Rev. Mat. Iberoam. (online first)
DOI 10.4171/RMI/1057

A polynomial Carleson operator along the paraboloid

Lillian B. Pierce and Po-Lam Yung

Abstract. In this work we extend consideration of the well-known polynomial Carleson operator to the setting of a Radon transform acting along the paraboloid in $\mathbb{R}^{n+1}$ for $n \geq 2$. Inspired by work of Stein and Wainger on the original polynomial Carleson operator, we develop a method to treat polynomial Carleson operators along the paraboloid via van der Corput estimates. A key new step in the approach of this paper is to approximate a related maximal oscillatory integral operator along the paraboloid by a smoother operator, which we accomplish via a Littlewood–Paley decomposition and the use of a square function. The most technical aspect then arises in the derivation of bounds for oscillatory integrals involving integration over lower-dimensional sets. The final theorem applies to polynomial Carleson operators with phase belonging to a certain restricted class of polynomials with no linear terms and whose homogeneous quadratic part is not a constant multiple of the defining function $|y|^2$ of the paraboloid in $\mathbb{R}^{n+1}$.

Dedicated to the memory of Elias M. Stein.

1. Introduction

A celebrated theorem of Carleson [1] proves an L^2 bound for the operator

$$f \to \sup_{\lambda \in \mathbb{R}} |T_\lambda f(x)|, \tag{1.1}$$

where for each $\lambda \in \mathbb{R}$,

$$T_\lambda f(x) = \text{p.v.} \int_{\mathbb{T}} f(x-y)\, e^{i\lambda y}\, \frac{dy}{y};$$

Mathematics Subject Classification (2010): 42B20, 43A50, 42B25, 44A12.
Keywords: Carleson operator, Radon transform, oscillatory integrals, square function, van der Corput estimates.

here by convention $\mathbb{T} = [-\pi, \pi]$. Carleson's work answered in the affirmative a long-standing question of Luzin on whether the Fourier series of an L^2 function must converge pointwise almost everywhere. Carleson's remarkable result inspired many generalizations: soon after, Hunt [4] proved L^p bounds for $1 < p < \infty$ for the Carleson operator (1.1), and Sjölin [10] introduced the Carleson operator to a higher dimensional setting by defining

$$T_\lambda f(x) = \int_{\mathbb{T}^n} f(x-y)\, e^{i\lambda\cdot y}\, K(y)\, dy, \tag{1.2}$$

for $\mathbb{T}^n = [-\pi, \pi]^n$ and an appropriate class of Calderón–Zygmund kernels K. Sjölin's work proved that

$$f \to \sup_{\lambda\in\mathbb{R}^n} |T_\lambda f(x)|$$

is a bounded operator on $L^p(\mathbb{T}^n)$ for all $1 < p < \infty$. Further landmark approaches to the Carleson operator were then developed by Fefferman [3] and Lacey and Thiele [5], and all together these works motivated the development of time-frequency analysis into a rich and active field of research.

This paper is inspired by a question of E. M. Stein, who asked if L^p bounds continue to hold when the linear phase $\lambda \cdot y$ in (1.2) is replaced by a real-valued polynomial on $\mathbb{R}^n$ of the form

$$\sum_{1\le|\alpha|\le d} \lambda_\alpha\, y^\alpha, \tag{1.3}$$

of fixed degree d. If $\lambda = (\lambda_\alpha)_{1\le|\alpha|\le d}$ is the set of real coefficients in (1.3), we denote this polynomial by $P_\lambda(y)$ and define

$$T_\lambda f(x) = \int_{\mathbb{R}^n} f(x-y)\, e^{iP_\lambda(y)}\, K(y)\, dy, \tag{1.4}$$

where K is again an appropriate Calderón–Zygmund kernel. Stein asked if one can show that for every $1 < p < \infty$ there exists a constant $A = A(p, n, d)$ such that

$$\big\| \sup_\lambda |T_\lambda f(x)| \big\|_{L^p(\mathbb{R}^n)} \le A\, \|f\|_{L^p(\mathbb{R}^n)} \tag{1.5}$$

for all $f \in L^p(\mathbb{R}^n)$, where the supremum is now taken over all sets of coefficients $\lambda = (\lambda_\alpha)_{1\le|\alpha|\le d}$ with each λ_α ranging over $\mathbb{R}$. This operator, now known as the polynomial Carleson operator, remains mysterious in many cases.

A convenient way to formulate the supremum in (1.5) is to define a stopping-time function $\lambda(x)$, which is taken to be any measurable function mapping $\mathbb{R}^n$ to the space of real coefficients for $P_\lambda(y)$. Then the desired bound (1.5) would be implied by a bound of the form

$$\|T_{\lambda(x)} f(x)\|_{L^p(\mathbb{R}^n)} \le A\, \|f\|_{L^p(\mathbb{R}^n)},$$

in which the norm A is independent of the choice of the stopping-time function $\lambda(x)$.

Along these lines, it was first shown by Ricci and Stein [9] that for any fixed polynomial $P(x, y)$ the operator

$$Tf(x) = \int_{\mathbb{R}^n} f(x-y)\, e^{iP(x,y)}\, K(y)\, dy$$

is bounded on $L^p(\mathbb{R}^n)$ for $1 < p < \infty$, with norm dependent only on the degree of $P(x, y)$ and not on the coefficients; this is the case of a polynomial stopping-time function, and hence does not imply the full inequality (1.5). In another direction, Stein [12] considered the specific case of $\mathbb{R}^1$ with purely quadratic phase polynomial $P_\lambda(y) = \lambda y^2$ and the corresponding operator

$$T_\lambda f(x) = \text{p.v.} \int_{\mathbb{R}} f(x-y)\, e^{iP_\lambda(y)}\, \frac{dy}{y}, \tag{1.6}$$

ultimately proving that

$$f \mapsto \sup_{\lambda \in \mathbb{R}} |T_\lambda f(x)|$$

is bounded on $L^p(\mathbb{R})$ for $1 < p < \infty$. The methods used in this case hinged upon an asymptotic for the Fourier transform of the kernel $e^{i\lambda y^2}/y$, which is not easily generalizable to higher powers in the phase, or to higher dimensions.

In 2001, Stein and Wainger [14] introduced a new approach to polynomial Carleson operators, based on van der Corput estimates for oscillatory integrals, as well as a Kolmogorov–Seliverstov stopping-time argument. Their result is as follows.

Theorem A. *Consider the operators T_λ as defined in (1.4), where $P_\lambda(y)$ is a real-valued polynomial with no linear terms, of the form*

$$P_\lambda(y) = \sum_{2 \le |\alpha| \le d} \lambda_\alpha\, y^\alpha. \tag{1.7}$$

Then for every $1 < p < \infty$ there exists a constant $A = A(p, n, d)$ such that

$$\big\| \sup_{\lambda} |T_\lambda f(x)| \big\|_{L^p(\mathbb{R}^n)} \le A\, \|f\|_{L^p(\mathbb{R}^n)}$$

for all $f \in L^p(\mathbb{R}^n)$, where the supremum is over all coefficients $\lambda = (\lambda_\alpha)_{2 \le |\alpha| \le d}$ of $P_\lambda(y)$.

The restriction that $P_\lambda(y)$ omits linear terms is inherent to methods of van der Corput type, as will be seen explicitly later.

Until very recently, bounds for the full polynomial Carleson operator remained unproved. Now, the case of the polynomial Carleson operator in dimension one has been resolved by Lie [6] and [7], who proved by time-frequency analysis methods that for

$$T_\lambda f(x) = \text{p.v.} \int_{\mathbb{T}} f(x-y)\, e^{iP_\lambda(y)}\, \frac{dy}{y}$$

with $P_\lambda(y)$ a phase polynomial including both linear and higher order terms, the corresponding Carleson operator is a bounded operator on L^p for $1 < p < \infty$ with

$$\big\| \sup_\lambda |T_\lambda f| \big\|_{L^p(\mathbb{T})} \leq A \, \|f\|_{L^p(\mathbb{T})}.$$

The case of the polynomial Carleson operator involving both linear and higher order terms in dimensions $n \geq 2$ remains open[1].

1.1. A new setting

In the work of this paper, we take the polynomial Carleson operator in a new direction by introducing Radon-type behavior. Suppose $\{P_\lambda(y)\}$ is a family of real-valued polynomials of $y \in \mathbb{R}^n$ indexed by a parameter λ. We assume all polynomials P_λ in the family are of degree at most d for a fixed positive integer d. For each λ in the parameter space, we define an operator T_λ, initially acting on functions f of Schwartz class, by

$$T_\lambda f(x,t) = \int_{\mathbb{R}^n} f(x-y, t-|y|^2)\, e^{iP_\lambda(y)} K(y)\, dy. \tag{1.8}$$

Thus T_λ integrates f along the paraboloid $(y, |y|^2) \subset \mathbb{R}^{n+1}$ against an oscillatory factor and a singular kernel. The kernel K is a Calderón–Zygmund kernel, that is, a tempered distribution agreeing with a C^1 function $K(x)$ for $x \neq 0$, such that it satisfies the differential inequalities

$$|\partial_x^\alpha K(x)| \leq A\, |x|^{-n-|\alpha|} \quad \text{for } 0 \leq |\alpha| \leq 1, \tag{1.9}$$

and such that $\hat{K}$ is an L^∞ function.

We then consider the polynomial Carleson operator of Radon type defined by

$$f \mapsto \sup_\lambda |T_\lambda f(x,t)|, \tag{1.10}$$

in which the supremum is taken over all λ in a suitable parameter space. When the parameter space consists of all possible coefficients $\lambda = (\lambda_\alpha)_{1 \leq |\alpha| \leq d}$ of real-valued polynomials $P_\lambda(y)$ of the form (1.3), it is reasonable to expect that the following a priori estimate should hold for f of Schwartz class:

$$\big\| \sup_\lambda |T_\lambda f| \big\|_{L^p(\mathbb{R}^{n+1})} \leq A \, \|f\|_{L^p(\mathbb{R}^{n+1})},$$

for every $1 < p < \infty$, where $A = A(p,n,d)$. Proof of a result of this type is expected to require inputs from multi-dimensional time-frequency analysis, and seems overly ambitious at this time.

Therefore we restrict ourselves to suprema over a narrower class of polynomial phases, and develop an approach for treating the operator (1.10) with van der Corput estimates. Such methods are inspired by the work of Stein and Wainger

[1] While this paper was in press, dimensions $n \geq 2$ were resolved by [15], [8].

in [14], and inherently require that the phase polynomial lack terms of certain lower orders. In the case of the operator (1.10) on the paraboloid, a first guess might be that van der Corput methods would require $P_\lambda(y)$ to lack both linear and quadratic terms; this may be seen intuitively by considering a model operator

$$Rf(x,t) = \int_{\mathbb{R}^n} f(x-y, t-|y|^2)\, \eta(y)\, e^{iP_\lambda(y)}\, dy, \tag{1.11}$$

where η is a smooth bump function of compact support. If the parameter λ is regarded as fixed for the moment, we may compute the Fourier multiplier of R to be

$$m(\xi,\theta) = \int_{\mathbb{R}^n} \eta(y)\, e^{iP_\lambda(y) - 2\pi i \xi\cdot y - 2\pi i\theta|y|^2}\, dy. \tag{1.12}$$

One may obtain an upper bound for $m(\xi,\theta)$ by applying van der Corput bounds to the oscillatory integral (1.12), as long as one has lower bounds for the coefficients of the phase $P_\lambda(y) - 2\pi\xi\cdot y - 2\pi\theta|y|^2$ as a polynomial in y; one would expect it to be difficult to obtain such lower bounds if certain linear or quadratic terms are present in P_λ. Of course, if $\lambda = \lambda(x,t)$ is a stopping-time function, we cannot compute a Fourier multiplier, so this is merely a heuristic.

Even considering the family $\{P_\lambda(y)\}$ of all phase polynomials that vanish to order at least 2 at the origin is still overly ambitious; the presence of the Radon transform in the operator poses further limitations on the number of degrees of freedom our method can allow in the family of polynomial phases. On the other hand, we will show that with a sufficiently sharp scalpel in hand, one may allow certain quadratic terms, as long as the degree 2 homogeneous portion of the phase polynomial is not precisely a nonzero multiple of $|y|^2$.

1.2. Statement of results

Our main result is the following:

Theorem 1.1. *Fix a dimension $n \geq 2$ and a degree $d \geq 2$. Let*

$$\mathscr{P} = \{p_2(y), p_3(y), \dots, p_d(y)\}$$

be a set of real-valued polynomials on $\mathbb{R}^n$, where each $p_j(y)$ is homogeneous of degree j, and $p_2(y) \neq C|y|^2$ for any nonzero constant C. For each $\lambda = (\lambda_2, \dots, \lambda_d) \in \mathbb{R}^{d-1}$, let

$$P_\lambda(y) = \sum_{m=2}^{d} \lambda_m\, p_m(y), \tag{1.13}$$

and let T_λ be defined as in (1.8). Then for every $1 < p < \infty$, we have an a priori inequality for functions f of Schwartz class:

$$\Big\| \sup_\lambda |T_\lambda f| \,\Big\|_{L^p(\mathbb{R}^{n+1})} \leq A\, \|f\|_{L^p(\mathbb{R}^{n+1})}, \tag{1.14}$$

where the supremum is taken over all $\lambda = (\lambda_2, \dots, \lambda_d) \in \mathbb{R}^{d-1}$, and the norm A may depend on p, n, d and the fixed polynomials $p_2, \dots, p_d$.

As a result of the a priori inequality, it follows from conventional limiting arguments that the Carleson operator extends to a bounded operator on L^p itself, for which the same bound (1.14) holds.

At its foundation, the idea of the proof is that for any value of the coefficient parameter λ one splits the integral defining T_λ in two parts: first, a part in which y is small enough relative to the coefficients of $P_\lambda(y)$ that $e^{iP_\lambda(y)}$ does not contribute significant oscillation; this part is compared to a maximal truncated singular Radon transform, which may be bounded on L^p relatively simply (see Theorem 5.1). The second part of T_λ comprises the region in which y is large enough relative to the coefficients of $P_\lambda(y)$ that $e^{iP_\lambda(y)}$ exhibits significant oscillation; this part is controlled via an auxiliary maximal oscillatory Radon transform, which is interesting in its own right. Precisely, our second main result is as follows.

Theorem 1.2. *Fix a dimension $n \geq 2$ and a degree $d \geq 2$. Let*

$$\mathscr{P} = \{p_2(y), p_3(y), \ldots, p_d(y)\}$$

be a set of real-valued polynomials on $\mathbb{R}^n$, where each $p_j(y)$ is homogeneous of degree j, and $p_2(y) \neq C|y|^2$ for any nonzero constant C. Let

$$\Lambda = \Lambda(\mathscr{P}) = \{2 \leq m \leq d \colon p_m(y) \text{ is not identically zero}\}. \tag{1.15}$$

For $\lambda = (\lambda_2, \ldots, \lambda_d) \in \mathbb{R}^{d-1}$, let $P_\lambda(y)$ be as in (1.13), and let

$$\|\lambda\| = \sum_{m \in \Lambda} |\lambda_m|. \tag{1.16}$$

Then one can define an operator ${}^{(\eta)}I_a^\lambda$ acting on f of Schwartz class by

$${}^{(\eta)}I_a^\lambda f(x,t) = \int_{\mathbb{R}^n} f(x-y, t-|y|^2)\, e^{iP_\lambda(y/a)} \frac{1}{a^n} \eta\Big(\frac{y}{a}\Big)\, dy, \tag{1.17}$$

for η a C^1 bump function supported in the unit ball on $\mathbb{R}^n$, $\lambda \in \mathbb{R}^{d-1}$, and $a > 0$.

Suppose furthermore that $\{\eta_k\}_{k \in \mathbb{Z}}$ is a family of bump functions supported in the unit ball $B_1(\mathbb{R}^n)$ with C^1 norm uniformly bounded by 1. Then there exists a fixed $\delta > 0$ such that for any Schwartz function f on $\mathbb{R}^{n+1}$ and any $r \geq 1$,

$$\Big\| \sup_{k \in \mathbb{Z}} \sup_{\substack{\lambda \in \mathbb{R}^{d-1} \\ r \leq \|\lambda\| < 2r}} |{}^{(\eta_k)}I_{2^k}^\lambda f(x,t)| \Big\|_{L^2(\mathbb{R}^{n+1})} \leq A\, r^{-\delta}\, \|f\|_{L^2(\mathbb{R}^{n+1})}, \tag{1.18}$$

in which the norm A may depend on n, d and the fixed polynomials $p_2, \ldots, p_d$.

In order to put this result in context, it is illustrative to note that for fixed k and λ, it is simple to bound ${}^{(\eta_k)}I_{2^k}^\lambda$ on L^2, since it has a bounded Fourier multiplier. Even after taking suprema over the parameters k and λ, the operator in question is bounded pointwise by the classical maximal Radon transform along the paraboloid, which acts on Schwartz functions f by

$$\mathcal{M}_{\mathrm{Rad}} f(x,t) = \sup_{a>0} \int_{\mathbb{R}^n} |f|(x-u, t-|u|^2) \frac{1}{a^n} \chi_{B_1}\Big(\frac{u}{a}\Big)\, du. \tag{1.19}$$

This is well known to be bounded on L^p for $1 < p \leq \infty$ when acting on functions of Schwartz class (see for instance Chapter 11 of [11]). Thus the key feature of Theorem 1.2 is the decay $r^{-\delta}$ of the norm of the operator as $r \to \infty$. We remark that it will be clear from the proof that Theorem 1.2 continues to hold for all $r \geq c$ for any fixed constant $c > 0$.

As summarized thus far, the basic structure of reducing the treatment of the Carleson operator to the family of oscillatory integral operators ${}^{(\eta_k)}I^\lambda_{2^k}$ is inspired by the work of Stein and Wainger in [14]. Their original approach was to apply a TT^* argument to ${}^{(\eta_k)}I^\lambda_{2^k}$ and then apply van der Corput bounds to the kernel of TT^* in order to show that TT^* could be majorized by certain maximal functions for which L^2 bounds were known. In our setting of Radon transforms, this approach is not sufficient, and in order to prove Theorem 1.2, we must diverge significantly from the original work of Stein and Wainger and develop a new strategy that allows us to control the presence of Radon-type behavior.

To motivate our new approach, we recall that the classical treatment of the maximal Radon transform on the paraboloid introduces a smoother version of the operator, as well as an accompanying square function. In our setting, we introduce a smoothed version of ${}^{(\eta)}I^\lambda_a$ defined by

$$(1.20)\qquad {}^{(\eta)}J^\lambda_a f(x,t) = \iint_{\mathbb{R}^{n+1}} f(x-y, t-z)\, e^{iP_\lambda(y/a)}\, \frac{1}{a^n}\, \eta\Big(\frac{y}{a}\Big)\, \frac{1}{a^2}\, \zeta\Big(\frac{z}{a^2}\Big)\, dy\, dz,$$

where ζ is a smooth function of compact support on $\mathbb{R}$, chosen such that $\int \zeta(s)ds = 1$. We also utilize a square function defined (roughly speaking) by

$$(1.21)\qquad S_r(f)(x,t) = \Big[\sum_{k=-\infty}^{\infty} \Big(\sup_{r \leq \|\lambda\| < 2r} |({}^{(\eta_k)}I^\lambda_{2^k} - {}^{(\eta_k)}J^\lambda_{2^k}) f(x,t)| \Big)^2 \Big]^{1/2}.$$

(The precise definition of $S_r(f)$ is given in equation (6.4).) Our result in Theorem 6.3 is that $S_r(f)$ is bounded on L^2, with a norm that decays as $r \to \infty$; this is the heart of proving Theorem 1.2.

1.3. Outline of the paper

Our approach succeeds by carefully intertwining the role of the square function, TT^* methods, and van der Corput estimates. This work ultimately entails significant technical details, thus Section 2 outlines the motivation for our approach, explaining why complications arise in the Radon case, and demonstrating why the tools we introduce allow us to circumvent these complications. We then assemble in Sections 3 and 4 certain preliminary results on van der Corput estimates and a Littlewood–Paley decomposition. In Section 5 we reduce the proof of Theorem 1.1 to Theorem 1.2. In Section 6, we define the relevant square function and reduce Theorem 1.2 to an L^2 bound for the square function. In Section 7 we prove Theorem 6.4, the main result that implies the boundedness of the square function, except for certain van der Corput estimates. These novel Van der Corput estimates are the most technical part of the paper, and for the sake of the reader we initially treat several examples in the case of dimension $n = 2$ in Section 8, reserving the fully general higher dimensional case for Section 9. Finally in Section 10 we complete the proof of several key propositions in general dimension $n \geq 2$. An appendix in Section 11 provides a proof of an auxiliary theorem on maximal truncated singular Radon transforms (Theorem 5.1). Readers interested in a first view of the key ideas may read Section 2 for a motivation of our approach, and then focus on Sections 5, 6, 7, and 8.

2. Anatomy of the proof

In the work of Stein and Wainger [14], the key oscillatory integral operator takes the form

$$I_a^\lambda f(x) = \int_{\mathbb{R}^n} f(x-y)\, e^{iP_\lambda(y/a)} \frac{1}{a^n}\, \eta\Big(\frac{y}{a}\Big)\, dy,$$

with η a C^1 bump function supported in the unit ball and $P_\lambda(y)$ a polynomial phase of the form (1.7), i.e., with no linear terms. Their analogue of Theorem 1.2 is obtained by using stopping-times to linearize the maximal aspect of the operator $f \mapsto \sup_{\lambda,a} |I_a^\lambda f(x)|$, so that this last operator is represented by

$$T : f \mapsto I_{a(x)}^{\lambda(x)} f(x),$$

where $a(x)$ and $\lambda(x)$ are arbitrary measurable stopping-time functions taking values of the form 2^k for $k \in \mathbb{Z}$ and $r \leq \|\lambda\| < 2r$, respectively. Stein and Wainger prove that if we define a kernel $K_{a_1,a_2}^{\lambda_1,\lambda_2}$ by

$$K_{a_1,a_2}^{\lambda_1,\lambda_2}(y) = \frac{1}{a_1^n}\frac{1}{a_2^n} \int_{\mathbb{R}^n} e^{iP_{\lambda_1}(\frac{y+z}{a_1}) - iP_{\lambda_2}(\frac{z}{a_2})}\, \eta\Big(\frac{y+z}{a_1}\Big)\, \eta\Big(\frac{z}{a_2}\Big)\, dz,$$

then $TT^*f(x)$ is given by

$$TT^*f(x) = \int_{\mathbb{R}^n} f(x-y)\, K_{a(x),a(x-y)}^{\lambda(x),\lambda(x-y)}(y)\, dy.$$

They then proceed to show, via a van der Corput estimate, that $K_{a_1,a_2}^{\lambda_1,\lambda_2}$ satisfies the following pointwise bound:

$$\begin{aligned} |K_{a_1,a_2}^{\lambda_1,\lambda_2}(x)| &\leq C\, r^{-\delta} \left(\frac{1}{a_1^n}\, \chi_{B_2}\Big(\frac{x}{a_1}\Big) + \frac{1}{a_2^n}\, \chi_{B_2}\Big(\frac{x}{a_2}\Big) \right) \\ &\quad + C \left(\frac{1}{a_1^n}\, \chi_{E_{\lambda_1}}\Big(\frac{x}{a_1}\Big) + \frac{1}{a_2^n}\, \chi_{E_{\lambda_2}}\Big(\frac{x}{a_2}\Big) \right), \end{aligned} \tag{2.1}$$

in which $B_2 = B_2(\mathbb{R}^n)$ is the ball of radius 2 and $E_{\lambda_1}, E_{\lambda_2}$ are certain exceptional subsets of $B_2(\mathbb{R}^n)$ with small measure, namely $|E_{\lambda_1}|, |E_{\lambda_2}| \leq r^{-\delta}$. An (L^2, L^2) norm for TT^* with appropriate decay in r can then be obtained by comparison to certain maximal functions, via a clever bilinear argument.

A straightforward generalization of this argument appears to fail in the setting of Theorem 1.2. Indeed, it is known that stopping-time arguments do not work well when Radon transforms are involved, and in order to illustrate this breakdown and to motivate how we proceed instead, we briefly consider the following simple model of a maximal Radon transform (without oscillatory factor) along a paraboloid in $\mathbb{R}^3$:

$$\mathcal{M}f(x,t) = \sup_{a>0} \int_{y\in\mathbb{R}^2} |f|(x-y, t-|y|^2)\, \frac{1}{a^2}\, \eta\Big(\frac{y}{a}\Big)\, dy. \tag{2.2}$$

Here f is a function of $(x,t) \in \mathbb{R}^2 \times \mathbb{R}$, and η is a fixed bump function supported in the unit ball in $\mathbb{R}^2$ such that $\eta = 1$ in a neighborhood of the origin. This operator is of course known to be bounded on $L^2(\mathbb{R}^3)$ by standard results about maximal Radon transforms along a paraboloid. However, for the sake of illustration we will attempt an alternative proof using stopping-times and TT^*, and we will see what goes wrong. This will make it clear how this approach would fail for the operator considered in Theorem 1.2.

To begin our (doomed) alternative approach, we fix a non-negative function f and pick any measurable stopping-time $a(x,t)$ taking positive real values, and consider the linear operator

$$Tf(x,t) = \int_{y\in\mathbb{R}^2} f(x-y, t-|y|^2)\, \frac{1}{a(x,t)^2}\, \eta\Big(\frac{y}{a(x,t)}\Big)\, dy;$$

then $TT^*f(x,t)$ can be written as

$$TT^*f(x,t) = \int_{y\in\mathbb{R}^2}\int_{z\in\mathbb{R}^2} f(x-y+z, t-|y|^2+|z|^2)\, \frac{1}{a_1^2}\, \eta\Big(\frac{y}{a_1}\Big)\, \frac{1}{a_2^2}\, \eta\Big(\frac{z}{a_2}\Big)\, dy\, dz, \tag{2.3}$$

where a_1, a_2 stand for the stopping-time functions

$$a_1 := a(x,t), \quad a_2 := a(x-y+z, t-|y|^2+|z|^2).$$

In the region in the (y,z)-plane where $a_1 \geq a_2$ we would sequentially make the change of variables $(y,z) \mapsto (u,\tau,\sigma)$, where

$$u = y - z, \quad \tau = \frac{u\cdot z}{|u|} = \frac{u_1 z_1 + u_2 z_2}{|u|}, \quad \sigma = \frac{u_2 z_1 - u_1 z_2}{|u|}.$$

On the other hand, in the region in the (y,z)-plane where $a_1 \leq a_2$ we would make the change of variables $(y,z) \mapsto (u,\tau,\sigma)$, where

$$u = z - y, \quad \tau = \frac{u\cdot y}{|u|} = \frac{u_1 y_1 + u_2 y_2}{|u|}, \quad \sigma = \frac{u_2 y_1 - u_1 y_2}{|u|}.$$

After this process, and a trivial integration with respect to σ, we would have

$$\begin{aligned} |TT^*f(x,t)| \lesssim & \int_{u\in\mathbb{R}^2}\int_{\tau\in\mathbb{R}} f(x-u, t-|u|^2-2|u|\tau)\, \frac{1}{b_1^2}\, \chi_{B_2}\Big(\frac{u}{b_1}\Big)\, \frac{1}{b_2}\, \chi_{B_2}\Big(\frac{\tau}{b_2}\Big)\, du\, d\tau \\ & + \int_{u\in\mathbb{R}^2}\int_{\tau\in\mathbb{R}} f(x+u, t+|u|^2+2|u|\tau)\, \frac{1}{b_3^2}\, \chi_{B_2}\Big(\frac{u}{b_3}\Big)\, \frac{1}{b_4}\, \chi_{B_2}\Big(\frac{\tau}{b_4}\Big)\, du\, d\tau, \end{aligned}$$

where

$$b_1 := a(x,t), \quad b_2 := a(x-u, t-|u|^2 - 2|u|\tau),$$
$$b_3 := a(x+u, t+|u|^2+2|u|\tau), \quad b_4 := a(x,t).$$

The two terms on the right-hand side above are similar; for the sake of illustration we will focus on the first term, which we temporarily denote by $I(x,t)$. Let $\mathcal{M}_{\mathrm{Rad}}$ denote the usual maximal Radon transform along the paraboloid (defined in (1.19)) and let $\mathcal{M}^{(3)}$ denote a one-dimensional maximal function in the t-variable alone. Then one could hope that the term $I(x,t)$ is bounded by

$$|I(x,t)| \le C \int_{u\in\mathbb{R}^2} \mathcal{M}^{(3)} f(x-u, t-|u|^2)\, \frac{1}{b_1^2}\, \chi_{B_2}\Big(\frac{u}{b_1}\Big)\, du \le C\, \mathcal{M}_{\mathrm{Rad}}\, \mathcal{M}^{(3)} f(x,t).$$

Unfortunately the first inequality above fails: the stopping-time function $b_2 = a(x-u, t-|u|^2-2|u|\tau)$ depends not only on x, t and u, but also on τ. This leads to a first crucial failure: indeed, if $b := b(\tau)$ is an unknown function of τ, then

$$\int_{\tau\in\mathbb{R}} F(t-\tau)\, \frac{1}{b}\, \chi_{B_1}\Big(\frac{\tau}{b}\Big)\, d\tau$$

is not necessarily bounded by the Hardy–Littlewood maximal function of F. In such situations, it can be advantageous to dualize the statements, and integrate $I(x,t)$ against a test function $g(x,t)$. Applying this idea and then changing variables $t \mapsto t + 2|u|\tau$ leads to the representation of

$$\int_{\mathbb{R}^3} I(x,t)\, g(x,t)\, dx\, dt$$

as

$$\int_{(x,t)\in\mathbb{R}^3} \int_{u\in\mathbb{R}^2} \int_{\tau\in\mathbb{R}} f(x-u, t-|u|^2) \frac{1}{c_1^2} \chi_{B_2}(\frac{u}{c_1}) \frac{1}{c_2} \chi_{B_2}(\frac{\tau}{c_2}) g(x, t+2|u|\tau)\, du\, d\tau dx dt,$$

where now the stopping-time parameters are

$$c_1 := a(x, t+2|u|\tau), \quad c_2 := a(x-u, t-|u|^2).$$

It is true now that c_2 is independent of τ, so one would hope that the above is bounded by

$$\int_{(x,t)\in\mathbb{R}^3} \int_{u\in\mathbb{R}^2} f(x-u, t-|u|^2)\, \frac{1}{c_1^2}\, \chi_{B_2}\Big(\frac{u}{c_1}\Big)\, \mathcal{M}^{(3)} g(x,t)\, du\, dx\, dt;$$

unfortunately this is still not true, since now c_1 depends on τ, so that any integration with respect to τ would also be forced to consider c_1.

The failure of such TT^* arguments for maximal Radon transforms (even without Carleson-type behavior) demonstrates the need for a different approach. In the case of the maximal Radon transform $\mathcal{M}f(x,t)$ considered in (2.2), a well-known successful strategy is to construct a smoother variant of the average over the paraboloid, and to compare the maximal Radon transform with this smoother maximal function via a square function. (One may see this for example in [13] and §1.2 of Chapter 11 in [11].) More precisely, the maximal Radon transform $\mathcal{M}f(x,t)$ is comparable (up to a constant) to $\sup_{k\in\mathbb{Z}} A_k f(x,t)$ where

$$A_k f(x,t) := \int_{\mathbb{R}^2} f(x-y, t-|y|^2)\, 2^{-2k}\, \eta(2^{-k}y)\, dy.$$

One then compares it to a smoother variant, namely $\sup_{k\in\mathbb{Z}} B_k f(x,t)$, where

$$B_k f(x,t) := \int_{\mathbb{R}^2} \int_{\mathbb{R}} f(x-y, t-s) 2^{-2k}\, \eta(2^{-k}y)\, 2^{-2k}\, \zeta(2^{-2k}s)\, ds\, dy,$$

in which ζ is a smooth bump function supported on $[-1,1] \subset \mathbb{R}$, such that $\int_{\mathbb{R}} \zeta(s)ds = 1$.

The operator $\sup_{k\in\mathbb{Z}} B_k f$ is known to be bounded on L^2 (e.g. by covering arguments), and the difference $\sup_{k\in\mathbb{Z}} |A_k f - B_k f|$ is then trivially dominated by the square function

$$\Big(\sum_{k\in\mathbb{Z}} |A_k f - B_k f|^2\Big)^{1/2}.$$

The boundedness of the square function on L^2 can be proved using methods related to the Fourier transform. (The condition $\int_{\mathbb{R}} \zeta(s)ds = 1$ is used to guarantee the vanishing of the multiplier of $A_k - B_k$ near the origin.)

We now return to the setting of this paper, and in particular the oscillatory integral operator considered in Theorem 1.2. In order to prove Theorem 1.2, we will use the square function S_r defined in (1.21) to control the supremum over k in the L^2 bound (1.18). Since morally speaking $I^\lambda_{2^k} - J^\lambda_{2^k}$ only 'sees' a function of 'frequency' 2^k, an appropriate Littlewood–Paley decomposition will allow us to reduce to a situation in which k is fixed. We then need to control the supremum over λ: this is handled using a TT^* argument and stopping-times (in the phase only), which ironically are now the key to success, once the square function has been applied. Roughly speaking, within the square function the presence of the inconvenient stopping-times normalizing the supports of the bump functions have been removed, and so TT^* and stopping-times may be used in order to apply van der Corput estimates. But now we encounter another difficulty, and it is here that we need to restrict ourselves to the special class of polynomial phases considered in Theorem 1.2, namely those that are in the span of the fixed polynomials $p_2(y), \dots, p_d(y)$ provided in the theorem hypotheses. This restriction, which appears as well in the statement of Theorem 1.1, has its origin in the proof of Theorem 1.2.

Again for the sake of illustration, let us suppose we are more ambitious, and wish to take a supremum over all real polynomial phases of degree $\leq d$ with no linear or quadratic terms. We fix our attention for the moment on the case of dimension $n = 2$ and let

$$\mathcal{Q}_d = \{\text{real polynomials of degree} \leq d \text{ on } \mathbb{R}^2\},$$
$$\mathcal{Q}^*_d = \{Q \in \mathcal{Q}_d \text{ that vanishes to order} \geq 2 \text{ at the origin}\}.$$

Let Γ_d^* be the set of all coefficients of polynomials in $\mathcal{Q}_d^*$, so that any element $\lambda \in \Gamma_d^*$ takes the form $\lambda = (\lambda_\alpha)_{3\leq|\alpha|\leq d}$. For each $\lambda \in \Gamma_d^*$, we may define a polynomial in $\mathcal{Q}_d^*$ by

$$Q_\lambda(y) = \sum_{3\leq|\alpha|\leq d} \lambda_\alpha y^\alpha,$$

with coefficient norm $\|\lambda\| = \sum_\alpha |\lambda_\alpha|$. We may now accordingly define $T_\lambda f$ and ${}^{(\eta)}I_a^\lambda f$ by (1.8) and (1.17) respectively, with the phase $Q_\lambda(y)$ in place of $P_\lambda(y)$. If we now want to prove that the polynomial Carleson operator $\sup_{\lambda\in\Gamma_d^*} |T_\lambda f|$ is bounded on $L^2(\mathbb{R}^3)$, we are led to consider the maximal oscillatory operator given by

$$\sup_{k\in\mathbb{Z}} \sup_{\substack{\lambda\in\Gamma_d^* \\ r\leq\|\lambda\|\leq 2r}} \left|{}^{(\eta)}I_{2^k}^\lambda f\right|,$$

for which we must show that there exists a $\delta > 0$ such that its L^2 norm is $\leq C r^{-\delta}$ for all $r \geq 1$. Since a square function argument will essentially allow one to handle the supremum over k, we will temporarily fix $k = 0$ and consider only the supremum over λ.

