that

$$\Big\|\Big\|\rho\Big(\frac{\cdot-x}{R}\Big)G_i\Big\|_{L^2_x}\Big\|_{L^2_t(I)}\lesssim\Big\|\Big(\sum_{\mathfrak{q}}\Big\|\rho\Big(\frac{\cdot-x}{R}\Big)G_{i,\mathfrak{q}}\Big\|_2^2\Big)^{\frac12}\Big\|_{L^2_t(I)}.$$

Combining this with the above inequality, we get

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L^2_t(I)}\Big\|_{L^{p/k}(B(x,R))}\lesssim\sigma^{-C_\varepsilon}R^{\varepsilon}\prod_{i=1}^{k}R^{-d(\frac12-\frac1p)}\Big\|\Big|\rho\Big(\frac{\cdot-x}{R}\Big)\Big|\Big(\sum_{\mathfrak{q}}\|G_{i,\mathfrak{q}}\|^2_{L^2_t(I)}\Big)^{\frac12}\Big\|_2.$$

Now Hölder's inequality gives the desired estimate (101). ☐

As an application of Corollary 3.11 we obtain the following.

Corollary 3.12. *Let $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$, $\eta\in\mathcal{E}(N)$, $0<\delta\ll\tilde\sigma\ll\sigma$, and $S_\delta=S_\delta(\psi,\eta)$ be defined by* (74). *Let Π be a k-plane which contains the origin. Suppose* (82) *holds whenever $\xi_i\in\operatorname{supp}\hat f_i+O(\tilde\sigma)$, $i=1,2,\dots,k$, and*

$$\boldsymbol{n}(\operatorname{supp}\hat f_i)\subset\Pi+O(\tilde\sigma),\quad i=1,2,\dots,k.\tag{102}$$

Let $\{\mathfrak{q}\}$, $\mathfrak{q}\in\frac12 I^d$, be the collection of dyadic cubes of side length l, $\tilde\sigma<l\le2\tilde\sigma$. Define $f_{i,\mathfrak{q}}$ by $\mathcal{F}(f_{i,\mathfrak{q}})=\chi_{\mathfrak{q}}\mathcal{F}(f_i)$. Then, if $2k/(k-1)\le p\le2$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon>0$ there is an $N=N(\varepsilon)$ such that

$$\Big\|\prod_{i=1}^{k}S_\delta f_i\Big\|_{L^{p/k}(B(x,1/\tilde\sigma))}\lesssim\sigma^{-C_\varepsilon}\tilde\sigma^{-\varepsilon}\prod_{i=1}^{k}\Big\|\Big(\sum_{\mathfrak{q}}|S_\delta f_{i,\mathfrak{q}}|^2\Big)^{\frac12}\rho_{B(x,1/\tilde\sigma)}\Big\|_{L^p}$$

holds uniformly for ψ and η.

This follows from Corollary 3.11. Indeed, it suffices to check that

$$G_i=\rho(\tilde\sigma(\cdot-x))\phi\big((D_d-\psi(D',t))/\sigma\big)f_i$$

satisfies the assumption of Corollary 3.11 with $\delta=\tilde\sigma$ as long as $\sigma\ll\tilde\sigma$. This is clear because

$$\widehat{G}_i(\cdot,t)=\tilde\sigma^{-d}\Big(e^{i\langle\cdot,x\rangle}\rho\Big(\frac{\cdot}{\tilde\sigma}\Big)\Big)*\Big(\phi\Big(\frac{\tau-\psi(\zeta,t)}{\sigma}\Big)\hat f_i\Big).$$

Proof of Proposition 3.10. The argument here is similar to the proof of Proposition 3.6. The estimate for $p=2$ follows from Hölder's inequality and Plancherel's theorem. So, by interpolation it is sufficient to show (100) for $p=2k/(k-1)$.

Let us set $R = \frac{1}{\delta} \gg 1$ and we may set $x = 0$. As usual we start with the assumption that, for $0 < \delta \ll \sigma$,

$$\Big\| \prod_{i=1}^{k} \|G_i\|_{L_t^2(I)} \Big\|_{L^{p/k}(B(0,R))} \le C R^{\alpha} R^{-\frac{d}{2}} \prod_{i=1}^{k} \|G_i\|_{L^2_{x,t}} \tag{103}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ whenever $G_1, \dots, G_k$ satisfy (81), (82) and (99). By (84) and Hölder's inequality, (103) is true for some large α. As before it is sufficient to show that (103) implies for any $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\| \prod_{i=1}^{k} \|G_i\|_{L_t^2(I)} \Big\|_{L^{p/k}(B(0,R))} \le C \sigma^{-\kappa} R^{\frac{\alpha}{2} + c\varepsilon} R^{-\frac{d}{2}} \prod_{i=1}^{k} \|G_i\|_{L^2_{x,t}}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$. Then iteration of this implication gives the desired estimate (100).

Fix $z \in \mathbb{R}^d$ and consider $\rho_{B(z,\sqrt{R})} G_1(\cdot, t), \dots, \rho_{B(z,\sqrt{R})} G_k(\cdot, t)$. Then it is clear from (81) and (99) that $\operatorname{supp} \mathcal{F}(\rho_{B(z,\sqrt{R})} G_i(\cdot, t)) \subset \Gamma^t + O(R^{-\frac{1}{2}})$ and

$$\boldsymbol{n}(\operatorname{supp} \mathcal{F}(\rho_{B(z,\sqrt{R})} G_i(\cdot, t))) \subset \Pi + O(R^{-\frac{1}{2}}).$$

Also, since $\delta \ll \sigma$, (82) holds if $\xi_i \in \operatorname{supp} \mathcal{F}(\rho_{B(z,\sqrt{R})} G_i(\cdot, t))$. Hence, by the assumption (103) we get

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}} \lesssim R^{\frac{\alpha}{2}} R^{-\frac{d}{4}} \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_{x,t}}. \tag{104}$$

Now we proceed in the same way as in the proof of Proposition 3.6, and we keep using the same notations. As before, let $\{\boldsymbol{q}\}$ be the collection of dyadic cubes (hence essentially disjoint) of side length $\sim R^{-\frac{1}{2}}$ such that $I^d = \bigcup \boldsymbol{q}$. We decompose the function $G_i(\cdot, t)$ into $G_{i,\boldsymbol{q}}(\cdot, t)$, which is defined by (89), and get (91), which is clear. Then, combining (91) and (104), we have

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}} \le C R^{\frac{\alpha}{2}} R^{-\frac{d}{4}} \prod_{i=1}^{k} \Big\| \Big(\sum_{\boldsymbol{q}} |\rho_{B(z,\sqrt{R})} G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}}.$$

Then this gives

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}} \lesssim R^{\frac{\alpha}{2} - \frac{d}{4}} \prod_{i=1}^{k} \Big\| \chi_{B(z,R^{1/2+\varepsilon})} \Big(\sum_{\boldsymbol{q}} |G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}} + \mathcal{E}, \tag{105}$$

where $\mathcal{E} = R^{-M} \prod_{i=1}^{k} \|G_i\|_{L^2_{x,t}}$ for any large M.

We also denote by $(\mathrm{N}^t)^{-1}$ (defined from $\mathrm{N}^t(I^{d-1})$ to I^{d-1}) the inverse of $\mathrm{N}^t : \Gamma^t \to \mathbb{S}^{d-1}$ which is well defined because $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$. Since $\partial_t \psi \in (1 - \varepsilon_\circ, 1 + \varepsilon_\circ)$, there is an interval $I_{i,\boldsymbol{q}}$ of length $CR^{-\frac{1}{2}}$ such that $G_{i,\boldsymbol{q}}(\cdot, t) = 0$ if $t \notin I_{i,\boldsymbol{q}}$; see (90). As in the proof of Proposition 3.6 we partition $I_{i,\boldsymbol{q}}$ into intervals $I^l_{i,\boldsymbol{q}} = [t_l, t_{l+1}]$, $l = 1, \dots, l_0$, of side length $\sim R^{-1}$. Since the Fourier transform of $G_i(\cdot, t)$ is supported in $\Gamma^{t_l} + O(\delta)$ if $t \in I^l_{i,\boldsymbol{q}} = [t_l, t_{l+1}]$ and the normal vectors are confined in $\Pi + O(\delta)$, it follows that

$$\operatorname{supp} \mathcal{F}(G_{i,\boldsymbol{q}}(\cdot, t)) \subset \Gamma^{t_l}(\delta) \cap ((\mathrm{N}^{t_l})^{-1}(\Pi) + O(\delta)), \quad t \in [t_l, t_{l+1}].$$

Fix t_l, and let us set

$$\xi_{i,\boldsymbol{q}}^{t_l} = (\zeta_{i,\boldsymbol{q}}^{t_l}, \tau_{i,\boldsymbol{q}}^{t_l}) \in ((\mathrm{N}^{t_l})^{-1}(\Pi) \cap \Gamma^{t_l}) \cap (\operatorname{supp} \mathcal{F}(G_{i,\boldsymbol{q}}(\cdot, t_l)) + O(\delta)).$$

