Singular integrals and maximal operators related to Carleson's theorem and curves in the plane

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Joris Roos
aus
Quakenbrück

Bonn 2017

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Christoph Thiele
2. Gutachter: Prof. Dr. Herbert Koch
Tag der Promotion: 14.07.2017
Erscheinungsjahr: 2017

Abstract

In this thesis we study several different operators that are related to Carleson's theorem and curves in the plane.

An interesting open problem in harmonic analysis is the study of analogues of Carleson's operator that feature integration along curves. In that context it is natural to ask whether the established methods of time-frequency analysis carry over to an anisotropic setting. We answer that question and also provide certain partial bounds for the Carleson operator along monomial curves using entirely different methods. Another line of results in this thesis concerns maximal operators and Hilbert transforms along variable curves in the plane. These are related to Carleson-type operators via a partial Fourier transform in the second variable. A central motivation for studying these operators stems from Zygmund's conjecture on differentiation along Lipschitz vector fields. One of our results can be understood as proving a curved variant of this conjecture.

This thesis consists of five chapters.

In Chapter 0 we explain some of the historical background and the open problems that motivated the work in this thesis and give a summary of the main results.

In Chapter 1 we prove a weak $(2, 2)$ bound for an anisotropic analogue of Carleson's operator associated with a smooth multiplier. We also discuss some of the obstructions for proving boundedness of a Carleson operator along curves and in particular, the modulation symmetries of the parabolic Carleson operator.

In Chapter 2 we prove L^p bounds for an anisotropic variant of the bilinear Hilbert transform. The bilinear Hilbert transform is intimately related to Carleson's operator as hinted by their shared modulation symmetry.

In Chapter 3 we prove certain partial bounds for the Carleson operator along monomial curves. The content of this chapter is taken from a joint work with Shaoming Guo, Lillian Pierce and Po-Lam Yung which will appear in the Journal of Geometric Analysis.

In Chapter 4 we study maximal operators and Hilbert transforms along variable curves in the plane. The content of this chapter is taken from a joint work with Shaoming Guo, Jonathan Hickman and Victor Lie which will appear in the Proceedings of the London Mathematical Society.

Acknowledgments

First and foremost I am deeply indebted to my faculty advisor, Prof. Christoph Thiele. I greatly admire his brilliant mathematical intuition and am thankful for his enormous patience. He has introduced me to the fascinating world of real harmonic analysis and successfully guided me through times of frustration and long periods of lacking motivation without ever getting impatient with the sometimes very slow progressing of my work. During our weekly meetings he was always ready and willing to discuss any mathematical question that I asked him and could always offer a plethora of helpful comments and ideas. I am grateful for and amazed at the large amount of time that he is willing to invest for every one of his students despite what I imagine to be an extremely busy schedule.

Next I would like to thank my older academic brother and friend Shaoming Guo who has made a huge impact on my development during the last two years of my graduate studies and really was in many aspects like a second advisor to me. I continue to be astonished by his sharp intuition for mathematical concepts and admire his immense energy and passion for mathematics. He has introduced me to many of the methods and problems that define my first research interests. This is also reflected in the fact that a good portion of the content of this thesis consists of two projects that we have worked on and completed together. I look forward to continuing our joint work on the many other projects that we have.

I would like to express my gratitude to Po-Lam Yung for our collaboration, numerous very valuable discussions and for inviting me to Hong Kong during a decisive time in my last year of studies. I am thankful for his generosity and kindness and look up to him for the pinpoint accuracy and precision of his thinking.

I thank Lillian Pierce for collaboration and insightful conversations. I appreciate and admire her persistent pursuit of absolute perfection and her uncanny ability to make even the most dry technical discussion sound eloquent.

My sincere gratitude also goes to Victor Lie who has made a lasting impact on me during the final year of my studies. I am grateful for his generosity with ideas, our collaboration and for hosting me at Purdue University.

I also thank Jonathan Hickman for collaboration and a lot of fun moments that we had at MSRI.

Moreover, I would like to express my appreciation for all of the members of the Bonn Harmonic Analysis and PDE group with many of which I share a lot of fun memories and have enjoyed countless inspiring conversations. I thank the Studienstiftung for financial support and creating lots of opportunities to meet and make friends with many wonderful and interesting people. I also thank the Bonn International Graduate School for generous travel support. I am grateful to my friends and family in Wildeshausen and surrounding area, Bonn, Indiana and the rest of the world for their unconditional support throughout the years.