Thus we restrict our attention to trying to prove

$$\Big\| \sup_{\substack{\lambda\in\Gamma_d^* \\ r\leq\|\lambda\|\leq 2r}} \left|{}^{(\eta)}I_1^\lambda f\right| \Big\|_{L^2(\mathbb{R}^3)} \leq C r^{-\delta} \|f\|_{L^2(\mathbb{R}^3)},$$

for some $\delta > 0$. Motivated by Stein and Wainger [14], a natural approach is to combine stopping-times with a TT^* argument: let $\lambda(x,t) = (\lambda_\alpha(x,t))_{3\leq|\alpha|\leq d}$ be a measurable stopping-time function taking values in Γ_d^*, and consider

$$Tf(x,t) = {}^{(\eta)}I_1^{\lambda(x,t)} f(x,t).$$

Then $TT^*f(x,t)$ can be written as

$$\begin{aligned} &TT^*f(x,t) \\ &\quad= \int_{u\in\mathbb{R}^2}\int_{\tau\in\mathbb{R}} f(x-u, t-|u|^2-2|u|\tau) K_\sharp^{\lambda(x,t),\lambda(x-u,t-|u|^2-2|u|\tau)}(u,\tau)\,du\,d\tau, \end{aligned} \tag{2.4}$$

where for ν, $\mu \in \Gamma_d^*$ we define the kernel

$$K_\sharp^{\nu,\mu}(u,\tau) := \int_{\sigma\in\mathbb{R}} e^{iP_\nu(u+z)-iP_\mu(z)}\,\eta(u+z)\,\eta(z)\,d\sigma,$$

in which $z = (z_1, z_2)$ is a function of (u,τ,σ) defined implicitly by the relations

$$\tau = \frac{u\cdot z}{|u|} = \frac{u_1z_1 + u_2z_2}{|u|}, \qquad \sigma = \frac{u_2z_1 - u_1z_2}{|u|}. \tag{2.5}$$

Using van der Corput estimates, we can show the existence of a small $\delta > 0$ such that the following holds: for any $r \geq 1$, and for each ν, $\mu \in \Gamma_d^*$ with $r \leq \|\nu\|$, $\|\mu\| \leq 2r$, there exists a "bad" (but small) set of u, denoted $G^{\mu,\nu} \subset B_2(\mathbb{R}^2)$, and for each $u \in B_2(\mathbb{R}^2)$ there exists a "bad" (but small) set of τ, denoted $F_u^{\nu,\mu} \subset B_2(\mathbb{R})$, such that

$$|G^{\nu,\mu}| \leq C r^{-\delta}, \quad |F_u^{\nu,\mu}| \leq C r^{-\delta},$$

and

$$|K_\sharp^{\nu,\mu}(u,\tau)| \leq C\,[r^{-\delta}\,\chi_{B_2}(u)\chi_{B_2}(\tau) + \chi_{G^{\nu,\mu}}(u)\chi_{B_2}(\tau) + \chi_{B_2}(u)\chi_{F_u^{\nu,\mu}}(\tau)]. \tag{2.6}$$

At first glance this looks as good as the bound (2.1) that is the keystone of Stein and Wainger's work. Yet in fact (2.6) actually fails to be effective since the small sets $G^{\nu,\mu}$ and $F_u^{\nu,\mu}$ depend simultaneously on ν and μ, and this is just as deleterious as the simultaneous appearance of a_1 and a_2 in formula (2.3), again because of the presence of Radon-type behavior. More explicitly, if we apply the estimate (2.6) for $K_\sharp^{\nu,\mu}$ in (2.4), it would show that $TT^*f(x,t)$ is bounded by a sum of three terms, of which we single out the term

$$\int_{u\in\mathbb{R}^2}\int_{\tau\in\mathbb{R}} |f|(x-u,t-|u|^2-2|u|\tau)\chi_{G^{\lambda(x,t),\lambda(x-u,t-|u|^2-2|u|\tau)}}(u)\chi_{B_2}(\tau)\,du\,d\tau.$$

We would like to say that this is bounded by a concatenation of maximal functions, including Radon-type adaptations of the "small set maximal functions" used by Stein and Wainger, however, this is not true, since when one integrates in τ, one must remember that the small set $G^{\lambda(x,t),\lambda(x-u,t-|u|^2-2|u|\tau)}$ depends also on τ.

We circumvent all of these difficulties by intertwining the methods of square functions, TT^*, and stopping-times. First, we eliminate the presence of two independent stopping-times a_1, a_2 by passing to the square function (1.21); note that inside each term of $S_r(f)$, the scaling parameters are fixed and equal: $a_1 = a_2 = 2^k$. (Here it is important that we are able to restrict the supremum over the scaling factors a to a supremum over a countable set of scaling factors $a = 2^k$ for $k \in \mathbb{Z}$, rather than over all $a > 0$, as in the work of Stein and Wainger.) Second, we apply a TT^* argument to each fixed summand within the square function, and apply van der Corput estimates to extract decay in r from the kernel of TT^*. (We note that our requirement that $n \geq 2$ arises here, since in the case $n = 1$, the kernel of TT^* does not take the form of an integral, and hence will not admit van der Corput estimates.) In particular, we are able to construct small bad sets G^ν and F_u^ν that depend on only ν but not on μ by an argument that explicitly uses the assumption that the polynomial phase $P_\lambda(y)$ belongs to the restricted class specified in Theorems 1.1 and 1.2.

Finally, as we have mentioned already, in order to estimate the square function S_r (in particular, to carry out the sum over k in (1.21)), we need to know that $(I_a^\lambda - J_a^\lambda)f$ sees only the part of f that has frequency a. To make this precise, we need to introduce a family of Littlewood–Paley projections Δ_j and to obtain some almost orthogonality for $(I_a^\lambda - J_a^\lambda)\Delta_j$ if a is very different from 2^j. This turns out to be rather tricky, especially when 2^j is much smaller than a; in that case one would like to write Δ_j as a derivative, and integrate by parts, but then a derivative may land on the singular kernel of I_a^λ. Since I_a^λ is a Radon transform that involves an integration over a lower dimensional submanifold, if the derivative is not tangential to the submanifold over which the integration takes place, then integration by parts does not work. It turns out that one must both define a correct Littlewood–Paley projection and use a TT^* argument in order to carry the argument out rigorously. See Section 7.4.1 for details, including a comment at the end of the same section on our choice of Littlewood–Paley projection.

3. Preliminary lemmas: Van der Corput estimates

We begin by recording several estimates of van der Corput type as stated in [14]. Let

$$Q_\lambda(x) = \sum_{0\leq|\alpha|\leq d} \lambda_\alpha \, x^\alpha$$

be a polynomial of degree d in $\mathbb{R}^m$, with coefficients $\lambda_\alpha \in \mathbb{R}$. Note that for the moment we allow constant, linear, quadratic, and higher order terms in Q_λ. Let

$$\|\lambda\| = \sum_{1\leq|\alpha|\leq d} |\lambda_\alpha| \tag{3.1}$$

denote the isotropic norm of the coefficients of non-constant terms. Let

$$[\![\lambda]\!] = \sum_{0\leq|\alpha|\leq d} |\lambda_\alpha| = \|\lambda\| + |Q_\lambda(0)| \tag{3.2}$$

denote the isotropic norm, including the constant term. We now quote Proposition 2.1 of [14].

Lemma 3.1. *For any C^1 function ψ defined on the unit ball $B_1(\mathbb{R}^m)$ such that $\|\psi\|_{C^1} \leq 1$, and for any convex subset $\Omega \subseteq B_1(\mathbb{R}^m)$, we have*

$$\Big| \int_\Omega e^{iQ_\lambda(x)} \, \psi(x) \, dx \Big| \leq C \, \|\lambda\|^{-1/d}, \tag{3.3}$$

where the constant C depends on the dimension m and the degree d of Q_λ, but not otherwise on Q_λ, ψ or Ω.

Note that this upper bound is of course independent of the constant term in $Q_\lambda(x)$. The result of the lemma continues to hold with the unit ball $B_1(\mathbb{R}^m)$ replaced by a Euclidean ball of any other fixed radius, such as $B_2(\mathbb{R}^m)$.

We will also require a lemma estimating the measure of the set where the real-valued polynomial Q_λ takes small values.

Lemma 3.2. *Let Q_λ be a polynomial as above (possibly including a nonzero constant term). For every $\rho > 0$,*

$$|\{x \in B_1(\mathbb{R}^m) : |Q_\lambda(x)| \leq \rho\}| \leq C \, \rho^{1/d} \, \|\lambda\|^{-1/d}, \tag{3.4}$$

where the constant C depends only on the dimension m and the degree d of Q_λ.

This lemma is a slight variant of Proposition 2.2 of [14]; in [14] they assumed that $Q_\lambda(y)$ has no constant term, but their proof carries over to the present case.

We will also require an improvement of this lemma when the polynomial Q_λ has a constant term that is large. (The estimate of the above lemma does not see the constant term of Q_λ, by definition of $\|\lambda\|$.) We prove:

Lemma 3.3. *Let Q_λ be a polynomial as above (possibly including a nonzero constant term). For every $\rho > 0$,*

$$|\{x \in B_1(\mathbb{R}^m) : |Q_\lambda(x)| \leq \rho\}| \leq C' \rho^{1/d} [\![\lambda]\!]^{-1/d}, \tag{3.5}$$

where the constant C' depends only on the dimension m and the degree d of Q_λ.

We note that both Lemma 3.2 and 3.3 continue to hold if $B_1(\mathbb{R}^m)$ is replaced by any other Euclidean ball of fixed radius, such as $B_2(\mathbb{R}^m)$.

To prove Lemma 3.3, first suppose that $\|\lambda\| \geq [\![\lambda]\!]/4$. Then Lemma 3.2 shows

$$|\{x \in B_1 : |Q_\lambda(x)| \leq \rho\}| \leq C \rho^{1/d} \|\lambda\|^{-1/d} \leq C\, 4^{1/d} \rho^{1/d} [\![\lambda]\!]^{-1/d}.$$

Next, we suppose that $\|\lambda\| < [\![\lambda]\!]/4$ so that necessarily

$$|Q_\lambda(0)| = [\![\lambda]\!] - \|\lambda\| \geq \frac{3}{4} [\![\lambda]\!]. \tag{3.6}$$

In this case we crucially use the fact that the set we consider lies inside a ball of fixed radius; for $x \in B_1$, (3.6) implies that

$$\begin{aligned} |Q_\lambda(x)| &\geq |Q_\lambda(0)| - \Big| \sum_{1 \leq |\alpha| \leq d} \lambda_\alpha x^\alpha \Big| \geq |Q_\lambda(0)| - \sum_{1 \leq |\alpha| \leq d} |\lambda_\alpha| \\ &= |Q_\lambda(0)| - \|\lambda\| \geq \frac{1}{2} [\![\lambda]\!]. \end{aligned} \tag{3.7}$$

In particular, if $[\![\lambda]\!] > 2\rho$, this implies the set $\{x \in B_1 : |Q_\lambda(x)| \leq \rho\}$ is empty and hence (3.5) trivially holds. On the other hand, if $[\![\lambda]\!] \leq 2\rho$ then we can prove (3.5) directly. We need only note that in this case,

$$\left(\frac{\rho}{[\![\lambda]\!]}\right)^{1/d} \geq \left(\frac{1}{2}\right)^{1/d}.$$

Thus we may apply the trivial bound:

$$|\{x \in B_1 : |Q_\lambda(x)| < \rho\}| \leq |B_1| \leq |B_1|\, 2^{1/d} \left(\frac{\rho}{[\![\lambda]\!]}\right)^{1/d}.$$

This establishes (3.5) with the constant $C' = \max\{C4^{1/d}, |B_1|2^{1/d}\}$.

3.1. Novel Van der Corput estimates for kernels

We now state the novel van der Corput estimates we will apply to bound the kernels of various operators of the form TT^*; these kernel estimates are in fact the heart of the paper.

Proposition 3.4 ($K_\flat^{\nu,\mu}$ van der Corput, $n \geq 2$)**.** *Fix any dimension $n \geq 1$, and any degree $d \geq 2$. Suppose $\nu = (\nu_\alpha)_{2 \leq |\alpha| \leq d}$ denotes the coefficients of a real-valued polynomial*

$$Q_\nu(y) = \sum_{2 \leq |\alpha| \leq d} \nu_\alpha\, y^\alpha$$

on $\mathbb{R}^n$ *that has no linear terms, and similarly for* $\mu = (\mu_\alpha)_{2\le|\alpha|\le d}$*. Let*

$$\|\nu\| = \sum_{2\le|\alpha|\le d} |\nu_\alpha|, \quad \|\mu\| = \sum_{2\le|\alpha|\le d} |\mu_\alpha|.$$

Given a C^1 *function* $\Psi(u,z)$ *supported on* $B_2(\mathbb{R}^n)\times B_1(\mathbb{R}^n)$*, define*

$$K_\flat^{\nu,\mu}(u) = \int_{\mathbb{R}^n} e^{iQ_\nu(u+z)-iQ_\mu(z)}\,\Psi(u,z)\,dz.$$

Suppose furthermore that

$$\|\Psi\|_{C^1} := \sup_{(u,z)\in B_2(\mathbb{R}^n)\times B_1(\mathbb{R}^n)} (|\Psi(u,z)| + |\nabla_z\Psi(u,z)|) \le 1.$$

Then there exists a small constant $\delta > 0$ *(depending only on* d*), such that the following holds: if* ν*,* μ *satisfy*

$$r \le \|\nu\|, \|\mu\| \le 2r$$

for some $r \ge 1$*, then there exists a small measurable set* $G^\nu \subset B_2(\mathbb{R}^n)$ *depending on* ν *(but not on* μ *nor* Ψ*), with*

$$|G^\nu| \le C\,r^{-\delta},$$

such that

$$|K_\flat^{\nu,\mu}(u)| \le C\left(r^{-\delta}\,\chi_{B_2}(u) + \chi_{G^\nu}(u)\right),$$

where in each case C *depends only on* n, d*.*

This is effectively a result of Stein and Wainger (namely Lemma 4.1 in [14], in the case $h = 1$, in their notation). It follows from a clever application of Lemmas 3.1 and 3.2, and crucially uses the assumption that $Q_\lambda(y)$ lacks linear terms.

In the Radon setting, we require a more elaborate version; we state here a result in dimension $n = 2$ that we prove in Section 8. We reserve the more technical statement of the result for general dimensions $n \ge 2$ for Proposition 9.1 in Section 9.

Proposition 3.5 ($K_\sharp^{\nu,\mu}$ van der Corput, $n = 2$)**.** *Fix the dimension* $n = 2$ *and a degree* $d \ge 2$*. Let*

$$\mathscr{P} = \{p_2(y), p_3(y), \ldots, p_d(y)\}$$

be a set of real-valued polynomials on $\mathbb{R}^2$*, where each* $p_j(y)$ *is homogeneous of degree* j*, and* $p_2(y) \ne C|y|^2$ *for any nonzero constant* C*. Let*

$$\Lambda = \Lambda(\mathscr{P}) = \{2 \le m \le d \colon p_m(y) \not\equiv 0\}.$$

For $\nu = (\nu_2, \ldots, \nu_d) \in \mathbb{R}^{d-1}$*, let*

$$P_\nu(y) = \sum_{m=2}^{d} \nu_m\, p_m(y) \quad \textit{and} \quad \|\nu\| = \sum_{m\in\Lambda} |\nu_m|,$$

and define $P_\mu(y)$ *and* $\|\mu\|$ *similarly for* $\mu = (\mu_2, \ldots, \mu_d) \in \mathbb{R}^{d-1}$*.*

Given a C^1 function $\Psi(u,z)$ supported on $B_2(\mathbb{R}^2)\times B_1(\mathbb{R}^2)$, define

$$K_\sharp^{\nu,\mu}(u,\tau) = \int_{\mathbb{R}} e^{iP_\nu(u+z)-iP_\mu(z)}\,\Psi(u,z)\,d\sigma,$$

where the z in the integral is defined implicitly in terms of u,τ,σ by

$$\tau = \frac{u_1z_1+u_2z_2}{|u|}, \qquad \sigma = \frac{-u_1z_2+u_2z_1}{|u|}. \tag{3.8}$$

Suppose furthermore that

$$\|\Psi\|_{C^1(\mathbb{R})} := \sup_{(u,z)\in B_2(\mathbb{R}^2)\times B_1(\mathbb{R}^2)} \left(|\Psi(u,z)| + |\frac{\partial}{\partial\sigma}\Psi(u,z)|\right) \le 1.$$

Then there exists a small constant $\delta>0$ (depending only on d) such that the following holds: if μ, ν satisfy

$$r \le \|\nu\|, \|\mu\| \le 2r$$

for some $r\ge 1$, then there exists a small set $G^\nu \subset B_2(\mathbb{R}^2)$, and for each $u\in B_2(\mathbb{R}^2)$ a small set $F_u^\nu \subset B_1(\mathbb{R})$, such that

$$|G^\nu| \le C r^{-\delta}, \quad |F_u^\nu| \le C r^{-\delta} \quad \textit{for all } u\in B_2(\mathbb{R}^2),$$

and

$$|K_\sharp^{\nu,\mu}(u,\tau)| \le C\left(r^{-\delta}\chi_{B_2}(u)\chi_{B_1}(\tau) + \chi_{G^\nu}(u)\chi_{B_1}(\tau) + \chi_{B_2}(u)\chi_{F_u^\nu}(\tau)\right). \tag{3.9}$$

The choices of the small sets G^ν and F_u^ν are independent of both μ and Ψ, and the constants C in the above upper bounds depend only on d and the fixed set of polynomials $p_2,\ldots,p_d$.

The significance of this result relative to that of Proposition 3.4 for $K_\flat^{\nu,\mu}$ when $n=2$ is that the integral $K_\sharp^{\nu,\mu}$ is now only a one-dimensional integral, yet we are still able to capture decay in r, as long as the polynomial phase occurring in $K_\sharp^{\nu,\mu}$ is of a more restrictive form than in $K_\flat^{\nu,\mu}$.

4. Preliminary lemmas: Littlewood–Paley decomposition

We now give an explicit construction of a Littlewood–Paley decomposition we will employ to bound the square function (1.21) on L^2. We start with a function $\chi\in C_c^\infty([-1,1])$ such that χ is identically 1 on a small neighborhood of the origin. We then set

$$\Phi(\tau) = \chi(\tau/4) - \chi(\tau),$$

so that Φ is a smooth function with support in $[-4,4]$ that vanishes in a small neighborhood of the origin. We observe that if we set $\Phi_j(\tau) = \Phi(2^{2j}\tau)$, then for any $\tau\neq 0$,

$$\sum_{j\in\mathbb{Z}} \Phi_j(\tau) = 1,$$

by a telescoping argument. We now define a Schwartz function Δ on $\mathbb{R}$ by

$$\Delta(t) = \check{\Phi}(t).$$

We also define corresponding scaled versions by $\Delta_j(t) = \check{\Phi}_j(t)$, so that

$$\Delta_j(t) = 2^{-2j}\Delta(2^{-2j}t).$$

We observe that

$$\int_{\mathbb{R}} \Delta(t)dt = \int_{\mathbb{R}} \Delta_j(t)\, dt = \Phi(0) = 0, \quad \text{for any } j. \tag{4.1}$$

Given any function $f \in \mathcal{S}(\mathbb{R}^{n+1})$ we define

$$\Delta_j f(x,t) = \int f(x, t-s)\, \Delta_j(s)\, ds,$$

so that $\Delta_j f$ is also of Schwartz class. Note that we use Δ_j to denote both the kernel and the operator itself. Also each Δ_j extends to a bounded linear operator on $L^2(\mathbb{R}^{n+1})$. Finally, we have defined the Littlewood–Paley decomposition so that it is compatible with parabolic scalings $(x,t) \mapsto (\delta x, \delta^2 t)$.

The following is a standard result, which follows from an easy application of the Fourier transform, and Fubini's theorem.

Proposition 4.1. *For all $f \in L^2(\mathbb{R}^{n+1})$, the partial sums $\sum_{j=-N}^{N} \Delta_j f$ converge to f in $L^2(\mathbb{R}^{n+1})$ norm. Moreover,*

$$\Big\| \Big(\sum_{j=-\infty}^{\infty} |\Delta_j f|^2 \Big)^{1/2} \Big\|_{L^2(\mathbb{R}^{n+1})} \leq C\, \|f\|_{L^2(\mathbb{R}^{n+1})}. \tag{4.2}$$

We will require an additional Littlewood–Paley decomposition in order to gain a reproducing property. Fix a smooth function $\tilde{\Phi}$ with compact support in $[-8,8]$ such that $\tilde{\Phi} \equiv 1$ on the support of Φ and $\tilde{\Phi} \equiv 0$ in a small neighborhood of the origin. Correspondingly, we define the scaled version $\tilde{\Phi}_j = \tilde{\Phi}(2^{2j}\tau)$ and the kernels $\tilde{\Delta}_j$ and associated operators $\tilde{\Delta}_j$ as above. In particular $\Phi\tilde{\Phi} = \Phi$ and thus, as operators,

$$\Delta_j \tilde{\Delta}_j = \Delta_j. \tag{4.3}$$

We summarize the additional simple properties we require as follows.

Proposition 4.2. *For all $f \in L^2$,*

$$f = \sum_{j=-\infty}^{\infty} \Delta_j \tilde{\Delta}_j f \tag{4.4}$$

where the convergence of the sum on the right-hand side is taken in the L^2 sense. In addition,

$$\Big\| \Big(\sum_{j=-\infty}^{\infty} |\tilde{\Delta}_j f|^2 \Big)^{1/2} \Big\|_{L^2(\mathbb{R}^{n+1})} \leq C\, \|f\|_{L^2(\mathbb{R}^{n+1})} \tag{4.5}$$

and

$$\Big\|\Big(\sum_{j=-\infty}^{\infty}|\Delta_j\tilde{\Delta}_j f|^2\Big)^{1/2}\Big\|_{L^2(\mathbb{R}^{n+1})}\leq C\,\|f\|_{L^2(\mathbb{R}^{n+1})}. \tag{4.6}$$

The first and third properties are simple consequences of (4.3) and Proposition 4.1; the second property follows from a similar argument to that needed for (4.2).

4.1. Convolutions

Later on we will require a third type of Littlewood–Paley operator, defined by $\Delta_j * \Delta_j^-$, where $\Delta_j^-(x) := \Delta_j(-x)$. Precisely, we define

$$\underline{\Delta}_j(t)=\int \Delta_j(w+t)\,\Delta_j(w)\,dw,$$

so that $\underline{\Delta}_j$ is also a Schwartz function. We note the following properties of $\underline{\Delta}_j$:

Lemma 4.3. *For all $j\in\mathbb{Z}$,*

$$\underline{\Delta}_j(t)=\frac{1}{a^2}\,\underline{\Delta}_{j-k}\Big(\frac{t}{a^2}\Big)\quad \text{if } a=2^k, \tag{4.7}$$

and

$$\int_{\mathbb{R}}\underline{\Delta}_j(t)\,dt=0. \tag{4.8}$$

In addition, the $\underline{\Delta}_j$ are in $L^1(\mathbb{R})$ with uniform norm, independent of j. Finally, for $\psi(t)=1/(1+t^2)$ we set

$$\psi_j(t)=2^{-2j}\,\psi(2^{-2j}t) \tag{4.9}$$

to be the L^1-normalization compatible with parabolic dilations. Then the derivative $(\underline{\Delta}_j)'$ satisfies

$$|(\underline{\Delta}_j)'(t)|\leq C\,2^{-2j}\,\psi_j(t). \tag{4.10}$$

The first three properties of $\underline{\Delta}_j$ are all simple consequences of the definition of Δ_j. Also, note that by dilation invariance, it suffices to establish (4.10) when $j=0$. In that case, the desired estimate reduces to $|\underline{\Delta}_0'(t)|\leq C/(1+t^2)$, which follows simply from the fact that $\underline{\Delta}_0$ is Schwartz.

In addition, we will use the following consequence of the mean value theorem.

Lemma 4.4. *Let $\psi(t)=1/(1+t^2)$ and set $\psi_j(t)=2^{-2j}\psi(2^{-2j}t)$. For each $j\geq 0$, for any $|\xi|\leq 2$, we have*

$$|\underline{\Delta}_j(t+\xi)-\underline{\Delta}_j(t)|\leq C\,2^{-2j}\,|\xi|\,\psi_j(t).$$

In addition, we note that ψ_j is a non-negative integrable function on $\mathbb{R}$, with L^1 norm uniformly bounded, independent of j.

Certainly, the mean value theorem applies to the Schwartz function $\underline{\Delta}_j$, showing that for any fixed t and any $|\xi| \leq 2$ we have

$$|\underline{\Delta}_j(t+\xi) - \underline{\Delta}_j(t)| \leq |\xi| \, |(\underline{\Delta}_j)'(t+\xi_0)|,$$

for some $|\xi_0| \leq |\xi| \leq 2$. We now apply the estimate (4.10) to conclude that

$$|\underline{\Delta}_j(t+\xi) - \underline{\Delta}_j(t)| \leq C\, 2^{-2j} \, |\xi| \, \psi_j(t+\xi_0).$$

It then remains to observe that

$$\psi_j(t+\xi_0) \leq C\, \psi_j(t)$$

for all $t \in \mathbb{R}$, $|\xi_0| \leq 2$, as may be easily verified.

4.2. Antiderivatives

Finally, we record for later use the following fact about antiderivatives of Δ_j and $\underline{\Delta}_j$.

Lemma 4.5. *There exist Schwartz functions $\widetilde{\Delta}$ and $\widetilde{\underline{\Delta}}$ on $\mathbb{R}$ such that upon setting $\widetilde{\Delta}_j(t) = 2^{-2j}\widetilde{\Delta}(2^{-2j}t)$ and $\widetilde{\underline{\Delta}}_j(t) = 2^{-2j}\widetilde{\underline{\Delta}}(2^{-2j}t)$ we have for every $j \in \mathbb{Z}$*

$$\Delta_j(t) = 2^{2j}\Big(\frac{d}{dt}\widetilde{\Delta}_j\Big)(t), \quad \underline{\Delta}_j(t) = 2^{2j}\Big(\frac{d}{dt}\widetilde{\underline{\Delta}}_j\Big)(t). \tag{4.11}$$

In particular, $\widetilde{\Delta}_j$ and $\widetilde{\underline{\Delta}}_j$ are uniformly in $L^1(\mathbb{R})$.

Again by dilation invariance, one could reduce to the case $j = 0$. The lemma would then follow from the following claim: if F is a Schwartz function with $\int_{\mathbb{R}} F(t)dt = 0$, then there exists another Schwartz function $\tilde{F}$ such that $\tilde{F}'(t) = F(t)$. To prove this claim, let

$$\widetilde{F}(t) = \int_{-\infty}^{t} F(\tau)\, d\tau.$$

Then it is immediate that $F(t) = \frac{d}{dt}\widetilde{F}(t)$, and we need only verify that $\widetilde{F}(t)$ is a Schwartz function. To see that $\widetilde{F}$ exhibits rapid decay as $t \to -\infty$, we use the fact that F has rapid decay, so that

$$|\widetilde{F}(t)| \leq C \int_{-\infty}^{t} \frac{1}{1+|\tau|^N}\, d\tau \leq C\, \frac{1}{1+|t|^{N-1}},$$

as $t \to -\infty$, for arbitrary large N. By the assumed fact that F integrates to zero, we can also use the alternative representation

$$\widetilde{F}(t) = -\int_{t}^{\infty} F(\tau)\, d\tau,$$

which similarly allows us to conclude that $\widetilde{F}$ exhibits rapid decay as $t \to +\infty$. Finally, we note that $\widetilde{F}(t)$ is infinitely differentiable and its derivatives exhibit rapid decay, since $(\widetilde{F})' = F$ and F is Schwartz. Thus $\widetilde{F}$ is Schwartz.

5. The proof of Theorem 1.1

In this section, we show how to deduce our main result, Theorem 1.1, from Theorem 1.2. Suppose we are given $d-1$ real polynomials $p_2(y), \dots, p_d(y)$ on $\mathbb{R}^n$, where each p_j is homogeneous of degree j, and $p_2(y) \neq C|y|^2$ for any nonzero constant C. We set $\Lambda = \{2 \le m \le d : p_m(y) \not\equiv 0\}$. For $\lambda = (\lambda_2, \dots, \lambda_d) \in \mathbb{R}^{d-1}$, we define $P_\lambda(y)$ as in (1.13) by

$$P_\lambda(y) = \sum_{m=2}^{d} \lambda_m \, p_m(y),$$

and $T_\lambda f$ as in (1.8) by

$$T_\lambda f(x,t) = \int_{\mathbb{R}^n} f(x-y, t-|y|^2) \, e^{iP_\lambda(y)} \, K(y) \, dy.$$

We want to show that for all $1 < p < \infty$ and all Schwartz functions f on $\mathbb{R}^{n+1}$,

$$\Big\| \sup_{\lambda \in \mathbb{R}^{d-1}} |T_\lambda f| \Big\|_{L^p(\mathbb{R}^{n+1})} \le A_p \, \|f\|_{L^p(\mathbb{R}^{n+1})}.$$

To do so, note that in the integral defining $T_\lambda f(x,t)$, if y is sufficiently small (with respect to λ) then $P_\lambda(y)$ is approximately zero and $e^{iP_\lambda(y)}$ can be approximated by 1. To make this precise, recall we have already defined a homogeneous norm $\|\lambda\|$ on $\mathbb{R}^{d-1}$ in (1.16). However, what is more relevant here is a non-isotropic norm on $\mathbb{R}^{d-1}$, which we define by

$$N(\lambda) = \sum_{m \in \Lambda} |\lambda_m|^{1/m}$$

for $\lambda \in \mathbb{R}^{d-1}$. The key is that $|P_\lambda(y)| \le 1$ whenever $|y| \le cN(\lambda)$, where c is some fixed constant dependent only on the dimension, the degree d, and the fixed polynomials $p_2, \dots, p_d$. We will make use of this observation very soon.

The Calderón–Zygmund kernel K admits a decomposition (see Chapter XIII of [11]) as

$$K(x) = \sum_{j=-\infty}^{\infty} 2^{-nj} \, \phi_j(2^{-j}x)$$

where each ϕ_j has the following properties:

(i) ϕ_j is a C^1 function with support in $1/4 < |x| \le 1$,

(ii) $|\partial_x^\alpha \phi_j(x)| \le C$ for $0 \le |\alpha| \le 1$ for some constant C that is uniform in j,

(iii) $\int_{\mathbb{R}^n} \phi_j(x) dx = 0$ for every j.

This allows us to decompose K as in [14]: precisely, given $\lambda \in \mathbb{R}^{d-1}$, we split K as

$$K = K_\lambda^+ + K_\lambda^- = \sum_{2^j < 1/N(\lambda)} K_j + \sum_{2^j \ge 1/N(\lambda)} K_j,$$

where $K_j(x) = 2^{-nj}\phi_j(2^{-j}x)$. Accordingly we split T_λ into $T_\lambda = T_\lambda^+ + T_\lambda^-$, where (respectively)

$$T_\lambda^\pm f(x,t) = \int_{\mathbb{R}^n} f(x-y, t-|y|^2)\, e^{iP_\lambda(y)} K_\lambda^\pm(y)\, dy.$$

To prove Theorem 1.1 it is sufficient to bound the L^p norms of $\sup_\lambda |T_\lambda^+ f|$ and $\sup_\lambda |T_\lambda^- f|$ individually. In the support of T_λ^-, where $2^j N(\lambda) \le 1$, the phase $P_\lambda(y)$ will not cause significant oscillation, and we will aim to remove the oscillatory factor and bound the remaining operator by maximal truncated singular Radon transforms. In the support of T_λ^+, we would expect the phase $P_\lambda(y)$ to contribute significant oscillation; this portion of the operator leads to the operators ${}^{(\eta_k)}I_{2^k}^\lambda$ and ${}^{(\eta_k)}J_{2^k}^\lambda$ defined in (1.17) and (1.20), and the square function (1.21).

5.1. Bounding T_λ^-

We first outline the treatment of T_λ^-. Note that $K_\lambda^-(y)$ is visibly supported where $|y| \le N(\lambda)^{-1}$, and in fact agrees precisely with K_λ if $|y| \le (4N(\lambda))^{-1}$. We may replace the oscillatory factor $e^{iP_\lambda(y)}$ by 1 with moderate error, since

$$\begin{aligned}|e^{iP_\lambda(y)} - 1| &\le c \sum_{2\le m\le d} |\lambda_m|\, |p_m(y)| \le c' \sum_{m\in\Lambda} N(\lambda)^m\, |y|^m \\ &= c' \sum_{2\le m\le d} (N(\lambda)|y|)^m \le c'' N(\lambda)\, |y|.\end{aligned}$$

Here the constant c' depends on the coefficients of the fixed set of polynomials $p_2, \dots, p_d$, and we have used the fact that for every $2 \le m \le d$, $|\lambda_m| \le N(\lambda)^m$, followed by the assumption that $|y|N(\lambda) \le 1$. It follows from this and the usual bound (1.9) for K that

$$\begin{aligned}T_\lambda^- f(x,t) &= \int_{|y|\le 1/N(\lambda)} K_\lambda^-(y) f(x-y, t-|y|^2)\, dy \\ &\quad + O\Big(N(\lambda) \int_{|y|\le 1/N(\lambda)} |y|^{-n+1} |f|(x-y, t-|y|^2)\, dy\Big). \qquad (5.1)\end{aligned}$$

Taking the supremum over λ, the first term on the right-hand side is dominated by a truncated maximal singular Radon transform; the second is dominated by a maximal function along the paraboloid.

Precisely, we define the operators and state the results we require, beginning with the truncated maximal singular Radon transform. Define a singular Radon transform along the paraboloid, initially acting on functions of Schwartz class, by

$$\mathcal{H}f(x,t) = \int f(x-y, t-|y|^2)\, K(y)\, dy, \qquad (5.2)$$

where $K(y)$ is a Calderón–Zygmund kernel on $\mathbb{R}^n$. Then $\mathcal{H}$ is known to be a bounded operator on L^p (see for example §4.5 of Chapter 11 of [11], or the earlier case for curves given in [13]). For each $\varepsilon > 0$, let

$$\mathcal{H}_\varepsilon f(x,t) = \int_{|y|>\varepsilon} f(x-y, t-|y|^2)\, K(y)\, dy$$

be a truncation of $\mathcal{H}$. The result we require is as follows.

Theorem 5.1. *For every $1 < p < \infty$, we have an a priori inequality for f of Schwartz class:*

$$\Big\| \sup_{\varepsilon>0} |\mathcal{H}_\varepsilon f| \Big\|_{L^p(\mathbb{R}^{n+1})} \leq A_p \, \|f\|_{L^p(\mathbb{R}^{n+1})}. \tag{5.3}$$

As usual, by a limiting argument, it follows that the operator $\sup_{\varepsilon>0} |\mathcal{H}_\varepsilon f|$ extends to a bounded operator acting on functions in L^p, satisfying the same bound (5.3). Theorem 5.1 can be deduced from a result of Duoandikoetxea and Rubio de Francia [2]; for completeness we adapt their proof to our setting, and provide the details in an appendix in Section 11.

It follows from Theorem 5.1 that the supremum over λ of the first term on the right-hand side of (5.1) is bounded on L^p, since

$$\begin{aligned} &\Big\| \sup_\lambda \int_{|y|\leq 1/N(\lambda)} K_\lambda^-(y) f(x-y, t-|y|^2)\, dy \Big\|_{L^p(\mathbb{R}^{n+1})} \\ &\qquad\qquad\qquad \leq \|\mathcal{H} f\|_{L^p(\mathbb{R}^{n+1})} + \Big\| \sup_\lambda |\mathcal{H}_\lambda f| \Big\|_{L^p(\mathbb{R}^{n+1})}. \end{aligned}$$

We now turn to the second term on the right-hand side of (5.1), which we compare to $\mathcal{M}_{\mathrm{Rad}} f(x,t)$, where $\mathcal{M}_{\mathrm{Rad}}$ is the maximal Radon transform along the paraboloid defined in (1.19), which is well-known to be bounded on $L^p(\mathbb{R}^{n+1})$ for $1 < p < \infty$. To make this comparison precise, let $\chi_{A_{2^l}}(y)$ denote the characteristic function of the annulus $2^{l-1} < |y| \leq 2^l$ and $\chi_{B_{2^l}}(y)$ the characteristic function for the ball of radius 2^l. Then

$$\begin{aligned} \sup_\lambda N(\lambda) &\int_{|y|\leq 1/N(\lambda)} |y|^{-n+1} |f|(x-y, t-|y|^2)\, dy \\ &= \sup_\lambda N(\lambda) \sum_{2^l \leq N(\lambda)^{-1}} \int |f|(x-y, t-|y|^2) |y|^{-n+1} \chi_{A_{2^l}}(y)\, dy \\ &\leq \sup_\lambda N(\lambda) \sum_{2^l \leq N(\lambda)^{-1}} 2^{-(l-1)(n-1)} \int |f|(x-y, t-|y|^2) \chi_{B_{2^l}}(y)\, dy \\ &\leq \sup_\lambda N(\lambda) N(\lambda)^{-1} \mathcal{M}_{\mathrm{Rad}} f(x,t) \leq \mathcal{M}_{\mathrm{Rad}} f(x,t). \end{aligned}$$

In total, we may conclude that

$$\Big\| \sup_\lambda |T_\lambda^- f| \Big\|_{L^p(\mathbb{R}^{n+1})} \leq A \, \|f\|_{L^p(\mathbb{R}^{n+1})},$$

for every $1 < p < \infty$, with A depending only on p, n, d and the fixed polynomials $p_2, \ldots, p_d$, as desired.