(As before, we may assume that this set is nonempty, otherwise the associated function $G_{i,\boldsymbol{q}}^l$ is equal to 0. See below.) Let $v_1, \dots, v_{k-1}$ be an orthonormal basis for the tangent space $T_{\xi_{i,\boldsymbol{q}}^{t_l}}((\mathrm{N}^{t_l})^{-1}(\Pi))$ at $\xi_{i,\boldsymbol{q}}^{t_l}$, and $u_1, \dots, u_{d-k}$ be a set of orthonormal vectors such that $\{\mathrm{N}^{t_l}(\xi_{i,\boldsymbol{q}}^{t_l}), v_1, \dots, v_{k-1}, u_1, \dots, u_{d-k}\}$ forms an orthonormal basis for $\mathbb{R}^d$. Let us set

$$\begin{aligned}\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l} = \big\{\xi : |(\xi - \xi_{i,\boldsymbol{q}}^{t_l}) \cdot \mathrm{N}^{t_l}(\xi_{i,\boldsymbol{q}}^{t_l})| \le C\delta,\ |(\xi - \xi_{i,\boldsymbol{q}}^{t_l}) \cdot v_i| \le C\sqrt{\delta},\ i = 1, \dots, k-1,& \\ \text{and } |(\xi - \xi_{i,\boldsymbol{q}}^{t_l}) \cdot u_i| \le C\delta,\ i = 1, \dots, d-k\big\},& \\ \boldsymbol{P}_{i,\boldsymbol{q}}^{t_l} = \big\{\xi : |\xi \cdot \mathrm{N}^{t_l}(\xi_{i,\boldsymbol{q}}^{t_l})| \le C,\ |\xi \cdot v_i| \le C\sqrt{\delta},\ i = 1, \dots, k-1, \text{ and } |\xi \cdot u_i| \le C,\ i = 1, \dots, d-k\big\}&\end{aligned}$$

with a sufficiently large $C > 0$. Then $\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot, t)),\ t \in [t_l, t_{l+1}]$ is supported in $\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l}$.

The rest of proof is similar to that of Proposition 3.6, so we shall be brief. Let $\mathfrak{m}_{i,\boldsymbol{q}}^{t_l}$ be a smooth function naturally adapted to $\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l}$ such that $\mathfrak{m}_{i,\boldsymbol{q}}^{t_l} \sim 1$ on $\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l}$ and $\mathcal{F}^{-1}(\mathfrak{m}_{i,\boldsymbol{q}}^{t_l})$ is supported in $R\boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}$. This can be done by using ρ and composing it with an appropriate affine map; for example, see (94). As before we define $G_{i,\boldsymbol{q}}^l(\cdot, t)$ by (95) and let $K_{i,\boldsymbol{q}}^{t_l} = \mathcal{F}^{-1}(\mathfrak{m}_{i,\boldsymbol{q}}^{t_l})$ so that $G_{i,\boldsymbol{q}}^l(\cdot, t) = G_{i,\boldsymbol{q}}^l(\cdot, t) * K_{i,\boldsymbol{q}}^{t_l}$ if $t \in I_{i,\boldsymbol{q}}^l$. Hence,

$$\sum_{\boldsymbol{q}} G_{i,\boldsymbol{q}} = \sum_{\boldsymbol{q},l} G_{i,\boldsymbol{q}}^l(\cdot, t) * K_{i,\boldsymbol{q}}^{t_l}, \quad |K_{i,\boldsymbol{q}}^{t_l}| \lesssim |R\boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}|^{-1} \chi_{R\boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}}.$$

Let us set $\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l} = R^{1+\varepsilon}\boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}$. Hence, from the same lines of inequalities as in (96) and repeating an argument similar to that in the proof of Proposition 3.6 we have, for $x \in B(y, R^{\frac{1}{2}+\varepsilon})$,

$$\prod_{i=1}^k \Big(\sum_{\boldsymbol{q}} \|G_{i,\boldsymbol{q}}\|_{L_t^2(I)}^2(x)\Big) \lesssim R^{c\varepsilon} \prod_{i=1}^k \sum_{\boldsymbol{q},i} \|G_{i,\boldsymbol{q}}^l(\cdot, t)\|_{L^2(I_{\boldsymbol{q}}^l)}^2 * \bigg(\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}|}\bigg)(y).$$

Now, we use the lines of argument from (23) to (26), and combine this with (105) to get

$$\begin{aligned}&\Big\|\prod_{i=1}^k \|G_i\|_{L_t^2(I)}\Big\|_{L^{2/(k-1)}(B(0,R))} \\ &\qquad \lesssim R^{c\varepsilon+\frac{\alpha}{2}} \bigg\|\prod_{i=1}^k \bigg(\sum_{\boldsymbol{q},l} \|G_{i,\boldsymbol{q}}^l(\cdot, t)\|_{L^2(I_{\boldsymbol{q}}^l)}^2 * \bigg(\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}|}\bigg)\bigg)^{\frac{1}{2}}\bigg\|_{L^{2/(k-1)}(B(0,2R))} + \mathcal{E}.\end{aligned}$$

Since $\sum_{\boldsymbol{q},l} \big\| \|\widetilde{G}_{i,\boldsymbol{q}}\|_{L_t^2(I_{\boldsymbol{q}}^l)}\big\|_2^2 \sim \sum_{\boldsymbol{q}} \big\|\|G_{i,\boldsymbol{q}}\|_{L_t^2(I_{\boldsymbol{q}})}\big\|_2^2 \sim \|G_i\|_{L_{x,t}^2}$, the proof is completed if we show

$$\bigg\|\prod_{i=1}^k \bigg(\sum_{\boldsymbol{q},l} f_{\boldsymbol{q},l} * \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}|}\bigg)\bigg\|_{L^{2/(k-1)}(B(0,2R))} \le CR^{c\varepsilon}\sigma^{-1}R^{-d} \prod_{i=1}^k \Big(\sum_{\boldsymbol{q},l} \|f_{\boldsymbol{q},l}\|_1\Big).$$

Finally, to show the above inequality we may repeat the argument in the last part in the proof of Proposition 2.11. In fact, we need only to show the associated Kakeya estimate; for example, see (28) and (97). Using the coordinates $(u, v) \times \Pi \times \Pi^\perp = \mathbb{R}^d$ as before, it is sufficient to show that the longer sides of $\boldsymbol{P}^{t_l}_{i,\boldsymbol{q}}$ are transverse to Π. More precisely, if $\varepsilon_\circ$ is sufficiently small and N is large enough, there exists a constant $c > 0$, independent of $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, such that, for

$$w \in \big(T_{\xi^{t_l}_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \operatorname{span}\{\mathrm{N}(\xi^{t_l}_{i,\boldsymbol{q}})\}\big)^{\perp}, \tag{106}$$

(29) holds. Since $\psi(\zeta,t) = \frac{1}{2}|\zeta|^2 + t + \mathcal{R}$ with $\|\mathcal{R}\|_{C^N(I^d\times I)} \le \varepsilon_\circ$, by the same perturbation argument it is sufficient to consider $\psi(\zeta,t) = \frac{1}{2}|\zeta|^2 + t$. For this case (29) clearly holds for w satisfying (106) because translation by t doesn't have any effect. The same argument works without modification. □

3E. ***Multiscale decomposition for $S_\delta f$.*** In this section we obtain a multiscale decomposition for the square function, which is to be combined with multilinear square-function estimates to prove Proposition 3.2. This will be carried out in a way similar to how we obtained the decomposition in Section 2, though we need to take care of the additional t-average.

Let $0 < \varepsilon_\circ \ll 1$, $1 \ll N$, $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$, and S_δ be given by (74). Let N^t, $\boldsymbol{n}$ be given by Definition 3.5. Let $\kappa = \kappa(\varepsilon_\circ, N)$ be the number given in Proposition 3.4 so that (77) holds whenever $0 < \varepsilon \le \kappa$, $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, and $\eta \in \mathcal{E}(N)$. As before, let $\sigma_1, \dots, \sigma_m$, and $M_1, \dots, M_m$ be dyadic numbers such that

$$\delta \ll \sigma_{d-1} \ll \cdots \ll \sigma_1 \ll \min(\kappa, 1), \quad M_i = \frac{1}{\sigma_i}. \tag{107}$$

We assume that f is Fourier supported in $\frac{1}{2}I^d$. We keep using the same notation as in Section 2E. In particular, $\{\mathfrak{q}^i\}$, $\{\mathfrak{Q}^i\}$ are the collection of (closed) dyadic intervals of side length $2\sigma_i$, $2M_i$, respectively, so that (37) and (40) holds.