I thank Polona Durcik for being my accomplice and best friend during all the ups and downs in these exciting and adventurous years of our graduate studies in Bonn.

Dedicated to my grandparents

Contents

Chapter 0

Introduction

A fundamental problem in harmonic analysis is to determine under what circumstances the Fourier series of a function on the unit circle converges pointwisely. While it is relatively easy to see that Fourier series of continuously differentiable functions converge uniformly, a final answer for the case of continuous functions has eluded mathematicians for decades. A breakthrough in the study of this problem was made by Lennart Carleson in 1966 [Car66], when he proved that Fourier series of square-integrable functions converge almost everywhere. This resolved a previously long-standing conjecture of Luzin. Carleson's proof introduced a number of new ideas and techniques to the field. Carleson's theorem was extended to L^p functions (where $1 < p < \infty$) by Hunt [Hun68]. An example due to Kolmogorov [Kol26] shows that there exist L^1 functions whose Fourier series diverge everywhere. From a modern perspective, the object that lies at the heart of the proof of Carleson's theorem is the *Carleson maximal operator* (also referred to as Carleson operator)[1]:

$$\mathscr{C}f(x) = \sup_{N\in\mathbb{R}} \left| \int_{-\infty}^{N} \widehat{f}(\xi) e^{ix\xi} d\xi \right| .$$

The boundedness properties of this operator are intimately tied to almost everywhere convergence of the Fourier inversion integral. For instance, it is easy to see that the L^2 bound

$$\|\mathscr{C}f\|_2 \leq C\|f\|_2$$

implies almost everywhere convergence of the Fourier inversion integral for $L^2(\mathbb{R})$ functions. This follows by approximating the $L^2(\mathbb{R})$ function by test functions and then using the boundedness of the Carleson operator to control the error term.

A good way to gain insight into the nature of a given operator is to look at its symmetries. For this purpose it is convenient to express the Carleson operator as follows:

$$\mathscr{C}f(x) = \sup_{N\in\mathbb{R}} \left|\mathrm{M}_N \mathcal{H} \mathrm{M}_{-N} f(x,y)\right| , \tag{1}$$

where $\mathrm{M}_\xi f(x) = e^{ix\xi} f(x)$ and $\mathcal{H}$ denotes the Fourier projection to the left half line in frequency. That is, $\widehat{\mathcal{H}f} = \widehat{f}\mathbf{1}_{(-\infty,0)}$. The operator $\mathcal{H}$ can be written as a linear combination of the Hilbert transform and the identity operator. From (1) it is evident

[1]We state the operator over the real numbers. Bounds carry over to the unit circle by a well-known transference principle (see [KT80]). The use of the letter N stresses the historical origin of the Carleson operator in the theory of Fourier series.

that the Carleson operator not only commutes with translations and dilations (which are the symmetries that characterize the Hilbert transform), but also hosts an additional symmetry, which is *modulation invariance*. To be precise, we have

$$\mathscr{C}(\mathrm{M}_\xi f) = \mathscr{C} f$$

for all $\xi \in \mathbb{R}$. Since modulation acts by translation in frequency, this property causes the failure of many classical tools such as Calderón-Zygmund theory and Littlewood-Paley decomposition, that rely on the origin being a distinguished point in frequency. A different kind of analysis is required for the study of such objects. The appropriate techniques were pioneered by Carleson and are today referred to as *time-frequency analysis*. C. Fefferman [Fef73] provided an alternative proof of Carleson's theorem. His proof is based on a careful decomposition of the Carleson operator, while Carleson's proof was based on decomposing the input function f. Finally, Lacey and Thiele [LT00] have introduced yet another, simplified method of proof, which elegantly exploits the underlying symmetries and decomposes both the operator and the function at the same time. Their method is inspired by their groundbreaking work on the bilinear Hilbert transform [LT97b], [LT99], which led to the resolution of Calderón's conjecture. The bilinear Hilbert transform shares the property of modulation invariance with Carleson's operator which is the reason that both operators are susceptible to the same time-frequency methods.

A number of variants and generalizations of Carleson's theorem have been considered in the literature many of which incorporate new interesting aspects into the analysis. These include higher dimensions, polynomial phase functions and Radon-type behavior. The results contained in this thesis touch on all three of these aspects.