5.2. Bounding T_λ^+

We next consider T_λ^+, which we recall is defined by

$$T_\lambda^+ f(x,t) = \sum_{2^j \geq 1/N(\lambda)} \int_{\mathbb{R}^n} f(x-y, t-|y|^2)\, e^{iP_\lambda(y)}\, 2^{-nj}\, \phi_j(2^{-j}y)\, dy \tag{5.4}$$

for $\lambda \in \mathbb{R}^{d-1}$ and f of Schwartz class. The key here is to write each term of the sum in the form of the operator ${}^{(\eta)}I_a^\lambda$, for some suitable bump function η, scaling parameter a, and coefficient parameter λ; we will then use Theorem 1.2 to obtain bounds for the L^p norms that are summable in j. To do so, recall the auxiliary operator defined in (1.17) by

$${}^{(\eta)}I_a^\lambda f(x,t) = \int_{\mathbb{R}^n} f(x-y, t-|y|^2)\, e^{iP_\lambda(y/a)}\, \frac{1}{a^n}\, \eta\Big(\frac{y}{a}\Big)\, dy \tag{5.5}$$

for $\lambda \in \mathbb{R}^{d-1}$. We then introduce a non-isotropic scaling

$$2^j \circ \lambda = (2^{jm}\lambda_m)_m,$$

where $\lambda = (\lambda_m)_{2 \leq m \leq d}$. Then $P_{2^j\circ\lambda}(y/2^j) = P_\lambda(y)$ since $P_\lambda(y) = \sum_{m=2}^d \lambda_m\, p_m(y)$ and the polynomials $p_m(y)$ are homogeneous of degree m. We also note that the norm $N(\lambda)$ we introduced earlier is homogeneous with respect to this dilation:

$$N(2^j \circ \lambda) = 2^j N(\lambda). \tag{5.6}$$

Now upon choosing the bump function $\eta = \phi_j$ and the scaling $a = 2^j$, then the j-th summand in (5.4) is precisely ${}^{(\eta)}I_a^{2^j\circ\lambda}$; that is to say,

$$T_\lambda^+ f(x,t) = \sum_{2^j \geq 1/N(\lambda)} {}^{(\phi_j)}I_{2^j}^{2^j\circ\lambda} f(x,t). \tag{5.7}$$

In order to bound the operators ${}^{(\phi_j)}I_{2^j}^{2^j\circ\lambda}$ and hence T_λ^+ on L^p, we will interpolate the key L^2 result of Theorem 1.2 with the following trivial lemma:

Lemma 5.2. *Let $\{\eta_k\}_{k\in\mathbb{Z}}$ be a family of C^1 bump functions with uniform bound $\|\eta_k\|_{C^1} \leq 1$. For any $1 < p \leq \infty$ and any Schwartz function f on $\mathbb{R}^{n+1}$,*

$$\Big\| \sup_{\substack{\lambda\in\mathbb{R}^{d-1}\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I_{2^k}^\lambda f| \Big\|_{L^p(\mathbb{R}^{n+1})} \leq A\, \|f\|_{L^p(\mathbb{R}^{n+1})},$$

with $A = A(p,n)$.

This lemma is simply a trivial consequence of the fact that ${}^{(\eta_k)}I_{2^k}^\lambda f(x,t)$ is majorized pointwise almost everywhere by the maximal Radon transform $\mathcal{M}_{\mathrm{Rad}} f(x,t)$ along the paraboloid. The precise interpolation result we now require is:

Corollary 5.3. *Under the hypotheses of Theorem 1.2, for any $1 < p < \infty$ there exists $\delta = \delta(p) > 0$ such that for all $r \geq 1$ and f of Schwartz class,*

$$\Big\| \sup_{\substack{N(\lambda) \geq r \\ k \in \mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} f(x,t)| \Big\|_{L^p(\mathbb{R}^{n+1})} \leq A\, r^{-\delta} \|f\|_{L^p(\mathbb{R}^{n+1})}, \tag{5.8}$$

where the norm A depends only on p, n, d and the fixed polynomials $p_2, \ldots, p_d$.

The L^2 case of this statement follows from Theorem 1.2 immediately upon recalling that when $N(\lambda) \geq 1$, then $N(\lambda) \leq c\|\lambda\|$ for some constant c dependent only on the dimension and the degree d, so that $\{\lambda : N(\lambda) \geq r\} \subset \{\lambda : c\|\lambda\| \geq r\}$. Thus for any $r \geq 1$,

$$\Big\| \sup_{\substack{N(\lambda) \geq r \\ k \in \mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} f(x,t)| \Big\|_{L^2} \leq \Big\| \sup_{\substack{c\|\lambda\| \geq r \\ k \in \mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} f(x,t)| \Big\|_{L^2}.$$

An application of Theorem 1.2 then shows that

$$\begin{aligned} \Big\| \sup_{\substack{\|\lambda\| \geq r/c \\ k \in \mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} f(x,t)| \Big\|_{L^2} &\leq \sum_{2^s \geq r/c} \Big\| \sup_{\substack{2^s \leq \|\lambda\| < 2^{s+1} \\ k \in \mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} f(x,t)| \Big\|_{L^2} \\ &\leq A \sum_{2^s \geq r/c} 2^{-s\delta} \|f\|_{L^2} \leq A'\, r^{-\delta} \|f\|_{L^2}, \end{aligned} \tag{5.9}$$

which proves (5.8) in the case $p = 2$. Once we have the crucial L^2 case of (5.8), we may interpolate it with the L^p bound (without decay in r) of Lemma 5.2 to conclude that for any $1 < p < \infty$ there exists a positive $\delta(p) > 0$ for which (5.8) holds.

We may now treat the operator T^+_λ. By (5.7),

$$\sup_\lambda |T^+_\lambda f(x,t)| \leq \sup_\lambda \sum_{2^j \geq 1/N(\lambda)} |{}^{(\phi_j)}I^{2^j \circ \lambda}_{2^j} f(x,t)|. \tag{5.10}$$

Now we make the key step that dissociates the coefficients of the phase from the scaling factor of the kernel by noting that we may bound (5.10) by

$$\begin{aligned} \sup_\lambda |T^+_\lambda f(x,t)| &\leq \sup_\lambda \sum_{N(2^j \circ \lambda) \geq 1} \sup_{k \in \mathbb{Z}} |{}^{(\phi_k)}I^{2^j \circ \lambda}_{2^k} f(x,t)| \\ &\leq \sum_{l=0}^{\infty} \sup_{\substack{2^l \leq N(\lambda') \leq 2^{l+1} \\ k \in \mathbb{Z}}} |{}^{(\phi_k)}I^{\lambda'}_{2^k} f(x,t)|. \end{aligned} \tag{5.11}$$

Here we have used the homogeneity relation (5.6). Taking L^p norms for any $1 < p < \infty$, we may then apply Theorem 1.2 in the form of its consequence (5.8) to conclude that

$$\begin{aligned} \Big\| \sup_\lambda |T^+_\lambda f(x,t)| \Big\|_{L^p} &\leq \sum_{l=0}^{\infty} \Big\| \sup_{\substack{N(\lambda') \geq 2^l \\ k \in \mathbb{Z}}} |{}^{(\phi_k)}I^{\lambda'}_{2^k} f(x,t)| \Big\|_{L^p} \\ &\leq A \sum_{l=0}^{\infty} 2^{-\delta l} \|f\|_{L^p} \leq A \|f\|_{L^p}, \end{aligned}$$

thus proving Theorem 1.1.

6. The proof of Theorem 1.2

6.1. A smoother operator

To prove Theorem 1.2, we require a smoother variant of ${}^{(\eta)}I_a^\lambda$, which we denote by ${}^{(\eta)}J_a^\lambda$. Since in treating the operator ${}^{(\eta)}J_a^\lambda$ one can allow more general polynomial phases, we will first state a result for a more general operator denoted ${}^{(\eta)}\tilde{J}_a^\lambda$, and then specialize to the case we need.

Theorem 6.1. *Fix a dimension $n \geq 1$, and a degree $d \geq 2$. Let*

$$Q_\lambda(y) = \sum_{2 \leq |\alpha| \leq d} \lambda_\alpha \, y^\alpha$$

be a real-valued polynomial on $\mathbb{R}^n$ with no linear terms, and write $\lambda = (\lambda_\alpha)_{2 \leq |\alpha| \leq d}$ for its coefficient parameter. Also write Ω_d for the set of all such coefficients λ. For $\lambda \in \Omega_d$, define

$$\|\lambda\| = \sum_{2 \leq \alpha \leq d} |\lambda_\alpha|.$$

Fix a C^1 bump function η supported in $B_1(\mathbb{R}^n)$, and another C^1 bump function ζ supported in $B_1(\mathbb{R})$, with $\|\eta\|_{C^1}, \|\zeta\|_{C^1} \leq 1$. Then for $\lambda \in \Omega_d$, and $a > 0$, one can define an operator $\tilde{J}_a^\lambda$, acting on Schwartz functions on $\mathbb{R}^{n+1}$, by

$$\tilde{J}_a^\lambda f(x,t) = \int_{\mathbb{R}^{n+1}} f(x-y, t-z) \, e^{iQ_\lambda(y/a)} \, \frac{1}{a^n} \, \eta\Big(\frac{y}{a}\Big) \, \frac{1}{a^2} \, \zeta\Big(\frac{z}{a^2}\Big) \, dy \, dz.$$

Furthermore, there exists some fixed $\delta > 0$ such that for any any $r \geq 1$,

$$\Big\| \sup_{\substack{k \in \mathbb{Z}, \lambda \in \Omega_d, \\ r \leq \|\lambda\| < 2r}} |\tilde{J}_{2^k}^\lambda f(x,t)| \Big\|_{L^2(\mathbb{R}^{n+1})} \leq A \, r^{-\delta} \, \|f\|_{L^2(\mathbb{R}^{n+1})},$$

in which the norm A depends on n, d.

In fact, this is effectively Theorem 1 of Stein and Wainger [14]. The original stopping-time argument of Stein and Wainger, together with our Proposition 3.4, is sufficient to treat $\tilde{J}_a^\lambda$ since it does not exhibit Radon-type behavior. However, for precision we note that our statement of Theorem 6.1 differs from Theorem 1 of [14] in that the operator $\tilde{J}_a^\lambda$ includes a product of bump functions $a^{-n}\eta(y/a)a^{-2}\zeta(z/a^2)$ scaled according to parabolic dilations, and the phase function $P_\lambda(y)$ is independent of z. These are merely cosmetic differences, which Stein and Wainger's argument can handle with only minute changes.

In addition, we remark that in Theorem 6.1 we could actually have replaced the bump function η by a one-parameter family of bump functions η_k, as long as the η_k are uniformly C^1 supported in the unit ball. This follows already from Stein–Wainger's original argument. To see this, one need only note that the key bound of Corollary 4.1 in [14] depends only on the C^1 norm of the bump function, which is uniformly bounded by assumption. (For more details, see also Section 8, where the same phenomenon occurs with respect to the operator ${}^{(\eta_k)}I_{2^k}^\lambda$.)

In our application, we only require a consequence of Theorem 6.1 in the special setting of the restricted class of polynomial phases considered in Theorem 1.2. Thus we fix once and for all polynomials $p_2(y), \ldots, p_d(y)$ with $p_2(y) \neq C|y|^2$ and define for $\lambda = (\lambda_2, \ldots, \lambda_d) \in \mathbb{R}^{d-1}$ the polynomial $P_\lambda(y)$ as in (1.13). Define also $\|\lambda\|$ as in (1.16), and fix a C^1 function ζ supported on $[-1,1] \subset \mathbb{R}$, with $\|\zeta\|_{C^1} \leq 1$, and more importantly

$$\int_{\mathbb{R}} \zeta(s) ds = 1.$$

From now on, for any $\lambda \in \mathbb{R}^{d-1}$, $a > 0$, any C^1 bump function η supported in the unit ball, and any Schwartz function f on $\mathbb{R}^{n+1}$, we define

$${}^{(\eta)}J_a^\lambda f(x,t) = \int_{\mathbb{R}^{n+1}} f(x-y, t-z)\, e^{iP_\lambda(y/a)}\, \frac{1}{a^n}\, \eta\Big(\frac{y}{a}\Big)\, \frac{1}{a^2}\, \zeta\Big(\frac{z}{a^2}\Big)\, dy\, dz.$$

We will also fix a one-parameter family of bump functions $\{\eta_k\}_{k\in\mathbb{Z}}$ on $\mathbb{R}^n$ that are all supported in $B_1(\mathbb{R}^n)$ and have C^1 norms uniformly bounded by 1. Then Theorem 6.1 implies the existence of a small $\delta > 0$ such that for any Schwartz function f on $\mathbb{R}^{n+1}$ and any $r \geq 1$,

$$\Big\| \sup_{\substack{k\in\mathbb{Z}, \lambda\in\mathbb{R}^{d-1}, \\ r \leq \|\lambda\| < 2r}} |{}^{(\eta_k)}J_{2^k}^\lambda f(x,t)| \Big\|_{L^2(\mathbb{R}^{n+1})} \leq A\, r^{-\delta}\, \|f\|_{L^2(\mathbb{R}^{n+1})}. \tag{6.1}$$

Here in fact the strength of Theorem 6.1 allows the norm to depend only on n, d, but in general we will also allow norms to depend on the set of polynomials $p_2, \ldots, p_d$, which are fixed once and for all.

In order to compare ${}^{(\eta)}J_a^\lambda$ to ${}^{(\eta)}I_a^\lambda$, it is helpful to display the singular support of the operator ${}^{(\eta)}I_a^\lambda$ more explicitly by rewriting the operator in the form

$${}^{(\eta)}I_a^\lambda f(x,t) = \int_{\mathbb{R}^n} \int_{\mathbb{R}} f(x-y, t-z)\, e^{iP_\lambda(y/a)}\, a^{-n}\, \eta_a\Big(\frac{y}{a}\Big)\, \delta_{z=|y|^2}\, dy\, dz,$$

where δ is the Dirac delta function. Temporarily, let I_a^λ and J_a^λ denote the kernels of the respective operators; then the key property of this construction is that for every fixed $y \in \mathbb{R}^n$, every $\lambda \in \mathbb{R}^{d-1}$ and every $a > 0$,

$$\int_{\mathbb{R}} (I_a^\lambda(y,z) - J_a^\lambda(y,z))\, dz = 0. \tag{6.2}$$

We will exploit this property later.

6.2. An approximation argument

In order to define the relevant square function we will use to pass from the singular operator I_a^λ to the smoother operator J_a^λ, it will be convenient to introduce the Littlewood–Paley decomposition constructed in Section 4. For any $N \geq 1$, and any Schwartz functions f, we let

$$L_N f = \sum_{j=-N}^{N} \Delta_j \tilde{\Delta}_j f.$$

The function $L_N f$ is then in the Schwartz class, by our observation in Section 4. In order to avoid convergence issues, we will use the finite sum $L_N f$ in order to approximate f in $L^2(\mathbb{R}^{n+1})$; all bounds will be independent of N, and morally speaking readers may regard $L_N f$ as representing f.

In order to prove Theorem 1.2, from this point on we fix a sequence $\{\eta_k\}_{k\in\mathbb{Z}}$ of C^1 bump functions with $\|\eta_k\|_{C^1} \leq 1$. We observe that trivially, for any Schwartz function f and any fixed N,

$$\sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} f| \leq \sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} L_N f| + \sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} (f - L_N f)|. \tag{6.3}$$

(From now on, we write $\|\lambda\| \approx r$ as a shorthand for $r \leq \|\lambda\| \leq 2r$ when $\lambda \in \mathbb{R}^{d-1}$.) We bound the second term on the right-hand side of (6.3) using the following trivial proposition.

Proposition 6.2. *For any Schwartz function g on $\mathbb{R}^{n+1}$,*

$$\Big\| \sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} g| \Big\|_{L^2} \leq A\, \|g\|_{L^2}.$$

This follows immediately from the pointwise estimate

$$\sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} g| \leq A\, \mathcal{M}_{\mathrm{Rad}}(g).$$

Note that Proposition 6.2 is extremely weak since it lacks the decay factor $r^{-\delta}$, but it is acceptable when applied to the Schwartz class function $f - L_N f$, which may be taken to have arbitrarily small L^2 norm, since $L_N f$ converges to f in L^2 norm, by (4.4).

As a result, in order to prove Theorem 1.2, it now suffices to prove that there exists a $\delta > 0$ such that

$$\Big\| \sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} L_N f| \Big\|_{L^2} \leq A\, r^{-\delta}\, \|f\|_{L^2},$$

where all constants are independent of N.

6.3. Introduction of the square function

Now we rigorously define the square function $S_r(f)$ acting on a Schwartz function f on $\mathbb{R}^{n+1}$ by

$$S_r(f) = \Big(\sum_{k\in\mathbb{Z}} \Big(\sup_{\|\lambda\|\approx r} |({}^{(\eta_k)}I^{\lambda}_{2^k} - {}^{(\eta_k)}J^{\lambda}_{2^k}) L_N f| \Big)^2 \Big)^{1/2}. \tag{6.4}$$

Then

$$\sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}I^{\lambda}_{2^k} L_N f| \leq \sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}J^{\lambda}_{2^k} L_N f| + S_r(f). \tag{6.5}$$

We bound the first term on the right-hand side of (6.5) directly by applying Theorem 6.1 and the inequality $\|L_N(f)\|_{L^2} \leq C\,\|f\|_{L^2}$ provided by (4.6), to conclude that

$$\Big\| \sup_{\substack{\|\lambda\|\approx r\\ k\in\mathbb{Z}}} |{}^{(\eta_k)}J^{\lambda}_{2^k} L_N f| \Big\|_{L^2} \leq C\,r^{-\delta}\,\|f\|_{L^2}.$$

The main result for the square function $S_r(f)$ is:

Theorem 6.3. *Let $\{\eta_k\}_k$ be a one-parameter family of bump functions supported in the unit ball with C^1 norm uniformly bounded by 1. Then there exists some fixed $\delta > 0$ such that for any Schwartz function f on $\mathbb{R}^{n+1}$ and any $r \geq 1$,*

$$\|S_r(f)\|_{L^2(\mathbb{R}^{n+1})} \leq A\,r^{-\delta}\,\|f\|_{L^2(\mathbb{R}^{n+1})},$$

uniformly in N.

We will derive this as a consequence of the following key inequality.

Theorem 6.4. *Let $\{\eta_k\}_k$ be as in Theorem 6.3. Then there exist positive constants $\varepsilon_0 > 0$ and $\delta_0 > 0$ such that for any Schwartz function F on $\mathbb{R}^{n+1}$, any $j, k \in \mathbb{Z}$, and any $r \geq 1$,*

$$\Big\| \sup_{\|\lambda\|\approx r} |({}^{(\eta_k)}I^{\lambda}_{2^k} - {}^{(\eta_k)}J^{\lambda}_{2^k})\Delta_j F| \Big\|_{L^2} \leq A\,r^{-\delta_0}\,2^{-\varepsilon_0|j-k|}\,\|F\|_{L^2}.$$

In both Theorems 6.3 and 6.4, and throughout the computations leading to these theorems, the norms may depend on n, d and the fixed polynomials $p_2, \ldots, p_d$.

That Theorem 6.4 is sufficient to prove the boundedness on L^2 of the square function $S_r(f)$ may be seen as follows. Temporarily define $F_j = \tilde{\Delta}_j f$ and

$$A_{j,k}(F) = \sup_{\|\lambda\|\approx r} |({}^{(\eta_k)}I^{\lambda}_{2^k} - {}^{(\eta_k)}J^{\lambda}_{2^k})\,\Delta_j F|.$$

Fix an $\varepsilon > 0$ that satisfies $\varepsilon < \varepsilon_0$, where ε_0 is the constant given in Theorem 6.4. Then

$$\begin{aligned}
S_r(f) &= \Big(\sum_{k\in\mathbb{Z}} \Big(\sup_{\|\lambda\|\approx r} |({}^{(\eta_k)}I^{\lambda}_{2^k} - {}^{(\eta_k)}J^{\lambda}_{2^k}) \sum_{|j|\leq N} \Delta_j F_j|\Big)^2\Big)^{1/2}\\
&\leq \Big(\sum_{k\in\mathbb{Z}} \Big(\sum_{|j|\leq N} \sup_{\|\lambda\|\approx r} |({}^{(\eta_k)}I^{\lambda}_{2^k} - {}^{(\eta_k)}J^{\lambda}_{2^k})\,\Delta_j F_j|\Big)^2\Big)^{1/2}\\
&= \Big(\sum_{k\in\mathbb{Z}} \Big(\sum_{|j|\leq N} A_{j,k}(F_j)\Big)^2\Big)^{1/2}\\
&= \Big(\sum_{k\in\mathbb{Z}} \Big(\sum_{|j|\leq N} 2^{-\varepsilon|j-k|}\,2^{\varepsilon|j-k|}\,A_{j,k}(F_j)\Big)^2\Big)^{1/2}\\
&\leq \Big(\sum_{k\in\mathbb{Z}} \Big(\sum_{j\in\mathbb{Z}} 2^{-2\varepsilon|j-k|}\Big)\Big(\sum_{|j|\leq N} 2^{2\varepsilon|j-k|}(A_{j,k}(F_j))^2\Big)\Big)^{1/2}\\
&\leq C\Big(\sum_{k\in\mathbb{Z}} \sum_{|j|\leq N} 2^{2\varepsilon|j-k|}\big(\sup_{\|\lambda\|\approx r} |({}^{(\eta_k)}I^{\lambda}_{2^k} - {}^{(\eta_k)}J^{\lambda}_{2^k})\,\Delta_j F_j|\big)^2\Big)^{1/2}.
\end{aligned}$$

Thus, taking L^2 norms and applying Theorem 6.4, we have

$$\|S_r(f)\|_{L^2}^2 \le C r^{-2\delta_0} \sum_{k\in\mathbb{Z}} \sum_{|j|\le N} 2^{2\varepsilon|j-k|}\, 2^{-2\varepsilon_0|j-k|}\, \|F_j\|_{L^2}^2.$$

Now using the fact that $\varepsilon < \varepsilon_0$, we may sum first in k (to obtain a constant coefficient) and then in j, using the property (4.5) to obtain

$$\|S_r(f)\|_{L^2}^2 \le C r^{-2\delta_0} \|f\|_{L^2}^2,$$

independent of N, thus proving Theorem 6.3 and hence Theorem 1.2. The remainder of the paper focuses on proving Theorem 6.4.

7. Proof of Theorem 6.4 for the difference operator $I^\lambda_{2^k} - J^\lambda_{2^k}$

7.1. Division into Cases 1 and 2

We recall that Theorem 6.4 claims that there exists a positive constant $\varepsilon_0 > 0$ and a constant $\delta_0 > 0$ such that for any Schwartz function F on $\mathbb{R}^{n+1}$, any $j, k \in \mathbb{Z}$, and any $r \ge 1$,

$$\Big\| \sup_{\|\lambda\|\approx r} |({}^{(\eta_k)}I^\lambda_{2^k} - {}^{(\eta_k)}J^\lambda_{2^k})\Delta_j F| \Big\|_{L^2} \le A\, r^{-\delta_0}\, 2^{-\varepsilon_0|j-k|}\, \|F\|_{L^2}. \tag{7.1}$$

Note that now that k is fixed, the bump function η_k is fixed, hence we will call the bump function η and omit the superscript η_k on the operators for simplicity.

We will divide the proof of (7.1) into two cases: Case 1, when $j \ge k$, and Case 2, when $j < k$. The main propositions in these cases are the following.

Proposition 7.1 (Case 1: $j \ge k$)**.** *There exists a positive constant $\delta_0 > 0$ such that for any Schwartz function F on $\mathbb{R}^{n+1}$, $r \ge 1$ and $j, k \in \mathbb{Z}$ with $j \ge k$,*

$$\Big\| \sup_{\|\lambda\|\approx r} |(I^\lambda_{2^k} - J^\lambda_{2^k})\Delta_j F| \Big\|_{L^2} \le A\, r^{-\delta_0}\, 2^{-(j-k)}\, \|F\|_{L^2}.$$

Proposition 7.2 (Case 2: $j < k$)**.** *There exists a positive constant $\delta_0 > 0$ and a positive constant $\varepsilon_0 > 0$ such that for any Schwartz function F on $\mathbb{R}^{n+1}$, $r \ge 1$ and $j, k \in \mathbb{Z}$ with $j < k$,*

$$\Big\| \sup_{\|\lambda\|\approx r} |I^\lambda_{2^k} \Delta_j F| \Big\|_{L^2} \le A\, r^{-\delta_0}\, 2^{\varepsilon_0(j-k)}\, \|F\|_{L^2} \tag{7.2}$$

$$\Big\| \sup_{\|\lambda\|\approx r} |J^\lambda_{2^k} \Delta_j F| \Big\|_{L^2} \le A\, r^{-\delta_0}\, 2^{\varepsilon_0(j-k)}\, \|F\|_{L^2}. \tag{7.3}$$

Combining Propositions 7.1 and 7.2 immediately proves Theorem 6.4. Note that in each proposition there are two types of decay present: the first is decay in r, which indicates the size of the coefficient parameter λ controlling the phase, and the second is with respect to $|j-k|$. In order to extract decay in r we will apply a TT^* argument with stopping-times. In order to extract decay with respect to $|j-k|$, we will use either the cancellation property (6.2) of the kernels of $I^\lambda_{2^k}$ and $J^\lambda_{2^k}$, or the cancellation property (4.1) of the Δ_j kernel in order to perform an integration by parts and pull out the desired factor.

Schematically, we proceed as follows. To prove Proposition 7.1 for Case 1 ($j \geq k$), we will use the cancellation property (6.2) for $I^\lambda_{2^k} - J^\lambda_{2^k}$ in order to place a derivative on the Δ_j factor, thus enabling us to pull out a factor of 2^{-j}, at the cost of a factor of 2^k. We will simultaneously perform a TT^* argument in order to extract the decay factor $r^{-\delta}$.

To prove Proposition 7.2 for Case 2 ($j < k$), we no longer need to exploit the cancellation property (6.2), and therefore it is convenient to separate the terms via the trivial upper bound

$$\Big\| \sup_{\|\lambda\| \approx r} |(I^\lambda_{2^k} - J^\lambda_{2^k})\Delta_j F| \Big\|_{L^2} \leq \Big\| \sup_{\|\lambda\| \approx r} |I^\lambda_{2^k} \Delta_j F| \Big\|_{L^2} + \Big\| \sup_{\|\lambda\| \approx r} |J^\lambda_{2^k} \Delta_j F| \Big\|_{L^2},$$

and treat the two terms on the right-hand side separately. We will bound each of these terms by interpolation between two bounds of quite different flavors. On the one hand we will prove the following.

Proposition 7.3 (Case 2: $j < k$)**.** *There exists a positive constant $\delta_0 > 0$ such that for any Schwartz function F on $\mathbb{R}^{n+1}$, $r \geq 1$ and $j, k \in \mathbb{Z}$ with $j < k$,*

$$\Big\| \sup_{\|\lambda\| \approx r} |I^\lambda_{2^k} \Delta_j F| \Big\|_{L^2} \leq A\, r^{-\delta_0} \, \|F\|_{L^2} \tag{7.4}$$

$$\Big\| \sup_{\|\lambda\| \approx r} |J^\lambda_{2^k} \Delta_j F| \Big\|_{L^2} \leq A\, r^{-\delta_0} \, \|F\|_{L^2}. \tag{7.5}$$

This we prove via TT^* arguments; we are allowed to focus purely on extracting decay from the oscillation of the exponential factor in the kernel, since we do not require any decay with respect to j, k. On the other hand, we will prove that:

Proposition 7.4 (Case 2: $j < k$)**.** *For any Schwartz function F on $\mathbb{R}^{n+1}$, $r \geq 1$ and $j, k \in \mathbb{Z}$ with $j < k$,*

$$\Big\| \sup_{\|\lambda\| \approx r} |I^\lambda_{2^k} \Delta_j F| \Big\|_{L^2} \leq A\, r^{1/2}\, 2^{(j-k)} \, \|F\|_{L^2} \tag{7.6}$$

$$\Big\| \sup_{\|\lambda\| \approx r} |J^\lambda_{2^k} \Delta_j F| \Big\|_{L^2} \leq A\, 2^{(j-k)} \, \|F\|_{L^2}. \tag{7.7}$$

Note that in Proposition 7.4 we are even willing to lose by a factor of $r^{1/2}$, as long as we extract a decay factor $2^{(j-k)}$. Intuitively, we use the cancellation property (4.1) that $\int_{\mathbb{R}} \Delta_j(\theta) d\theta = 0$ in order to write Δ_j as a derivative, and integrate by parts to place a derivative on the kernel of $I^\lambda_{2^k}$ or $J^\lambda_{2^k}$. In the case of $J^\lambda_{2^k}$, it is then a straightforward matter to extract a factor of 2^{j-k}. But in the case of $I^\lambda_{2^k}$, the situation is more complicated since the kernel of $I^\lambda_{2^k}$ is supported on the paraboloid and is not differentiable in all directions. However, we are able to proceed with a version of this argument by adapting a TT^* argument, which allows us to differentiate in a way that contributes only an allowable singularity. We then lose by a factor of r in the bound for $\|TT^*\|_{L^2}$, which comes from differentiating the oscillatory factor $e^{iP_\lambda(y)}$. (We reiterate that in this case the TT^* formulation aids in differentiating, but is not required in order to extract cancellation from the oscillatory phase.)

Taking the geometric mean of the bounds in Propositions 7.3 and 7.4, namely

$$\Big\| \sup_{\|\lambda\|\approx r} |I_{2^k}^{\lambda} \Delta_j F| \Big\|_{L^2} \leq (r^{-\delta_0})^{\theta} \, (r^{1/2} \, 2^{(j-k)})^{1-\theta} \, \|F\|_{L^2},$$

with a choice of θ sufficiently close to 1 (such that $\delta_0\theta > (1-\theta)/2$), finally yields the desired result of Proposition 7.2.

7.2. Derivation of the generic kernel

We now set the scene for proving Propositions 7.1, 7.3 and 7.4. First, in Proposition 7.1, fix $j \geq k$ and temporarily set $a = 2^k$ (for notational convenience). Let $\lambda\colon (x,t) \mapsto (\lambda_2, \dots, \lambda_d)$ be any measurable stopping-time function from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{d-1}$. We define an operator T acting on Schwartz functions f by setting

$$Tf(x,t) = (I_a^{\lambda(x,t)} - J_a^{\lambda(x,t)})\Delta_j f(x,t). \tag{7.8}$$

Now we apply the method of TT^*. Explicitly, $Tf(x,t)$ may be expressed as

$$\iint_{\mathbb{R}^{n+2}} f(x-y, t-s-\theta) e^{iP_{\lambda(x,t)}(\frac{y}{a})} \frac{1}{a^n} \eta(\frac{y}{a}) \Big(\delta_{s=|y|^2} - \frac{1}{a^2}\zeta\Big(\frac{s}{a^2}\Big)\Big)\Delta_j(\theta)\, dy\, ds\, d\theta.$$

Formally changing variables $\theta \mapsto \theta - s$, we may then express $Tf(x,t)$ as

$$\iint_{\mathbb{R}^{n+2}} f(x-y, t-\theta) e^{iP_{\lambda(x,t)}(\frac{y}{a})} \frac{1}{a^n}\, \eta\Big(\frac{y}{a}\Big) \Big(\delta_{s=|y|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{s}{a^2}\Big)\Big)\Delta_j(\theta - s)\, dy\, ds\, d\theta.$$

Hence the operator T^* acts on functions f of Schwartz class by

$$\begin{aligned} T^*f(x,t) = \iint_{\mathbb{R}^{n+2}} & f(x+z, t+\omega)\, e^{-iP_{\lambda(x+z,t+\omega)}(z/a)}\, \frac{1}{a^n}\, \eta\Big(\frac{z}{a}\Big) \\ & \cdot \Big(\delta_{\xi=|z|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{\xi}{a^2}\Big)\Big)\, \Delta_j(\omega - \xi)\, dz\, d\xi\, d\omega. \end{aligned}$$

It follows that

$$\begin{aligned} TT^*f(x,t) = \iint_{\mathbb{R}^{2n+4}} & f(x-y+z, t-\theta+\omega)\, e^{iP_{\lambda(x,t)}(\frac{y}{a}) - iP_{\lambda(x-y+z,t-\theta+\omega)}(\frac{z}{a})} \\ & \cdot \frac{1}{a^n}\,\eta\Big(\frac{y}{a}\Big)\, \frac{1}{a^n}\,\eta\Big(\frac{z}{a}\Big)\Big(\delta_{s=|y|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{s}{a^2}\Big)\Big)\Big(\delta_{\xi=|z|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{\xi}{a^2}\Big)\Big) \\ & \cdot \Delta_j(\theta - s)\Delta_j(\omega - \xi)\, dy\, ds\, dz\, d\xi\, d\theta\, d\omega. \end{aligned}$$

Changing variables by setting $u = y - z$ and letting $\theta \mapsto \theta + \omega$, this becomes

$$\begin{aligned} TT^*f(x,t) = \iint_{\mathbb{R}^{2n+4}} & f(x-u, t-\theta)\, e^{iP_{\lambda(x,t)}(\frac{u+z}{a}) - iP_{\lambda(x-u,t-\theta)}(\frac{z}{a})}\, \frac{1}{a^n} \\ & \cdot \eta\Big(\frac{u+z}{a}\Big)\, \frac{1}{a^n}\,\eta\Big(\frac{z}{a}\Big)\Big(\delta_{s=|u+z|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{s}{a^2}\Big)\Big)\Big(\delta_{\xi=|z|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{\xi}{a^2}\Big)\Big) \\ & \cdot \Delta_j(\theta - s + \omega)\Delta_j(\omega - \xi)\, du\, ds\, dz\, d\xi\, d\theta\, d\omega. \end{aligned}$$

The integral in ω we recall defines the kernel $\underline{\Delta}_j$ constructed in Section 4.1; note that

$$\underline{\Delta}_j(\theta - s + \xi) := \int_{\mathbb{R}} \Delta_j(\theta - s + \omega)\Delta_j(\omega - \xi)\, d\omega.$$

Then

$$\begin{aligned} TT^* f(x,t) = \iint_{\mathbb{R}^{2n+3}} & f(x-u, t-\theta)\, e^{iP_{\lambda(x,t)}(\frac{u+z}{a}) - iP_{\lambda(x-u,t-\theta)}(\frac{z}{a})} \\ & \cdot \frac{1}{a^n}\,\eta\Big(\frac{u+z}{a}\Big)\,\frac{1}{a^n}\,\eta\Big(\frac{z}{a}\Big)\Big(\delta_{s=|u+z|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{s}{a^2}\Big)\Big)\Big(\delta_{\xi=|z|^2} - \frac{1}{a^2}\,\zeta\Big(\frac{\xi}{a^2}\Big)\Big) \\ & \cdot \underline{\Delta}_j(\theta - s + \xi)\, du\, ds\, dz\, d\xi\, d\theta. \end{aligned} \tag{7.9}$$

Now recalling that $a = 2^k$, we see that $\underline{\Delta}_j(\omega) = \frac{1}{a^2}\underline{\Delta}_{j-k}\left(\frac{\omega}{a^2}\right)$ by (4.7). So we have

$$TT^* f(x,t) = \iint_{\mathbb{R}^{n+1}} f(x-u, t-\theta)\, \frac{1}{a^{n+2}}\, \mathcal{K}^{\lambda(x,t),\lambda(x-u,t-\theta)}\Big(\frac{u}{a}, \frac{\theta}{a^2}\Big)\, du\, d\theta, \tag{7.10}$$

where for each ν, $\mu \in \mathbb{R}^{d-1}$ we define the kernel

$$\begin{aligned} \mathcal{K}^{\nu,\mu}(u,\theta) = \iint_{\mathbb{R}^{n+2}} & e^{iP_\nu(u+z) - iP_\mu(z)}\eta(u+z)\eta(z) \\ & \cdot\ \big(\delta_{s=|u+z|^2} - \zeta(s)\big)\,\big(\delta_{\xi=|z|^2} - \zeta(\xi)\big)\,\underline{\Delta}_{j-k}(\theta - s + \xi)\, dz\, d\xi\, ds. \end{aligned} \tag{7.11}$$

Estimating this kernel is now the main focus of proving Proposition 7.1.

Next, in order to set the stage for proving Propositions 7.3 and 7.4, we fix $a = 2^k$ and fix a measurable stopping-time function $\lambda(x,t)\colon \mathbb{R}^{n+1} \to \mathbb{R}^{d-1}$ and define two linear operators T_1 and T_2 respectively by

$$\begin{aligned} T_1 f(x,t) &= I_a^{\lambda(x,t)} \Delta_j f(x,t) \\ T_2 f(x,t) &= J_a^{\lambda(x,t)} \Delta_j f(x,t). \end{aligned}$$

Then for each of $i = 1, 2$ we may compute as above that

$$T_i T_i^* f(x,t) = \iint_{\mathbb{R}^{n+1}} f(x-u, t-\theta)\, \frac{1}{a^{n+2}}\ {}^{(i)}\mathcal{K}^{\lambda(x,t),\lambda(x-u,t-\theta)}\Big(\frac{u}{a}, \frac{\theta}{a^2}\Big)\, du\, d\theta, \tag{7.12}$$

where for ν, $\mu \in \mathbb{R}^{d-1}$ the respective kernels are given by

$$\begin{aligned} {}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta) = \iint_{\mathbb{R}^{n+2}} & e^{iP_\nu(u+z) - iP_\mu(z)}\, \eta(u+z)\, \eta(z)\, \delta_{s=|u+z|^2}\, \delta_{\xi=|z|^2} \\ & \cdot\ \underline{\Delta}_{j-k}(\theta - s + \xi)\, dz\, d\xi\, ds \end{aligned} \tag{7.13}$$

and

$$\begin{aligned} {}^{(2)}\mathcal{K}^{\nu,\mu}(u,\theta) = \iint_{\mathbb{R}^{n+2}} & e^{iP_\nu(u+z) - iP_\mu(z)}\eta(u+z)\eta(z)\, \zeta(s)\, \zeta(\xi) \\ & \cdot\ \underline{\Delta}_{j-k}(\theta - s + \xi)\, dz\, d\xi\, ds. \end{aligned} \tag{7.14}$$

One then needs to estimate these kernels to complete the proofs of Propositions 7.3 and 7.4; we return to this in Section 7.4.