3E1. *Decomposition by normal vector sets.* Let $\{\theta^i\}$ be a discrete subset of $\mathbb{S}^{d-1}$ whose elements are separated by distance $\sim \sigma_i$. Let $\mathfrak{d}^i$ be disjoint subsets of $\{\mathfrak{q}^i\}$ which satisfies, for some θ^i,

$$\mathfrak{d}^i \subset \{\mathfrak{q}^i : \operatorname{dist}(\boldsymbol{n}(\mathfrak{q}^i), \theta^i) \le C\sigma_i\} \tag{108}$$

and

$$\bigcup_{\mathfrak{d}^i} \mathfrak{d}^i = \{\mathfrak{q}^i\}, \quad i = 1, \dots, m. \tag{109}$$

Obviously, such a partitioning of $\{\mathfrak{q}^i\}$ is possible. Disjointness between $\mathfrak{d}^i$ will be useful later for decomposing the square function. Then we also define an auxiliary operator by

$$\mathfrak{S}_{\mathfrak{d}^i} f = \Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} |S_\delta f_{\mathfrak{q}^i}|^2\Big)^{\frac{1}{2}}.$$

As before, $\mathfrak{d}^i$, $\mathfrak{d}^i_*$, $\mathfrak{d}^i_j$, and $\mathfrak{d}^i_{j*}$ denote the elements in $\{\mathfrak{d}^i\}$ for the rest of this section.

Definition 3.13. We define $\boldsymbol{n}(\mathfrak{d}^i)$ to be a vector† $\theta \in \{\theta^i\}$ such that $\operatorname{dist}(\boldsymbol{n}(\mathfrak{q}^i), \theta) \le C\sigma_i$ whenever $\mathfrak{q}^i \in \mathfrak{d}^i$. Particularly, we may set $\boldsymbol{n}(\mathfrak{d}^i) = \theta^i$ if (108) holds.

†Possibly, there is more than one θ. In that case we simply choose one of them. Ambiguity of the definition does not cause any problem in what follows.

Since the map N^t is injective for each t, the elements of $\mathfrak{d}^i$ are contained in an $O(\sigma_i)$ neighborhood of the curve $\{\xi : \boldsymbol{n}(\xi) = \theta^i\}$ with $\theta^i = \boldsymbol{n}(\mathfrak{d}^i)$. From (72) we observe that for any interval J of length σ_i there are as many as $O(1)$ $\mathfrak{q}^i \in \mathfrak{d}^i$ such that $\phi\big((D_d - \psi(D', t))/\delta\big) f_{\mathfrak{q}^i} \neq 0$ if $t \in J$. Hence, dividing I intervals of length $\sim \sigma_i$ and taking integration in t we see that

$$S_\delta\Bigg(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} f_{\mathfrak{q}^i}\Bigg) \lesssim \Bigg(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} |S_\delta f_{\mathfrak{q}^i}|^2\Bigg)^{\frac{1}{2}} = \mathfrak{S}_{\mathfrak{d}^i} f \tag{110}$$

with the implicit constant independent of $\mathfrak{d}^i$. Since $S_\delta f \le \sum_{\mathfrak{d}^i} S_\delta\big(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} f_{\mathfrak{q}^i}\big)$, $i = 1, \dots, m$, we also have

$$S_\delta f \lesssim \sum_{\mathfrak{d}^i}\Bigg(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} |S_\delta f_{\mathfrak{q}^i}|^2\Bigg)^{\frac{1}{2}} = \sum_{\mathfrak{d}^i} \mathfrak{S}_{\mathfrak{d}^i} f. \tag{111}$$

3E2. *σ_1-scale decomposition.* Decomposition at this stage is similar to that of T_δ in Section 2. So, we shall be brief. Fix $x \in \mathbb{R}^d$ and let $\mathfrak{d}^1_* \in \{\mathfrak{d}^1\}$ such that

$$\mathfrak{S}_{\mathfrak{d}^1_*} f(x) = \max_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x).$$

Considering the cases $\sum_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x) \le 100^d \mathfrak{S}_{\mathfrak{d}^1_*} f(x)$ and $\sum_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x) > 100^d \mathfrak{S}_{\mathfrak{d}^1_*} f(x)$ separately, we have

$$\begin{aligned} S_\delta f(x) &\lesssim \sum_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x) \lesssim \mathfrak{S}_{\mathfrak{d}^1_*} f(x) + \sigma_1^{1-d} \max_{\substack{\mathfrak{d}^1 \\ |\boldsymbol{n}(\mathfrak{d}^1_*) - \boldsymbol{n}(\mathfrak{d}^1)| \gtrsim \sigma_1}} (\mathfrak{S}_{\mathfrak{d}^1_*} f(x) \mathfrak{S}_{\mathfrak{d}^1} f(x))^{\frac{1}{2}} \\ &\lesssim \mathfrak{S}_{\mathfrak{d}^1_*} f(x) + \sigma_1^{1-d} \max_{\substack{\mathfrak{d}^1_1, \mathfrak{d}^1_2 \\ |\boldsymbol{n}(\mathfrak{d}^1_1) - \boldsymbol{n}(\mathfrak{d}^1_2)| \gtrsim \sigma_1}} (\mathfrak{S}_{\mathfrak{d}^1_1} f(x) \mathfrak{S}_{\mathfrak{d}^1_2} f(x))^{\frac{1}{2}}. \end{aligned}$$

Since $\#\mathfrak{d}^i \lesssim \sigma_1^{-1}$ and $\mathfrak{S}_{\mathfrak{d}^1_1} f \mathfrak{S}_{\mathfrak{d}^1_2} f = \big(\sum_{\mathfrak{q}^1_1 \in \mathfrak{d}^1_1, \mathfrak{q}^1_2 \in \mathfrak{d}^1_2} (S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2})^2\big)^{\frac{1}{2}}$,

$$S_\delta f(x) \lesssim \sigma_1^{\frac{1}{p} - \frac{1}{2}} \Bigg(\sum_{\mathfrak{q}^1 \in \mathfrak{d}^1_*} |S_\delta f_{\mathfrak{q}^1}|^p\Bigg)^{\frac{1}{p}} + \sigma_1^{-C}\Bigg(\sum_{\substack{\mathfrak{d}^1_1, \mathfrak{d}^1_2 \\ |\boldsymbol{n}(\mathfrak{d}^1_1) - \boldsymbol{n}(\mathfrak{d}^1_2)| \gtrsim \sigma_1}} (S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2})^{\frac{p}{2}}\Bigg)^{\frac{1}{p}}.$$

Taking the L^p norm on both side of the inequality yields

$$\|S_\delta f\|_p \lesssim \sigma_1^{\frac{1}{p} - \frac{1}{2}} \Bigg(\sum_{\mathfrak{q}^1} \|S_\delta f_{\mathfrak{q}^1}\|_p^p\Bigg)^{\frac{1}{p}} + \sigma_1^{-C}\Bigg(\sum_{\mathfrak{q}^1_1, \mathfrak{q}^1_2 \text{ trans}} \|S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}}^{\frac{p}{2}}\Bigg)^{\frac{1}{p}}.$$

Hence, using Proposition 3.4 and Lemma 2.6, we have

$$\|S_\delta f\|_p \lesssim \sigma_1^{\frac{2}{p}} B_p(\sigma_1^{-2}\delta)\|f\|_p + \sigma_1^{-C} \max_{\mathfrak{q}^1_1, \mathfrak{q}^1_2 \text{ trans}} \|S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}}^{\frac{1}{2}}. \tag{112}$$

We proceed to decompose those terms appearing in the bilinear expression.

3E3. *σ_k-scale decomposition, $k \ge 2$.* Fixing σ, for $l \in \sigma^{-1}\mathbb{Z}^d$, let A_l and τ_l be given by (32). The following is a slight modification of Lemma 2.13.