7.3. Proof of Proposition 7.1 (Case $j \geq k$)

To prove Proposition 7.1, it is sufficient to prove that if $a = 2^k$ and

$$\lambda = (\lambda_2, \dots, \lambda_d)\colon \mathbb{R}^{n+1} \to \mathbb{R}^{d-1}$$

is a measurable stopping-time satisfying $\|\lambda(x,t)\| \approx r$, then for Tf defined as in (7.8),

$$\|TT^* f\|_{L^2} \leq A^2\, r^{-2\delta_0}\, 2^{-2\varepsilon_0 |j-k|}\, \|f\|_{L^2}. \tag{7.15}$$

We split the kernel (7.11) of TT^* into $\mathbf{I} + \mathbf{II}$, where $\mathbf{I}$ gives the contribution of $\delta_{s=|u+z|^2}$, and $\mathbf{II}$ gives the contribution of $\zeta(s)$.

7.3.1. The term I ($n = 2$). Explicitly, we evaluate the delta function in $\mathbf{I}$ and express $\mathbf{I}$ as the integral

$$\iint_{\mathbb{R}^{n+1}} e^{iP_\nu(u+z) - iP_\mu(z)} \eta(u+z)\eta(z) \left(\delta_{\xi=|z|^2} - \zeta(\xi)\right) \underline{\Delta}_{j-k}(\theta - |u+z|^2 + \xi)\, dz\, d\xi.$$

We split this integral into two terms, one coming from $\delta_{\xi=|z|^2}$, and the other coming from $\zeta(\xi)$. In the first term we evaluate the delta function to replace ξ by $|z|^2$, and then reintroduce a trivial integration in ξ via the property $1 = \int \zeta(\xi + |z|^2)\, d\xi$. In the second term we change variables: $\xi \mapsto \xi + |z|^2$. Grouping terms, we get

$$\begin{aligned} \mathbf{I} = \iint_{\mathbb{R}^{n+1}} & e^{iP_\nu(u+z) - iP_\mu(z)}\, \eta(u+z)\, \eta(z)\, \zeta(\xi + |z|^2) \\ & \cdot \left(\underline{\Delta}_{j-k}(\theta - |u+z|^2 + |z|^2) - \underline{\Delta}_{j-k}(\theta - |u+z|^2 + |z|^2 + \xi)\right)\, dz\, d\xi. \end{aligned} \tag{7.16}$$

This is now the first point where we restrict our attention temporarily to the $(2+1)$-dimensional case, that is the $n = 2$ case, in order to present the main ideas to the reader without unnecessary technical complications. We will return to analyze this term in general dimension $n \geq 2$ in Section 10.1. Meanwhile, we will continue to write $\mathbb{R}^n$ with the understanding that $n = 2$, in order to aid the reader conceptually.

We would like to isolate an oscillatory integral within (7.16) that is independent of the Littlewood–Paley factors $\underline{\Delta}_{j-k}$. To this end, we note that $|u+z|^2 - |z|^2 = |u|^2 + 2u \cdot z$. This motivates defining a new variable $\tau = u \cdot z / |u|$. In the case $n = 2$, we would thus make the change of variables $z \mapsto (\tau, \sigma)$, where

$$\tau = \frac{u \cdot z}{|u|}, \quad \sigma = \frac{u_2 z_1 - u_1 z_2}{|u|}. \tag{7.17}$$

Note that this change of variables is simply a rotation, and hence has unit Jacobian. In particular, since the support of η restricts to $|z| \leq 1$, we also have $|\tau| \leq 1$, $|\sigma| \leq 1$. (In the case $n \geq 2$ we will need to make a more general change of variables, and we postpone this discussion until Section 9.) After this change of variables, we have

$$|\mathbf{I}| \le \int_{\mathbb{R}^2} |K_\sharp^{\nu,\mu}(u,\tau;\xi)|\chi_{B_1}(\tau) \cdot \left|\underline{\Delta}_{j-k}(\theta-|u|^2-2|u|\tau)-\underline{\Delta}_{j-k}(\theta-|u|^2-2|u|\tau+\xi)\right| d\tau\, d\xi$$

where

$$K_\sharp^{\nu,\mu}(u,\tau;\xi)=\int_{\mathbb{R}^{n-1}} e^{iP_\nu(u+z)-iP_\mu(z)}\,\eta(u+z)\,\eta(z)\,\zeta(\xi+|z|^2)\,d\sigma;$$

here z is defined implicitly by u,τ,σ. From the supports of η and ζ, we see that $K_\sharp^{\nu,\mu}(u,\tau;\xi)$ has ξ support in $[-2,2]$. Thus it follows from the mean-value theorem, as recorded in Lemma 4.4, that in absolute value $\mathbf{I}$ is dominated by

$$C\,2^{-2(j-k)}\int_{\mathbb{R}^2}|K_\sharp^{\nu,\mu}(u,\tau;\xi)|\chi_{B_2}(\xi)2^{-2(j-k)}|\xi|\psi_{j-k}(\theta-|u|^2-2|u|\tau)\chi_{B_1}(\tau)\,d\tau\,d\xi,$$

where we recall from (4.9) that $\psi_{j-k}(t)$ is a non-negative integrable function on $\mathbb{R}^1$ with L^1 norm uniformly bounded, independent of $j-k$. We also note that due to the support of η, $K_\sharp^{\nu,\mu}$ naturally has u support inside $B_2(\mathbb{R}^n)$.

We now apply Proposition 3.5 (which also specifies $n=2$) to bound $K_\sharp^{\nu,\mu}$ and conclude that there exists $\delta>0$ and a small set $G^\nu\subset B_2(\mathbb{R}^n)$ with $|G^\nu|\le C r^{-\delta}$, and for each $u\in B_2(\mathbb{R}^n)$ a small set $F_u^\nu\subset B_1(\mathbb{R})$ with $|F_u^\nu|\le C r^{-\delta}$, such that

$$|K_\sharp^{\nu,\mu}(u,\tau;\xi)|\le C\left[r^{-\delta}\chi_{B_2}(u)\chi_{B_1}(\tau)+\chi_{G^\nu}(u)\chi_{B_1}(\tau)+\chi_{B_2}(u)\chi_{F_u^\nu}(\tau)\right]. \tag{7.18}$$

Moreover, these estimates are uniform in ξ, as the small sets do not depend on ξ, and neither do the bounds. Hence after applying this in $\mathbf{I}$ and integrating trivially in ξ we obtain

$$|\mathbf{I}|\le C\,2^{-2(j-k)}\int_{\mathbb{R}}\left(r^{-\delta}\chi_{B_2}(u)\chi_{B_1}(\tau)+\chi_{G^\nu}(u)\chi_{B_1}(\tau)+\chi_{B_2}(u)\chi_{F_u^\nu}(\tau)\right) \cdot \psi_{j-k}(\theta-|u|^2-2|u|\tau)\,d\tau. \tag{7.19}$$

We now plug this into (7.10), recalling that $\nu=\lambda(x,t)$, and see that the contribution of this part of the kernel to $TT^*f(x,t)$ is thus bounded by

$$2^{-2(j-k)}\iint_{\mathbb{R}^{n+1}}|f|(x-u,t-\theta)\int_{\mathbb{R}}\frac{1}{a^n}\Big(r^{-\delta}\chi_{B_2}\left(\frac{u}{a}\right)\chi_{B_1}(\tau) + \chi_{G^{\lambda(x,t)}}\left(\frac{u}{a}\right)\chi_{B_1}(\tau)+\chi_{B_2}\left(\frac{u}{a}\right)\chi_{F^{\lambda(x,t)}_{(\frac{u}{a})}}(\tau)\Big) \cdot\frac{1}{a^2}\psi_{j-k}\left(\frac{\theta-|u|^2-2a|u|\tau}{a^2}\right)d\tau\,du\,d\theta. \tag{7.20}$$

We now recognize this as a (non-maximal) averaging operator, with a fixed normalization $a=2^k$, to which we will apply the following simple lemma.

Lemma 7.5. *Let $\chi(x,y)$ be an integrable function on $\mathbb{R}^m \times \mathbb{R}^m$ such that there are constants C_0 and λ for which*

$$\Big\| \sup_{x\in\mathbb{R}^m} |\chi(x,y)| \Big\|_{L^1(\mathbb{R}^m(dy))} \leq C_0 \tag{7.21}$$

and

$$\sup_{x\in\mathbb{R}^m} \|\chi(x,y)\|_{L^1(\mathbb{R}^m(dy))} \leq \lambda. \tag{7.22}$$

Then for any $1 \leq p \leq \infty$ there exists a constant $C = C(C_0, p)$ such that for any $f \in L^p(\mathbb{R}^m)$,

$$\Big\| \int_{\mathbb{R}^m} f(x-y)\,\chi(x,y)\,dy \Big\|_{L^p(\mathbb{R}^m(dx))} \leq C\,\lambda^{1/p'}\,\|f\|_{L^p(\mathbb{R}^m)}, \tag{7.23}$$

where p' is the conjugate exponent to p.

We delay the proof momentarily, and apply the lemma to (7.20) with the function $\chi((x,t),(u,\theta))$ defined to be

$$\int_{\mathbb{R}} \frac{1}{a^n}\Big(r^{-\delta}\,\chi_{B_2}\Big(\frac{u}{a}\Big)\,\chi_{B_1}(\tau) + \chi_{G^{\lambda(x,t)}}\Big(\frac{u}{a}\Big)\,\chi_{B_1}(\tau) + \chi_{B_2}\Big(\frac{u}{a}\Big)\,\chi_{F^{\lambda(x,t)}_{(\frac{u}{a})}}(\tau)\Big) \cdot \frac{1}{a^2}\,\psi_{j-k}\Big(\frac{\theta-|u|^2-2a|u|\tau}{a^2}\Big)\,d\tau.$$

To verify condition (7.21), we note that since for all choices of (x,t) we have the small sets $G^{\lambda(x,t)} \subset B_2(\mathbb{R}^n)$ and $F^{\lambda(x,t)}_{(u/a)} \subset B_1(\mathbb{R})$, then

$$\sup_{(x,t)\in\mathbb{R}^{n+1}} |\chi((x,t),(u,\theta))| \leq \int_{\mathbb{R}} \frac{1}{a^n}\,\chi_{B_2}\Big(\frac{u}{a}\Big)\,\chi_{B_1}(\tau)\frac{1}{a^2}\,\psi_{j-k}\Big(\frac{\theta-|u|^2-2a|u|\tau}{a^2}\Big)\,d\tau.$$

Taking the $L^1(du\,d\theta)$ norm of both sides by integrating first in θ, we obtain

$$\Big\| \sup_{(x,t)\in\mathbb{R}^{n+1}} |\chi((x,t),(u,\theta))| \Big\|_{L^1(du\,d\theta)} \leq C,$$

independent of j,k.

To verify (7.22), for each fixed $(x,t) \in \mathbb{R}^n$ we compute the $L^1(du\,d\theta)$ norm of $\chi((x,t),(u,\theta))$ by integrating successively in θ,τ, and u, in that order. Then by the small measure of the sets $G^{\lambda(x,t)}$ and $F^{\lambda(x,t)}_{(u/a)}$ (and in the first term, the factor $r^{-\delta}$ out front), we have

$$\sup_{(x,t)\in\mathbb{R}^{n+1}} \|\chi((x,t),(u,\theta))\|_{L^1(du\,d\theta)} \leq C\,r^{-\delta},$$

independent of j,k. Thus by Lemma 7.5 we may conclude that the L^2 norm of the portion of TT^*f considered in (7.20) is bounded above by

$$A\,2^{-2(j-k)}\,r^{-\delta/2}, \tag{7.24}$$

which is sufficient for Proposition 7.1.

Finally, we briefly prove Lemma 7.5. We first observe that under the hypothesis (7.21) we have a trivial pointwise bound for all $x \in \mathbb{R}^m$:

$$\Big| \int_{\mathbb{R}^m} f(x-y)\chi(x,y)\,dy \Big| \le \|f\|_{L^\infty(\mathbb{R}^m)} \sup_{x\in\mathbb{R}^m} \int_{\mathbb{R}^m} |\chi(x,y)|\,dy \le \lambda \|f\|_{L^\infty(\mathbb{R}^m)},$$

which proves the desired bound for $p = \infty$. Now for $1 \le p \le \infty$, under the hypothesis (7.22) we also have the bound

$$\begin{aligned} \Big\| \int_{\mathbb{R}^m} f(x-y)\chi(x,y)\,dy \Big\|_{L^p(\mathbb{R}^m(dx))} & \\ &\le \Big\| \int_{\mathbb{R}^m} |f|(x-y)\Big(\sup_{z\in\mathbb{R}^m} |\chi(z,y)|\Big)\,dy \Big\|_{L^p(\mathbb{R}^m(dx))} \\ &\le \int_{\mathbb{R}^m} \|f\|_{L^p(\mathbb{R}^m)}\Big(\sup_{z\in\mathbb{R}^m} |\chi(z,y)|\Big)\,dy \le C_0 \|f\|_{L^p(\mathbb{R}^m)}. \end{aligned} \tag{7.25}$$

Interpolation between this bound and the L^∞ bound gives the desired result of the lemma. We remark that Lemma 7.5 replaces the so-called small set maximal functions used by Stein and Wainger in [14]. In our setting, we still require the ability to track the L^2 norm relative to the size of the small exceptional sets, but we no longer have true maximal functions since the normalization factor $a = 2^k$ is now fixed. (Here we again see the advantage of working inside the square function.)

7.3.2. The term II (general $n \ge 2$). Next we look at **II**, which is the contribution to the kernel (7.11) from $\zeta(s)$; we may now treat all dimensions $n \ge 2$ with no additional difficulty. Up to a sign, the term **II** can be expressed as

$$\iint_{\mathbb{R}^{n+2}} e^{iP_\nu(u+z)-iP_\mu(z)}\eta(u+z)\eta(z)\left(\delta_{\xi=|z|^2} - \zeta(\xi)\right)\zeta(s)\underline{\Delta}_{j-k}(\theta - s + \xi)\,dz\,d\xi\,ds.$$

We split this into two terms, one coming from $\delta_{\xi=|z|^2}$, another coming from $\zeta(\xi)$. In the first one we integrate the delta function to replace ξ by $|z|^2$, and change variables $s \mapsto s + |z|^2$. Then we write $1 = \int \zeta(\xi + |z|^2)\,d\xi$ to reintroduce a trivial integration in ξ. In the second term, we change variables $s \mapsto s + |z|^2$, $\xi \mapsto \xi + |z|^2$. We then get

$$\begin{aligned} \mathbf{II} = \iint_{\mathbb{R}^{n+2}} & e^{iP_\nu(u+z)-iP_\mu(z)}\eta(u+z)\eta(z)\zeta(\xi+|z|^2)\zeta(s+|z|^2) \\ &\cdot \left(\underline{\Delta}_{j-k}(\theta - s) - \underline{\Delta}_{j-k}(\theta - s + \xi)\right)\,dz\,ds\,d\xi. \end{aligned}$$

Hence

$$|\mathbf{II}| \le \int_{\mathbb{R}^2} |K_\flat^{\lambda(x,t),\lambda(x-u,t-\theta)}(u;\xi,s)|\,\big|\underline{\Delta}_{j-k}(\theta - s) - \underline{\Delta}_{j-k}(\theta - s + \xi)\big|\,ds\,d\xi,$$

where

$$K_\flat^{\nu,\mu}(u;\xi,s) = \int_{\mathbb{R}^n} e^{iP_\nu(u+z)-iP_\mu(z)}\,\eta(u+z)\,\eta(z)\,\zeta(\xi+|z|^2)\,\zeta(s+|z|^2)\,dz.$$

Note that $K_\flat^{\nu,\mu}(u;\xi,s)$ has compact ξ and s support in $B_2(\mathbb{R})$, and moreover is an n-dimensional integral over the full variable z. We again apply the mean value theorem in the form of Lemma 4.4, concluding that

$$|\mathbf{II}| \leq 2^{-2(j-k)} \int_{\mathbb{R}^2} |K_\flat^{\nu,\mu}(u;\xi,s)|\, \chi_{B_2}(\xi)\, \chi_{B_2}(s)\, \psi_{j-k}(\theta - s)\, ds\, d\xi,$$

where ψ_{j-k} is a non-negative integrable function with L^1 norm uniformly bounded, independent of $j-k$. Now we apply Proposition 3.4 (general $n \geq 2$) to bound $K_\flat^{\nu,\mu}$, so that there exists $\delta > 0$ and a small set $G^\nu \subset B_2(\mathbb{R}^n)$ with

$$|G^\nu| \leq C r^{-\delta}$$

such that

$$|K_\flat^{\nu,\mu}(u;\xi,s)| \leq C\left[r^{-\delta}\chi_{B_2}(u) + \chi_{G^\nu}(u)\right].$$

This estimate is uniform in ξ and s, so plugging this back into $\mathbf{II}$ and integrating trivially in ξ, we get

$$|\mathbf{II}| \leq C\, 2^{-2(j-k)} \left(r^{-\delta}\chi_{B_2}(u) + \chi_{G^\nu}(u)\right) \int_{\mathbb{R}} \psi_{j-k}(\theta - s)\, \chi_{B_2}(s)\, ds.$$

Here for notational convenience, we will temporarily set

$$\tilde{\psi}_{j-k}(\theta) = \int_{\mathbb{R}} \psi_{j-k}(\theta - s)\, \chi_{B_2}(s)\, ds$$

which is itself an integrable function of θ, with L^1 norm uniformly bounded independent of $j-k$. Recalling that $\nu = \lambda(x,t)$, the contribution of the portion $\mathbf{II}$ of the kernel to TT^*f in (7.10) is then bounded by

$$C\, 2^{-2(j-k)} \iint_{\mathbb{R}^{n+1}} |f(x-u, t-\theta)|\, \frac{1}{a^n}\left(r^{-\delta}\chi_{B_2}\left(\frac{u}{a}\right) + \chi_{G^{\lambda(x,t)}}\left(\frac{u}{a}\right)\right) \cdot \frac{1}{a^2}\tilde{\psi}_{j-k}\left(\frac{\theta}{a^2}\right) d\theta\, du. \tag{7.26}$$

We again call upon Lemma 7.5, this time with the choice

$$\chi((x,t),(u,\theta)) = \frac{1}{a^n}\left(r^{-\delta}\chi_{B_2}\left(\frac{u}{a}\right) + \chi_{G^{\lambda(x,t)}}\left(\frac{u}{a}\right)\right)\frac{1}{a^2}\tilde{\psi}_{j-k}\left(\frac{\theta}{a^2}\right).$$

We may verify (7.21) by noting that

$$\sup_{(x,t)\in\mathbb{R}^{n+1}} |\chi((x,t),(u,\theta))| \leq \frac{1}{a^n}\chi_{B_2}\left(\frac{u}{a}\right)\frac{1}{a^2}\tilde{\psi}_{j-k}\left(\frac{\theta}{a^2}\right),$$

and the right-hand side is uniformly in $L^1(du\,d\theta)$ with a bounded norm. We may verify (7.22) by noting that because of the factor $r^{-\delta}$ out front of the first term of χ, and the small measure of the set $G^{\lambda(x,t)}$ in the second term of χ, we have

$$\sup_{(x,t)\in\mathbb{R}^{n+1}} \|\chi((x,t),(u,\theta))\|_{L^1(du\,d\theta)} \leq C r^{-\delta}.$$

Hence by Lemma 7.5, the L^2 norm of the portion of TT^*f contributed by (7.26) is bounded above by $\leq Cr^{-\delta/2}2^{-2(j-k)}$. Combining this with (7.24), we have proved that when $j \geq k$,

$$\|TT^*f\|_{L^2} \leq Cr^{-\delta/2}\, 2^{-2(j-k)}\, \|f\|_{L^2},$$

which proves (7.15) and hence Proposition 7.1 with $\delta_0 = \delta/4$, in dimension $n = 2$.

To conclude the proof of Proposition 7.1, it remains to prove (7.15) in dimensions $n > 2$; this requires us to make a slightly different change of variables in estimating (7.16), and that is taken up in Section 10.1.

7.4. Proof of Propositions 7.3 and 7.4 (Case $j < k$)

In proving Propositions 7.3 and 7.4, we will need to prove (7.4) and (7.6) for I_a^λ, and (7.5) and (7.7) for J_a^λ. Below in Section 7.4.1, we will first prove (7.4) and (7.6) in dimension $n = 2$, since in this dimension the proof will be slightly cleaner. The case $n > 2$ is then deferred to Section 10.2. Then in Section 7.4.2, we prove (7.5) and (7.7), which we can do without any additional difficulty in general dimensions $n \geq 2$.

7.4.1. Proof of (7.4) and (7.6) ($n = 2$). Suppose now we are in dimension $n = 2$. For $j < k$, we will prove (7.4) of Proposition 7.3 by showing that if $\lambda = (\lambda_2, \ldots, \lambda_d)\colon \mathbb{R}^{n+1} \to \mathbb{R}^{d-1}$ is a measurable stopping-time satisfying $\|\lambda(x,t)\| \approx r$, then $T_1 f := I_{2^k}^{\lambda(x,t)} f(x,t)$ satisfies

$$\|T_1T_1^* f(x,t)\|_{L^2} \leq Cr^{-\delta/2}\|f\|_{L^2},$$

for some small $\delta > 0$.

We recall from (7.13) that the kernel relevant to $T_1T_1^*$ is

$${}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta) = \int_{\mathbb{R}^n} e^{iP_\nu(u+z) - iP_\mu(z)}\, \eta(u+z)\, \eta(z)\, \underline{\Delta}_{j-k}(\theta - |u+z|^2 + |z|^2)\, dz,$$

where $\nu = \lambda(x,t), \mu = \lambda(x-u, t-\theta)$. We make the change of variables $z \mapsto (\tau, \sigma)$ as defined in (7.17), so that

$$ {}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta) = \int_{\mathbb{R}} K_\sharp^{\nu,\mu}(u,\tau)\, \underline{\Delta}_{j-k}(\theta - |u|^2 - 2|u|\tau)\, d\tau, \tag{7.27}$$

where

$$K_\sharp^{\nu,\mu}(u,\tau) = \int_{\mathbb{R}^{n-1}} e^{iP_\nu(u+z) - iP_\mu(z)}\, \eta(u+z)\, \eta(z)\, d\sigma.$$

We apply the bound of Proposition 3.5 ($n = 2$) to $K_\sharp^{\nu,\mu}$ to conclude that

$$\begin{aligned} |{}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta)| \leq C \int_{\mathbb{R}} & \big(r^{-\delta}\, \chi_{B_2}(u)\chi_{B_1}(\tau) + \chi_{G^\nu}(u)\chi_{B_1}(\tau) + \chi_{B_2}(u)\chi_{F_u^\nu}(\tau)\big) \\ & \cdot\ \underline{\Delta}_{j-k}(\theta - |u|^2 - 2|u|\tau)\, d\tau. \end{aligned} \tag{7.28}$$

Inserting this kernel bound into $T_1T_1^*$ via (7.12), we see that

$$|T_1T_1^*f(x,t)| \leq \int_{\mathbb{R}^{n+1}} |f|(x-u,t-\theta)\,\frac{1}{a^{n+2}}\left|\,{}^{(1)}\mathcal{K}^{\lambda(x,t),\lambda(x-u,t-\theta)}\left(\frac{u}{a},\frac{\theta}{a^2}\right)\right| du\,d\theta$$

$$\leq C\int_{\mathbb{R}^{n+1}} |f|(x-u,t-\theta)\,\chi((x,t),(u,\theta))\,du\,d\theta, \tag{7.29}$$

where we have defined $\chi((x,t),(u,\theta))$ to be the function

$$\chi((x,t),(u,\theta)) = \int_{\mathbb{R}} \Big(r^{-\delta}\,\frac{1}{a^n}\,\chi_{B_2}\left(\frac{u}{a}\right)\chi_{B_1}(\tau) + \frac{1}{a^n}\,\chi_{G^{\lambda(x,t)}}\left(\frac{u}{a}\right)\chi_{B_1}(\tau)$$
$$+ \frac{1}{a^n}\,\chi_{B_2}\left(\frac{u}{a}\right)\chi_{F^{\lambda(x,t)}_{(u/a)}}(\tau)\Big)\left|\frac{1}{a^2}\,\underline{\Delta}_{j-k}\left(\frac{\theta-|u|^2-2a|u|\tau}{a^2}\right)\right| d\tau.$$

Since $\underline{\Delta}_{j-k}$ is uniformly in L^1 independent of $j-k$ (Lemma 4.3), we may use the same argument that we applied to (7.20) to show via Lemma 7.5 that $T_1T_1^*f$ has L^2 norm majorized by $C\,r^{-\delta/2}$, which proves (7.4) with $\delta_0=\delta/4$.

Next we prove (7.6) of Proposition 7.4. To do so, we still want to use the method of TT^*, not because we want to extract decay from the phase, but because we want to introduce the variables u and τ and integrate by parts in τ to pick up a decay of $2^{2(j-k)}$. So again let $\lambda=(\lambda_2,\dots,\lambda_d)\colon \mathbb{R}^{n+1}\to\mathbb{R}^{d-1}$ be a measurable stopping-time satisfying $\|\lambda(x,t)\|\approx r$, and define T_1 as before. We return to equation (7.27), from whence we note that since $\int \underline{\Delta}_{j-k}(\tau)\,d\tau=0$, we may use (4.11) to write

$$\underline{\Delta}_{j-k}(\theta-2|u|\tau-|u|^2) = -\frac{2^{2(j-k)}}{2|u|}\frac{d}{d\tau}\left[\underline{\widetilde{\Delta}}_{j-k}(\theta-2|u|\tau-|u|^2)\right], \tag{7.30}$$

for the Schwartz function $\underline{\widetilde{\Delta}}_{j-k}$ we constructed in Lemma 4.5. Thus by integration by parts,

$${}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta) = \frac{2^{2(j-k)}}{2|u|}\int_{\mathbb{R}} \partial_\tau K_\sharp^{\nu,\mu}(u,\tau)\,\underline{\widetilde{\Delta}}_{j-k}(\theta-2|u|\tau-|u|^2)\,d\tau.$$

We now note that since $K_\sharp^{\nu,\mu}(u,\tau)$ is supported where $|u|\leq 2$, $|\tau|\leq 1$ and is a smooth function of τ,

$$|\partial_\tau K_\sharp^{\nu,\mu}(u,\tau)| \leq C\,r\,\chi_{B_2}(u)\,\chi_{B_1}(\tau),$$

where the factor of r comes from bringing down coefficients of size $\|\nu\|,\|\mu\|\approx r$ when differentiating the phase $P_\nu(u+z)-P_\mu(z)$ with respect to τ. Hence

$$|{}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta)| \leq C\,\frac{r\,2^{2(j-k)}}{2|u|}\,\chi_{B_2}(u)\int_{\mathbb{R}} \chi_{B_1}(\tau)\,\underline{\widetilde{\Delta}}_{j-k}(\theta-2|u|\tau-|u|^2)\,d\tau.$$

Thus using (7.12) again, in total

$$|T_1T_1^*f(x,t)| \leq \iint_{\mathbb{R}^{n+1}} |f|(x-u,t-\theta)\,\frac{1}{a^{n+2}}\left|\,{}^{(1)}\mathcal{K}^{\lambda(x,t),\lambda(x-u,t-\theta)}\left(\frac{u}{a},\frac{\theta}{a^2}\right)\right| du\,d\theta$$

$$\leq C\,r\,2^{2(j-k)}\iint_{\mathbb{R}^{n+1}} |f|(x-u,t-\theta)\,\chi((x,t),(u,\theta))\,du\,d\theta, \tag{7.31}$$

where we have defined

$$\chi((x,t),(u,\theta)) = \int_{\mathbb{R}} \Big(\frac{2|u|}{a}\Big)^{-1} \frac{1}{a^n}\,\chi_{B_2}\Big(\frac{u}{a}\Big)\,\chi_{B_1}(\tau)\,\frac{1}{a^2}\,\underline{\widetilde{\Delta}}_{j-k}\Big(\frac{\theta}{a^2} - 2a|u|\frac{\tau}{a^2} - \Big|\frac{u}{a}\Big|^2\Big)\,d\tau.$$

To bound the averaging operator (7.31), we only need to bound

$$\Big\| \sup_{(x,t)\in\mathbb{R}^{n+1}} |\chi((x,t),(u,\theta))| \Big\|_{L^1(dud\theta)}$$

and then proceed as in (7.25). Now

$$\begin{aligned}
&\Big\| \sup_{(x,t)\in\mathbb{R}^{n+1}} |\chi((x,t),(u,\theta))| \Big\|_{L^1(du\,d\theta)} \\
&= \int_{\mathbb{R}^{n+2}} \Big(\frac{2|u|}{a}\Big)^{-1} \frac{1}{a^n}\,\chi_{B_2}\Big(\frac{u}{a}\Big)\,\chi_{B_1}(\tau)\,\frac{1}{a^2}\,\underline{\widetilde{\Delta}}_{j-k}\Big(\frac{\theta}{a^2} - 2a|u|\frac{\tau}{a^2} - \Big|\frac{u}{a}\Big|^2\Big)\,d\tau\,du\,d\theta \\
&\le C \int_{\mathbb{R}^n} \Big(\frac{2|u|}{a}\Big)^{-1} \frac{1}{a^n}\,\chi_{B_2}\Big(\frac{u}{a}\Big)\,du \cdot \int_{\mathbb{R}} \chi_{B_1}(\tau)\,d\tau \le C'.
\end{aligned}$$

Here we have used the facts that $\underline{\widetilde{\Delta}}_{j-k}$ is uniformly in L^1 (Lemma 4.5), and that $|u|^{-1}$ is locally integrable in $\mathbb{R}^n$ for $n \ge 2$. It follows that

$$\|T_1 T_1^* f\|_{L^2(\mathbb{R}^{n+1})} \le C\,r\,2^{2(j-k)}, \tag{7.32}$$

as desired, in the case $n = 2$.

We pause momentarily to remark on the necessity of applying the Littlewood–Paley decomposition carefully. First, it was necessary to compute $T_1T_1^*$ in order to exploit the cancellation property $\int_{\mathbb{R}} \Delta(\theta)\,d\theta = 0$. Indeed, suppose we computed the kernel of T_1 directly rather than taking $T_1T_1^*$; then for example when $0 = j < k$ we would observe that $T_1 f(x,t)$ is given by

$$T_1 f(x,t) = \iint_{\mathbb{R}^{n+1}} f(x-u,t-\theta) e^{iP_{\lambda(x,t)}(u/2^k)}\,\frac{1}{2^{nk}}\,\eta\Big(\frac{u}{2^k}\Big)\,\Delta(\theta - |u|^2)\,du\,d\theta.$$

It is not clear how one can gain a factor 2^{-k} from this, from the cancellation property of Δ directly.

Next, suppose we had chosen a different Littlewood–Paley projection, say an $(n+1)$-dimensional projection of the form

$$P_j f(x,t) = \iint_{\mathbb{R}^{n+1}} f(x-y,t-s)\,\frac{1}{2^{(n+2)j}}\,P\Big(\frac{y}{2^j},\frac{s}{2^{2j}}\Big)\,dy\,ds,$$

with an appropriate function P. With this projection in mind, if we then set $T_1 f(x,t) = I_a^{\lambda(x,t)} P_j f(x,t)$ instead, one would have difficulty gaining decay in $|j-k|$ when one tries integrating by parts in the kernel of $T_1T_1^*$. For instance, in the case $0 = j < k$ the kernel of $T_1T_1^*$ can be read off via

$$\begin{aligned}
T_1T_1^* f(x,t) = \int_{\mathbb{R}^{n+1}} f(x-v,t-\theta) \int_{\mathbb{R}^{n+1}} \int_{\mathbb{R}} & e^{iP_{\lambda(x,t)}(\frac{u+z}{a}) - iP_{\lambda(x-v,t-\theta)}(\frac{z}{a})} \\
& \cdot \eta\Big(\frac{u+z}{a}\Big)\,\eta\Big(\frac{z}{a}\Big)\,(P*P)(v-u,\theta-|u|^2-2u\cdot z)\,du\,dz\,dv\,d\theta.
\end{aligned} \tag{7.33}$$

Assume $\int_{\mathbb{R}^{n+1}} P(y,s)\,dyds = 0$. Then one can write $P(y,s)$, or indeed $(P*P)(y,s)$, as

$$\frac{\partial \tilde{P}^{(1)}}{\partial y_1}(y,s) + \cdots + \frac{\partial \tilde{P}^{(n)}}{\partial y_n}(y,s) + \frac{\partial \tilde{P}^{(n+1)}}{\partial s}(y,s),$$

where $\tilde{P}^{(1)}, \dots, \tilde{P}^{(n+1)}$ are some Schwartz functions. If we apply the above expression for $(P*P)(y,s)$, then $(P*P)(v-u, \theta - |u|^2 - 2u\cdot z)$ can be expressed as

$$\Big[\frac{\partial \tilde{P}^{(1)}}{\partial y_1} + \cdots + \frac{\partial \tilde{P}^{(n)}}{\partial y_n} + \frac{\partial \tilde{P}^{(n+1)}}{\partial s}\Big](v-u, \theta - |u|^2 - 2u\cdot z).$$

It is not clear how one can plug this into (7.33), and integrate by parts to gain a factor $a^{-1} = 2^{-k}$. So the choice of a good Littlewood–Paley projection here is very important; our choice Δ_j works for this particular purpose.

7.4.2. Proof of (7.5) and (7.7) (general $n \geq 2$). We now turn to (7.5) of Proposition 7.3, for which we will show that if $\lambda = (\lambda_2, \dots, \lambda_d)\colon \mathbb{R}^{n+1} \to \mathbb{R}^{d-1}$ is a measurable stopping-time satisfying $\|\lambda(x,t)\| \approx r$, then for $T_2 f := J_{2^k}^{\lambda(x,t)} f(x,t)$,

$$\|T_2 T_2^* f(x,t)\|_{L^2} \leq C r^{-2\delta_0} \|f\|_{L^2}$$

for some small $\delta_0 > 0$. Here we can work in general dimension $n \geq 2$. We recall from (7.14) that the kernel relevant to $T_2T_2^*$ is

$$\begin{aligned} {}^{(2)}\mathcal{K}^{\nu,\mu}(u,\theta) &= \int_{\mathbb{R}^{n+2}} e^{iP_\nu(u+z) - iP_\mu(z)} \eta(u+z)\eta(z)\zeta(s)\zeta(\xi)\underline{\Delta}_{j-k}(\theta - s + \xi)\,dz\,d\xi\,ds \\ &= \int_{\mathbb{R}^2} K_\flat^{\nu,\mu}(u)\zeta(s)\zeta(\xi)\underline{\Delta}_{j-k}(\theta - s + \xi)\,d\xi\,ds, \end{aligned}$$

where

$$K_\flat^{\nu,\mu}(u) = \int_{\mathbb{R}^n} e^{iP_\nu(u+z) - iP_\mu(z)} \eta(u+z)\eta(z)\,dz.$$

We apply Proposition 3.4 and conclude that there exists a $\delta > 0$ and a small set $G^\nu \subset B_2(\mathbb{R}^n)$ with $|G^\nu| \leq r^{-\delta}$ so that

$$|{}^{(2)}\mathcal{K}^{\nu,\mu}(u,\theta)| \leq C(r^{-\delta}\chi_{B_2}(u) + \chi_{G^\nu}(u)) \int_{\mathbb{R}^2} \big|\zeta(s)\zeta(\xi)\underline{\Delta}_{j-k}(\theta - s + \xi)\big|\,d\xi\,ds.$$

This kernel bound shows that $|T_2T_2^* f(x,t)|$ is bounded above by

$$\begin{aligned} C \iint_{\mathbb{R}^{n+1}} |f|(x-u, t-\theta)\Big|\frac{1}{a^{n+2}}\, {}^{(2)}\mathcal{K}^{\lambda(x,t),\lambda(x-u,t-\theta)}\Big(\frac{u}{a}, \frac{\theta}{a^2}\Big)\Big|\,du\,d\theta \\ \leq C \iint_{\mathbb{R}^{n+1}} |f|(x-u, t-\theta)\,\chi((x,t),(u,\theta))\,du\,d\theta \end{aligned}$$

where we define $\chi((x,t),(u,\theta))$ to be

$$\int_{\mathbb{R}^2} \frac{1}{a^n}\Big(r^{-\delta}\chi_{B_2}\Big(\frac{u}{a}\Big) + \chi_{G^{\lambda(x,t)}}\Big(\frac{u}{a}\Big)\Big)\zeta(s)\zeta(\xi)\frac{1}{a^2}\Big|\underline{\Delta}_{j-k}\Big(\frac{\theta}{a^2} - s + \xi\Big)\Big|\,d\xi\,ds.$$

As in previous arguments, we may check that χ satisfies the hypotheses of Lemma 7.5, that is,

$$\Big\| \sup_{(x,t)\in\mathbb{R}^{n+1}} |\chi((x,t),(u,\theta))| \Big\|_{L^1(du\,d\theta)} \leq C$$

and

$$\sup_{(x,t)\in\mathbb{R}^{n+1}} \|\chi((x,t),(u,\theta))\|_{L^1(du\,d\theta)} \leq C r^{-\delta},$$

since ζ is in L^1 and $\underline{\Delta}_{j-k}$ is uniformly in L^1, independent of $j-k$ (Lemma 4.3). Applying Lemma 7.5, we may conclude that

$$\|T_2 T_2^* f\|_{L^2} \leq C r^{-\delta/2} \|f\|_{L^2},$$

which proves (7.5) with $\delta_0 = \delta/4$.

Next we prove (7.7) of Proposition 7.4. We want to bound the operator $f \mapsto \sup_{\|\lambda\|\simeq r} |J_a^\lambda \Delta_j f|$ on L^2, where $a = 2^k$. Here we will use the fact that $\int_{\mathbb{R}} \Delta_j(\theta)\, d\theta = 0$ in order to place a derivative onto the kernel of $J_{2^k}^\lambda$. We will not require a stopping time, nor a TT^* argument, since we are not seeking decay with respect to r and no singular support on the paraboloid is present. We first isolate the kernel. Note that J_a^λ is a convolution operator, whose convolution kernel is a (parabolic) dilation of $e^{iP_\lambda(y)}\eta(y)\zeta(s)$ by a. Also, Δ_j is a convolution operator, whose convolution kernel is a (parabolic) dilation of $\delta_{y=0}\Delta(s)$ by 2^j, which is the same as the (parabolic) dilation of $\delta_{y=0}\Delta_{j-k}(s)$ by $a = 2^k$. Hence

$$J_a^\lambda \Delta_j f(x,t) = \iint_{\mathbb{R}^{n+1}} f(x-y, t-s) \frac{1}{a^{n+2}} L^\lambda\Big(\frac{y}{a}, \frac{s}{a^2}\Big)\, dy\, ds,$$

where $L^\lambda(y,s)$ is the convolution of $e^{iP_\lambda(y)}\eta(y)\zeta(s)$ with $\delta_{y=0}\Delta_{j-k}(s)$, namely

$$L^\lambda(y,s) = e^{iP_\lambda(y)}\eta(y) \int_{\mathbb{R}} \zeta(s-\theta)\, \Delta_{j-k}(\theta)\, d\theta.$$

We now use the fact that $\int \Delta_{j-k}(\theta)\, d\theta = 0$ to apply Lemma 4.5 and write

$$L^\lambda(y,s) = 2^{2(j-k)} e^{iP_\lambda(y)}\eta(y) \int_{\mathbb{R}} \zeta(s-\theta) \Big(\frac{d}{d\theta}\widetilde{\Delta}_{j-k}\Big)(\theta)\, d\theta,$$

with the antiderivative $\widetilde{\Delta}_{j-k}$ provided by Lemma 4.5. Then after integration by parts,

$$L^\lambda(y,s) = -2^{2(j-k)}\, e^{iP_\lambda(y)}\eta(y) \int_{\mathbb{R}} \frac{d}{d\theta}\zeta(s-\theta)\, \widetilde{\Delta}_{j-k}(\theta)\, d\theta,$$

so that after applying the trivial bound to η we obtain

$$|L^\lambda(y,s)| \leq C\, 2^{2(j-k)} L(y,s),$$

say, where we define

$$L(y,s) = \chi_{B_1}(y)\, |\zeta' * \widetilde{\Delta}_{j-k}|(s).$$

We note that this bound is independent of λ, and as a result,

$$\sup_\lambda |J^\lambda_{2^k} \Delta_j f| \leq C\, 2^{2(j-k)} |f| * \Big(\frac{1}{a^{n+2}} L\Big(\frac{\cdot}{a}, \frac{\cdot}{a^2}\Big)\Big).$$

We now need only recall that $\|L\|_{L^1(\mathbb{R}^{n+1})} \leq C$ since $\widetilde{\Delta}_{j-k}$ is uniformly in L^1 (Lemma 4.5). Thus by a further application of Young's inequality,

$$\| \sup_\lambda |J^\lambda_{2^k} \Delta_j f| \, \|_{L^2(\mathbb{R}^{n+1})} \leq A\, 2^{2(j-k)} \, \|f\|_{L^2(\mathbb{R}^{n+1})},$$

as required.

7.5. Summary

The work of this section has completed the proof of Propositions 7.1, 7.3 and 7.4, except for the following:

(a) estimate of the term **I** in (7.16) in dimensions $n > 2$;

(b) proof of (7.4) and (7.6) in dimensions $n > 2$; and

(c) two key oscillatory integral bounds for the kernels $K^{\nu,\mu}_\flat$ and $K^{\nu,\mu}_\sharp$.

The last point (c) occupies the main body of the next two sections. The first two points (a) and (b) are then quickly resolved in Sections 10.1 and 10.2 respectively.

8. Van der Corput estimates for kernels: Part I

We have now reached the heart of the matter: the van der Corput estimates for the kernels arising in the TT^* arguments throughout Section 7. These estimates break into two cases: kernels of the form $K^{\nu,\mu}_\flat$, which arise from operators not involving Radon-type behavior, and kernels of the form $K^{\nu,\mu}_\sharp$, which arise from operators exhibiting Radon-type behavior along the paraboloid.

8.1. Proof of the van der Corput estimate for $K^{\nu,\mu}_\flat$

The van der Corput estimate Proposition 3.4 for the kernel $K^{\nu,\mu}_\flat$ is implied by Lemma 4.1 of [14] (in the case $h = 1$ in their notation). For completeness, we briefly recall the proof here. We recall that $r \leq \|\nu\|, \|\mu\| \leq 2r$. We consider the terms in the phase $Q_\nu(u+z) - Q_\mu(z)$ of $K^{\nu,\mu}_\flat(u)$ that are linear in z; precisely, these terms contribute

$$\sum_{1 \leq j \leq n} Q^{(j)}_\nu(u)\, z_j$$

to the phase, where we set

$$Q^{(j)}_\nu(u) = \sum_{2 \leq |\alpha| \leq d} \nu_\alpha\, \alpha_j\, u^{\alpha - e_j}.$$

(Here we note in particular that there is no contribution from $Q_\mu(z)$ since the original phase function is assumed to have no linear terms.) Thus by Lemma 3.1,

$$|K_\flat^{\nu,\mu}(u)| \leq c\Big(\sum_{j=1}^{n}|Q_\nu^{(j)}(u)|\Big)^{-1/d}, \tag{8.1}$$

for a constant c depending only on n, d. It remains to show that the sum of linear coefficients in (8.1) is bounded below by a small power of r, except possibly for u belonging to some small exceptional set. We fix $\rho > 0$, to be specified later, and define

$$G^\nu = \Big\{u \in B_2(R^n) : \sum_{j=1}^{n}|Q_\nu^{(j)}(u)| < \rho\Big\}.$$

Then for $u \in B_2 \setminus G^\nu$,

$$|K_\flat^{\nu,\mu}(u)| \leq c\rho^{-1/d}.$$

Moreover, $G^\nu \subset \cap_j\{u \in B_2 : |Q_\nu^{(j)}(u)| < \rho\}$. Hence by Lemma 3.2,

$$|G^\nu| \leq c\rho^{1/d} \min_{1\leq j\leq n}\Bigg(\sum_{\substack{2\leq|\alpha|\leq d\\ |\alpha-e_j|\geq 1}}|\nu_\alpha|\alpha_j\Bigg)^{-1/d} \simeq c\rho^{1/d}\Bigg(\sum_{j=1}^{n}\sum_{\substack{2\leq|\alpha|\leq d\\ |\alpha-e_j|\geq 1}}|\nu_\alpha|\alpha_j\Bigg)^{-1/d},$$

for a constant c depending only on n, d. Here the key observation is that because there are no linear terms in the original phase polynomial, for every α such that $\nu_\alpha \neq 0$ there is at least one j for which $|\alpha - e_j| \geq 1$, so it follows that

$$\sum_{j=1}^{n}\sum_{\substack{2\leq|\alpha|\leq d\\ |\alpha-e_j|\geq 1}}|\nu_\alpha|\alpha_j \geq C\sum_{2\leq|\alpha|\leq d}|\nu_\alpha| \geq Cr.$$

Thus $|G^\nu| \leq C(r/\rho)^{-1/d}$ for a constant C depending only on n, d, and so choosing $\rho = r^{\delta_0}$ with $\delta_0 < 1$ completes the proof of the proposition, with $\delta = \min(\frac{1-\delta_0}{d}, \frac{\delta_0}{d})$.

8.2. Van der Corput estimates for $K_\sharp^{\nu,\mu}$: strategy and examples

We now turn to the more challenging case of the kernel $K_\sharp^{\nu,\mu}$, which arises in TT^* estimates involving Radon-type behavior; for this we recall that our goal (at least in the case $n = 2$) is to prove Proposition 3.5 (and later its higher dimensional analogue Proposition 9.1). But in order to aid the reader, rather than proving Proposition 3.5 immediately, we first prove the desired bound for $K_\sharp^{\nu,\mu}$ in three example cases in dimension $n = 2$; these examples already illustrate the main difficulties. We will then develop in Section 9 the full n-dimensional argument that will prove both Proposition 3.5 and Proposition 9.1 at once.

Recall that in dimension $n = 2$, we have fixed a set $\mathscr{P}$ of $d-1$ homogeneous polynomials on $\mathbb{R}^2$ with real coefficients, say

$$p_m(y) = \sum_{|\alpha|=m} c_\alpha y^\alpha, \quad \text{for } m = 2, \ldots, d, \tag{8.2}$$

where the coefficients c_α are fixed once and for all. We assume that $p_2(y) \neq C|y|^2$ for any nonzero constant C. We will also write $p_1(y) \equiv 0$ for convenience (i.e. we set $c_\alpha = 0$ whenever $|\alpha| = 1$). Furthermore, for $\nu = (\nu_2, \dots, \nu_d)$, $\mu = (\mu_2, \dots, \mu_d)$, we have

$$K_\sharp^{\nu,\mu}(u,\tau) = \int_{\mathbb{R}} e^{iP_\nu(u+z) - iP_\mu(z)}\, \Psi(u,z)\, d\sigma,$$

where

$$P_\nu(y) = \sum_{m=2}^{d} \nu_m p_m(y), \quad P_\mu(y) = \sum_{m=2}^{d} \mu_m p_m(y), \tag{8.3}$$

and $\Psi(u,z)$ is a C^1 function supported on $B_2(\mathbb{R}^2) \times B_1(\mathbb{R})$ with $\|\Psi\|_{C^1} \leq 1$. Here z is defined implicitly in terms of u, τ, σ by (7.17), which we now write in the form

$$z_1 = \frac{u_1\tau + u_2\sigma}{|u|}, \quad z_2 = \frac{u_2\tau - u_1\sigma}{|u|}. \tag{8.4}$$

Hence $K_\sharp^{\nu,\mu}$ is an oscillatory integral in σ, whose phase is a polynomial in σ, and with the trivial bound

$$|K_\sharp^{\nu,\mu}(u,\tau)| \leq C\,\chi_{B_2}(u)\,\chi_{B_1}(\tau).$$

For fixed u, τ, in order to apply a van der Corput estimate to $K_\sharp^{\nu,\mu}(u,\tau)$ and deduce a bound at (u,τ) of the form

$$|K_\sharp^{\nu,\mu}(u,\tau)| \leq A\, r^{-\delta'} \quad \text{for some } \delta' > 0,$$

we must show that within the phase $P_\nu(u+z) - P_\mu(z)$ (regarded as a polynomial in σ), the coefficient of at least one monomial σ^l with $1 \leq l \leq d$ is bounded below by r^δ for some $\delta > 0$. The coefficients of this polynomial (with respect to σ) are functions of u, τ. Thus our strategy is to show that for "most" u, τ, at least one such coefficient with respect to the variable σ is large; the remaining exceptional u, τ are shown to belong to a sufficiently small exceptional set.

It is convenient to define the following notation to indicate the norm of the coefficients of $P_\nu(u+z) - P_\mu(z)$ as a polynomial in σ:

$$\|P_\nu(u+z) - P_\mu(z)\|_\sigma := \sum_{1 \leq l \leq d} |C[\sigma^l](u,\tau)|,$$

where $C[\sigma^l](u,\tau)$ is defined to be the coefficient of σ^l when one expands $P_\nu(u+z) - P_\mu(z)$ as a polynomial in σ. We also note that given a polynomial of τ, say $R(\tau)$, we will let $[\![R]\!]_\tau$ denote the norm of the coefficients (including constant term) of R as a polynomial in τ; note that this norm was considered before in (3.2). Analogously, for a polynomial $W(u)$, we will let $[\![W]\!]_u$ denote the norm of the coefficients (including constant term) of W as a polynomial in u.

Fix a small positive real number $0 < \varepsilon_1 < 1$. Then for a given (u,τ), if there is an index l for which

$$|C[\sigma^l](u,\tau)| \geq r^{\varepsilon_1}, \tag{8.5}$$

then

$$\|P_\nu(u+z) - P_\mu(z)\|_\sigma \geq r^{\varepsilon_1}, \tag{8.6}$$

so that by Lemma 3.1 we would already know that

$$|K_\sharp^{\nu,\mu}(u,\tau)| \leq c r^{-\varepsilon_1/d}, \tag{8.7}$$

for a constant c depending only on n, d, which would be sufficient for our purposes. Our strategy is to reduce to considering the set of (u, τ) for which

$$|C[\sigma^l](u,\tau)| \leq r^{\varepsilon_1} \quad \text{for all } l = 3, \ldots, d, \tag{8.8}$$

and then deduce certain information about the coefficient of σ or σ^2. At this point we recall that the natural "enemy" is a phase component that is precisely a multiple of $|y|^2$, that is, the defining function of the paraboloid. Thus we will say that any homogeneous polynomial of the form $C|y|^{2k}$ for some $k \geq 1$ is *parabolic*; polynomials that are not a multiple of a power of $|y|^2$ will be termed *non-parabolic*.

8.3. Case A example: no parabolic term is present

In our first example we consider the case where $\mathscr{P} = \{p_3(y), p_4(y)\}$ with $p_3(y) = y_1^3$ and $p_4(y) = y_1^4$. Then the phase polynomial is given by

$$P_\lambda(y) = \lambda_4 y_1^4 + \lambda_3 y_1^3.$$

We will use downward induction to reduce our consideration to the coefficient of σ in $P_\nu(u+z) - P_\mu(z)$.

We note that by the relations (8.4) giving z_1, z_2 in terms of σ, τ, we may easily compute

$$u_1 + z_1 = \frac{u_1}{|u|}(|u| + \tau) + \frac{u_2}{|u|}\sigma, \quad \text{and} \quad u_2 + z_2 = \frac{u_2}{|u|}(|u| + \tau) - \frac{u_1}{|u|}\sigma.$$

It will also be helpful for later use to note that $|z|^2 = \tau^2 + \sigma^2$, while $|u + z|^2 = (|u| + \tau)^2 + \sigma^2$.

We recall that by assumption $r \leq \|\nu\|, \|\mu\| \leq 2r$. We compute explicitly that

$$P_\nu(u+z) - P_\mu(z) = \sum_{l=0}^{4} C[\sigma^l](u,\tau)\, \sigma^l$$

where

$$C[\sigma^4](u,\tau) = (\nu_4 - \mu_4)\frac{u_2^4}{|u|^4}$$

$$C[\sigma^3](u,\tau) = \nu_4 4(|u| + \tau)\frac{u_1 u_2^3}{|u|^4} + \nu_3 \frac{u_2^3}{|u|^3} - \mu_4 4\tau \frac{u_1 u_2^3}{|u|^4} - \mu_3 \frac{u_2^3}{|u|^3}$$

$$C[\sigma^2](u,\tau) = \nu_4 6(|u| + \tau)^2 \frac{u_1^2 u_2^2}{|u|^4} + \nu_3 3(|u| + \tau)\frac{u_1 u_2^2}{|u|^3} - \mu_4 6\tau^2 \frac{u_1^2 u_2^2}{|u|^4} - \mu_3 3\tau \frac{u_1 u_2^2}{|u|^3}$$

$$C[\sigma](u,\tau) = \nu_4 4(|u| + \tau)^3 \frac{u_1^3 u_2}{|u|^4} + \nu_3 3(|u| + \tau)^2 \frac{u_1^2 u_2}{|u|^3} - \mu_4 4\tau^3 \frac{u_1^3 u_2}{|u|^4} - \mu_3 3\tau^2 \frac{u_1^2 u_2}{|u|^3}.$$

(We may clearly disregard the terms that are constant with respect to σ.)

Our strategy is to use downward induction to eliminate the presence of $\mu = (\mu_3, \mu_4)$ by approximating $\nu_4 - \mu_4$ and $\nu_3 - \mu_3$ (up to small error) by functions independent of μ_3, μ_4. We will then be able to represent the coefficient of σ independently of μ_3, μ_4 (up to small error) and show that except for a small set of u, τ the coefficient of σ is sufficiently large (that is, at least a small power of r).

Precisely, we proceed as follows. Fix $0 < \varepsilon_1 < 1$. We may assume that $|C[\sigma^4](u,\tau)| \leq r^{\varepsilon_1}$, since otherwise (8.6) and hence (8.7) would already be known. Under this assumption,

$$\left|(\nu_4 - \mu_4)\frac{u_2^4}{|u|^4}\right| \leq r^{\varepsilon_1}. \tag{8.9}$$

We may next assume that $|C[\sigma^3](u,\tau)| \leq r^{\varepsilon_1}$, otherwise (8.7) would already be known. We re-write $C[\sigma^3](u,\tau)$ as

$$C[\sigma^3](u,\tau) = (\nu_3 - \mu_3)\,\frac{u_2^3}{|u|^3} + (\nu_4 - \mu_4)\,4\tau\frac{u_1u_2^3}{|u|^4} + \nu_4\,4\,|u|\,\frac{u_1u_2^3}{|u|^4},$$

and under the assumption that this is $O(r^{\varepsilon_1})$ we see that

$$(\nu_3 - \mu_3)\frac{u_2^3}{|u|^3} = -(\nu_4 - \mu_4)\,4\tau\,\frac{u_1u_2^3}{|u|^4} - \nu_4\,4\,|u|\,\frac{u_1u_2^3}{|u|^4} + O(r^{\varepsilon_1}). \tag{8.10}$$

We then multiply through by $u_2/|u|$ and apply (8.9) to deduce that

$$(\nu_3 - \mu_3)\,\frac{u_2^4}{|u|^4} = -\nu_4\,4\,|u|\,\frac{u_1u_2^4}{|u|^5} + O(r^{\varepsilon_1}). \tag{8.11}$$

In this example it is unnecessary to consider the coefficient of σ^2; instead we turn immediately to the coefficient of σ, which we re-write as

$$\begin{aligned} C[\sigma](u,\tau) &= (\nu_4 - \mu_4)\,4\tau^3\,\frac{u_1^3u_2}{|u|^4} + (\nu_3 - \mu_3)\,3\tau^2\,\frac{u_1^2u_2}{|u|^3} \\ &\quad + \nu_4\,4(|u|^3 + 3|u|\tau^2 + 3|u|^2\tau)\frac{u_1^3u_2}{|u|^4} + \nu_3\,3(|u|^2 + 2|u|\tau)\frac{u_1^2u_2}{|u|^3}. \end{aligned} \tag{8.12}$$

Note that we have isolated all presence of $\mu = (\mu_3, \mu_4)$ into the first two terms. We multiply both sides of the identity by $u_2^3/|u|^3$ and apply (8.9) and (8.11) to the first two terms on the right-hand side to see that

$$C[\sigma](u,\tau)\,\frac{u_2^3}{|u|^3} = R^{(1)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}),$$

where $R^{(1)}_{\nu,u}(\tau)$, defined by

$$-\nu_4 \cdot 12|u|\tau^2\frac{u_1^3u_2^4}{|u|^7} + \nu_4 \cdot 4(|u|^3 + 3|u|\tau^2 + 3|u|^2\tau)\frac{u_1^3u_2^4}{|u|^7} + \nu_3 \cdot 3(|u|^2 + 2|u|\tau)\frac{u_1^2u_2^4}{|u|^6},$$

is a polynomial in τ with coefficients dependent only on ν, u (and independent of μ). We will show that $R^{(1)}_{\nu,u}(\tau)$ is large for almost all u, τ.

The contribution to $R^{(1)}_{\nu,u}(\tau)$ that is constant with respect to τ may be written as $|u|^{-4}W^{(1)}_\nu(u)$ where we define

$$W^{(1)}_\nu(u) = 4\nu_4 u_1^3 u_2^4 + 3\nu_3 u_1^2 u_2^4.$$

(Note that this is independent of μ.) We fix ε_2 with $\varepsilon_1 < \varepsilon_2 < 1$ and define an exceptional set

$$G^\nu = \{u \in B_2(\mathbb{R}^2) : |W^{(1)}_\nu(u)| \le r^{\varepsilon_2}\}.$$

For $u \in B_2(\mathbb{R}^2) \setminus G^\nu$, we set

$$F^\nu_u = \{\tau \in B_1(\mathbb{R}) : |R^{(1)}_{\nu,u}(\tau)| \le C_0 r^{\varepsilon_1}\}$$

for a sufficiently large constant C_0; for $u \in G^\nu$ we set $F^\nu_u = \emptyset$.

For all $(u,\tau) \in B_2(\mathbb{R}^2) \times B_1(\mathbb{R})$ with $u \notin G^\nu$, $\tau \notin F^\nu_u$, we have

$$|R^{(1)}_{\nu,u}(\tau)| \ge C_0 r^{\varepsilon_1},$$

so that for such (u,τ) we have

$$\begin{aligned}\|P_\nu(u+z) - P_\mu(z)\|_\sigma \ge |C[\sigma](u,\tau)| &\ge \left|\frac{u_2^3}{|u|^3}\, C[\sigma](u,\tau)\right| \\ &\ge |R^{(1)}_{\nu,u}(\tau)| - O(r^{\varepsilon_1}) \ge C_0 r^{\varepsilon_1} - O(r^{\varepsilon_1}) \ge r^{\varepsilon_1},\end{aligned}$$

if C_0 is sufficiently large. It then follows from the van der Corput estimate of Lemma 3.1 that

$$|K^{\nu,\mu}_\sharp(u,\tau)| \le C r^{-\varepsilon_1/4} \quad \text{if } u \notin G^\nu \text{ and } \tau \notin F^\nu_u,$$

for a constant C depending only on n, d. (In more general arguments, we will at this stage obtain a constant C that can depend on the fixed coefficients of the set of polynomials $p_2, \dots, p_d$.)

On the other hand, the exceptional sets are small. Indeed G^ν is the set of $u \in B_2(\mathbb{R}^2)$ where $W^{(1)}_\nu(u)$ is small; since the polynomial $W^{(1)}_\nu(u)$ is visibly a sum of two monomials of distinct total degrees, we see that

$$[\![W^{(1)}_\nu(u)]\!]_u \ge |\nu_3| + |\nu_4| \ge r,$$

which implies by Lemma 3.3 that

$$|G^\nu| \le C\left(\frac{r^{\varepsilon_2}}{r}\right)^{1/4} = C r^{-(1-\varepsilon_2)/4}.$$

Furthermore, if $u \in B_2 \setminus G^\nu$, then using the notation $[\![\cdot]\!]_\tau$ to denote the isotropic norm of the coefficients of $R^{(1)}_{\nu,u}(\tau)$ as a polynomial in τ, we have

$$[\![R^{(1)}_{\nu,u}(\tau)]\!]_\tau \ge |u|^{-4}|\,W^{(1)}_\nu(u)| \ge C\,|W^{(1)}_\nu(u)| \ge C r^{\varepsilon_2}.$$

Thus if $u \in B_2 \setminus G^\nu$, it follows from Lemma 3.3 that

$$|F^\nu_u| \le C\left(\frac{C_0 r^{\varepsilon_1}}{r^{\varepsilon_2}}\right)^{1/2} \le C' r^{-(\varepsilon_2-\varepsilon_1)/2}.$$

This suffices to prove Proposition 3.5 in this particular case.

8.4. Case B1 example: a parabolic term is present and $p_2(y) \equiv 0$

For our second example we consider the case where $\mathscr{P} = \{p_4(y)\}$, with $p_4(y) = |y|^4$. In this case

$$P_\lambda(y) = \lambda_4 |y|^4,$$

a purely parabolic polynomial.

In this case we will see that it suffices to reduce our consideration to the coefficient of σ^2. We recall that by assumption $r \leq \|\nu\|, \|\mu\| \leq 2r$. We use the identities $|u+z|^2 = (|u|+\tau)^2 + \sigma^2$ and $|z|^2 = \tau^2 + \sigma^2$ to compute explicitly that

$$P_\nu(u+z) - P_\mu(z) = \sum_{l=0}^{4} C[\sigma^l](u,\tau)\sigma^l$$

where

$$C[\sigma^4](u,\tau) = \nu_4 - \mu_4, \qquad C[\sigma^2](u,\tau) = \nu_4 2(|u|+\tau)^2 - \mu_4 2\tau^2,$$
$$C[\sigma^3](u,\tau) = 0, \qquad C[\sigma](u,\tau) = 0.$$

Here clearly we cannot reduce our consideration to the coefficient of σ, which is identically zero; we instead we use downward induction to eliminate the presence of $\nu_4 - \mu_4$ in the coefficient of σ^2.

Fix $0 < \varepsilon_1 < 1$. We may assume that $|C[\sigma^4](u,\tau)| \leq r^{\varepsilon_1}$, otherwise (8.6) and hence (8.7) would already be known. Under this assumption,

$$|\nu_4 - \mu_4| \leq r^{\varepsilon_1}. \tag{8.13}$$

We next consider the coefficient of σ^2, which we re-write via (8.13) as

$$C[\sigma^2](u,\tau) = (\nu_4 - \mu_4)2\tau^2 + \nu_4 2(|u|^2 + 2|u|\tau) = R^{(2)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}).$$

where

$$R^{(2)}_{\nu,u}(\tau) = \nu_4 2(|u|^2 + 2|u|\tau).$$

We note that the constant term in $R^{(2)}_{\nu,u}(\tau)$ with respect to τ is

$$W^{(2)}_\nu(u) = 2\nu_4 |u|^2.$$

We fix ε_2 with $\varepsilon_1 < \varepsilon_2 < 1$ and set

$$G^\nu = \{u \in B_2(\mathbb{R}^2) : |W^{(2)}_\nu(u)| \leq r^{\varepsilon_2}\},$$

and for $u \in B_2 \setminus G^\nu$ we set

$$F^\nu_u = \{\tau \in B_1(\mathbb{R}) : |R^{(2)}_{\nu,u}(\tau)| \leq C_0 r^{\varepsilon_1}\}$$

for some large absolute constant C_0 to be determined later. If $u \in G^\nu$, we set $F^\nu_u = \emptyset$. For all $(u,\tau) \in B_2(\mathbb{R}^2) \times B_1(\mathbb{R})$ with $u \notin G^\nu$, $\tau \notin F^\nu_u$, we have

$$|R^{(2)}_{\nu,u}(\tau)| \geq C_0 r^{\varepsilon_1},$$

so that for such (u,τ)

$$\|P_\nu(u+z) - P_\mu(z)\|_\sigma \geq |C[\sigma^2](u,\tau)| \geq |R^{(2)}_{\nu,u}(\tau)| - O(r^{\varepsilon_1}) \geq C_0 r^{\varepsilon_1} - O(r^{\varepsilon_1}) \geq r^{\varepsilon_1},$$

if C_0 is sufficiently large. It then follows from the van der Corput estimate of Lemma 3.1 that

$$|K_\sharp^{\nu,\mu}(u,\tau)| \leq C\,r^{-\varepsilon_1/4} \quad \text{if } u \notin G^\nu \text{ and } \tau \notin F_u^\nu.$$

It simply remains to verify that the exceptional sets are small. Since G^ν is the set of $u \in B_2$ where $W_\nu^{(2)}(u)$ is small and the polynomial $W_\nu^{(2)}(u)$ is visibly a homogeneous polynomial with coefficient ν_4, we see that

$$[\![W_\nu^{(2)}(u)]\!]_u \geq |\nu_4| \geq r,$$

which implies by Lemma 3.3 that

$$|G^\nu| \leq C\left(\frac{r^{\varepsilon_2}}{r}\right)^{1/2} = C\,r^{-(1-\varepsilon_2)/2}.$$

Furthermore, if $u \in B_2 \setminus G^\nu$, then

$$[\![R_{\nu,u}^{(2)}(\tau)]\!]_\tau \geq |W_\nu^{(2)}(u)| \geq r^{\varepsilon_2}.$$

Thus if $u \in B_2 \setminus G^\nu$, it follows from Lemma 3.3 that

$$|F_u^\nu| \leq C\left(\frac{C_0\,r^{\varepsilon_1}}{r^{\varepsilon_2}}\right) \leq C' r^{-(\varepsilon_2-\varepsilon_1)}.$$

This suffices to prove Proposition 3.5 in the case under consideration.

8.5. Case B2 example: a parabolic term is present and $p_2(y) \not\equiv 0$

For our third and final example we consider $\mathscr{P} = \{p_4(y), p_2(y)\}$ where $p_4(y) = |y|^4$, $p_2(y) = y_1^2$. Then

$$P_\lambda(y) = \lambda_4\,|y|^4 + \lambda_2\,y_1^2.$$

This requires a hybrid argument that considers the coefficients of both σ and σ^2. We again recall that $r \leq \|\nu\|, \|\mu\| \leq 2r$ and compute that

$$P_\nu(u+z) - P_\mu(z) = \sum_{l=0}^{4} C[\sigma^l](u,\tau)\,\sigma^l$$

where

$$\begin{aligned}
C[\sigma^4](u,\tau) &= \nu_4 - \mu_4,\\
C[\sigma^3](u,\tau) &= 0,\\
C[\sigma^2](u,\tau) &= 2\nu_4(|u|+\tau)^2 + \nu_2\,\frac{u_2^2}{|u|^2} - 2\mu_4\tau^2 - \mu_2\,\frac{u_2^2}{|u|^2},\\
C[\sigma](u,\tau) &= 2\nu_2(|u|+\tau)\,\frac{u_1u_2}{|u|^2} - 2\mu_2\tau\,\frac{u_1u_2}{|u|^2}.
\end{aligned}$$

Fix $0 < \varepsilon_1 < 1$. We may assume that $|C[\sigma^4](u,\tau)| \leq r^{\varepsilon_1}$ since otherwise (8.7) would be known, and deduce that

$$|\nu_4 - \mu_4| \leq r^{\varepsilon_1}. \tag{8.14}$$

We re-write the coefficient of σ^2 as

$$C[\sigma^2](u,\tau) = (\nu_4 - \mu_4)2\tau^2 + \nu_4 2(|u|^2 + 2|u|\tau) + (\nu_2 - \mu_2)\frac{u_2^2}{|u|^2}.$$

We apply (8.14) to conclude that

$$C[\sigma^2](u,\tau) = \nu_4 2(|u|^2 + 2|u|\tau) + (\nu_2 - \mu_2)\frac{u_2^2}{|u|^2} + O(r^{\varepsilon_1}). \tag{8.15}$$

This still includes dependence on $\nu_2 - \mu_2$, which we will eliminate by considering the coefficient of σ.

We may assume that $|C[\sigma](u,\tau)| \leq r^{\varepsilon_1}$, since otherwise (8.7) would be known; this allows us to conclude that

$$(\nu_2 - \mu_2)2\tau\frac{u_1u_2}{|u|^2} = -\nu_2|u|\frac{u_1u_2}{|u|^2} + O(r^{\varepsilon_1}). \tag{8.16}$$

We multiply (8.15) by $2\tau\frac{u_1}{|u|}$ and apply (8.16) to deduce that

$$C[\sigma^2](u,\tau)\cdot 2\tau\frac{u_1}{|u|} = R^{(2)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}),$$

where we define

$$R^{(2)}_{\nu,u}(\tau) = \nu_4 \cdot 4\tau\frac{u_1}{|u|}(|u|^2 + 2|u|\tau) - \nu_2|u|\Big(\frac{u_1u_2^2}{|u|^3}\Big).$$

Our assumption is that $\|\nu\| = |\nu_4| + |\nu_2| \approx r$. We now must break into two further cases, depending on whether $|\nu_4| \geq r/2$ or $|\nu_2| \geq r/2$. In the first case, we single out the coefficient of the linear term in τ in $R^{(2)}_{\nu,u}(\tau)$, which takes the form $|u|^{-1}W^{(2,1)}_\nu(u)$, where we define the polynomial

$$W^{(2,1)}_\nu(u) = 4\nu_4u_1|u|^2 = 4\nu_4(u_1^3 + u_1u_2^2).$$

We fix ε_2 with $\varepsilon_1 < \varepsilon_2 < 1$ and set

$$G^\nu = \{u \in B_2(\mathbb{R}^2) : |W^{(2,1)}_\nu(u)| \leq r^{\varepsilon_2}\},$$

and for $u \in B_2 \setminus G^\nu$ we set

$$F^\nu_u = \{\tau \in B_1(\mathbb{R}) : |R^{(2)}_{\nu,u}(\tau)| \leq C_0r^{\varepsilon_1}\}$$

for some large absolute constant C_0 to be determined. If $u \in G^\nu$, we set $F^\nu_u = \emptyset$. For all $(u,\tau) \in B_2(\mathbb{R}^2) \times B_1(\mathbb{R})$ with $u \notin G^\nu$, $\tau \notin F^\nu_u$,

$$|R^{(2)}_{\nu,u}(\tau)| \geq C_0r^{\varepsilon_1},$$

so that for such (u,τ) we have

$$\begin{aligned}\|P_\nu(u+z) - P_\mu(z)\|_\sigma \geq |C[\sigma^2](u,\tau)| &\gtrsim \Big|C[\sigma^2](u,\tau)\cdot 2\tau\frac{u_1}{|u|}\Big| \\ &\geq |R^{(2)}_{\nu,u}(\tau)| - O(r^{\varepsilon_1}) \geq C_0r^{\varepsilon_1} - O(r^{\varepsilon_1}) \geq r^{\varepsilon_1},\end{aligned}$$

if C_0 is sufficiently large. It then follows from the van der Corput estimate of Lemma 3.1 that

$$|K^{\nu,\mu}_\sharp(u,\tau)| \leq Cr^{-\varepsilon_1/4} \quad \text{if } u \notin G^\nu \text{ and } \tau \notin F^\nu_u.$$

It simply remains to verify that the exceptional sets are small. Since G^ν is the set of $u \in B_2$ where $W_\nu^{(2,1)}(u)$ is small, and

$$[\![W_\nu^{(2)}(u)]\!]_u \geq |\nu_4| \geq r/2,$$

we see by Lemma 3.3 that

$$|G^\nu| \leq C\left(\frac{r^{\varepsilon_2}}{r}\right)^{1/2} = C\,r^{-(1-\varepsilon_2)/2}.$$

Furthermore, if $u \in B_2 \setminus G^\nu$, then

$$[\![R_{\nu,u}^{(2)}(\tau)]\!]_\tau \geq |u|^{-1}\,|W_\nu^{(2,1)}(u)| \geq C\,|W_\nu^{(2,1)}(u)| \geq C\,r^{\varepsilon_2}.$$

Thus if $u \in B_2 \setminus G^\nu$, it follows from Lemma 3.3 that

$$|F_u^\nu| \leq C\left(\frac{C_0\,r^{\varepsilon_1}}{r^{\varepsilon_2}}\right)^{1/2} \leq C' r^{-(\varepsilon_2-\varepsilon_1)/2}.$$

In the remaining case when $|\nu_2| \geq r/2$, we single out the coefficient of the constant term in τ in $R_{\nu,u}^{(2)}(\tau)$, setting

$$W_\nu^{(2,0)}(u) = -\nu_2\,u_1\,u_2^2,$$

and proceed analogously. This suffices to prove Proposition 3.5 in the case under consideration.

9. Van der Corput estimates for kernels: Part II

9.1. Partition of unity and change of variables

We now state and prove a general version of Proposition 3.5 for $K_\sharp^{\nu,\mu}$ in all dimensions $n \geq 2$; for this we first require some notation. In the case $n = 2$, we were motivated to make the change of variables in (7.17) because setting $\tau = u \cdot z/|u|$ captures the behavior of z in the expression

$$|u+z|^2 - |z|^2 = |u|^2 + 2u\cdot z = |u|^2 + 2|u|\tau. \tag{9.1}$$

We continue to define $\tau = u \cdot z/|u|$ in the case of general dimension, but for $n \geq 3$ there is no longer a unique choice of σ orthogonal to τ, and we must be more careful.

We now set some notation. For any variable $u \in \mathbb{R}^n$, let $u^{(j)}$ denote the variable in $\mathbb{R}^{n-1}$ that omits the j-th coordinate of u. Similarly, for any multi-index $\alpha \in \mathbb{Z}^n_{\geq 0}$ we let $\alpha^{(j)} \in \mathbb{Z}^{n-1}_{\geq 0}$ denote the multi-index omitting α_j. Given coordinates $(u,z) \in \mathbb{R}^n \times \mathbb{R}^n \setminus \{(0,0)\}$, we may fix a coordinate $1 \leq l \leq n$ where $u_l \neq 0$, and make the change of variables $z \mapsto (\tau,\sigma) \in \mathbb{R} \times \mathbb{R}^{n-1}$ defined by

$$\tau = \frac{u\cdot z}{|u|} \tag{9.2}$$

$$\sigma = \frac{u^{(l)}\tau - |u|z^{(l)}}{u_l}. \tag{9.3}$$

Correspondingly, z is defined implicitly in terms of (τ, σ) by the relations

$$z^{(l)} = \frac{u^{(l)}\tau - u_l\sigma}{|u|} \tag{9.4}$$

$$z_l = \frac{u_l\tau + u^{(l)} \cdot \sigma}{|u|}. \tag{9.5}$$

The intuition behind these choices is as follows: τ again captures the behavior of z in the relation (9.1), while σ is defined so that (9.4), which is the higher dimensional analogue of the equation defining z_2 in (8.4), holds. Indeed, the relation (9.4) specifies σ uniquely, and one can then verify that (9.5), which is analogous to the equation defining z_1 in (8.4), continues to hold. By explicit computation, the Jacobian associated to this change of variables is $(|u_l|/|u|)^{n-2}$.