Lemma 3.14. *Let $\mathfrak{d}$ be a subset of $\{\mathfrak{q}^i\}$. Set $\mathfrak{S}_\mathfrak{d} f = \big(\sum_{\mathfrak{q}^i \in \mathfrak{d}} |S_\delta f_{\mathfrak{q}^i}|^2\big)^{\frac12}$, and set*

$$[\mathfrak{S}_\mathfrak{d} f] = \sum_{l \in M_i \mathbb{Z}^d} A_l^{\frac12} \mathfrak{S}_\mathfrak{d}(\tau_l f), \quad |[\mathfrak{S}_\mathfrak{d} f]| = \sum_{l,l' \in M_i \mathbb{Z}^d} (A_l A_{l'})^{\frac12} \mathfrak{S}_\mathfrak{d}(\tau_{(l+l')} f).$$

If $x, x_0 \in \mathfrak{Q}^i$, the following inequality holds with the implicit constants independent of $\mathfrak{d}$:

$$\mathfrak{S}_\mathfrak{d} f(x) \lesssim [\mathfrak{S}_\mathfrak{d} f](x_0) \lesssim |[\mathfrak{S}_\mathfrak{d} f]|(x). \tag{113}$$

Proof. Note that $\mathfrak{q}^i$ is a cube of side length $2\sigma_i$. Since $x, x_0 \in \mathfrak{Q}^i$, using (35) and the Cauchy–Schwarz inequality, we get

$$\left|\phi\Big(\frac{D_d - \psi(D',t)}{\delta}\Big) f_{\mathfrak{q}^i}(x)\right|^2 \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_l \left|\phi\Big(\frac{\eta(D,t)(D_d - \psi(D',t))}{\delta}\Big) \tau_l f_{\mathfrak{q}^i}(x_0)\right|^2.$$

Integrating in t we get

$$(S_\delta f_{\mathfrak{q}^i}(x))^2 \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_l (S_\delta(\tau_l f_{\mathfrak{q}^i})(x_0))^2. \tag{114}$$

Summation in $\mathfrak{q}^i \in \mathfrak{d}$ gives

$$\Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta f_{\mathfrak{q}^i}(x))^2\Big)^{\frac12} \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_l^{\frac12} \Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta(\tau_l f_{\mathfrak{q}^i})(x_0))^2\Big)^{\frac12},$$

from which we get the first inequality of (113). By interchanging the roles of x and x_0 in (114) and summation in $\mathfrak{q}^i \in \mathfrak{d}$, it follows that

$$\sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta(\tau_l f_{\mathfrak{q}^i})(x_0))^2 \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_{l'} \sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta(\tau_{(l+l')} f_{\mathfrak{q}^i})(x))^2.$$

Putting this in the right-hand side of the above inequality and repeating the same argument, we get the second inequality of (113). □

Now we have the bilinear decomposition (112) on which we build a higher degree of multilinear decomposition.

3E4. *From k-transversal to $(k+1)$-transversal, $2 \le k \le m$.* Let us be given cubes $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ of side length σ_{k-1} which satisfy (55). Though we use the same notation as in the multiplier-estimate case, it should be noted that the normal vector field $\boldsymbol{n}$ is defined on $I^{d-1} \times CI$ (see Definition 3.5). As before, we denote by $\{\mathfrak{q}_i^k\}$ the collection of dyadic cubes of side length σ_k contained in $\mathfrak{q}_i^{k-1}$, see (56), which are partitioned into the subsets of $\{\mathfrak{d}_i^k\}$ so that

$$\bigcup_{\mathfrak{d}_i^k} \Big(\bigcup_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} \mathfrak{q}_i^k\Big) = \mathfrak{q}_i^{k-1}, \quad i = 1, \dots, k.$$

So, we can write

$$\prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}} f_{\mathfrak{q}_i^{k-1}}\Big) = \prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{d}_i^k} \sum_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)$$

and recall the definition

$$\mathfrak{S}_{\mathfrak{d}_i^k} F_{\mathfrak{q}_i^{k-1}} := \Big(\sum_{\mathfrak{q}_i^k \in \mathfrak{d}_j^k} |S_\delta F_{\mathfrak{q}_i^k}|^2 \Big)^{\frac{1}{2}}.$$

Fix $\mathfrak{Q}^k$ and let x_0 be the center of $\mathfrak{Q}^k$. Let $\mathfrak{d}_{i*}^k \in \{\mathfrak{d}_i^k\}$ be an angular partition such that

$$\mathfrak{S}_{\mathfrak{d}_{i*}^k} f_{\mathfrak{q}_i^{k-1}}(x_0) = \max_{\mathfrak{d}_i^k} \mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}(x_0).$$

Let us set

$$\bar{\Lambda}_i^k = \big\{ \mathfrak{d}_i^k : [\mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}](x_0) > (\sigma_k)^{kd} \max_{1 \le j \le k} [\mathfrak{S}_{\mathfrak{d}_{j*}^k} f_{\mathfrak{q}_j^{k-1}}](x_0) \big\}, \quad 1 \le i \le k. \tag{115}$$

We split the sum to get

$$\prod_{i=1}^k S_\delta \Big(\sum_{\mathfrak{d}_i^k} \sum_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} f_{\mathfrak{q}_i^k} \Big) \le \prod_{i=1}^k S_\delta \Big(\sum_{\mathfrak{d}_i^k \in \bar{\Lambda}_i^k} \sum_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} f_{\mathfrak{q}_i^k} \Big) + \sum_{(\mathfrak{d}_1^k, \dots, \mathfrak{d}_k^k) \notin \prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k S_\delta \Big(\sum_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} f_{\mathfrak{q}_i^k} \Big). \tag{116}$$

Thus, if $x \in \mathfrak{Q}^k$, by (113) and (110) the second term in the right-hand side is bounded by

$$\begin{aligned}
\sum_{(\mathfrak{d}_1^k, \dots, \mathfrak{d}_k^k) \notin \prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k S_\delta \Big(\sum_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} f_{\mathfrak{q}_i^k} \Big)(x) &\lesssim \sum_{(\mathfrak{d}_1^k, \dots, \mathfrak{d}_k^k) \notin \prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k [\mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}](x_0) \\
&\lesssim \big(\max_{1 \le j \le k} [\mathfrak{S}_{\mathfrak{d}_{j*}^k} f_{\mathfrak{q}_j^{k-1}}](x_0) \big)^k \\
&\lesssim \big(\max_{1 \le j \le k} [\mathfrak{S}_{\mathfrak{d}_{j*}^k} f](x_0) \big)^k \lesssim \big(\max_{\mathfrak{d}^k} |[\mathfrak{S}_{\mathfrak{d}^k} f]|(x) \big)^k.
\end{aligned} \tag{117}$$

Here $\{\mathfrak{d}^k\} = \bigcup_{1 \le i \le k} \{\mathfrak{d}_i^k\}$ and the third inequality follows from the definition of $\mathfrak{S}_{\mathfrak{d}_j^k} f$ because $\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}$. Since (117) holds for each $\mathfrak{Q}^k$, integrating over all $\mathfrak{Q}^k$, using Lemma 3.14, Proposition 3.4 and Lemma 2.6, we get

$$\begin{aligned}
\Big\| \sum_{(\mathfrak{d}_1^k, \dots, \mathfrak{d}_k^k) \notin \prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k S_\delta \Big(\sum_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} f_{\mathfrak{q}_i^k} \Big) \Big\|_{\frac{p}{k}}^{\frac{1}{k}} &\lesssim \big\| \max_{\mathfrak{d}^k} |[\mathfrak{S}_{\mathfrak{d}^k} f]| \big\|_p \lesssim \sup_h \big\| \max_{\mathfrak{d}^k} \mathfrak{S}_{\mathfrak{d}^k}(\tau_h f) \big\|_p \\
&\lesssim \sup_h \Big(\sum_{\mathfrak{d}^k} \| \mathfrak{S}_{\mathfrak{d}^k}(\tau_h f) \|_p^p \Big)^{\frac{1}{p}} \\
&\lesssim \sup_h \sigma_k^{(\frac{1}{p} - \frac{1}{2})} \Big(\sum_{\mathfrak{q}_i^k} \| S_\delta \tau_h f_{\mathfrak{q}_i^k} \|_p^p \Big)^{\frac{1}{p}} \\
&\lesssim \sigma_k^{\frac{2}{p}} B_p(\sigma_k^{-2} \delta) \| f \|_p.
\end{aligned} \tag{118}$$

The second-to-last inequality follows from the definition of $\mathfrak{S}_{\mathfrak{d}^k} f$ and Hölder's inequality since there are as many as $O(\sigma_k^{-1})$ $\mathfrak{q}^k \subset \mathfrak{d}^k$.

We note that vectors $\boldsymbol{n}(\mathfrak{d}_{1*}^k), \dots, \boldsymbol{n}(\mathfrak{d}_{k*}^k)$ are linearly independent because $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal. We also denote by $\Pi_*^k = \Pi_*^k(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)$ the k-plane spanned by the vectors $\boldsymbol{n}(\mathfrak{d}_{1*}^k), \dots, \boldsymbol{n}(\mathfrak{d}_{k*}^k)$. Let us set

$$\bar{\mathfrak{N}} = \bar{\mathfrak{N}}(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k) = \{ \mathfrak{d}^k : \mathrm{dist}(\boldsymbol{n}(\mathfrak{d}^k), \Pi_*^k) \le C\sigma_k \}.$$

We split the sum and use the triangle inequality so that

$$\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) \le \prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k\\ \mathfrak{d}_i^k\in\overline{\mathfrak{N}}}}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) + \sum_{\substack{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k\\ \mathfrak{d}_i^k\notin\overline{\mathfrak{N}}\text{ for some } i}} \prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big). \tag{119}$$