We choose a partition of unity

$$1 = \sum_{l=1}^{n} W_l(s) \tag{9.6}$$

for $s \in \mathbb{S}^{n-1} \subseteq \mathbb{R}^n$ such that for each $1 \le l \le n$, $W_l \in C_c^\infty(\mathbb{S}^{n-1})$ and $W_l(s)$ is supported where

$$\left| \frac{s_l}{s} \right| \ge c_0 \tag{9.7}$$

for some fixed $c_0 > 0$.

In general, given an integral of the form

$$\mathcal{L}(u) = \int_{\mathbb{R}^n} L(u, z)\, dz$$

with $u \in \mathbb{R}^n$ and $L(u, z)$ supported where $|u| \le 2$, $|z| \le 1$, we will decompose this as $\mathcal{L} = \sum_{l=1}^{n} \mathcal{L}_l$ where

$$\mathcal{L}_l(u) = W_l\Big(\frac{u}{|u|}\Big) \int_{\mathbb{R}^n} L(u, z)\, dz.$$

For each $1 \le l \le n$, within the integral defining $\mathcal{L}_l$ we will make the change of variables defined by (9.4) and (9.5) according to the l-th coordinate, in order to obtain

$$\mathcal{L}_l(u) = \Big(\frac{|u_l|}{|u|}\Big)^{n-2} W_l\Big(\frac{u}{|u|}\Big) \int_{\mathbb{R}} \int_{\mathbb{R}^{n-1}} \tilde{L}(u, \tau; \sigma)\, d\sigma\, d\tau,$$

in which the function $\tilde{L}(u, \tau; \sigma)$ is implicitly defined in terms of $L(u, z)$ by (9.4) and (9.5).

In the case of $\mathbb{R}^2 \times \mathbb{R}$, τ and σ enjoyed the convenient relation that $|z|^2 = |\tau|^2 + |\sigma|^2$, which allowed us to draw the conclusion that $|\tau|, |\sigma| \le 1$ as long as $|z| \le 1$. This identity no longer need hold in higher dimensions; instead one can compute that according to the change of variables (9.4) and (9.5),

$$|z|^2 = z_l^2 + |z^{(l)}|^2 = |\tau|^2 + \frac{|u^{(l)}|^2 |\sigma|^2 \cos^2\theta_l + u_l^2 |\sigma|^2}{|u|^2}, \tag{9.8}$$

where θ_l is the angle between $u^{(l)}$ and σ in $\mathbb{R}^{n-1}$. In the case of $\mathbb{R}^2 \times \mathbb{R}$, this angle always had $\cos\theta = \pm 1$, so we trivially obtained the upper bound $|\sigma| \leq 1$ for the range of integration in σ, but in higher dimensions this may not be the case. Throwing away the non-negative cosine term and $|\tau|^2$, and using only $|z| \leq 1$, (9.8) trivially yields the bound

$$|\sigma| \leq \frac{|u|}{|u_l|}. \tag{9.9}$$

By construction, this is a bounded region, because of the key restriction provided by $W_l(u/|u|)$, which requires that $|u_l/u| \geq c_0 > 0$; this is our motivation for the partition of unity.

We will deploy this partition of unity and change of variables in two places: to treat the term **I** in Section 7.3.1, and to treat the kernel ${}^{(1)}\mathcal{K}^{\nu,\mu}$ in Section 7.4.1. In each case, we will require a corresponding van der Corput estimate for a family of oscillatory integrals $K^{\nu,\mu}_{\sharp,l}$ for $1 \leq l \leq n$, each corresponding to a different component in the partition of unity and the relevant change of variables. We state the necessary bounds, a generalization of Proposition 3.5, below.

Proposition 9.1 ($K^{\nu,\mu}_{\sharp}$ van der Corput, general $n \geq 2$)**.** *Fix a dimension $n \geq 2$ and a degree $d \geq 2$. Let $\mathscr{P} = \{p_2(y), p_3(y), \ldots, p_d(y)\}$ be a set of real-valued polynomials on $\mathbb{R}^n$, where each $p_j(y)$ is homogeneous of degree j, and $p_2(y) \neq C|y|^2$ for any nonzero constant C. Let $\Lambda = \Lambda(\mathscr{P}) = \{2 \leq m \leq d \colon p_m(y) \not\equiv 0\}$. For $\nu = (\nu_2, \ldots, \nu_d) \in \mathbb{R}^{d-1}$, let*

$$P_\nu(y) = \sum_{m=2}^{d} \nu_m p_m(y) \quad \text{and} \quad \|\nu\| = \sum_{m\in\Lambda} |\nu_m|,$$

and define $P_\mu(y)$ and $\|\mu\|$ similarly for $\mu = (\mu_2, \ldots, \mu_d) \in \mathbb{R}^{d-1}$.

Recall the partition of unity given by W_l for $1 \leq l \leq n$. Given a C^1 function $\Psi(u, z)$ supported on $B_2 \times B_1 \subset \mathbb{R}^n \times \mathbb{R}^n$, define for each $1 \leq l \leq n$ the integral

$$K^{\nu,\mu}_{\sharp,l}(u, \tau) = \Big(\frac{|u_l|}{|u|}\Big)^{n-2} W_l\Big(\frac{u}{|u|}\Big) \int_{\mathbb{R}^{n-1}} e^{iP_\nu(u+z) - iP_\mu(z)}\, \Psi(u, z)\, d\sigma,$$

where z is defined implicitly in terms of u, τ, σ by

$$\tau = \frac{u \cdot z}{|u|} \tag{9.10}$$

$$\sigma = \frac{u^{(l)}\tau - |u|z^{(l)}}{u_l}. \tag{9.11}$$

Suppose furthermore that

$$\|\Psi\|_{C^1(\mathbb{R})} := \sup_{(u,z)\in B_2(\mathbb{R}^n)\times B_1(\mathbb{R}^n)} \Big(|\Psi(u, z)| + \Big|\frac{\partial}{\partial\sigma}\Psi(u, z)\Big|\Big) \leq 1.$$

Then there exists a small constant $\delta > 0$ (depending only on d) such that the following holds: if ν, μ satisfy

$$r \leq \|\nu\|, \|\mu\| \leq 2r$$

for some $r \geq 1$*, then there exists a small set* $G^\nu \subset B_2(\mathbb{R}^n)$*, and for each* $u \in B_2(\mathbb{R}^n)$ *a small set* $F_u^\nu \subset B_1(\mathbb{R})$*, such that*

$$|G^\nu| \leq C r^{-\delta}, \quad |F_u^\nu| \leq C r^{-\delta} \quad \text{for all } u \in B_2(\mathbb{R}^n), \tag{9.12}$$

and such that for all $1 \leq l \leq n$*,*

$$|K_{\sharp,l}^{\nu,\mu}(u,\tau)| \leq C\left(r^{-\delta}\chi_{B_2}(u)\chi_{B_1}(\tau) + \chi_{G^\nu}(u)\chi_{B_1}(\tau) + \chi_{B_2}(u)\chi_{F_u^\nu}(\tau)\right). \tag{9.13}$$

The choices of the small sets G^ν *and* F_u^ν *are independent of both* μ *and the amplitude* Ψ*; the constants* C *in the above upper bounds depend on* n, d *and the fixed set of polynomials* $p_2, \ldots, p_d$*.*

We note that due to the support of $W_l(u/|u|)$ we have the upper bound (9.9) for the support of the integral in σ, leading to the trivial bound

$$|K_{\sharp,l}^{\nu,\mu}(u,\tau)| \leq C\, c_0^{-(n-1)}\, \chi_{B_2}(u)\, \chi_{B_1}(\tau), \tag{9.14}$$

for the finite nonzero constant c_0. Thus the import of Proposition 9.1 is to extract decay in r under the hypotheses of the proposition.

9.2. Computing the phase

We now prove Proposition 9.1 in full generality. We will construct for each index $1 \leq l \leq n$ a pair of exceptional sets G^ν and F_u^ν; taking the union over n of these sets will clearly give exceptional sets that work for all l simultaneously and still satisfy the small measure conditions (9.12). As it will be notationally convenient, we will focus on the case $l = n$, but the argument we present does not depend on this choice in any way beyond notation. In particular, from now on, τ and σ will refer to the definitions (9.10) and (9.11) with the choice $l = n$.

We recall the polynomials P_ν, P_μ defined as in Proposition 9.1, and compute the phase:

Lemma 9.2. *The phase* $P_\nu(u+z) - P_\mu(z)$ *of* $K_{\sharp,n}^{\nu,\mu}(u,\tau)$ *is a polynomial in* $\sigma \in \mathbb{R}^{n-1}$*, which we will denote by*

$$P_\nu(u+z) - P_\mu(z) = \sum_{0 \leq |\beta| \leq d} C[\sigma^\beta](u,\tau)\,\sigma^\beta, \tag{9.15}$$

where for each multi-index β*, if* $|\beta| = l$ *then the coefficient* $C[\sigma^\beta](u,\tau)$ *is given by*

$$C[\sigma^\beta](u,\tau) = \sum_{m=l}^{d} \left(\nu_m(|u|+\tau)^{m-l} - \mu_m \tau^{m-l}\right) B_{m,\beta}\Big(\frac{u}{|u|}\Big), \tag{9.16}$$

in which

$$B_{m,\beta}(w) = \sum_{|\alpha|=m} c_\alpha\, A_{\alpha,\beta}(w). \tag{9.17}$$

Here $A_{\alpha,\beta}(w)$ *is a polynomial in* w *that is homogeneous of degree* $|\alpha|$*, and the coefficients* c_α *are fixed once and for all by the choice of the polynomials* $\mathscr{P} = \{p_m : m = 2, \ldots, d\}$*.*

To prove Lemma 9.2, we define a family of polynomials $A_{\alpha,\beta}(w)$ acting on $w \in B_1(\mathbb{R}^n)$ and parametrized by multi-indices $\alpha \in \mathbb{Z}^n_{\geq 0}$ and $\beta \in \mathbb{Z}^{n-1}_{\geq 0}$ (note the differing dimensions). For any multi-index $\alpha \in \mathbb{Z}^n_{\geq 0}$, we specify a polynomial $A_{\alpha,\beta}$ for each $\beta \in \mathbb{Z}^{n-1}_{\geq 0}$ with $|\beta| \leq |\alpha|$ by the defining relation

$$\sum_{|\beta|\leq|\alpha|} A_{\alpha,\beta}(w)\sigma^\beta = (w_n + w^{(n)} \cdot \sigma)^{\alpha_n} (w^{(n)} - w_n\sigma)^{\alpha^{(n)}}. \tag{9.18}$$

It is clear by inspection that with this definition, $A_{\alpha,\beta}$ is a polynomial in w homogeneous of degree $|\alpha|$. (Moreover, if $|\beta| \geq |\alpha|$ then $A_{\alpha,\beta}$ is the zero polynomial.) Note as well that these polynomials are defined purely in terms of combinatorial coefficients, and are independent of the fixed polynomials p_m and of stopping-times. For example, in the case of dimension $n = 2$, we represent the one-dimensional index $\beta \in \mathbb{Z}_{\geq 0}$ by l and compute that for each $1 \leq l \leq |\alpha|$,

$$A_{\alpha,l}(w) := \sum_{\substack{0\leq j\leq\alpha_1 \\ 0\leq k\leq\alpha_2 \\ j+k=l}} (-1)^j \binom{\alpha_1}{j}\binom{\alpha_2}{k} w_1^{\alpha_1-j+k}\, w_2^{\alpha_2-k+j}. \tag{9.19}$$

This is visibly a homogeneous polynomial in $w \in \mathbb{R}^2$ of degree $|\alpha|$. In arbitrary dimensions, it is too cumbersome to perform an explicit binomial expansion, so we use the implicit definition (9.18) instead.

We note that one may also generalize the relation (9.18) to

$$\sum_{|\beta|\leq|\alpha|} A_{\alpha,\beta}(w)\, T^{|\alpha|-|\beta|}\sigma^\beta = (w_n T + w^{(n)} \cdot \sigma)^{\alpha_n}\, (w^{(n)}T - w_n\sigma)^{\alpha^{(n)}}, \tag{9.20}$$

for a generic variable $T \in \mathbb{R}$. To proceed with the proof of Lemma 9.2, we use the relations (9.4) and (9.5) to compute that

$$u_n + z_n = \frac{u_n}{|u|}(|u| + \tau) + \frac{u^{(n)} \cdot \sigma}{|u|},$$
$$u^{(n)} + z^{(n)} = \frac{u^{(n)}}{|u|}(|u| + \tau) - \frac{u_n}{|u|}\sigma.$$

We may now expand $P_\nu(u + z)$ as the expression

$$\sum_{m=2}^{d} \nu_m \sum_{|\alpha|=m} c_\alpha \Big(\frac{u_n}{|u|}(|u| + \tau) + \frac{u^{(n)}}{|u|} \cdot \sigma\Big)^{\alpha_n} \Big(\frac{u^{(n)}}{|u|}(|u| + \tau) - \frac{u_n}{|u|}\,\sigma\Big)^{\alpha^{(n)}}.$$

Applying (9.20) with $w = u/|u|$ and $T = (|u| + \tau)$ we now see that

$$\begin{aligned} P_\nu(u + z) &= \sum_{m=2}^{d} \nu_m \sum_{|\alpha|=m} c_\alpha \Big(\sum_{|\gamma|\leq|\alpha|} A_{\alpha,\gamma}\Big(\frac{u}{|u|}\Big)(|u| + \tau)^{|\alpha|-|\gamma|}\,\sigma^\gamma\Big) \\ &= \sum_{0\leq|\beta|\leq d} \Big(\sum_{m\geq|\beta|} \nu_m \sum_{|\alpha|=m} c_\alpha\, A_{\alpha,\beta}\Big(\frac{u}{|u|}\Big)(|u| + \tau)^{|\alpha|-|\beta|}\Big)\,\sigma^\beta. \end{aligned}$$

We compute similarly that $P_\mu(z)$ can be written as

$$\begin{aligned} P_\mu(z) &= \sum_{m=2}^{d} \mu_m \sum_{|\alpha|=m} c_\alpha \Big(\frac{u_n}{|u|}\tau + \frac{u^{(n)}}{|u|}\sigma\Big)^{\alpha_n} \Big(\frac{u^{(n)}}{|u|}\tau - \frac{u_n}{|u|}\sigma\Big)^{\alpha^{(n)}} \\ &= \sum_{0\le|\beta|\le d} \Big(\sum_{m\ge|\beta|} \mu_m \sum_{|\alpha|=m} c_\alpha A_{\alpha,\beta}\Big(\frac{u}{|u|}\Big)\tau^{|\alpha|-|\beta|}\Big)\sigma^\beta \end{aligned} \tag{9.21}$$

with the same functions $A_{\alpha,\beta}$, and this completes the proof of Lemma 9.2.

9.3. Preliminary properties of $A_{\alpha,\beta}$ and $B_{m,\beta}$

We summarize the key properties of the polynomials $A_{\alpha,\beta}$, which we will prove in Section 9.7.

Lemma 9.3 (Properties of $A_{\alpha,\beta}$).

(1) *Suppose that $|\beta| = 1$, so that for some $1 \le j \le n-1$ we may write $\beta = e_j$ where $e_j = (0, \dots, 1, \dots, 0)$ is the j-th unit vector. Then for any $|\alpha| \ge 1$,*

$$A_{\alpha,e_j}(w) = \alpha_n w^{\alpha+e_j-e_n} - \alpha_j w^{\alpha-e_j+e_n}.$$

(2) *If $\beta = 2e_j$ with $1 \le j \le n-1$, then for any $|\alpha| \ge 2$,*

$$A_{\alpha,2e_j}(w) = \binom{\alpha_n}{2} w^{\alpha+2e_j-2e_n} - \alpha_n \alpha_j w^\alpha + \binom{\alpha_j}{2} w^{\alpha-2e_j+2e_n}.$$

(3) *If $|\beta| = |\alpha|$, then $A_{\alpha,\beta}(\omega)$ is a monomial,*

$$A_{\alpha,\beta}(w) = C_{\alpha,\beta}(-1)^{|\alpha^{(n)}|} w_n^{|\alpha^{(n)}|} (w^{(n)})^{\beta-\alpha^{(n)}},$$

for a non-negative combinatorial constant $C_{\alpha,\beta}$. Moreover, given any multi-index $\alpha \in \mathbb{Z}^n_{\ge 0}$ with $|\alpha| \ge 1$, there exists a multi-index $\beta \in \mathbb{Z}^{n-1}_{\ge 0}$ with $|\beta| = |\alpha|$ such that the coefficient $C_{\alpha,\beta}$ is nonzero.

We also record the key properties of the polynomials $B_{m,\beta}(w)$ which we require; these are proved in Sections 9.8 to 9.10. While these appear to be simple properties of a combinatorial nature, they encode the advantages of the restricted class of polynomials we consider, and lie at the heart of the bound for $K_\sharp^{\nu,\mu}$. As before, we let

$$\Lambda = \{2 \le m \le d\colon p_m(y) \not\equiv 0\}.$$

We recall that for $m \in \Lambda$, $p_m(y)$ is parabolic if $p_m(y) = C|y|^m$ for some constant $C \neq 0$, and non-parabolic otherwise.

Lemma 9.4 (Properties of $B_{m,\beta}$).

(1) *The polynomial $B_{m,\beta}$ depends only on p_m and β (and not on any other p_j, $j \neq m$). In particular, if $p_m \equiv 0$, then $B_{m,\beta} \equiv 0$ for all β.*

(2) *There is a constant C_B depending only on the degree d, the dimension n, and the fixed coefficients c_α such that for all $1 \le m \le d$ and all $1 \le |\beta| \le d$,*

$$|B_{m,\beta}(w)| \le C_B, \quad \text{for all } |w| \le 1.$$

(3) *$B_{m,\beta}(w)$ is a homogeneous polynomial in w of degree m.*

(4) *If $m \in \Lambda$, then there exists some $|\beta| = m$ such that $B_{m,\beta}$ is not the zero polynomial.*

(5) *If $m \in \Lambda$ and $p_m(y)$ is parabolic, then $B_{m,\beta}(w)$ is the zero polynomial for all β with $|\beta|$ odd.*

(6) *If $m \in \Lambda$ and $p_m(y)$ is not parabolic, then there exists β with $|\beta| = 1$ such that $B_{m,\beta}(w)$ is not the zero polynomial.*

(7) *If $m \in \Lambda$ and $p_m(y)$ is parabolic, then there exists some β with $|\beta| = 2$ such that $B_{m,\beta}(w)$ is not the zero polynomial.*

We now demonstrate how to derive the $K^{\nu,\mu}_{\sharp,n}$ bound of Proposition 9.1 from these properties. As in the specific examples we considered in Section 8, the general strategy of the proof will be to understand when $|C[\sigma^\beta](u,\tau)|$ is small, since if $|C[\sigma^\beta](u,\tau)|$ is large for some $1 \leq |\beta| \leq d$, then we may bound the kernel $K^{\nu,\mu}_{\sharp,n}$ by a van der Corput estimate.

For each $1 \leq |\beta| \leq d$, We re-write the expression (9.16) defining $C[\sigma^\beta](u,\tau)$ for $|\beta| = l$ as

$$
\begin{aligned}
C[\sigma^\beta](u,\tau) &= \sum_{m=l}^{d} (\nu_m - \mu_m)\,\tau^{m-l}\, B_{m,\beta}\Big(\frac{u}{|u|}\Big) \\
&\quad + \sum_{m=l+1}^{d} \nu_m((|u|+\tau)^{m-l} - \tau^{m-l})\, B_{m,\beta}\Big(\frac{u}{|u|}\Big). \qquad (9.22)
\end{aligned}
$$

We first use a downward induction process to eliminate the presence of $\mu = (\mu_2,\dots,\mu_d)$ in the coefficient $C[\sigma^\beta](u,\tau)$ that we hope to show is large. The expression (9.22) makes clear that for each $1 \leq |\beta| \leq d$, $C[\sigma^\beta](u,\tau)$ depends only on μ_m with $m \geq |\beta|$. Our strategy thus relies on the fact that if $|C[\sigma^\beta](u,\tau)|$ is small for all $|\beta| \geq l_0$, then for all $l \geq l_0$ one can essentially rewrite μ_l in terms of $\nu = (\nu_2,\dots,\nu_d)$, u and τ. Note also that $C[\sigma^\beta](u,\tau)$ is a polynomial of degree at most $d - |\beta|$ in τ.

We now make precise the process of inductively eliminating the presence of μ-coefficients. We recall that we have by assumption $r \leq \|\nu\|, \|\mu\| \leq 2r$. We will consider three cases, motivated by the key examples we considered in Section 8. Case A: a non-parabolic term dominates, namely there exists $m \in \Lambda$, with $p_m(y)$ not parabolic, such that $|\nu_m| \simeq r$; Case B: a parabolic term dominates, namely there exists $m \in \Lambda$ with $p_m(y)$ parabolic, such that $|\nu_m| \simeq r$; we then further subdivide this into case B1 when $p_2(y) \equiv 0$ and case B2 when $p_2(y) \not\equiv 0$.

9.4. Case A: a non-parabolic term dominates

In this case we aim to show that there exists a multi-index $\beta = \beta^*$ with $|\beta^*| = 1$ such that $|C[\sigma^{\beta^*}](u,\tau)|$ is large for "most" u and τ.

Fix $0 < \varepsilon_1 < \varepsilon_2 < 1$. If $|C[\sigma^\beta](u,\tau)| \geq r^{\varepsilon_1}$ for some $2 \leq |\beta| \leq d$, then $|K^{\nu,\mu}_{\sharp,n}(u,\tau)| \leq C r^{-\varepsilon_1/d}$ by the van der Corput estimate of Lemma 3.1, as desired (with constant C dependent on n, d). So we may assume that

$$|C[\sigma^\beta](u,\tau)| \le r^{\varepsilon_1} \quad \text{for all } 2 \le |\beta| \le d. \tag{9.23}$$

Applying the assumption (9.23) to (9.22) shows that for each $2 \le l \le d$, for every $|\beta| = l$ we have

$$\begin{aligned}(\nu_l - \mu_l) B_{l,\beta}\Big(\frac{u}{|u|}\Big) = &- \sum_{m=l+1}^{d} (\nu_m - \mu_m)\, \tau^{m-l}\, B_{m,\beta}\Big(\frac{u}{|u|}\Big) \\ &- \sum_{m=l+1}^{d} \nu_m((|u|+\tau)^{m-l} - \tau^{m-l})\, B_{m,\beta}\Big(\frac{u}{|u|}\Big) + O(r^{\varepsilon_1}).\end{aligned} \tag{9.24}$$

We now use Statement 4 of Lemma 9.4 to conclude that for each $2 \le l \le d$ with $l \in \Lambda$, there exists some β with $|\beta| = l$ for which $B_{l,\beta}$ is not the zero polynomial. For each $l \in \Lambda$ we will pick such a distinguished index β and denote it by $\beta(l)$. We thus obtain for each $l \in \Lambda$ a relation (9.24) specialized to the distinguished index $\beta(l)$ that will allow us to express $\nu_l - \mu_l$ in terms of $\nu_m - \mu_m$ for $m > l$, and some harmless terms involving only u, τ and $\nu = (\nu_2, \dots, \nu_d)$.

By a downward induction argument on l, we conclude that for all $2 \le l \le d$ we may write

$$(\nu_l - \mu_l) \prod_{\substack{j \ge l \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) = S^{(l)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}), \tag{9.25}$$

where $S^{(l)}_{\nu,u}(\tau)$ is a polynomial in τ whose coefficients depend only on ν and u (but not μ).

We now feed this information into an analysis of $C[\sigma^\beta](u,\tau)$ for indices $|\beta| = 1$. We fix any index β with $|\beta| = 1$ and see that in (9.22), $C[\sigma^\beta](u,\tau)$ may be expressed as

$$\begin{aligned}C[\sigma^\beta](u,\tau) = &\sum_{m=2}^{d} (\nu_m - \mu_m)\, \tau^{m-1}\, B_{m,\beta}\Big(\frac{u}{|u|}\Big) \\ &+ \sum_{m=2}^{d} \nu_m((|u|+\tau)^{m-1} - \tau^{m-1})\, B_{m,\beta}\Big(\frac{u}{|u|}\Big).\end{aligned} \tag{9.26}$$

Note in particular that the first sum begins with $m = 2$, since $p_1(y) \equiv 0$ and so we have $B_{1,\beta} \equiv 0$ for all β; thus all terms in the first sum are linear or higher order with respect to τ. We will use (9.25) to eliminate the presence of μ in the first sum in (9.26), and then we will single out the constant terms with respect to τ in the resulting polynomial (which will come only from the second sum on the right-hand side of (9.26)). To proceed with this plan, we multiply (9.26) through by the polynomial

$$\prod_{\substack{j \ge 2 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big), \tag{9.27}$$

and substitute (9.25) wherever possible. (Recall that by construction, (9.27) is not identically zero.) One then concludes that

$$C[\sigma^\beta](u,\tau)\prod_{\substack{j\geq 2\\ j\in\Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) = R^{(\beta)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}), \tag{9.28}$$

where $R^{(\beta)}_{\nu,u}(\tau)$ is a polynomial in τ of degree $\leq d-1$ whose coefficients depend only on ν and u (but not μ). (The superscript β reflects that this comes from the coefficient $C[\sigma^\beta](u,\tau)$.) In fact, we will write $R^{(\beta)}_{\nu,u}$ as a constant term in τ, plus higher powers of τ. The constant term in τ arise from the contribution from the second term of (9.26) only, so

$$R^{(\beta)}_{\nu,u}(\tau) = \sum_{m=2}^{d} \nu_m\,|u|^{m-1}\,B_{m,\beta}\Big(\frac{u}{|u|}\Big)\prod_{\substack{j\geq 2\\ j\in\Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) + R^{(\beta,1)}_{\nu,u}(\tau),$$

where the first term is constant with respect to τ, and $R^{(\beta,1)}_{\nu,u}(\tau)$ is a polynomial in τ with no constant term and with coefficients that depend only on u,ν (and not μ). Let $W^{(\beta)}_\nu(u)$ be the polynomial in u defined by

$$W^{(\beta)}_\nu(u) := \sum_{m=2}^{d} \nu_m\,B_{m,\beta}(u)\prod_{\substack{j\geq 2\\ j\in\Lambda}} B_{j,\beta(j)}(u). \tag{9.29}$$

We note that since $B_{m,\beta}$ is homogeneous of degree m, the degree of $W^{(\beta)}_\nu(u)$ with respect to u is at most

$$s_0 = d + \sum_{\substack{j\geq 2\\ j\in\Lambda}} j. \tag{9.30}$$

Moreover, we see that

$$R^{(\beta)}_{\nu,u}(\tau) = |u|^{-s_1}\,W^{(\beta)}_\nu(u) + R^{(\beta,1)}_{\nu,u}(\tau), \quad \text{where} \quad s_1 = 1 + \sum_{\substack{j\geq 2\\ j\in\Lambda}} j.$$

Recall that in the current case, we assumed that there exists $m\in\Lambda$, say m_0, with $p_{m_0}(y)$ not parabolic, such that $|\nu_{m_0}|\simeq r$. For that m_0, we know that there exists an index β with $|\beta|=1$ such that $B_{m_0,\beta}(u)$ is not identically zero by Statement (6) of Lemma 9.4; we will denote this distinguished β by β^*; this choice of a distinguished index β^* depends only on the original choice of polynomial p_{m_0}. We now define our exceptional sets, with respect to the fixed β^* with $|\beta^*|=1$ determined above. Recall that $0<\varepsilon_1<\varepsilon_2<1$ and set

$$G^\nu := \{u\in B_2(\mathbb{R}^n)\colon |W^{(\beta^*)}_\nu(u)|\leq r^{\varepsilon_2}\},$$

and for $u\in B_2\setminus G^\nu$, let

$$F^\nu_u := \{\tau\in B_1(\mathbb{R})\colon |R^{(\beta^*)}_{\nu,u}(\tau)|\leq C_0\,r^{\varepsilon_1}\}$$

for some large absolute constant C_0 to be determined. Also define $F_u^\nu := \emptyset$ if $u \in G^\nu$. Then for all $(u,\tau) \in B_2(\mathbb{R}^n) \times B_1(\mathbb{R})$ with $u \notin G^\nu$, $\tau \notin F_u^\nu$, we have

$$|R_{\nu,u}^{(\beta^*)}(\tau)| \geq C_0 r^{\varepsilon_1},$$

so that for such (u,τ), we conclude from (9.28) that

$$\begin{aligned} |C[\sigma^{\beta^*}](u,\tau)| &\geq C_B^{-|\Lambda|} \left| C[\sigma^{\beta^*}](u,\tau) \prod_{\substack{j\geq 2\\ j\in\Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big)\right| \\ &\geq C_B^{-|\Lambda|}\big(|R_{\nu,u}^{(\beta^*)}(\tau)| - O(r^{\varepsilon_1})\big) \geq C_B^{-|\Lambda|}\big(C_0 r^{\varepsilon_1} - O(r^{\varepsilon_1})\big) \geq C_B^{-|\Lambda|} r^{\varepsilon_1} \end{aligned}$$

if C_0 is sufficiently large. It follows from the van der Corput estimate of Lemma 3.1 that

$$|K_{\sharp,n}^{\nu,\mu}(u,\tau)| \leq C r^{-\varepsilon_1/d} \quad \text{if } u \notin G^\nu \text{ and } \tau \notin F_u^\nu,$$

for some fixed constant C dependent only on n, d, and the initial choice of the polynomials $p_2, \ldots, p_d$.

On the other hand, $K_{\sharp,n}^{\nu,\mu}(u,\tau)$ is supported on $B_2(\mathbb{R}^n) \times B_1(\mathbb{R})$, and is bounded by some uniform constant C (see (9.14)). Hence to complete the proof of Proposition 9.1 in this case, we only need to show that G^ν and F_u^ν are sets of small measures. Now G^ν is the set of $u \in B_2$ where $W_\nu^{(\beta^*)}(u)$ is small, and in (9.29), $W_\nu^{(\beta^*)}(u)$ is represented as a sum of homogeneous polynomials of different total degrees. In particular, the coefficient ν_{m_0} for which $|\nu_{m_0}| \simeq r$ appears in $W_\nu^{(\beta^*)}(u)$ as the coefficient of a homogeneous polynomial that is by construction not identically zero, and which has different total degree from all other terms in $W_\nu^{(\beta^*)}(u)$. Thus we see that

$$[\![W_\nu^{(\beta^*)}]\!]_u \geq C|\nu_{m_0}| \geq C r,$$

which implies by Lemma 3.3 that

$$|G^\nu| \leq C\Big(\frac{r^{\varepsilon_2}}{r}\Big)^{1/s_0} = C r^{-(1-\varepsilon_2)/s_0},$$

since the degree of $W_\nu^{(\beta^*)}(u)$ is at most s_0. (Here, and below, the constant C depends on the polynomials $B_{m,\beta}$, and hence on n, d and the fixed polynomials $p_2, \ldots, p_d$.) Furthermore, if $u \in B_2 \setminus G^\nu$, then

$$[\![R_{\nu,u}^{(\beta^*)}]\!]_\tau \geq |u|^{-s_1}\, |W_\nu^{(\beta^*)}(u)| \geq C|W_\nu^{(\beta^*)}(u)| \geq C r^{\varepsilon_2}.$$

Thus if $u \in B_2 \setminus G^\nu$, then by Lemma 3.3,

$$|F_u^\nu| \leq C\Big(\frac{C_0 r^{\varepsilon_1}}{r^{\varepsilon_2}}\Big)^{1/(d-1)} = C r^{-(\varepsilon_2-\varepsilon_1)/(d-1)}.$$

The same inequality is clearly true if $u \in G^\nu$, since then $F_u^\nu = \emptyset$. Thus we have $|F_u^\nu|$ being small for all $u \in B_2$. This concludes the proof of Proposition 9.1 in the case under consideration.

9.5. Case B1: a parabolic term dominates and $p_2(y) \equiv 0$

In Case B there exists $m \in \Lambda$ with $p_m(y)$ parabolic such that $|\nu_m| \simeq r$. We recall the division into two subcases: Case B1, in which $p_2(y) \equiv 0$, and Case B2, in which $p_2(y) \not\equiv 0$.

We first consider Case B1. Similarly to Case A, we fix $0 < \varepsilon_1 < \varepsilon_2 < 1$ and start by assuming without loss of generality that $|C[\sigma^\beta](u,\tau)| \leq r^{\varepsilon_1}$ for all $3 \leq |\beta| \leq d$. For each $3 \leq l \leq d$ we choose (via Statement (4) of Lemma 9.4) a distinguished index $\beta(l)$ with $|\beta(l)| = l$ such that $B_{l,\beta(l)}$ is not the zero polynomial. For each pair $(l, \beta(l))$ we use the relation (9.24) to provide an expression for $(\nu_l - \mu_l)B_{l,\beta(l)}$ which we then feed into a downward induction argument, with the result that for all $3 \leq l \leq d$,

$$(\nu_l - \mu_l) \prod_{\substack{j \geq l \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) = S^{(l)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}), \tag{9.31}$$

where $S^{(l)}_{\nu,u}(\tau)$ is a polynomial in τ whose coefficients depend only on ν and u (but not μ). We now feed this information into an analysis of $C[\sigma^\beta](u,\tau)$ for some $|\beta| = 2$ to be chosen precisely later.

For now, fix any β with $|\beta| = 2$. Since $p_2(y) \equiv 0$, we have $B_{2,\beta} \equiv 0$ for all β, so by (9.22), $C[\sigma^\beta](u,\tau)$ reduces to

$$\begin{aligned} C[\sigma^\beta](u,\tau) = &\sum_{m=3}^{d} (\nu_m - \mu_m)\, \tau^{m-2}\, B_{m,\beta}\Big(\frac{u}{|u|}\Big) \\ &+ \sum_{m=3}^{d} \nu_m ((|u| + \tau)^{m-2} - \tau^{m-2})\, B_{m,\beta}\Big(\frac{u}{|u|}\Big). \end{aligned} \tag{9.32}$$

In particular, since the first sum starts with $m \geq 3$, all terms from the first sum are linear or higher order with respect to τ. The idea is to use (9.31) to eliminate the role of μ in the first sum, and then to consider the constant term with respect to τ coming from the second sum. To proceed with this, we multiply (9.32) through by $\prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}(u/|u|)$, and substitute (9.31) wherever possible. One then concludes that

$$C[\sigma^\beta](u,\tau) \prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) = R^{(\beta)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}), \tag{9.33}$$

where $R^{(\beta)}_{\nu,u}(\tau)$ is a polynomial in τ whose coefficients depend only on ν and u (but not μ). (The superscript β again reflects that this comes from the coefficient $C[\sigma^\beta](u,\tau)$.) In fact,

$$R^{(\beta)}_{\nu,u}(\tau) = \sum_{m=3}^{d} \nu_m\, |u|^{m-2}\, B_{m,\beta}\Big(\frac{u}{|u|}\Big) \prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) + R^{(\beta,1)}_{\nu,u}(\tau),$$

where the first term is constant with respect to τ and $R^{(\beta,1)}_{\nu,u}(\tau)$ is a polynomial in τ with no constant term and with coefficients dependent only on ν, u (and not μ).

Let $W_\nu^{(\beta)}(u)$ be the polynomial in u defined by

$$W_\nu^{(\beta)}(u) := \sum_{m=3}^{d} \nu_m\, B_{m,\beta}(u) \prod_{\substack{j\geq 3\\ j\in\Lambda}} B_{j,\beta(j)}(u), \tag{9.34}$$

which we note has degree at most

$$s_2 = d + \sum_{\substack{j\geq 3\\ j\in\Lambda}} j.$$

We also recall that each $B_{j,\beta(j)}(u)$ is not identically zero whenever $j \in \Lambda$, because we have chosen $\beta(j)$ using Statement (4) of Lemma 9.4. Then we can rewrite

$$R_{\nu,u}^{(\beta)}(\tau) = |u|^{-s_3}\, W_\nu^{(\beta)}(u) + R_{\nu,u}^{(\beta,1)}(\tau), \quad \text{where} \quad s_3 = 2 + \sum_{\substack{j\geq 3\\ j\in\Lambda}} j.$$

Recall that in the current case, we assumed that there exists $m \in \Lambda$, say m_0, with $p_{m_0}(y)$ parabolic such that $|\nu_{m_0}| \simeq r$. Since $p_2(y) \equiv 0$, we have $2 \notin \Lambda$, so the m_0 above cannot be 2; furthermore, by Statement (7) of Lemma 9.4, for this m_0, there exists a β with $|\beta| = 2$ such that $B_{m,\beta}(u)$ is not identically zero. We will denote this distinguished index by β^*; the choice of β^* depends only on p_{m_0}. We now define our exceptional sets as follows: recall that $0 < \varepsilon_1 < \varepsilon_2 < 1$ and set

$$G^\nu := \{u \in B_2(\mathbb{R}^n) \colon |W_\nu^{(\beta^*)}(u)| \leq r^{\varepsilon_2}\},$$

and for $u \in B_2 \setminus G^\nu$, let

$$F_u^\nu := \{\tau \in B_1(\mathbb{R}) \colon |R_{\nu,u}^{(\beta^*)}(\tau)| \leq C_0\, r^{\varepsilon_1}\}$$

for some large absolute constant C_0 to be determined. Also define $F_u^\nu := \emptyset$ if $u \in G^\nu$. Then for all $(u,\tau) \in B_2(\mathbb{R}^n) \times B_1(\mathbb{R})$ with $u \notin G^\nu$, $\tau \notin F_u^\nu$, we have

$$|R_{\nu,u}^{(\beta^*)}(\tau)| \geq C_0\, r^{\varepsilon_1},$$

so for such (u,τ), we conclude from (9.33) that

$$\begin{aligned} |C[\sigma^{\beta^*}](u,\tau)| &\geq C_B^{-|\Lambda|} \left| C[\sigma^{\beta^*}](u,\tau) \prod_{\substack{j\geq 3\\ j\in\Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) \right| \\ &\geq C_B^{-|\Lambda|} \big(|R_{\nu,u}^{(\beta^*)}(\tau)| - O(r^{\varepsilon_1}) \big) \geq C_B^{-|\Lambda|} \big(C_0\, r^{\varepsilon_1} - O(r^{\varepsilon_1}) \big) \geq C_B^{-|\Lambda|} r^{\varepsilon_1} \end{aligned}$$

if C_0 is sufficiently large. It follows from the van der Corput estimate of Lemma 3.1 that

$$|K_{\sharp,n}^{\nu,\mu}(u,\tau)| \leq C\, r^{-\varepsilon_1/d} \quad \text{if } u \notin G^\nu \text{ and } \tau \notin F_u^\nu.$$

On the other hand, we will now show that G^ν and F_u^ν are sets of small measures. First, G^ν is the set of $u \in B_2$ where $W_\nu^{(\beta^*)}(u)$ is small, and in (9.34), $W_\nu^{(\beta^*)}(u)$ is represented as a sum of homogeneous polynomials of different total degrees; in particular ν_{m_0} appears as the coefficient of a homogeneous polynomial in $W_\nu^{(\beta^*)}(u)$

that is by construction not identically zero, and has different total degree than all other terms in $W_\nu^{(\beta^*)}(u)$. Thus in the current case,

$$[\![W_\nu^{(\beta^*)}]\!]_u \geq C\,|\nu_{m_0}| \geq C\,r,$$

which implies by Lemma 3.3 that

$$|G^\nu| \leq C\left(\frac{r^{\varepsilon_2}}{r}\right)^{1/s_2} = C\,r^{-(1-\varepsilon_2)/s_2},$$

where the degree of $W^{(\beta^*)}$ is at most s_2. Furthermore, if $u \in B_2 \setminus G^\nu$, then

$$[\![R_{\nu,u}^{(\beta^*)}]\!]_\tau \geq |u|^{-s_3}\,|W_\nu^{(\beta^*)}(u)| \geq C\,|W_\nu^{(\beta^*)}(u)| \geq C\,r^{\varepsilon_2}.$$

Thus if $u \in B_2 \setminus G^\nu$, then by Lemma 3.3,

$$|F_u^\nu| \leq C\left(\frac{C_0\,r^{\varepsilon_1}}{r^{\varepsilon_2}}\right)^{1/d} = C\,r^{-(\varepsilon_2-\varepsilon_1)/d}.$$

The same inequality is clearly true if $u \in G^\nu$. Thus we have $|F_u^\nu|$ being small for all $u \in B_2$. This concludes the proof of Proposition 9.1 in the case under consideration.

9.6. Case B2: a parabolic term dominates and $p_2(y) \not\equiv 0$

In this case, we will need to consider the coefficients of two terms $\sigma^{\beta_1^*}$ and $\sigma^{\beta_2^*}$ for some $|\beta_1^*| = 1$ and $|\beta_2^*| = 2$ to be chosen precisely later. We fix $0 < \varepsilon_1 < \varepsilon_2 < 1$ and suppose first that $|C[\sigma^\beta](u,\tau)| \geq r^{\varepsilon_1}$ for some $1 \leq |\beta| \leq d$ with $|\beta| \neq 2$. Then $|K_{\sharp,n}^{\nu,\mu}(u,\tau)| \leq C\,r^{-\varepsilon_1/d}$ as desired. Thus we may assume that $|C[\sigma^\beta](u,\tau)| \leq r^{\varepsilon_1}$ for all β with $|\beta| = 1$ and all β with $3 \leq |\beta| \leq d$. From the latter set of conditions, we use the relations (9.24) for each $3 \leq l \leq d$ and an appropriate choice $\beta(l)$ such that $B_{l,\beta(l)}$ is not the zero polynomial, to provide expressions for $(\nu_l - \mu_l)B_{l,\beta(l)}$ in terms of $\nu_m - \mu_m$ for $m > l$. We then feed these expressions into a downward induction in l in order to show that for all $3 \leq l \leq d$,

$$(\nu_l - \mu_l) \prod_{\substack{j \geq l \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) = S_{\nu,u}^{(l)}(\tau) + O(r^{\varepsilon_1}), \tag{9.35}$$

where $S_{\nu,u}^{(l)}(\tau)$ is a polynomial in τ whose coefficients depend only on ν and u (but not μ).

Next for each $|\beta| = 1$, we apply the assumption that $|C[\sigma^\beta](u,\tau)| \leq r^{\varepsilon_1}$, to solve for $(\nu_2 - \mu_2)\tau B_{2,\beta}$ in (9.22). We conclude that for each β with $|\beta| = 1$ we have

$$\begin{aligned}(\nu_2 - \mu_2)\tau B_{2,\beta}\Big(\frac{u}{|u|}\Big) = &- \sum_{m=3}^{d} (\nu_m - \mu_m)\,\tau^{m-1}\,B_{m,\beta}\Big(\frac{u}{|u|}\Big) \\ &- \sum_{m=2}^{d} \nu_m((|u|+\tau)^{m-1} - \tau^{m-1})\,B_{m,\beta}\Big(\frac{u}{|u|}\Big) + O(r^{\varepsilon_1}).\end{aligned} \tag{9.36}$$

Now we recall that by Statement (6) of Lemma 9.4, if $m \in \Lambda$ and $p_m(y)$ is not parabolic, then there exists an index β with $|\beta| = 1$ such that $B_{m,\beta}$ is not the zero polynomial. By the assumption of our main theorem, we always have p_2 not parabolic, and since in the case under consideration $2 \in \Lambda$, there exists by Statement (6) of Lemma 9.4 an index $|\beta| = 1$, which we will denote by β_1^*, for which B_{2,β_1^*} is not the zero polynomial. We will use the relation (9.36) specialized to the choice $\beta = \beta_1^*$.

We multiply through (9.36), specialized to $\beta = \beta_1^*$, by the polynomial

$$\prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big),$$

use (9.35) to eliminate the presence of μ in the first sum on the right-hand side, and then group terms according to powers of τ. We note that the second term on the right-hand side of (9.36) contributes constant and linear terms with respect to τ, while the first sum contributes only terms that are at least order 2 with respect to τ. The result is that

$$\begin{aligned}(\nu_2 - \mu_2)\, \tau\, B_{2,\beta_1^*}\Big(\frac{u}{|u|}\Big) \prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) \\ = - \sum_{m=2}^{d} \nu_m\, |u|^{m-1} B_{m,\beta_1^*}\Big(\frac{u}{|u|}\Big) \prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) \\ - \tau \sum_{m=3}^{d} \nu_m\, (m-1)\, |u|^{m-2}\, B_{m,\beta_1^*}\Big(\frac{u}{|u|}\Big) \prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big) \\ + \tau^2\, S^{(2)}_{\nu,u}(\tau) + O(r^{\varepsilon_1}),\end{aligned} \tag{9.37}$$

where $S^{(2)}_{\nu,u}(\tau)$ is a polynomial in τ whose coefficients depend only on ν and u (but not μ).

We now feed this information into an analysis of $C[\sigma^\beta](u,\tau)$ for some β with $|\beta| = 2$ to be chosen later. For now we fix any β with $|\beta| = 2$ and recall that by definition, for this β

$$\begin{aligned}C[\sigma^\beta](u,\tau) = \sum_{m=2}^{d} (\nu_m - \mu_m)\, \tau^{m-2}\, B_{m,\beta}\Big(\frac{u}{|u|}\Big) \\ + \sum_{m=3}^{d} \nu_m((|u| + \tau)^{m-2} - \tau^{m-2})\, B_{m,\beta}\Big(\frac{u}{|u|}\Big).\end{aligned} \tag{9.38}$$

(Note that unlike Case B1, the first sum here begins with $m = 2$.) We want to use (9.35) and (9.37) to make the first sum on the right-hand side independent of μ: hence we multiply (9.38) through by

$$\tau\, B_{2,\beta_1^*}\Big(\frac{u}{|u|}\Big) \prod_{\substack{j \geq 3 \\ j \in \Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big),$$

and substitute (9.37) to treat the term including $(\nu_2-\mu_2)$ and (9.35) to treat terms including $(\nu_m-\mu_m)$ for all $m\geq 3$. One then concludes that

$$C[\sigma^\beta](u,\tau)\,\tau\,B_{2,\beta_1^*}\Big(\frac{u}{|u|}\Big)\prod_{\substack{j\geq 3\\ j\in\Lambda}}B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big)=R^{(\beta)}_{\nu,u}(\tau)+O(r^{\varepsilon_1}), \tag{9.39}$$

where $R^{(\beta)}_{\nu,u}(\tau)$ is a polynomial in τ whose coefficients depends only on ν and u (but not μ). (The superscript β reflects that this comes from $C[\sigma^\beta](u,\tau)$.)

In fact we may compute the coefficient of τ in $R^{(\beta)}_{\nu,u}(\tau)$ explicitly. We need only note that the term in $R^{(\beta)}_{\nu,u}(\tau)$ that is linear in τ comes from the terms in $C[\sigma^\beta](u,\tau)$ that are constant with respect to τ. Then we use the fact that the first term on the right-hand side of (9.38) contributes a constant with respect to τ with the $m=2$ summand; the second term on the right-hand side of (9.38) contributes a constant term in τ for each $3\leq m\leq d$. This shows that the coefficient of τ in $R^{(\beta)}_{\nu,u}(\tau)$ is given by

$$\begin{aligned}-\sum_{m=3}^{d}\nu_m\,(m-1)\,|u|^{m-2}\,B_{m,\beta_1^*}\Big(\frac{u}{|u|}\Big)B_{2,\beta}\Big(\frac{u}{|u|}\Big)\prod_{\substack{j\geq 3\\ j\in\Lambda}}B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big)\\ +\sum_{m=3}^{d}\nu_m\,|u|^{m-2}\,B_{m,\beta}\Big(\frac{u}{|u|}\Big)B_{2,\beta_1^*}\Big(\frac{u}{|u|}\Big)\prod_{\substack{j\geq 3\\ j\in\Lambda}}B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big).\end{aligned} \tag{9.40}$$

Using the homogeneity property of $B_{m,\beta}$, we see that (9.40) simplifies to

$$|u|^{-s_4}\Big(\prod_{\substack{j\geq 3\\ j\in\Lambda}}B_{j,\beta(j)}(u)\Big)\sum_{m=3}^{d}\nu_m[-(m-1)B_{m,\beta_1^*}(u)B_{2,\beta}(u)+B_{m,\beta}(u)B_{2,\beta_1^*}(u)]$$

where

$$s_4=4+\sum_{\substack{j\geq 3\\ j\in\Lambda}}j.$$

Let $W^{(\beta)}_\nu(u)$ be the polynomial in u defined by

$$\Big(\prod_{\substack{j\geq 3\\ j\in\Lambda}}B_{j,\beta(j)}(u)\Big)\sum_{m=3}^{d}\nu_m[-(m-1)B_{m,\beta_1^*}(u)B_{2,\beta}(u)+B_{m,\beta}(u)B_{2,\beta_1^*}(u)], \tag{9.41}$$

which has total degree at most

$$s_5=d+2+\sum_{\substack{j\geq 3\\ j\in\Lambda}}j.$$

Then we can conclude that

$$[\![R^{(\beta)}_{\nu,u}]\!]_\tau\geq |u|^{-s_4}\,|W^{(\beta)}_\nu(u)|\geq C\,|W^{(\beta)}_\nu(u)|.$$

Recall that in the current case, we assumed that there exists an $m \in \Lambda$, say m_0, with $p_{m_0}(y)$ parabolic, such that $|\nu_{m_0}| \simeq r$. Since we assumed $p_2(y) \neq C|y|^2$ for any $C \neq 0$ in Proposition 9.1, we know that $m_0 \geq 4$. Thus we can single out the contribution to $W_\nu^{(\beta)}(u)$ from m_0, which by (9.41) is

$$\nu_{m_0}\Bigg(\prod_{\substack{j\geq 3\\ j\in\Lambda}} B_{j,\beta(j)}(u)\Bigg)[-(m_0-1)B_{m_0,\beta_1^*}(u)B_{2,\beta}(u)+B_{m_0,\beta}(u)B_{2,\beta_1^*}(u)]. \tag{9.42}$$

It is clear that the total degree of this contribution is distinct from that of all other terms in $W_\nu^{(\beta)}(u)$. Our aim now is to show that we can pick a particular index β with $|\beta| = 2$ such that (9.42) is a nonzero polynomial with respect to u, so that in particular we can conclude that $W_\nu^{(\beta)}(u)$ contains a coefficient of size $|\nu_{m_0}| \simeq r$.

Since $m_0 \in \Lambda$ and p_{m_0} is parabolic we know by Statement (5) of Lemma 9.4 that $B_{m_0,\beta_1^*} \equiv 0$ since $|\beta_1^*| = 1$, which is odd. Thus the contribution (9.42) to $W_\nu^{(\beta)}(u)$ from m_0 is in fact precisely

$$\nu_{m_0}\Bigg(\prod_{\substack{j\geq 3\\ j\in\Lambda}} B_{j,\beta(j)}(u)\Bigg) B_{m_0,\beta}(u)\, B_{2,\beta_1^*}(u). \tag{9.43}$$

We recall that by construction, $B_{2,\beta_1^*} \not\equiv 0$ and $B_{j,\beta(j)}(u) \not\equiv 0$ for all $3 \leq j \leq d$, $j \in \Lambda$. Next, we note by Statement (7) of Lemma 9.4 that there exists β with $|\beta| = 2$ such that $B_{m_0,\beta}$ is not the zero polynomial; we will call this distinguished index β_2^*.

We now define our exceptional sets as follows: recall that $0 < \varepsilon_1 < \varepsilon_2 < 1$, and for the choice $\beta = \beta_2^*$ with $|\beta_2^*| = 2$ distinguished above, set

$$G^\nu := \{u \in B_2(\mathbb{R}^n) \colon |W_\nu^{(\beta_2^*)}(u)| \leq r^{\varepsilon_2}\},$$

and for $u \in B_2 \setminus G^\nu$, let

$$F_u^\nu := \{\tau \in B_1(\mathbb{R}) \colon |R_{\nu,u}^{(\beta_2^*)}(\tau)| \leq C_0 r^{\varepsilon_1}\}$$

for some large absolute constant C_0 to be determined. Also define $F_u^\nu := \emptyset$ if $u \in G^\nu$. Then for all $(u,\tau) \in B_2(\mathbb{R}^n) \times B_1(\mathbb{R})$ with $u \notin G^\nu$, $\tau \notin F_u^\nu$, we have

$$|R_{\nu,u}^{(\beta_2^*)}(\tau)| \geq C_0 r^{\varepsilon_1},$$

so for such (u,τ) we conclude by (9.39) that

$$\begin{aligned}|C[\sigma^{\beta_2^*}](u,\tau)| &\geq C_B^{-|\Lambda|}\left|C[\sigma^{\beta_2^*}](u,\tau)\tau B_{2,\beta_1^*}\Big(\frac{u}{|u|}\Big)\prod_{\substack{j\geq 3\\ j\in\Lambda}} B_{j,\beta(j)}\Big(\frac{u}{|u|}\Big)\right|\\ &\geq C_B^{-|\Lambda|}\left(|R_{\nu,u}^{(\beta_2^*)}(\tau)| - O(r^{\varepsilon_1})\right) \geq C_B^{-|\Lambda|}\left(C_0 r^{\varepsilon_1} - O(r^{\varepsilon_1})\right) \geq C_B^{-|\Lambda|} r^{\varepsilon_1}\end{aligned}$$

if C_0 is sufficiently large. It follows from the van der Corput estimate of Lemma 3.1 that

$$|K_{\sharp,n}^{\nu,\mu}(u,\tau)| \leq C r^{-\varepsilon_1/d} \quad \text{if } u \notin G^\nu \text{ and } \tau \notin F_u^\nu.$$

On the other hand, we will now show that G^ν and F^ν_u are sets of small measures. First, G^ν is the set of $u \in B_2$ where $W_\nu^{(\beta_2^*)}(u)$ is small, and in (9.41), $W_\nu^{(\beta_2^*)}(u)$ is represented as a sum of homogeneous polynomials of different total degrees. We have already noted above that the contribution to $W_\nu^{(\beta_2^*)}$ from m_0 is the term (9.43) specialized to $\beta = \beta_2^*$, and that this has total degree different from all other terms in $W_\nu^{(\beta_2^*)}$. We have also noted that by construction, the term (9.43) with $\beta = \beta_2^*$ is not the zero polynomial; thus in particular it contributes a term in the polynomial $W_\nu^{(\beta_2^*)}$ that has coefficient $|\nu_{m_0}| \simeq r$. We therefore see that

$$[\![W_\nu^{(\beta_2^*)}]\!]_u \geq C|\nu_{m_0}| \geq C r,$$

which implies by Lemma 3.3 that

$$|G^\nu| \leq C\left(\frac{r^{\varepsilon_2}}{r}\right)^{1/s_5} = C r^{-(1-\varepsilon_2)/s_5}.$$

Furthermore, if $u \in B_2 \setminus G^\nu$, then

$$[\![R_{\nu,u}^{(\beta_2^*)}]\!]_\tau \geq C|W_\nu^{(\beta_2^*)}(u)| \geq C r^{\varepsilon_2}.$$

Thus if $u \in B_2 \setminus G^\nu$, then by Lemma 3.3,

$$|F^\nu_u| \leq C\left(\frac{C_0 r^{\varepsilon_1}}{r^{\varepsilon_2}}\right)^{1/d} = C r^{-(\varepsilon_2-\varepsilon_1)/d}.$$

The same inequality is clearly true if $u \in G^\nu$. Thus we have $|F^\nu_u|$ being small for all $u \in B_2$. This concludes the proof of Proposition 9.1 in this final case.

9.7. Proof of Lemma 9.3 for $A_{\alpha,\beta}$

We now turn to the proof of the key properties of $A_{\alpha,\beta}$ given in Lemma 9.3. Recall the expansion (9.18) that defines $A_{\alpha,\beta}(w)$:

$$\sum_{|\beta|\leq|\alpha|} A_{\alpha,\beta}(w)\,\sigma^\beta = (w_n + w^{(n)}\cdot\sigma)^{\alpha_n}(w^{(n)} - w_n\sigma)^{\alpha^{(n)}}. \tag{9.44}$$

We recall that $w \in \mathbb{R}^n$ while $\sigma = (\sigma_1, \dots, \sigma_{n-1}) \in \mathbb{R}^{n-1}$, so that the multi-index α is n-dimensional while the multi-index β is $(n-1)$-dimensional.

To prove Statement (1), we use the fact that $|\beta| = 1$ to compute $A_{\alpha,\beta}$ explicitly. Suppose that $\beta = e_j$, where $e_j = (0, \dots, 1, \dots, 0)$ is the j-th unit vector, with $1 \leq j \leq n-1$. Then to compute $A_{\alpha,\beta}(w)$, we must pick out the coefficient of $\sigma^\beta = \sigma_j$ on the right-hand side of the expansion (9.44), namely

$$\alpha_n w_j w_n^{\alpha_n-1} w_1^{\alpha_1}\cdots w_{n-1}^{\alpha_{n-1}} + \alpha_j w_n^{\alpha_n} w_1^{\alpha_1}\cdots(-w_n)w_j^{\alpha_j-1}\cdots w_{n-1}^{\alpha_{n-1}}.$$

This coefficient of σ_j is the polynomial $A_{\alpha,e_j}(w)$ we seek; we write it more efficiently as

$$A_{\alpha,e_j}(w) = \alpha_n w^{\alpha+e_j-e_n} - \alpha_j w^{\alpha-e_j+e_n}.$$

(We pause here to note that if we were working with $K^{\nu,\mu}_{\sharp,l}(u,\tau)$ for any $1 \leq l \leq n$, we would consider all $1 \leq j \leq n$ with $j \neq l$ and the analogous expression would clearly be

$$A_{\alpha,e_j}(w) = \alpha_l\, w^{\alpha+e_j-e_l} - \alpha_j\, w^{\alpha-e_j+e_l}.$$

Thus although we appear to be privileging the n-th component, nothing in the proof depends more than notationally on this choice.) Statement (2) follows in a similar fashion from examining the expansion (9.44).

To prove Statement (3), fix $\alpha \in \mathbb{Z}^n_{\geq 0}$ and fix any $\beta \in \mathbb{Z}^{n-1}_{\geq 0}$ with $|\beta| = |\alpha|$. To single out the term $A_{\alpha,\beta}(w)$ from the left hand side of (9.44), we must simply pick out the coefficient of σ^β on the right-hand side. Expand the right-hand side as

$$(w_n + w_1\sigma_1 + \cdots + w_{n-1}\sigma_{n-1})^{\alpha_n}(w_1 - w_n\sigma_1)^{\alpha_1}\cdots(w_{n-1} - w_n\sigma_{n-1})^{\alpha_{n-1}}. \tag{9.45}$$

To prove that $A_{\alpha,\beta}$ is a monomial, we must verify that no more than one term in this expansion is of the form σ^β. Clearly, since $|\alpha| = |\beta|$, in order for the total degree of such a monomial to reach $|\beta|$, we must always choose α_j copies of the σ_j-factor in each of the last $n-1$ terms, so that the product of the last $n-1$ terms contributes

$$(-1)^{|\alpha^{(n)}|}\, w_n^{|\alpha^{(n)}|}\, \sigma^{\alpha^{(n)}}. \tag{9.46}$$

The remaining part of the σ^β monomial comes from the first factor, which we write as

$$\begin{aligned}(w_n + w_1\sigma_1 + \cdots + w_{n-1}\sigma_{n-1})^{\alpha_n} \\ = \sum_{\gamma_1+\cdots+\gamma_n=\alpha_n} \binom{\alpha_n}{\gamma_1,\ldots,\gamma_n}(w_1\sigma_1)^{\gamma_1}\cdots(w_{n-1}\sigma_{n-1})^{\gamma_{n-1}}(w_n)^{\gamma_n}.\end{aligned} \tag{9.47}$$

Here we are using the usual multinomial coefficient, defined for $k_1 + \cdots + k_n = k$ by

$$\binom{k}{k_1,\ldots,k_n} = \frac{k!}{k_1!\cdots k_n!}.$$

A multi-index $\gamma \in \mathbb{Z}^n_{\geq 0}$ in the sum (9.47) will contribute to the σ^β monomial (when combined with the second factor (9.46)) if and only if it solves the equations

$$|\gamma| = \alpha_n \tag{9.48}$$

$$\gamma^{(n)} + \alpha^{(n)} = \beta. \tag{9.49}$$

There is at most one solution to this system. Indeed, if $\alpha_j > \beta_j$ for any $1 \leq j \leq n-1$ then there is no solution γ, since (9.49) cannot be satisfied by any $\gamma^{(n)} \in \mathbb{Z}^{n-1}_{\geq 0}$, and in this case $A_{\alpha,\beta} \equiv 0$ (that is, the constant $C_{\alpha,\beta}$ is zero). But if

$$\alpha_j \leq \beta_j \quad \text{for all } 1 \leq j \leq n-1, \tag{9.50}$$

then we may solve (9.49) uniquely for $\gamma^{(n)}$. The condition (9.50) along with $|\beta| = |\alpha|$ also guarantees that

$$|\gamma^{(n)}| = |\beta - \alpha^{(n)}| = |\beta| - |\alpha^{(n)}| = \alpha_n,$$

which implies in (9.48) that $\gamma_n = 0$. We have shown that if (9.50) holds, there is precisely the unique solution $\gamma = (\beta - \alpha^{(n)}, 0)$ for the system (9.48)–(9.49) and consequently $A_{\alpha,\beta}$ is the monomial

$$A_{\alpha,\beta}(w) = C_{\alpha,\beta}\,(-1)^{|\alpha^{(n)}|}\,w_n^{|\alpha^{(n)}|}\,(w^{(n)})^{\beta-\alpha^{(n)}}$$

with combinatorial coefficient

$$C_{\alpha,\beta} = \binom{\alpha_n}{(\beta_1-\alpha_1),\ldots,(\beta_{n-1}-\alpha_{n-1}),0} \neq 0.$$

In particular, $A_{\alpha,\beta}$ is a monomial with nonzero coefficient for any β such that $|\beta| = |\alpha|$ and $\beta_j \geq \alpha_j$ for all $1 \leq j \leq n-1$, and otherwise is the zero polynomial. (Certainly there is at least one such β for every α.)

9.8. Proof of Lemma 9.4 for $B_{m,\beta}$, Statements (1)–(5)

Statements (1) and (2) are clear from the definition of $B_{m,\beta}$. Statement (3) is a consequence of the fact, already observed, that for each α, β, the polynomial $A_{\alpha,\beta}(w)$ is homogeneous of degree $|\alpha|$. To see Statement (4), we fix an $m \in \Lambda$ and recall that

$$B_{m,\beta}(w) = \sum_{|\alpha|=m} c_\alpha\, A_{\alpha,\beta}(w).$$

By Lemma 9.3, for each $|\alpha| = |\beta|$, $A_{\alpha,\beta}(w)$ is a monomial, and in particular we can write

$$B_{m,\beta}(w) = \sum_{|\alpha|=m} c_\alpha\, C_{\alpha,\beta}\, w_n^{|\alpha^{(n)}|}\,(w^{(n)})^{\beta-\alpha^{(n)}},$$

for non-negative combinatorial constants $C_{\alpha,\beta}$. For β fixed, each monomial

$$w_n^{|\alpha^{(n)}|}\,(w^{(n)})^{\beta-\alpha^{(n)}}$$

with $|\alpha| = |\beta|$ is distinct. Thus it suffices to show that there is some β with $|\beta| = m$ and some $|\alpha| = m$ such that $c_\alpha C_{\alpha,\beta} \neq 0$. We first note that since $m \in \Lambda$, there is some α with $|\alpha| = m$ such that $c_\alpha \neq 0$; we will call this $\alpha(m)$. It now suffices to find β with $|\beta| = m$ such that $C_{\alpha(m),\beta} \neq 0$; such a choice of β is in fact guaranteed by Lemma 9.3. This proves Statement (4).

We now turn to Statement (5). Suppose $p_m(y)$ is parabolic for some index $2 \leq m \leq d$, say $p_m(y) = C|y|^m$ for some nonzero constant C; in particular m must be even, so we write $m = 2k$. Recall that (9.15) holds for all choices of ν and μ; we will apply that equation to expand $C|z|^m$ alone by setting $\nu_j = 0$ for all j, $\mu_m = 1$, and $\mu_j = 0$ for all $j \neq m$. With these choices in (9.15), we see that

$$-\,C\,|z|^m = -\sum_{0\leq|\beta|\leq m} \tau^{m-|\beta|}\,B_{m,\beta}\Big(\frac{u}{|u|}\Big)\,\sigma^\beta, \tag{9.51}$$

in which z is defined implicitly by (9.4) and (9.5) (with index $l = n$).

On the other hand, if z is defined by (9.4) and (9.5), then

$$|z|^{2k} = (z_n^2 + |z^{(n)}|^2)^k = \frac{1}{|u|^{2k}} \left\{ \tau^2 |u|^2 + (u^{(n)} \cdot \sigma)^2 + u_n^2 |\sigma|^2 \right\}^k.$$

If we denote the right-hand side by $T(\sigma)$, say, it is clear by inspection that $T(\sigma)$ is an even polynomial in σ, that is to say $T(\sigma) = T(-\sigma)$. Now fix any multi-index β; the coefficient of σ^β in $T(\sigma)$ is of course $\partial^\beta T(\sigma)\big|_{\sigma=0}$. But since T is an even polynomial, we see that

$$\partial^\beta T(\sigma)\big|_{\sigma=0} = \partial^\beta (T(-\sigma))\big|_{\sigma=0} = (-1)^{|\beta|} \, (\partial^\beta T)(-\sigma)\big|_{\sigma=0} = (-1)^{|\beta|} \, (\partial^\beta T)(\sigma)\big|_{\sigma=0}$$

which shows that the coefficient of σ^β in $T(\sigma)$ must be zero whenever $|\beta|$ is odd. In the expansion (9.51) this shows that $\tau^{m-|\beta|} B_{m,\beta}(u/|u|)$ is identically zero in τ, u whenever $|\beta|$ is odd, which implies that $B_{m,\beta}(w)$ is the zero polynomial whenever $|\beta|$ is odd.

9.9. Proof of Lemma 9.4 for $B_{m,\beta}$, Statement (6)

We will prove that if $m \in \Lambda$ and $p_m(y)$ is not parabolic, then there exists β with $|\beta| = 1$ such that $B_{m,\beta}$ is not the zero polynomial. In fact, it is easier to prove the equivalent statement that if $m \in \Lambda$ and $B_{m,\beta} \equiv 0$ for all $|\beta| = 1$, then p_m is parabolic. We recall that being parabolic puts a constraint on the relationships between the coefficients c_α (fixed once and for all) in the definition $p_m(y) = \sum_{|\alpha|=m} c_\alpha y^\alpha$. Precisely, if $p_m(y) = C|y|^m$ for $m = 2k$, then

$$C\,|y|^{2k} = C\,(y_1^2 + \cdots + y_n^2)^k = C \sum_{k_1+k_2+\cdots+k_n=k} \binom{k}{k_1, k_2, \ldots, k_n} y_1^{2k_1} y_2^{2k_2} \cdots y_n^{2k_n}.$$

Thus we see for $m = 2k \in \Lambda$, $p_{2k}(y)$ being parabolic is characterized the property that there exists a nonzero constant C such that for any partition $k_1 + \cdots + k_n = k$ and corresponding multi-index $\alpha = (2k_1, \ldots, 2k_n)$, the coefficient c_α must satisfy

$$c_\alpha = C \binom{k}{k_1, k_2, \ldots, k_n},$$

and for all α with $|\alpha| = 2k$ that have an odd entry, $c_\alpha = 0$.

We now fix any $m \in \Lambda$ and assume that $B_{m,\beta} \equiv 0$ for all $|\beta| = 1$. We will show that p_m must be parabolic. Fix any β with $|\beta| = 1$, which we will denote by $\beta = e_j$ for some $1 \le j \le n-1$. We apply Statement (1) of Lemma 9.3 to compute that

$$B_{m,e_j}(w) = \sum_{|\alpha|=m} c_\alpha A_{\alpha,e_j}(w) = \sum_{|\alpha|=m} c_\alpha \, (\alpha_n w^{\alpha+e_j-e_n} - \alpha_j w^{\alpha-e_j+e_n}).$$

We now re-write this by grouping coefficients for each fixed monomial w^ρ, as

$$B_{m,e_j}(w) = \sum_{|\rho|=m} \left[c_{\rho+e_n-e_j} (\rho_n + 1) - c_{\rho-e_n+e_j} (\rho_j + 1) \right] w^\rho. \tag{9.52}$$

By notational convention, if $\rho_j = 0$ then $c_{\rho+e_n-e_j} = 0$ and if $\rho_n = 0$ then $c_{\rho-e_n+e_j} = 0$, so the corresponding terms do not actually appear in the expression above.

Under our assumption that $B_{m,e_j} \equiv 0$ for all $1 \leq j \leq n-1$, we see that for each j all the coefficients in (9.52) are identically zero, so that for all $|\rho| = m$ and all $1 \leq j \leq n-1$ we have

$$c_{\rho+e_n-e_j}(\rho_n + 1) - c_{\rho-e_n+e_j}(\rho_j + 1) = 0. \tag{9.53}$$

Again, if either ρ_j or ρ_n is zero, then the corresponding term does not appear in the expression above. First we note that since $(\rho_n + 1)$ and $(\rho_j + 1)$ will never vanish for multi-indices $\rho \in \mathbb{Z}^n_{\geq 0}$, (9.53) shows that the coefficients $c_{\rho+e_n-e_j}$ and $c_{\rho-e_n+e_j}$ are either both zero or both nonzero. Second, we note that (9.53) provides relations between coefficients with indices sharing a certain parity property. For any $m \geq 1$, let $\Gamma_{\text{odd}}(m)$ denote the set of multi-indices $\gamma \in \mathbb{Z}^n_{\geq 0}$ with $|\gamma| = m$ such that at least one coordinate of γ is odd. We will prove:

Lemma 9.5. *Let $m \in \Lambda$ and suppose that $B_{m,e_j} \equiv 0$ for all $1 \leq j \leq n-1$. Then $c_\gamma = 0$ for all $\gamma \in \Gamma_{\text{odd}}(m)$.*

Assume this for the moment. If m is odd, then every γ with $|\gamma| = m$ contains an odd entry, so as a consequence of Lemma 9.5, if $m \in \Lambda$ yet $B_{m,e_j} \equiv 0$ for all $1 \leq j \leq n-1$ then we must have m even. Thus assuming $m = 2k$ is even, we let $\Gamma_{\text{even}}(m)$ denote the set of multi-indices $\gamma \in \mathbb{Z}^n_{\geq 0}$ with $|\gamma| = m$ such that all entries in γ are even; in this case we will write $\gamma = (2\gamma_1, \ldots, 2\gamma_n)$. Then we will deduce from the identity (9.53) the following lemma.

Lemma 9.6. *Let $m \in \Lambda$ with $m = 2k$ for some $k \geq 1$ and suppose that $B_{m,e_j} \equiv 0$ for all $1 \leq j \leq n-1$. Let γ^* denote the element $(0, 0, \ldots, 2k) \in \Gamma_{\text{even}}(2k)$ and define the constant*

$$C := c_{\gamma^*}.$$

Then for all $\gamma = (2\gamma_1, \ldots, 2\gamma_n) \in \Gamma_{\text{even}}(2k)$, the coefficient c_γ of $p_{2k}(y)$ must satisfy

$$c_\gamma = C\binom{k}{\gamma_1, \ldots, \gamma_n}. \tag{9.54}$$

With Lemmas 9.5 and 9.6 in hand, we can deduce Statement (6) of Lemma 9.4 immediately. Indeed, we have already seen that if $m \in \Lambda$ but $B_{m,\beta} \equiv 0$ for all $|\beta| = 1$ then m must be even, and now for $m = 2k$ we may conclude

$$p_m(y) = \sum_{\alpha \in \Gamma_{\text{even}}(m)} c_\alpha y^\alpha = C \sum_{\substack{|\alpha|=2k \\ \alpha=(2\alpha_1,\ldots,2\alpha_n)}} \binom{k}{\alpha_1, \ldots, \alpha_n} y^{2\alpha_1} \cdots y^{2\alpha_n} = C|y|^{2k}.$$

From this we also deduce that C must be nonzero, since otherwise we would have $p_m \equiv 0$, which would contradict $m \in \Lambda$.

We now prove Lemmas 9.5 and 9.6. To prove Lemma 9.5 in a notationally economical fashion, we define the equivalence relation $\gamma \sim \gamma'$ to indicate that

$c_\gamma = 0$ if and only if $c_{\gamma'} = 0$. Suppose $\gamma \in \Gamma_{\text{odd}}(m)$ is given; we will prove first that $\gamma \sim \gamma'$ for some γ' with an entry equal to 1, and then that for any γ' with an entry equal to 1 we necessarily have $c_{\gamma'} = 0$.

For the first statement, we assume that γ is given with γ_j odd; there are two cases, depending on whether $1 \leq j \leq n-1$ or $j = n$. We first assume that $1 \leq j \leq n-1$, in which case we will observe that $\gamma \sim \gamma - 2\theta e_j + 2\theta e_n$ for any non-negative integer θ such that $2\theta \leq \gamma_j$. This follows from a simple induction argument; the base case $\theta = 0$ is clear. Assuming the induction hypothesis that $\gamma \sim \gamma - 2(\theta-1)e_j + 2(\theta-1)e_n$ for some $\theta \geq 1$, we apply (9.53) with the choice $\rho = \gamma - (2\theta-1)e_j + (2\theta-1)e_n$ to see that

$$\begin{aligned} c_{\rho+e_n-e_j}&(\rho_n+1) - c_{\rho-e_n+e_j}(\rho_j+1) \\ &= c_{\gamma+2\theta e_n - 2\theta e_j}(\gamma_n + 2\theta) - c_{\gamma+2(\theta-1)e_n-2(\theta-1)e_j}(\gamma_j - 2\theta + 2) = 0. \end{aligned}$$

Since $(\gamma_j - 2\theta + 2) \neq 0$ and $(\gamma_n + 2\theta) \neq 0$, it follows that $\gamma - 2(\theta-1)e_j + 2(\theta-1)e_n \sim \gamma - 2\theta e_j + 2\theta e_n$, which completes the induction.