For the k-tuples $(\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k)$ appearing in the second summation of the right-hand side, there is a $\mathfrak{d}_i^k$ for which $\boldsymbol{n}(\mathfrak{d}_i^k)$ is not contained in $\Pi_*^k+O(\sigma_k)$. In particular, suppose that $\boldsymbol{n}(\mathfrak{d}_1^k)\notin\Pi_*^k+O(\sigma_k)$. Then, by (113) and (115) we have

$$\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)(x) \lesssim \prod_{i=1}^{k}[\mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}](x_0) \le \sigma_k^{-C}\big([\mathfrak{S}_{\mathfrak{d}_1^k} f_{\mathfrak{q}_1^{k-1}}](x_0)\big)^{\frac{k}{k+1}} \prod_{i=1}^{k}\big([\mathfrak{S}_{\mathfrak{d}_{i*}^k} f_{\mathfrak{q}_i^{k-1}}](x_0)\big)^{\frac{k}{k+1}}.$$

Recall that $\mathrm{Vol}(\boldsymbol{n}(\xi_1),\boldsymbol{n}(\xi_2),\dots,\boldsymbol{n}(\xi_k))\gtrsim\sigma_1\cdots\sigma_{k-1}$ if $\xi_i\in\mathfrak{q}_i^{k-1}$, $i=1,\dots,k$. From the definition of $\overline{\mathfrak{N}}$ it follows that $\mathrm{dist}(\boldsymbol{n}(\mathfrak{q}^k),\Pi_*^k)\gtrsim\sigma_k$ if $\mathfrak{q}^k\in\mathfrak{d}^k$ and $\boldsymbol{n}(\mathfrak{d}^k)\notin\overline{\mathfrak{N}}$. Hence

$$\mathrm{Vol}\big(\boldsymbol{n}(\xi_1),\boldsymbol{n}(\xi_2),\dots,\boldsymbol{n}(\xi_k),\boldsymbol{n}(\xi_{k+1})\big)\gtrsim\sigma_1\cdots\sigma_k$$

if $\xi_i\in\mathfrak{q}_i^k$ and $\mathfrak{q}_i^k\in\mathfrak{d}_{i*}^k$, $i=1,\dots,k$, and $\xi_{k+1}\in\mathfrak{q}_{k+1}^k$ and $\mathfrak{q}_{k+1}^k\in\mathfrak{d}_1^k$. So these cubes are transversal. Since there are only $O(\sigma_k^{-C})$ σ_k-scale cubes, by (113) and Hölder's inequality

$$\begin{aligned}\prod_{i=1}^{k} S_\delta\Big(&\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)(x)\\ &\lesssim \sigma_k^{-C}\big(|[\mathfrak{S}_{\mathfrak{d}_1^k} f_{\mathfrak{q}_1^{k-1}}]|(x)\big)^{\frac{k}{k+1}}\prod_{i=1}^{k}\big(|[\mathfrak{S}_{\mathfrak{d}_{i*}^k} f_{\mathfrak{q}_i^{k-1}}]|(x)\big)^{\frac{k}{k+1}}\\ &\lesssim \sigma_k^{-C}\sum_{l_1,l_1',\dots,l_{k+1},l_{k+1}'\in M_k\mathbb{Z}^d}\prod_{i=1}^{k+1}\widetilde{A}_{l_i}\widetilde{A}_{l_i'}\Big(\sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\text{ trans}}\Big(\prod_{i=1}^{k+1}S_\delta(\tau_{(l_i+l_i')}f_{\mathfrak{q}_i^k})(x)\Big)^{\frac{p}{k+1}}\Big)^{\frac{k}{p}}.\end{aligned}$$

Here $\widetilde{A}_{l_i},\widetilde{A}_{l_i'}$ are rapidly decaying sequences. The same is true for any $\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k$ satisfying $\mathfrak{d}_i^k\in\bar{\Lambda}_i^k$, $1\le i\le k$, and $\mathfrak{d}_i^k\notin\overline{\mathfrak{N}}$ for some i and this holds regardless of $\mathfrak{Q}^k$. So, we have, for any x,

$$\begin{aligned}\sum_{\substack{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k\\ \mathfrak{d}_i^k\notin\overline{\mathfrak{N}}\text{ for some } i}}&\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)(x)\\ &\lesssim\sigma_k^{-C}\sum_{l_1,l_1',\dots,l_{k+1},l_{k+1}'}\prod_{i=1}^{k+1}\widetilde{A}_{l_i}\widetilde{A}_{l_i'}\Big(\sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\text{ trans}}\Big(\prod_{i=1}^{k+1}S_\delta(\tau_{(l_i+l_i')}f_{\mathfrak{q}_i^k})(x)\Big)^{\frac{p}{k+1}}\Big)^{\frac{k}{p}}.\end{aligned}\tag{120}$$

Since $\widetilde{A}_{l_i},\widetilde{A}_{l_i'}$ are rapidly decaying, taking the $L^{\frac{p}{k}}$ norm and a simple manipulation give

$$\Big\|\sum_{\substack{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k\\ \mathfrak{d}_i^k\notin\overline{\mathfrak{N}}\text{ for some } i}}\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)\Big\|_{\frac{p}{k}} \lesssim\sigma_k^{-C}\sup_{\tau_1,\dots,\tau_{k+1}}\max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\text{ trans}}\Big\|\prod_{i=1}^{k+1}S_\delta(\tau_i f_{\mathfrak{q}_i^k})\Big\|_{\frac{p}{k+1}}^{\frac{k}{k+1}}.\tag{121}$$

We now combine the inequalities (116), (117), (119), (120) to get

$$
\begin{aligned}
&\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{d}_i^k}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)\\
&\qquad\lesssim \big(\max_{\mathfrak{d}^k} |[\mathfrak{S}_{\mathfrak{d}^k} f]|(x)\big)^k + \chi_{\mathfrak{Q}^k}\prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k\\ \mathfrak{d}_i^k\in\bar{\mathfrak{N}}}}\ \sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)\\
&\qquad\qquad + \sigma_k^{-C}\sum_{l_1,l_1',\dots,l_{k+1},l_{k+1}'}\ \prod_{i=1}^{k+1}\widetilde{A}_{l_i}\widetilde{A}_{l_i'}\Big(\sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\ \text{trans}}\Big(\prod_{i=1}^{k+1} S_\delta(\tau_{(l_i+l_i')} f_{\mathfrak{q}_i^k})(x)\Big)^{\frac{p}{k+1}}\Big)^{\frac{k}{p}}.
\end{aligned}
$$

Here $\bar{\mathfrak{N}}$ depends on $\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k$. By taking the $\frac{1}{k}$-th power, integrating on $\mathbb{R}^d$ and using (118) and (121) we get

$$
\begin{aligned}
\Big\|\Big(\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\subset\mathfrak{q}_i^{k-1}} f_{\mathfrak{q}_i^k}\Big)\Big)^{\frac{1}{k}}\Big\|_p &\lesssim \sigma_k^{\frac{2}{p}} B_p(\sigma_k^{-2}\delta)\|f\|_p + \sigma_k^{-C}\sup_{\tau_1,\dots,\tau_{k+1}}\ \max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\ \text{trans}}\Big\|\prod_{i=1}^{k+1} S_\delta(\tau_i f_{\mathfrak{q}_i^k})\Big\|_{\frac{p}{k+1}}^{\frac{k}{k+1}}\\
&\quad+\Big(\sum_{\mathfrak{Q}^k}\Big\|\prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\\ [\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)}}\ \sum_{\substack{\mathfrak{q}_i^k\in\mathfrak{d}_i^k\\ \mathfrak{q}_i^k\subset\mathfrak{q}_i^{k-1}}} f_{\mathfrak{q}_i^k}\Big)\Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}}\Big)^{\frac{1}{p}}, \qquad (122)
\end{aligned}
$$

where $[\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)$ denotes a subset of $\bar{\mathfrak{N}}(\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)$ which depends on $\mathfrak{q}_1^{k-1},\dots,$ $\mathfrak{q}_k^{k-1},\mathfrak{Q}^k$.