With this fact in hand, given $\gamma \in \Gamma_{\text{odd}}(m)$ with $\gamma_j = 2k+1$ for some $k \geq 0$, we may conclude that $\gamma \sim \gamma' := \gamma - 2ke_j + 2ke_n$, where γ' has j-th coordinate equal to 1 by construction. If we instead had γ_j odd for $j = n$, we would similarly use induction to show that for any $1 \leq i \leq j-1$, $\gamma \sim \gamma - 2\theta e_n + 2\theta e_i$, for any non-negative integer θ such that $2\theta \leq \gamma_n$; then we would apply this to show $\gamma \sim \gamma'$ for some γ' with $\gamma'_n = 1$.

Next we show that any γ with an entry equal to 1 has $c_\gamma = 0$. Assume that γ is such that $\gamma_j = 1$ for some $1 \leq j \leq n$. There are again two cases to consider: $1 \leq j \leq n-1$ and $j = n$. In the case that $\gamma_j = 1$ for some $1 \leq j \leq n-1$, we apply (9.53) with the choice $\rho = \gamma + e_n - e_j$, to conclude that

$$c_\gamma = c_{\rho - e_n + e_j} = 0.$$

(Here we are using the fact that with our particular choice of ρ, the j-th entry of $\rho + e_n - e_j$ is negative, so that by convention the coefficient $c_{\rho+e_n-e_j} = 0$.) Similarly, in the case that $j = n$, we choose any $1 \leq i \leq n-1$ we like and apply (9.53) with the choice $\rho = \gamma - e_n + e_i$, to conclude that

$$c_\gamma = c_{\rho+e_n-e_i} = 0.$$

This completes the proof of Lemma 9.5.

In Lemma 9.6, we only consider $m = 2k$ even and indices $\gamma \in \Gamma_{\text{even}}(2k)$, which we denote by $\gamma = (2\gamma_1, \dots, 2\gamma_n)$. For $\gamma, \gamma' \in \Gamma_{\text{even}}(2k)$, we define the equivalence relation $\gamma \sim \gamma'$ to represent that c_γ satisfies (9.54) if and only if $c_{\gamma'}$ satisfies (9.54).

We note first of all that (9.54) holds true for γ^*, since

$$c_{\gamma^*} = C \binom{k}{0, \dots, 0, k} = C \frac{k!}{0! \cdots 0! k!} = C.$$

Thus it suffices to prove $\gamma^* \sim \gamma$ for any $\gamma \in \Gamma_{\text{even}}(2k)$. Thus we fix any $\gamma = (2\gamma_1, \dots, 2\gamma_n) \in \Gamma_{\text{even}}(2k)$ and note that

$$\gamma = \gamma^* + 2\gamma_1 e_1 + \cdots + 2\gamma_{n-1}e_{n-1} - (2k - 2\gamma_n)e_n = \gamma^* + \sum_{j=1}^{n-1} 2\gamma_j e_j - \left(\sum_{j=1}^{n-1} 2\gamma_j\right) e_n.$$

We will prove:

Lemma 9.7. *Given $\gamma \in \Gamma_{\text{even}}(2k)$, suppose $\gamma' = \gamma + 2\theta e_j - 2\theta e_n$ for some $1 \leq j \leq n-1$ and some non-negative integer θ with $2\theta \leq 2\gamma_n$. Then $\gamma \sim \gamma'$.*

With this lemma in hand, we see that

$$\begin{aligned}\gamma^* &\sim \gamma^* + 2\gamma_1 e_1 - 2\gamma_1 e_n \sim (\gamma^* + 2\gamma_1 e_1 - 2\gamma_1 e_n) + 2\gamma_2 e_2 - 2\gamma_2 e_n \\ &\sim \cdots \sim \gamma^* + \sum_{j=1}^{n-1} 2\gamma_j e_j - \Big(\sum_{j=1}^{n-1} 2\gamma_j\Big) e_n.\end{aligned}$$

Thus $\gamma^* \sim \gamma$, so that Lemma 9.6 will be proved as soon as we have proved Lemma 9.7.

We now prove Lemma 9.7 by induction. The statement clearly holds for the base case $\theta = 0$. We make the inductive hypothesis that $\gamma \sim \gamma + 2(\theta-1)e_j - 2(\theta-1)e_n$ for some integer $\theta \geq 1$, and show that this implies $\gamma \sim \gamma + 2\theta e_j - 2\theta e_n$. Here we may naturally assume we are in the case $2\theta \leq 2\gamma_n$. Without loss of generality we assume we are dealing with the case $j = 1$, for notational simplicity. Under the inductive hypothesis, c_γ satisfies (9.54) if and only if $c_{\gamma+2(\theta-1)e_1-2(\theta-1)e_n}$ also satisfies (9.54). We need only show that $c_{\gamma+2(\theta-1)e_1-2(\theta-1)e_n}$ satisfies (9.54) if and only if $c_{\gamma+2\theta e_1-2\theta e_n}$ satisfies (9.54).

By (9.53) applied with ρ chosen to be $\gamma + (2\theta-1)e_1 - (2\theta-1)e_n$, we see that

$$\begin{aligned}&c_{\rho+e_n-e_1}(\rho_n+1) - c_{\rho-e_n+e_1}(\rho_1+1) \\ &\qquad = c_{\gamma-2(\theta-1)e_n+2(\theta-1)e_1}(2\gamma_n - 2(\theta-1)) - c_{\gamma-2\theta e_n+2\theta e_1}(2\gamma_1+2\theta) = 0.\end{aligned}$$

Since $\gamma \in \Lambda$ by assumption, we have $c_\gamma \neq 0$, and so by the induction hypothesis, $c_{\gamma-2(\theta-1)e_n+2(\theta-1)e_1} \neq 0$. Thus we may write

$$\frac{c_{\gamma+2\theta e_1-2\theta e_n}}{c_{\gamma+2(\theta-1)e_1-2(\theta-1)e_n}} = \frac{(\gamma_n-(\theta-1))}{(\gamma_1+\theta)}. \tag{9.55}$$

It is now clear from (9.55) that

$$c_{\gamma+2\theta e_1-2\theta e_n} = C\frac{k!}{(\gamma_1+\theta)!\gamma_2!\cdots\gamma_{n-1}!(\gamma_n-\theta)!}$$

if and only if

$$c_{\gamma+2(\theta-1)e_1-2(\theta-1)e_n} = C\frac{k!}{(\gamma_1+(\theta-1))!\,\gamma_2!\cdots\gamma_{n-1}!\,(\gamma_n-(\theta-1))!}.$$

We may conclude that $\gamma + 2(\theta-1)e_1 - 2(\theta-1)e_n \sim \gamma + 2\theta e_1 - 2\theta e_n$, which proves Lemma 9.7. This completes the proof of Statement (6) of Lemma 9.4.

9.10. Proof of Lemma 9.4 for $B_{m,\beta}$, Statement (7)

We will prove that if $m \in \Lambda$ and $p_m(y)$ is parabolic, then there exists some β with $|\beta| = 2$ such that $B_{m,\beta}$ is not the zero polynomial. In fact, it suffices to consider

$\beta = 2e_j$ for any fixed $1 \le j \le n-1$; to fix ideas we choose $j=1$. We recall the definition

$$B_{m,2e_1}(w) = \sum_{|\alpha|=m} c_\alpha A_{\alpha,2e_1}(w).$$

By Statement (2) of Lemma 9.3, for each α with $|\alpha| = m$,

$$A_{\alpha,2e_1}(w) = \binom{\alpha_n}{2} w^{\alpha+2e_1-2e_n} - \alpha_n \alpha_1 w^\alpha + \binom{\alpha_1}{2} w^{\alpha-2e_1+2e_n},$$

with the understanding that any term with a multi-index with a negative entry does not actually appear. Thus after regrouping terms,

$$B_{m,2e_1}(w) = \sum_{|\gamma|=m} \left\{ c_{\gamma-2e_1+2e_n} \binom{\gamma_n+2}{2} - c_\gamma \gamma_n \gamma_1 + c_{\gamma+2e_1-2e_n} \binom{\gamma_1+2}{2} \right\} w^\gamma,$$

still with the understanding that any term involving c_α with a multi-index α with a negative entry does not appear. In particular, we see that that coefficient of w_1^m, that is w^γ with $\gamma = (m, 0, \ldots, 0)$, is precisely

$$c_{(m-2)e_1+2e_n}. \tag{9.56}$$

By assumption, $m \in \Lambda$ and p_m is parabolic, so that m is even and $c_\alpha \neq 0$ for all indices α such that $|\alpha| = 2k$ and all entries in α are even. Thus we see that (9.56) is a nonzero constant, and thus in particular $B_{m,2e_1}$ is not the zero polynomial. This suffices for Statement (7), and completes the proof of Lemma 9.4, and hence of Proposition 9.1.

10. Final treatment of Propositions 7.1, 7.3, and 7.4 in general dimension

10.1. Completing the proof of Proposition 7.1 (general $n \geq 2$)

With Proposition 9.1 in hand, we return briefly to the treatment of the term $\mathbf{I}$ in Section 7.3.1 for all dimensions $n \geq 2$. We recall that this term is represented as in (7.16) by

$$\mathbf{I} = \iint_{\mathbb{R}^{n+1}} e^{iP_\nu(u+z)-iP_\mu(z)} \eta(u+z)\eta(z)\zeta(\xi+|z|^2) \\ \cdot \left(\underline{\Delta}_{j-k}(\theta - |u+z|^2 + |z|^2) - \underline{\Delta}_{j-k}(\theta - |u+z|^2 + |z|^2 + \xi) \right) dz\, d\xi.$$

We would again like to isolate an oscillatory integral within this term that is independent of the $\underline{\Delta}_{j-k}$ factors. Regardless of the dimension, we are still motivated to define a new variable $\tau = (u \cdot z)/|u|$ in order to capture the behavior of z in $|u+z|^2 - |z|^2$. We fix the partition of unity $\sum_l W_l(s)$ given in (9.6) and write

$$\mathbf{I} = \sum_{l=1}^n \mathbf{I}_l$$

where for each $1 \leq l \leq n$,

$$\mathbf{I}_l = W_l\Big(\frac{u}{|u|}\Big) \iint_{\mathbb{R}^{n+1}} e^{iP_\nu(u+z)-iP_\mu(z)}\, \eta(u+z)\, \eta(z)\, \zeta(\xi+|z|^2) \cdot \big(\underline{\Delta}_{j-k}(\theta-|u+z|^2+|z|^2) - \underline{\Delta}_{j-k}(\theta-|u+z|^2+|z|^2+\xi)\big)\, dz\, d\xi.$$

For each index l we make the change of variables $z \mapsto (\tau, \sigma)$ relevant to the l-th coordinate as given in (9.2) and (9.3); under this transformation the term $\mathbf{I}_l$ becomes

$$\mathbf{I}_l = \Big(\frac{|u_l|}{|u|}\Big)^{n-2} W_l\Big(\frac{u}{|u|}\Big) \int_{\mathbb{R}^2} K^{\nu,\mu}_{\sharp,l}(u,\tau;\xi)\big(\underline{\Delta}_{j-k}(\theta-|u|^2-2|u|\tau) - \underline{\Delta}_{j-k}(\theta-|u|^2-2|u|\tau+\xi)\big)\, d\tau\, d\xi,$$

where

$$K^{\nu,\mu}_{\sharp,l}(u,\tau;\xi) = \int_{\mathbb{R}^{n-1}} e^{iP_\nu(u+z)-iP_\mu(z)}\, \eta(u+z)\, \eta(z)\, \zeta(\xi+|z|^2)\, d\sigma.$$

Here $z^{(l)} \in \mathbb{R}^{n-1}$, $z_l \in \mathbb{R}$ are implicitly defined in terms of $u \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}^{n-1}$, $\tau \in \mathbb{R}$ by (9.4) and (9.5). As previously observed in (9.9), the range of integration for σ is in the compact set $|\sigma| \leq c_0$, where c_0 is the absolute constant specified in (9.7). Moreover, $K^{\nu,\mu}_{\sharp,l}(u,\tau;\xi)$ has support where $u \in B_2(\mathbb{R}^n)$, $|\xi| \leq 2$, $|\tau| \leq 1$.

Applying the mean-value theorem to $\underline{\Delta}_{j-k}$, as recorded in Lemma 4.4, we have

$$|\mathbf{I}_l| \leq C\, 2^{-2(j-k)} \int_{\mathbb{R}^2} |K^{\nu,\mu}_{\sharp,l}(u,\tau;\xi)|\, |\xi| \chi_{B_2}(\xi)\, \psi_{j-k}(\theta-|u|^2-2|u|\tau)\, \chi_{B_1}(\tau)\, d\tau\, d\xi,$$

where $\psi_{j-k}(t)$, as defined in Lemma 4.4, is an L^1 dilation of $(1+t^2)^{-1}$, and is thus uniformly in L^1, independent of $j-k$. We now apply Proposition 9.1 to bound $K^{\nu,\mu}_{\sharp,l}$ and conclude that there exists $\delta > 0$ and a small set $G^\nu \subset B_2(\mathbb{R}^n)$ with $|G^\nu| \leq C r^{-\delta}$, and for each $u \in B_2(\mathbb{R}^n)$ a small set $F^\nu_u \subset B_1(\mathbb{R})$ with $|F^\nu_u| \leq C r^{-\delta}$, such that

$$(10.1)\quad |K^{\nu,\mu}_{\sharp,l}(u,\tau;\xi)| \leq C\left[r^{-\delta}\, \chi_{B_2}(u)\chi_{B_1}(\tau) + \chi_{G^\nu}(u)\chi_{B_1}(\tau) + \chi_{B_2}(u)\chi_{F^\nu_u}(\tau)\right].$$

Moreover, these estimates are uniform in ξ and the index l, as the small sets do not depend on ξ or l, and neither do the bounds. Hence we may sum over $1 \leq l \leq n$ to obtain (for some constant C depending only on n, d and the fixed polynomials $p_2, \ldots, p_d$)

$$|\mathbf{I}| \leq C\, 2^{-2(j-k)} \int_{\mathbb{R}^2} \big(r^{-\delta}\, \chi_{B_2}(u)\chi_{B_1}(\tau) + \chi_{G^\nu}(u)\chi_{B_1}(\tau) + \chi_{B_2}(u)\chi_{F^\nu_u}(\tau)\big) \cdot \chi_{B_2}(\xi)\, \psi_{j-k}(\theta-|u|^2-2|u|\tau)\, d\tau\, d\xi.$$

This is the exact analogue of (7.19) and we may proceed via the argument used to treat (7.19) in Section 7.3.1. We conclude that the contribution of the term $\mathbf{I}$ to the kernel of $TT^*f(x,t)$ leads to an operator with L^2 norm bounded above by $C\, 2^{-2(j-k)} r^{-\delta/2}$, as in (7.24).

10.2. Completing the proof of (7.4) and (7.6) for I_a^λ (general $n \geq 2$)

We next briefly return to the treatment of the kernel ${}^{(1)}\mathcal{K}^{\nu,\mu}$ in Section 7.4.1, now treating the case of general dimension $n \geq 2$. We need to prove (7.4) and (7.6).

Assume $j < k$. We first prove (7.4). Recall from (7.13) that the kernel relevant to $T_1 T_1^*$ is

$$ {}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta) = \int_{\mathbb{R}^n} e^{iP_\nu(u+z)-iP_\mu(z)}\,\eta(u+z)\,\eta(z)\,\underline{\Delta}_{j-k}(\theta - |u+z|^2 + |z|^2)\,dz. $$

Again using the partition of unity (9.6), we write

$$ {}^{(1)}\mathcal{K}^{\nu,\mu}(u,\theta) = \sum_{l=1}^{n} {}^{(1)}\mathcal{K}_l^{\nu,\mu}(u,\theta), \tag{10.2} $$

with

$$ {}^{(1)}\mathcal{K}_l^{\nu,\mu}(u,\theta) = W_l\Big(\frac{u_l}{|u|}\Big) \int_{\mathbb{R}^n} e^{iP_\nu(u+z)-iP_\mu(z)}\,\eta(u+z)\,\eta(z)\,\underline{\Delta}_{j-k}(\theta - |u+z|^2 + |z|^2)\,dz. $$

For each index l we make the change of variables $z \mapsto (\tau,\sigma)$ defined with respect to the l-th coordinate in (9.2) and (9.3), so that

$$ {}^{(1)}\mathcal{K}_l^{\nu,\mu}(u,\theta) = \Big(\frac{|u_l|}{|u|}\Big)^{n-2} W_l\Big(\frac{u}{|u|}\Big) \int_{\mathbb{R}} K_{\sharp,l}^{\nu,\mu}(u,\tau)\,\underline{\Delta}_{j-k}(\theta - |u|^2 - 2|u|\tau)\,d\tau, \tag{10.3} $$

where

$$ K_{\sharp,l}^{\nu,\mu}(u,\tau) = \int_{\mathbb{R}^{n-1}} e^{iP_\nu(u+z)-iP_\mu(z)}\,\eta(u+z)\,\eta(z)\,d\sigma. $$

We apply the nontrivial bound of Proposition 9.1 to $K_{\sharp,l}^{\nu,\mu}$ and the trivial bound to $\underline{\Delta}_{j-k}$, to conclude that the analogue of (7.28) holds, uniformly in l. From here the analysis of $T_1 T_1^*$ proceeds as in (7.29), and we may conclude that this portion of the operator has norm bounded by $Cr^{-\delta/2}$.

We next prove (7.6). We again partition ${}^{(1)}\mathcal{K}_l^{\nu,\mu}(u,\theta)$ as in (10.2) with components given by (10.3) after the appropriate change of variables. We then use the identity (7.30) for $\underline{\Delta}_{j-k}$ as before, so that we may write each component as

$$ {}^{(1)}\mathcal{K}_l^{\nu,\mu}(u,\theta) = \frac{2^{2(j-k)}}{2|u|}\Big(\frac{|u_l|}{|u|}\Big)^{n-2} W_l\Big(\frac{u}{|u|}\Big) \cdot \int_{\mathbb{R}} \partial_\tau K_{\sharp,l}^{\nu,\mu}(u,\tau)\,\widetilde{\underline{\Delta}}_{j-k}(\theta - 2|u|\tau - |u|^2)\,d\tau, $$

with the Schwartz function $\widetilde{\underline{\Delta}}_{j-k}$ constructed in Lemma 4.5. We now note that since $K_{\sharp,l}^{\nu,\mu}(u,\tau)$ is supported where $|u| \leq 2$, $|\tau| \leq 1$ and is a smooth function of τ,

$$ |\partial_\tau K_{\sharp,l}^{\nu,\mu}(u,\tau)| \leq c_0^{-(n-1)}\, r\, \chi_{B_2}(u)\,\chi_{B_1}(\tau), $$

where c_0 is the absolute constant (9.7) coming from the restriction $|\sigma| \leq c_0^{-1}$, and the factor of r comes from bringing down coefficients of size $\|\nu\|, \|\mu\| \approx r$ when differentiating the phase $P_\nu(u+z) - P_\mu(z)$ with respect to τ. Hence

$$(10.4)\quad |{}^{(1)}\mathcal{K}_l^{\nu,\mu}(u,\theta)| \leq \frac{c_0^{-(n-1)} r\, 2^{2(j-k)}}{2|u|} \chi_{B_2}(u) \int \chi_{B_1}(\tau)\, \underline{\widetilde{\Delta}}_{j-k}(\theta - 2|u|\tau - |u|^2)\, d\tau,$$

uniformly in l. By the uniformity in l we see that we may sum the bounds provided by (10.4) over $1 \leq l \leq n$ so that a bound of the order (10.4) holds for the full kernel ${}^{(1)}\mathcal{K}^{\nu,\mu}$. Thus from here on the analysis of $T_1 T_1^*$ may proceed as from (7.31) to (7.32). This completes the proof.

11. Appendix: Proof of Theorem 5.1

Let $\gamma(t) = (t, |t|^2) \subset \mathbb{R}^{n+1}$ denote a parametrization of the paraboloid and define $\mathcal{H}_\lambda$ acting on Schwartz functions f on $\mathbb{R}^{n+1}$ by

$$(11.1)\qquad \mathcal{H}_\lambda f(x) = \int_{|t|>\lambda} f(x - \gamma(t))\, K(t)\, dt,$$

where K is a Calderón–Zygmund kernel. Our goal is to prove Theorem 5.1, which states that the maximal truncated singular Radon transform defined by

$$\mathcal{H}^* f(x) = \sup_{\lambda>0} |\mathcal{H}_\lambda f(x)|$$

is bounded on L^p for all $1 < p < \infty$. The following simple proof is based on the approach of [2].

We recall the non-truncated singular integral operator $\mathcal{H}$ along the paraboloid defined in (5.2) and the maximal Radon transform $\mathcal{M}_{\mathrm{Rad}}$ along the paraboloid, defined in (1.19), both of which are known to be bounded on L^p for $1 < p < \infty$. We first make the simple observation that if $\lambda, \lambda' > 0$ with $\lambda \leq \lambda' \leq 2\lambda$, then

$$|\mathcal{H}_{\lambda'} f| \lesssim |\mathcal{H}_\lambda f| + \mathcal{M}_{\mathrm{Rad}} f,$$

so that it suffices to prove that the operator

$$(11.2)\qquad f \mapsto \sup_{k\in\mathbb{Z}} |\mathcal{H}_{2^k} f|$$

is bounded on L^p for $1 < p < \infty$.

In preparation for the proof of this, observe that the Fourier multiplier $m(\xi)$ of $\mathcal{H}$ is given by

$$m(\xi) = \int_{\mathbb{R}^n} K(t)\, e^{-2\pi i \gamma(t)\cdot\xi}\, dt.$$

Similarly the Fourier multiplier $m_\lambda(\xi)$ of $\mathcal{H}_\lambda$ is given by

$$(11.3)\qquad m_\lambda(\xi) = \int_{|t|>\lambda} K(t)\, e^{-2\pi i \gamma(t)\cdot\xi}\, dt.$$

It will be convenient to define a non-isotropic dilation of $\xi = (\xi', \xi_{n+1}) \in \mathbb{R}^{n+1}$ by

$$\lambda \circ \xi = (\lambda \xi', \lambda^2 \xi_{n+1}).$$

Accordingly, we denote the non-isotropic norm of ξ by $\|\xi\| = |\xi'| + |\xi_{n+1}|^{1/2}$, and the non-isotropic ball of radius r by

$$\mathrm{Ball}_r = \{y : \|y\| \leq r\}.$$

We may also define a maximal function averaging over non-isotropic balls by

$$\mathcal{M}_{\mathrm{Ball}} f(x) = \sup_{a>0} \int f(x-y) \, \frac{1}{a^{n+2}} \, \chi_{\mathrm{Ball}_1}(a^{-1} \circ y) \, dy;$$

this is known to be a bounded operator on L^p for $1 < p \leq \infty$ (see [11], Chapter I).

Now let $\eta \in C_c^\infty(\mathbb{R}^{n+1})$ be a smooth bump function that is identically one for $\|\xi\| \leq 1/2$ and vanishes for $\|\xi\| \geq 1$. Define $\mathcal{R}_\lambda$ to be the operator with Fourier multiplier $(1 - \eta(\lambda \circ \xi)) m_\lambda(\xi)$; this may be compared pointwise to $\mathcal{H}_\lambda$ as follows (we defer the proof for the moment).

Lemma 11.1.

$$|\mathcal{H}_\lambda f| \lesssim |\mathcal{R}_\lambda f| + \mathcal{M}_{\mathrm{Ball}}(\mathcal{H} f) + \mathcal{M}_{\mathrm{Ball}} f.$$

As the last two terms are known to be bounded on L^p for $1 < p < \infty$, the proof of (11.2) reduces to showing that

$$f \mapsto \sup_{k \in \mathbb{Z}} |\mathcal{R}_{2^k} f| \quad \text{is bounded on } L^p \text{ for } 1 < p < \infty. \tag{11.4}$$

For each $l \in \mathbb{Z}$ we set

$$K^{(l)}(u) = \frac{1}{2^{ln}} K\Big(\frac{u}{2^l}\Big);$$

as is well known, each such kernel is also a Calderón–Zygmund kernel satisfying the same bounds as K, with constants uniform in l. We now define

$$\widetilde{m}^{(l)}(\xi) = \int_{1 \leq |t| \leq 2} K^{(l)}(t) \, e^{-2\pi i \gamma(t) \cdot \xi} \, dt. \tag{11.5}$$

Then for each $k \in \mathbb{Z}$ we may decompose the multiplier m_{2^k} defined in (11.3) as

$$m_{2^k}(\xi) = \sum_{j=0}^{\infty} \int_{2^{j+k} \leq |t| \leq 2^{j+k+1}} K(t) \, e^{-2\pi i \gamma(t) \cdot \xi} \, dt = \sum_{j=0}^{\infty} \widetilde{m}^{(j+k)}(2^{j+k} \circ \xi).$$

Correspondingly, $\mathcal{R}_{2^k}$ has the Fourier multiplier

$$\sum_{j=0}^{\infty} (1 - \eta(2^k \circ \xi)) \, \widetilde{m}^{(j+k)}(2^{j+k} \circ \xi),$$

so that if we let $T_{j,k}$ denote the operator with multiplier

$$(1 - \eta(2^k \circ \xi)) \, \widetilde{m}^{(j+k)}(2^{j+k} \circ \xi),$$

we have

$$\mathcal{R}_{2^k} = \sum_{j=0}^{\infty} T_{j,k}.$$

The proof of (11.4) will now follow quickly once we have the following two claims.

Lemma 11.2. *For each integer $j \geq 0$ and all $1 < p < \infty$,*

$$\| \sup_{k \in \mathbb{Z}} |T_{j,k}| \|_{L^p} \leq C_p \|f\|_{L^p}.$$

Lemma 11.3. *For each integer $j \geq 0$,*

$$\Big\| \sup_{k \in \mathbb{Z}} |T_{j,k}| \Big\|_{L^2} \leq \Big\| \Big(\sum_{k \in \mathbb{Z}} |T_{j,k} f|^2 \Big)^{1/2} \Big\|_{L^2} \leq C\, 2^{-j/2} \, \|f\|_{L^2}.$$

Assuming these facts for the moment, we note that

$$\sup_{k \in \mathbb{Z}} |\mathcal{R}_{2^k} f| \leq \sum_{j=0}^{\infty} \sup_{k \in \mathbb{Z}} |T_{j,k} f|. \tag{11.6}$$

But interpolating the results of Lemmas 11.2 and 11.3, we see that for all $1 < p < \infty$ there exists some $\varepsilon = \varepsilon(p) > 0$ such that

$$\Big\| \sup_{k \in \mathbb{Z}} |T_{j,k}| \Big\|_{L^p} \leq C\, 2^{-j\varepsilon} \|f\|_{L^p}.$$

Taking norms and adding these estimates in (11.6), we see that $\| \sup_{k \in \mathbb{Z}} |\mathcal{R}_{2^k} f| \|_{L^p}$ is finite for any $1 < p < \infty$, as desired. This completes the proof of Theorem 5.1, aside from the proofs of the three small lemmas, which we treat in the next section.

11.1. Proof of the lemmas

To prove Lemma 11.1, we note that certainly

$$m_\lambda(\xi) = (1 - \eta(\lambda \circ \xi))\, m_\lambda(\xi) + \eta(\lambda \circ \xi)\, m(\xi) + \eta(\lambda \circ \xi)\, (m_\lambda(\xi) - m(\xi)). \tag{11.7}$$

The first term on the right-hand side corresponds to $\mathcal{R}_\lambda$ by definition. The second term is the Fourier multiplier of the operator

$$f \mapsto \int \mathcal{H} f(x - y)\, \lambda^{-(n+2)}\, \check{\eta}(\lambda^{-1} \circ y)\, dy,$$

which may be majorized by $\mathcal{M}_{\mathrm{Ball}}(\mathcal{H} f)$, with a constant independent of λ. To handle the third term, we first note that we can write

$$m_\lambda(\xi) - m(\xi) = \int_{|t| \leq \lambda} K(t)\, e^{-2\pi i \gamma(t) \cdot \xi}\, dt = \int_{|t| \leq 1} K^{(\lambda^{-1})}(t)\, e^{-2\pi i \gamma(t) \cdot (\lambda \circ \xi)}\, dt,$$

which we denote by $m_1^{(\lambda^{-1})}(\lambda \circ \xi)$. (Here we are slightly abusing notation by letting $K^{(\rho)}$ denote $\rho^{-n} K(\rho^{-1} \cdot)$ for any parameter $\rho > 0$, not necessarily dyadic.)

Thus the third term in (11.7) is the dilation of the function

$$\eta(\xi)\, m_1^{(\lambda^{-1})}(\xi), \tag{11.8}$$

where $m_1^{(\lambda^{-1})}(\xi)$ is the Fourier transform of the compactly supported distribution defined (up to sign) for $(t,s) \in \mathbb{R}^{n+1}$ by

$$\chi_{|t|\leq 1}(t)\, K^{(\lambda^{-1})}(t)\, \delta_{s=|t|^2}.$$

The Fourier transform of this compactly supported distribution is a C^∞ function, and due to the uniformity of $K^{(\lambda^{-1})}$ in λ, the bound for the Fourier transform and all its derivative holds uniformly in λ. This verifies that (11.8) is a compactly supported C^∞ function (uniformly in λ), and hence we see that its inverse Fourier transform may be bounded on $\mathbb{R}^{n+1}$ by $C_N(1+\|x\|)^{-N}$ for any N (uniformly in λ). Thus upon dilating the last term in (11.7) by λ, multiplying with $\hat{f}(\xi)$, and taking the inverse Fourier transform, we see that this term may be bounded by $\mathcal{M}_{\mathrm{Ball}} f$, as claimed.

To prove Lemma 11.2, we let S_j denote the operator with multiplier $\widetilde{m}^{(j)}(2^j \circ \xi)$, that is,

$$S_j f(x) = \int_{2^j \leq |t| \leq 2^{j+1}} f(x-\gamma(t)) K(t)\, dt.$$

In particular, we may see immediately from the upper bound (1.9) for K that

$$|S_j f(x)| \lesssim \frac{1}{2^{nj}} \int_{2^j \leq |t| \leq 2^{j+1}} |f(x-\gamma(t))|\, dt \lesssim \mathcal{M}_{\mathrm{Rad}} f,$$

uniformly in j. In addition, by adapting the argument given above for Lemma 11.1, we see that uniformly in j,k,

$$|T_{j,k} f| \leq |S_{j+k} f| + \mathcal{M}_{\mathrm{Ball}}(S_{j+k} f) \lesssim \mathcal{M}_{\mathrm{Rad}} f + \mathcal{M}_{\mathrm{Ball}}(\mathcal{M}_{\mathrm{Rad}} f).$$

Thus we immediately obtain that for each $j \geq 0$ and any $1 < p < \infty$,

$$\Big\| \sup_{k\in\mathbb{Z}} |T_{j,k} f| \Big\|_{L^p} \leq C_p \|f\|_{L^p}.$$

To prove Lemma 11.3, it suffices by Plancherel's theorem to bound

$$\sum_{k\in\mathbb{Z}} \big|(1-\eta(2^k\circ\xi))\, \widetilde{m}^{(j+k)}(2^{j+k}\circ\xi)\big|^2$$

in L^∞. For this we will use the following van der Corput estimate:

$$\widetilde{m}^{(j)}(\xi) \leq \frac{C}{\|\xi\|^{1/2}}. \tag{11.9}$$

With this in hand, and applying the vanishing of $1-\eta$ near the origin, we conclude that

$$\sum_{k\in\mathbb{Z}} \big|(1-\eta(2^k\circ\xi))\,\widetilde{m}^{(j+k)}(2^{j+k}\circ\xi)\big|^2 \le \sum_{2^k\ge\|\xi\|^{-1}} |\widetilde{m}^{(j+k)}(2^{j+k}\circ\xi)|^2$$
$$\le \sum_{2^k\ge\|\xi\|^{-1}} \frac{C}{2^{j+k}\,\|\xi\|} = C2^{-j},$$

which is sufficient for Lemma 11.3. Finally, to prove (11.9) we recall that

$$\widetilde{m}^{(j)}(\xi) = \int_{1\le|t|\le 2} K^{(j)}(t)\, e^{-2\pi i(t\cdot\xi' + |t|^2\xi_{n+1})}\, dt.$$

If $|\xi'| \ge 8n|\xi_{n+1}|$, then the first derivative test gives

$$|\widetilde{m}^{(j)}(\xi)| \le |\xi'|^{-1},$$

uniformly in j. For indeed, without loss of generality we may assume that

$$|\xi_1'| \ge \frac{1}{n}\,|\xi'| \ge 8\,|\xi_{n+1}|.$$

Then the first partial with respect to t_1 of the phase is $-2\pi i(\xi_1' + 2t_1\xi_{n+1})$, which is bounded below in absolute value by $2\pi\cdot\frac{1}{2}|\xi_1'|$ if $|\xi_1'|\ge 8|\xi_{n+1}|$ and $1\le|t|\le 2$. Writing

$$e^{-2\pi i(t\cdot\xi'+|t|^2\xi_{n+1})} = \frac{\partial_{t_1}(e^{-2\pi i(t\cdot\xi'+|t|^2\xi_{n+1})})}{-2\pi i(\xi_1' + 2t_1\xi_{n+1})}$$

and integrating by parts once, our claim follows; note that the boundary terms also obey this bound. On the other hand, if $|\xi'|\le 8n|\xi_{n+1}|$ then the second derivative test applied to the integral in t_1 gives

$$|\widetilde{m}^{(j)}(\xi)| \le |\xi_{n+1}|^{-1/2};$$

see for example the Corollary of Proposition 2 of Chapter VIII in [11]. Combining these bounds proves (11.9).

Acknowledgements. The authors would like to thank Elias M. Stein for suggesting this problem and for a number of helpful conversations – as well as for many years of generous mentoring. The authors also thank Petru Mironescu and Jean van Schaftingen for a helpful discussion on stopping-times. The authors thank the Institute for Advanced Study, Princeton University, the University of Oxford, and the Hausdorff Center for Mathematics for providing pleasant working environments during a number of collaborative visits.

References

[1] Carleson, L.: On convergence and growth of partial sums of Fourier series. *Acta Math.* **116** (1966), 135–157.

[2] Duoandikoetxea, J. and Rubio de Francia, J. L.: Maximal and singular integral operators via Fourier transform estimate. *Invent. Math.* **84** (1986), no. 3, 541–561.

[3] Fefferman, C.: Pointwise convergence of Fourier series. *Ann. of Math. (2)* **98** (1973), 551–571.

[4] Hunt, R. A.: On the convergence of Fourier series. In *Orthogonal expansions and their continuous analogues (Proc. Conf., Edwardsville, Ill., 1967)*, 235–255. Southern Illinois Univ. Press, Carbondale, Ill, 1968.

[5] Lacey, M. and Thiele, C.: A proof of boundedness of the Carleson operator. *Math. Res. Lett.* **7** (2000), no. 4, 361–370.

[6] Lie, V.: The weak-L^2 boundedness of the quadratic Carleson operator. *Geom. Funct. Anal.* **19** (2009), no. 2, 457–497.

[7] Lie, V.: The polynomial Carleson operator. Preprint, arXiv: 1105.4504, 2017.

[8] Lie, V.: A note on the polynomial Carleson operator in higher dimensions. Preprint, arXiv: 1712.03092, 2017.

[9] Ricci, F. and Stein, E. M.: Harmonic analysis on nilpotent groups and singular integrals. I. Oscillatory integrals. *J. Funct. Anal.* **73** (1987), no. 1, 179–194.

[10] Sjölin, P.: Convergence almost everywhere of certain singular integrals and multiple Fourier series. *Ark. Mat.* **9** (1971), 65–90.

[11] Stein, E. M.: *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton Mathematical Series 43, Monographs in Harmonic Analysis III, Princeton University Press, Princeton, NJ, 1993.

[12] Stein, E. M.: Oscillatory integrals related to Radon-like transforms. Proceedings of the conference in honor of Jean-Pierre Kahane (Orsay, 1993), *J. Fourier Anal. Appl.* (1995), Special Issue, 535–551.

[13] Stein, E. M. and Wainger, S.: Problems in harmonic analysis related to curvature. *Bull. Amer. Math. Soc.* **84** (1978), no. 6, 1239–1295.

[14] Stein, E. M. and Wainger, S.: Oscillatory integrals related to Carleson's theorem. *Math. Res. Lett.* **8** (2001), no. 5–6, 789–800.

[15] Zorin-Kranich, P.: Maximal polynomial modulations of singular integrals. Preprint, arXiv: 1711.03524, 2018.

Received October 21, 2016; revised June 11, 2017. Published online March 5, 2019.

Lillian B. Pierce: Department of Mathematics, Duke University, 120 Science Drive, Durham NC 27708, USA.
E-mail: pierce@math.duke.edu

Po-Lam Yung: Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong.
E-mail: plyung@math.cuhk.edu.hk

Pierce has been partially supported by NSF DMS-1402121 and CAREER grant DMS-1652173, and earlier during this collaboration by NSF DMS-0902658, DMS-0635607, the Simonyi Fund at the Institute for Advanced Study and a Marie Curie Incoming International Fellowship at Oxford University, with additional support in Oxford via an EPSRC Developing Leaders grant. Yung was partially supported by NSF grant DMS-1201474, a Titchmarsh Fellowship at Oxford University, a Junior Research Fellowship from St. Hilda's College, and a direct grant for research from the Chinese University of Hong Kong (4053120).