3E5. *Multiscale decomposition.* For $k=2,\dots,m$, let us set

$$
\bar{\mathfrak{M}}^k f = \sup_{\tau_1,\dots,\tau_k}\ \max_{\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1}\ \text{trans}}\Big(\sum_{\mathfrak{Q}^k}\Big\|\prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\\ [\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)}}\ \sum_{\substack{\mathfrak{q}_i^k\in\mathfrak{d}_i^k\\ \mathfrak{q}_i^k\subset\mathfrak{q}_i^{k-1}}} \tau_i f_{\mathfrak{q}_i^k}\Big)\Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}}\Big)^{\frac{1}{p}}.
$$

Here $[\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)$ also depends on $\tau_1,\dots,\tau_k$ but this doesn't affect the overall bound. Starting from (112) we successively apply (122) to k-scale transversal products (given by $\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1}$ transversal). After decomposition up to the m-th scale we get

$$
\begin{aligned}
\|S_\delta f\|_p &\lesssim \sum_{k=1}^{m}\sigma_{k-1}^{-C}\sigma_k^{\frac{2}{p}} B_p(\sigma_k^{-2}\delta)\|f\|_p + \sum_{k=2}^{m}\sigma_{k-1}^{-C}\bar{\mathfrak{M}}^k f\\
&\quad+\sigma_m^{-C}\sup_{\tau_1,\dots,\tau_{m+1}}\ \max_{\mathfrak{q}_1^m,\dots,\mathfrak{q}_{m+1}^m\ \text{trans}}\Big\|\prod_{i=1}^{m+1} S_\delta\tau_i f_{\mathfrak{q}_i^m}\Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}}. \qquad (123)
\end{aligned}
$$

3F. ***Proof of Proposition 3.2.*** We may assume $d\ge 9$ since $p_s\ge 2(d+2)/d$ for $d<9$ and the sharp bound for $p\ge 2(d+2)/d$ is verified in [Lee et al. 2012]. So, we have $p_s(d)\ge 2(d-1)/(d-2)$. The proof is similar to that of Proposition 2.3. Let $\beta>0$ and we aim to show that $\mathcal{B}^\beta(s)\le C$ for $0<s\le 1$ if $p\ge p_s(d)$. We choose $\varepsilon>0$ such that $(100d)^{-1}\beta\ge\varepsilon$. Fix $\varepsilon_\circ>0$ and $N=N(\varepsilon)$ such that Corollaries 3.8, 3.11 and 3.12 hold uniformly for $\psi\in\bar{\mathfrak{G}}(\varepsilon_\circ,N)$.

Let $s < \delta \le 1$. Obviously, $(\sigma_k^{-2}\delta)^{\frac{d-2}{2}-\frac{d}{p}+\beta} B(\sigma_k^{-2}\delta) \le \mathcal{B}^\beta(s) + \sigma_k^{-C}$ because $s \le \sigma_k^{-2}\delta$ and $B(\delta) = B_p(\delta) \le C$ for $\delta \gtrsim 1$. Hence, it follows that

$$\sigma_k^{\frac{2}{p}} B(\sigma_k^{-2}\delta) \lesssim \sigma_k^{2(\frac{d-2}{2}-\frac{d-1}{p})+2\beta} \delta^{-\frac{d-2}{2}+\frac{d}{p}-\beta} (\mathcal{B}^\beta(s) + \sigma_k^{-C}). \tag{124}$$

We first consider the $(m+1)$-product in (123). By Corollary 3.8 we have, for $p \ge 2(m+1)/m$,

$$\sup_{\tau_1,\dots,\tau_{m+1}} \max_{\mathfrak{q}_1^m,\dots,\mathfrak{q}_{m+1}^m \text{ trans}} \Big\| \prod_{i=1}^{m+1} S_\delta \tau_i f_{\mathfrak{q}_i^m} \Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}} \le C_\varepsilon \sigma_m^{-C} \delta^{-\frac{d-2}{2}+\frac{d}{p}-\varepsilon} \|f\|_p. \tag{125}$$

For $\overline{\mathfrak{M}}^k$, as before we have two types of estimates. The first one follows from Corollary 3.8, while the second one is a consequence of the square-function estimates in Corollary 3.12. From the definition of $\overline{\mathfrak{M}}^k$, we note that $\mathfrak{q}_1^k, \mathfrak{q}_2^k, \dots, \mathfrak{q}_k^k$ are contained, respectively, in $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$, which are transversal. Hence, we have

$$\prod_{i=1}^k S_\delta \Bigg(\sum_{\mathfrak{d}_i^k \in [\overline{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)} \sum_{\substack{\mathfrak{q}_i^k \in \mathfrak{d}_i^k \\ \mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}}} \tau_i f_{\mathfrak{q}_i^k} \Bigg)(x) \le \sum_{\mathfrak{q}_1^k,\mathfrak{q}_2^k,\dots,\mathfrak{q}_k^k \text{ trans}} \prod_{i=1}^k S_\delta(\tau_i f_{\mathfrak{q}_i^k})(x).$$

Here "$\mathfrak{q}_1^k, \mathfrak{q}_2^k, \dots, \mathfrak{q}_k^k$ trans" means $\mathrm{Vol}(\boldsymbol{n}(\xi_1), \dots, \boldsymbol{n}(\xi_k)) \ge \sigma_1 \cdots \sigma_{k-1}$ provided $\xi_i \in \mathfrak{q}_i^k$, $i = 1, \dots, k$. Since there are as many as $O(\sigma_{k-1}^{-C})$ k-tuples $(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$ and the above holds regardless of $\mathfrak{Q}^k$, by Corollary 3.12 we have, for $p \ge 2k/(k-1)$,

$$\overline{\mathfrak{M}}^k f \lesssim \sigma_k^{-C} \sup_{\tau_1,\dots,\tau_k} \sum_{\mathfrak{q}_1^k,\mathfrak{q}_2^k,\dots,\mathfrak{q}_k^k \text{ trans}} \Big\| \prod_{i=1}^k S_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}} \lesssim \sigma_k^{-C} \delta^{-\frac{d-2}{2}+\frac{d}{p}-\varepsilon} \|f\|_p. \tag{126}$$

Estimates for $\overline{\mathfrak{M}}^k$ via Corollary 3.11. By fixing $\tau_1, \dots, \tau_k$, and $(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$ satisfying $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal, we first handle the integral over $\mathfrak{Q}^k$ which is in the definition of $\overline{\mathfrak{M}}^k$. For $i = 1, \dots, k$, set

$$f_i = \sum_{\mathfrak{d}_i^k \in [\overline{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)} \Bigg(\sum_{\substack{\mathfrak{q}_i^k \in \mathfrak{d}_i^k \\ \mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}}} \tau_i f_{\mathfrak{q}_i^k} \Bigg).$$

Since $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal, (82) holds with $\sigma = \sigma_1 \cdots \sigma_{k-1}$ whenever $\xi_i \in \operatorname{supp} \hat{f}_i + O(\sigma_k)$, $i = 1, 2, \dots, k$. Also note that $\boldsymbol{n}(\mathfrak{d}_1^k), \dots, \boldsymbol{n}(\mathfrak{d}_k^k) \subset \Pi_*^k(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)$. Hence, it follows that (102) holds with $\tilde{\sigma} = \sigma_k$. Let us set

$$\mathcal{Q}(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k) = \{\mathfrak{q}^k : \boldsymbol{n}(\mathfrak{q}^k) \in [\overline{\mathfrak{N}}](\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)\}.$$

Let write $\mathfrak{Q}^k = \mathfrak{q}(z, 1/\sigma_k)$. Then, by Corollary 3.12 we have, for $2 \le p \le 2k/(k-1)$,

$$\Big\| \Big(\prod_{i=1}^k S_\delta f_i \Big)^{\frac{1}{k}} \Big\|_{L^p(\mathfrak{Q}^k)}^p \lesssim \sigma_{k-1}^{-C_\varepsilon} \sigma_k^{-\varepsilon} \prod_{i=1}^k \Big\| \Big(\sum_{\substack{\mathfrak{q}_i^k \in \mathfrak{q}_{k-1}^i \\ \mathfrak{q}_i^k \in \mathcal{Q}(\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)}} |S_\delta \tau_i f_{\mathfrak{q}^k}|^2 \Big)^{\frac{1}{2}} \rho_{B(z,\frac{C}{\sigma_k})} \Big\|_{L^p}^{\frac{p}{k}}.$$

The dyadic cubes of side length σ_k in $\mathcal{Q}(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)$ are contained in an $O(\sigma_k)$-neighborhood of $\boldsymbol{n}^{-1}(\Pi_*^k)$ which is a smooth k-dimensional surface. Thus,

$$\#\{\mathfrak{q}_i^k \subset \mathfrak{q}_{k-1}^i : \mathfrak{q}_i^k \in \mathcal{Q}(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)\} \le C\sigma_k^{-k}.$$

Now, by Hölder's inequality we get

$$\Big\| \Big(\prod_{i=1}^k S_\delta f_i \Big)^{\frac{1}{k}} \Big\|_{L^p(\mathfrak{Q}^k)}^p \lesssim \sigma_{k-1}^{-C_\varepsilon} \sigma_k^{-\varepsilon - k(\frac{p}{2}-1)} \prod_{i=1}^k \Big\| \Big(\sum_{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}} |S_\delta \tau_i f_{\mathfrak{q}_i^k}|^p \Big)^{\frac{1}{p}} \rho_{\mathfrak{Q}^k} \Big\|_{L^p}^{\frac{p}{k}}.$$

Summation along $\mathfrak{Q}^k$ using the rapid decay of the Schwartz function ρ gives

$$\Big\| \Big(\prod_{i=1}^k S_\delta f_i \Big)^{\frac{1}{k}} \Big\|_{L^p} \lesssim \sigma_{k-1}^{-C_\varepsilon} \sigma_k^{-\varepsilon - k(\frac{1}{2}-\frac{1}{p})} \prod_{i=1}^k \Big\| \Big(\sum |S_\delta \tau_i f_{\mathfrak{q}_i^k}|^p \Big)^{\frac{1}{p}} \Big\|_p^{\frac{1}{k}}.$$

Hence, using Proposition 3.4, Lemma 2.6, and (124), for $2 \le p \le 2k/(k-1)$, we have

$$\begin{aligned} \Big\| \Big(\prod_{i=1}^k S_\delta f_i \Big)^{\frac{1}{k}} \Big\|_{L^p} &\lesssim \sigma_{k-1}^{-C} \sigma_k^{-\varepsilon - \frac{k-1}{2} + \frac{k+1}{p}} B(\sigma_k^{-2}\delta) \|f\|_p \\ &\lesssim \sigma_{k-1}^{-C} \delta^{-\frac{d-2}{2} + \frac{d}{p} - \beta} \sigma_k^{\beta + \frac{2d-k-3}{2} - \frac{2d-k-1}{p}} (\sigma_k^{-C} + \mathcal{B}^\beta(s)) \|f\|_p \\ &\lesssim \sigma_{k-1}^{-C} \delta^{-\frac{d-2}{2} + \frac{d}{p} - \beta} (\sigma_k^{-C} + \sigma_k^\alpha \mathcal{B}^\beta(s)) \|f\|_p \end{aligned}$$

with some $\alpha > 0$ if $p \ge 2(2d-k-1)/(2d-k-3)$. Here we have used $(100d)^{-1}\beta \ge \varepsilon$. We note that the right-hand side of the above is independent of $\tau_1, \dots, \tau_k$ and there are only $O(\sigma_{k-1}^{-C})$ many k-tuples $(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$ satisfying $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal. Thus, recalling the definition of $\overline{\mathfrak{M}^k f}$, we have for $2 \le p \le 2k/(k-1)$

$$\overline{\mathfrak{M}^k f} \lesssim \sigma_{k-1}^{-C} \delta^{-\frac{d-2}{2} + \frac{d}{p} - \beta} \big(\sigma_k^{-C} + \sigma_k^\alpha \mathcal{B}^\beta(s) \big) \|f\|_p$$

with some $\alpha > 0$ provided that $p \ge 2(2d-k-1)/(2d-k-3)$. Combining this and (126) we have, for some $\alpha > 0$,

$$\overline{\mathfrak{M}}^k f \le C\delta^{-\frac{d-2}{2} + \frac{d}{p} - \beta} \big(\sigma_k^{-C} + \sigma_k^\alpha \mathcal{B}^\beta(s) \big) \|f\|_p. \tag{127}$$

provided that $p \ge \min\big(2(2d-k-1)/(2d-k-3), 2k/(k-1)\big)$.

Closing induction. Let us set

$$p(m) = \max\Big(\max_{1 \le k \le m} \min\Big(\frac{2(2d-k-1)}{2d-k-3}, \frac{2k}{k-1} \Big), \frac{2(m+1)}{m} \Big).$$

Since $p \ge p_s > 2(d-1)/(d-2)$ and $(100d)^{-1}\beta \ge \varepsilon$, we have

$$\sigma_k^{\frac{2}{p}} B(\sigma_k^{-2}\delta) \lesssim \sigma_k^\alpha \delta^{-\frac{d-2}{2} + \frac{d}{p} - \beta} (\mathcal{B}^\beta(s) + \sigma_k^{-C})$$

for some $\alpha > 0$. Using (123), we combine the estimates (124), (125), and (127) to get

$$\|S_\delta f\|_p \le C \sum_{k=1}^m (\sigma_{k-1}^{-C} + \sigma_{k-1}^{-C} \sigma_k^\alpha \mathcal{B}^\beta(s)) \delta^{-\frac{d-2}{2} + \frac{d}{p} - \beta} \|f\|_p + C\sigma_m^{-C} \delta^{-\frac{d-2}{2} + \frac{d}{p} - \beta} \|f\|_p$$

for some $\alpha > 0$ as long as $p \geq p(m)$. The rest of proof is similar to that in Section 2F, so we intend to be brief. By using the stability of the estimates along $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$, multiplying by $\delta^{\frac{d-2}{2}-\frac{d}{p}+\beta}$ on both sides and taking the supremum along ψ, η and f, and taking the supremum along δ, $s < \delta \leq 1$, we get

$$\mathcal{B}^\beta(s) \leq C\left(\sum_{k=1}^m \sigma_{k-1}^{-C}\sigma_k^\alpha\right)\mathcal{B}^\beta(s) + C\sum_{k=1}^m \sigma_k^{-C}$$

for some $\alpha > 0$ provided that $p \geq p(m)$. Choosing $\sigma_1, \dots, \sigma_{m-1}$ such that $C\left(\sum_{k=1}^{m-1}\sigma_{k-1}^{-C}\sigma_k^\alpha\right) \leq \frac{1}{2}$ gives $\mathcal{B}^\varepsilon(\delta) \leq C\sigma_m^{-C}$ for $p \geq p(m)$. Therefore, to complete the proof we need only to check that the minimum of $p(m)$, $2 \leq m \leq d-1$, is p_s. This can be done by a simple computation.

Acknowledgements

S. Lee was partially supported by NRF (Republic of Korea) grant No. 2015R1A2A2A05000956. He would like thank Andreas Seeger for discussions on related problems.

Note in proof

The range of sharp L^p bound for the Bochner–Riesz means in Theorem 1.1 was recently improved by Guth, Hickman and Iliopoulou [Guth et al. 2017] for $d \geq 4$.

References

[Annoni 2017] M. Annoni, "Almost everywhere convergence for modified Bochner–Riesz means at the critical index for $p \geq 2$", *Pacific J. Math.* **286**:2 (2017), 257–275. MR Zbl

[Bejenaru 2016] I. Bejenaru, "Optimal multilinear restriction estimates for a class of surfaces with curvature", preprint, 2016. arXiv

[Bejenaru 2017] I. Bejenaru, "The optimal trilinear restriction estimate for a class of hypersurfaces with curvature", *Adv. Math.* **307** (2017), 1151–1183. MR Zbl

[Bennett 2014] J. Bennett, "Aspects of multilinear harmonic analysis related to transversality", pp. 1–28 in *Harmonic analysis and partial differential equations*, edited by P. Cifuentes et al., Contemp. Math. **612**, Amer. Math. Soc., Providence, RI, 2014. MR Zbl

[Bennett et al. 2006] J. Bennett, A. Carbery, and T. Tao, "On the multilinear restriction and Kakeya conjectures", *Acta Math.* **196**:2 (2006), 261–302. MR Zbl

[Bennett et al. 2015] J. Bennett, N. Bez, T. C. Flock, and S. Lee, "Stability of the Brascamp–Lieb constant and applications", preprint, 2015. arXiv

[Bourgain 1991a] J. Bourgain, "Besicovitch type maximal operators and applications to Fourier analysis", *Geom. Funct. Anal.* **1**:2 (1991), 147–187. MR Zbl

[Bourgain 1991b] J. Bourgain, "On the restriction and multiplier problems in $\mathbb{R}^3$", pp. 179–191 in *Geometric aspects of functional analysis (1989–90)*, edited by V. D. Milman and J. Lindenstrauss, Lecture Notes in Math. **1469**, Springer, 1991. MR Zbl

[Bourgain 1991c] J. Bourgain, "L^p-estimates for oscillatory integrals in several variables", *Geom. Funct. Anal.* **1**:4 (1991), 321–374. MR Zbl

[Bourgain 2013] J. Bourgain, "On the Schrödinger maximal function in higher dimension", *Tr. Mat. Inst. Steklova* **280** (2013), 53–66. In Russian; translated in *Proc. Steklov Inst. Math.* **280**:1 (2013), 46–60. MR Zbl

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

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

[Carbery 1983] A. Carbery, "The boundedness of the maximal Bochner–Riesz operator on $L^4(\mathbb{R}^2)$", *Duke Math. J.* **50**:2 (1983), 409–416. MR Zbl

[Carbery 1985] A. Carbery, "Radial Fourier multipliers and associated maximal functions", pp. 49–56 in *Recent progress in Fourier analysis* (El Escorial, 1983), edited by I. Peral and J. L. Rubio de Francia, North-Holland Math. Stud. **111**, North-Holland, Amsterdam, 1985. MR Zbl

[Carbery and Valdimarsson 2013] A. Carbery and S. I. Valdimarsson, "The endpoint multilinear Kakeya theorem via the Borsuk–Ulam theorem", *J. Funct. Anal.* **264**:7 (2013), 1643–1663. MR Zbl

[Carbery et al. 1984] A. Carbery, G. Gasper, and W. Trebels, "Radial Fourier multipliers of $L^p(\mathbb{R}^2)$", *Proc. Nat. Acad. Sci. U.S.A.* **81**:10 (1984), 3254–3255. MR

[Carbery et al. 1988] A. Carbery, J. L. Rubio de Francia, and L. Vega, "Almost everywhere summability of Fourier integrals", *J. London Math. Soc.* (2) **38**:3 (1988), 513–524. MR Zbl

[Carleson and Sjölin 1972] L. Carleson and P. Sjölin, "Oscillatory integrals and a multiplier problem for the disc", *Studia Math.* **44** (1972), 287–299. MR Zbl

[Christ 1985] M. Christ, "On almost everywhere convergence of Bochner–Riesz means in higher dimensions", *Proc. Amer. Math. Soc.* **95**:1 (1985), 16–20. MR Zbl

[Christ 1987] M. Christ, "Weak type endpoint bounds for Bochner–Riesz multipliers", *Rev. Mat. Iberoamericana* **3**:1 (1987), 25–31. MR Zbl

[Christ 1988] M. Christ, "Weak type $(1, 1)$ bounds for rough operators", *Ann. of Math.* (2) **128**:1 (1988), 19–42. MR Zbl

[Du et al. 2017] X. Du, L. Guth, and X. Li, "A sharp Schrödinger maximal estimate in $\mathbb{R}^2$", *Ann. of Math.* (2) **186**:2 (2017), 607–640. MR Zbl

[Fefferman 1970] C. Fefferman, "Inequalities for strongly singular convolution operators", *Acta Math.* **124** (1970), 9–36. MR Zbl

[Fefferman 1971] C. Fefferman, "The multiplier problem for the ball", *Ann. of Math.* (2) **94** (1971), 330–336. MR Zbl

[Fefferman 1973] C. Fefferman, "A note on spherical summation multipliers", *Israel J. Math.* **15** (1973), 44–52. MR Zbl

[Rubio de Francia 1985] J. L. Rubio de Francia, "A Littlewood–Paley inequality for arbitrary intervals", *Rev. Mat. Iberoamericana* **1**:2 (1985), 1–14. MR Zbl

[Guth 2010] L. Guth, "The endpoint case of the Bennett–Carbery–Tao multilinear Kakeya conjecture", *Acta Math.* **205**:2 (2010), 263–286. MR Zbl

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

[Guth 2016b] L. Guth, "Restriction estimates using polynomial partitioning, II", preprint, 2016. arXiv

[Guth et al. 2017] L. Guth, J. Hickman, and M. Iliopoulou, "Sharp estimates for oscillatory integral operators via polynomial partitioning", preprint, 2017. arXiv

[Hörmander 1973] L. Hörmander, "Oscillatory integrals and multipliers on FL^p", *Ark. Mat.* **11** (1973), 1–11. MR Zbl

[Lee 2004] S. Lee, "Improved bounds for Bochner–Riesz and maximal Bochner–Riesz operators", *Duke Math. J.* **122**:1 (2004), 205–232. MR Zbl

[Lee 2006a] S. Lee, "Linear and bilinear estimates for oscillatory integral operators related to restriction to hypersurfaces", *J. Funct. Anal.* **241**:1 (2006), 56–98. MR Zbl

[Lee 2006b] S. Lee, "On pointwise convergence of the solutions to Schrödinger equations in $\mathbb{R}^2$", *Int. Math. Res. Not.* **2006** (2006), art. id. 32597. MR Zbl

[Lee and Seeger 2015] S. Lee and A. Seeger, "On radial Fourier multipliers and almost everywhere convergence", *J. Lond. Math. Soc.* (2) **91**:1 (2015), 105–126. MR Zbl

[Lee and Vargas 2008] S. Lee and A. Vargas, "Sharp null form estimates for the wave equation", *Amer. J. Math.* **130**:5 (2008), 1279–1326. MR Zbl

[Lee and Vargas 2010] S. Lee and A. Vargas, "Restriction estimates for some surfaces with vanishing curvatures", *J. Funct. Anal.* **258**:9 (2010), 2884–2909. MR Zbl

[Lee and Vargas 2012] S. Lee and A. Vargas, "On the cone multiplier in $\mathbb{R}^3$", *J. Funct. Anal.* **263**:4 (2012), 925–940. MR Zbl

[Lee et al. 2012] S. Lee, K. M. Rogers, and A. Seeger, "Improved bounds for Stein's square functions", *Proc. Lond. Math. Soc.* (3) **104**:6 (2012), 1198–1234. MR Zbl

[Lee et al. 2013] S. Lee, K. M. Rogers, and A. Seeger, "On space-time estimates for the Schrödinger operator", *J. Math. Pures Appl.* (9) **99**:1 (2013), 62–85. MR Zbl

[Lee et al. 2014] S. Lee, K. M. Rogers, and A. Seeger, "Square functions and maximal operators associated with radial Fourier multipliers", pp. 273–302 in *Advances in analysis: the legacy of Elias M. Stein*, edited by C. Fefferman et al., Princeton Math. Ser. **50**, Princeton Univ. Press, 2014. MR Zbl

[Ou and Wang 2017] Y. Ou and H. Wang, "A cone restriction estimate using polynomial partitioning", preprint, 2017. arXiv

[Ramos 2016] J. Ramos, "The trilinear restriction estimate with sharp dependence on the transversality", preprint, 2016. arXiv

[Seeger 1986] A. Seeger, "On quasiradial Fourier multipliers and their maximal functions", *J. Reine Angew. Math.* **370** (1986), 61–73. MR Zbl

[Seeger 1996] A. Seeger, "Endpoint inequalities for Bochner–Riesz multipliers in the plane", *Pacific J. Math.* **174**:2 (1996), 543–553. MR Zbl

[Shayya 2017] B. Shayya, "Weighted restriction estimates using polynomial partitioning", *Proc. Lond. Math. Soc.* (3) **115**:3 (2017), 545–598. MR Zbl

[Stein 1958] E. M. Stein, "Localization and summability of multiple Fourier series", *Acta Math.* **100** (1958), 93–147. MR Zbl

[Stein 1986] E. M. Stein, "Oscillatory integrals in Fourier analysis", pp. 307–355 in *Beijing lectures in harmonic analysis* (Beijing, 1984), edited by E. M. Stein, Ann. of Math. Stud. **112**, Princeton Univ. Press, 1986. MR Zbl

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, 1993. MR Zbl

[Sunouchi 1967] G. Sunouchi, "On the Littlewood–Paley function g^* of multiple Fourier integrals and Hankel multiplier transformations", *Tôhoku Math. J.* (2) **19** (1967), 496–511. MR Zbl

[Tao 1996] T. Tao, "Weak-type endpoint bounds for Riesz means", *Proc. Amer. Math. Soc.* **124**:9 (1996), 2797–2805. MR Zbl

[Tao 1998] T. Tao, "The weak-type endpoint Bochner–Riesz conjecture and related topics", *Indiana Univ. Math. J.* **47**:3 (1998), 1097–1124. MR Zbl

[Tao 1999] T. Tao, "The Bochner–Riesz conjecture implies the restriction conjecture", *Duke Math. J.* **96**:2 (1999), 363–375. MR Zbl

[Tao 2002] T. Tao, "On the maximal Bochner–Riesz conjecture in the plane for $p < 2$", *Trans. Amer. Math. Soc.* **354**:5 (2002), 1947–1959. MR Zbl

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

[Tao and Vargas 2000a] T. Tao and A. Vargas, "A bilinear approach to cone multipliers, I: Restriction estimates", *Geom. Funct. Anal.* **10**:1 (2000), 185–215. MR Zbl

[Tao and Vargas 2000b] T. Tao and A. Vargas, "A bilinear approach to cone multipliers, II: Applications", *Geom. Funct. Anal.* **10**:1 (2000), 216–258. MR Zbl

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

[Temur 2014] F. Temur, "A Fourier restriction estimate for surfaces of positive curvature in $\mathbb{R}^6$", *Rev. Mat. Iberoam.* **30**:3 (2014), 1015–1036. MR Zbl

[Tomas 1975] P. A. Tomas, "A restriction theorem for the Fourier transform", *Bull. Amer. Math. Soc.* **81** (1975), 477–478. MR Zbl

[Wisewell 2005] L. Wisewell, "Kakeya sets of curves", *Geom. Funct. Anal.* **15**:6 (2005), 1319–1362. MR Zbl

[Wolff 2001] T. Wolff, "A sharp bilinear cone restriction estimate", *Ann. of Math.* (2) **153**:3 (2001), 661–698. MR Zbl

[Zhang 2015] R. Zhang, "The endpoint perturbed Brascamp–Lieb inequality with examples", preprint, 2015. arXiv

Received 4 Oct 2017. Accepted 12 Jan 2018.

SANGHYUK LEE: shklee@snu.ac.kr
School of Mathematical Sciences, Seoul National University, Seoul, South Korea

mathematical sciences publishers

Analysis & PDE
msp.org/apde

See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2018 is US $275/year for the electronic version, and $480/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS & PDE

Volume 11 No. 6 2018

2157-5045(2018)11:6;1-O