An extension of the Carleson-Hunt theorem to higher dimensions was given by Sjölin [Sjö71] who used methods from Carleson's original article [Car66]. Pramanik and Terwilleger [PT03] studied analogues of Carleson's theorem in higher dimensions by adapting the method of Lacey and Thiele. Also see [GTT04] for an extension of [PT03] to L^p for $p > 1$.

A further direction of inquiry introduced a polynomial phase to the Carleson operator. Fix a natural number d. Consider the *polynomial Carleson operator* defined by

$$\mathscr{C}_d f(x) = \sup_P \left| \int_{\mathbb{R}^n} f(x-y) e^{iP(y)} K(y) dy \right|, \tag{2}$$

where the supremum runs over all polynomials with real coefficients of degree not exceeding d and K is a Calderón-Zygmund kernel. Stein asked whether this operator satisfies any L^p bounds. The main difficulty in this problem is caused by the presence of polynomial modulation symmetries. More precisely, setting $\mathrm{M}_P f(x) = e^{iP(x)} f(x)$ for an arbitrary polynomial P of degree up to d, we have

$$\mathscr{C}_d \mathrm{M}_P f = \mathscr{C}_d f.$$

V. Lie [Lie09], [Lie11] was able to answer this question[2] for the case $n = 1$ (with $K(x) = p.v.\frac{1}{x}$) using a delicate extension of Fefferman's time-frequency approach combined with oscillatory integral techniques. The case $n > 1$ is still open.

[2]Like Fefferman, Lie was working over the unit circle, but again the results can be transferred to $\mathbb{R}$.

Stein and Wainger [SW01] used a very different approach based only on TT^* and oscillatory integral estimates to obtain an L^2 bound for a variant of the operator (2) where the supremum is restricted to polynomials lacking linear terms. Their surprisingly simple and elegant proof works for all $n \geq 1$ and does not require any tools from time-frequency analysis. The reason that this seemingly small modification of the operator has such a dramatic effect is that it causes all modulation symmetries to disappear.

Pierce and Yung [PY15] have recently introduced a variant of the polynomial Carleson operator featuring Radon-type behavior by integrating along a paraboloid. Specifically, they studied the operator

$$f \mapsto \sup_P \left| \int_{\mathbb{R}^n} f(x-t, y-|t|^2) e^{iP(t)} K(t) dt \right|, \quad (x,y) \in \mathbb{R}^n \times \mathbb{R}, \tag{3}$$

where $n \geq 2$, K is a Calderón-Zygmund kernel and the supremum goes over all polynomials P of degree not exceeding d which are lacking linear as well as certain types of quadratic terms. They were able to obtain L^p bounds for $1 < p < \infty$ for this operator. Their method of proof is based on TT^* techniques in the spirit of Stein and Wainger [SW01]. As a consequence of the restriction in the supremum, the operator does not have any modulation symmetries. Moreover, the method does not work in the two-dimensional case corresponding to $n = 1$ in (3).

A natural object in this context that guided a lot of the work for this thesis is the Carleson operator along the monomial curve (t, t^d) in the plane.

Open Problem 1. *Fix an integer $d \geq 2$ and let*

$$H_d f(x,y) = p.v. \int_{\mathbb{R}} f(x-t, y-t^d) \frac{dt}{t}, \quad (x,y) \in \mathbb{R}^2,$$

Does the operator

$$\mathscr{C}^{(d)} f = \sup_{\xi \in \mathbb{R}^2} |\mathrm{M}_{-\xi} H_d \mathrm{M}_\xi f|$$

satisfy any L^p bounds?

(We continue to use the notation $\mathrm{M}_\xi f(x) = e^{ix\cdot\xi} f(x)$ for modulations in $\mathbb{R}^n$. Here $\cdot$ denotes the Euclidean inner product.)

The operator $\mathscr{C}^{(d)}$ features an *anisotropic dilation symmetry.* That is, setting

$$\mathrm{D}_\lambda f(x,y) = f(\lambda x, \lambda^d y)$$

for $\lambda > 0$, we have

$$\mathscr{C}^{(d)} \mathrm{D}_\lambda f = \mathrm{D}_\lambda \mathscr{C}^{(d)} f.$$

Thus it is natural to ask whether the techniques of time-frequency analysis carry over to the anisotropic setting. This turns out to be the case. The purpose of Chapters 1 and 2 is to develop and showcase the tools of time-frequency analysis in the anisotropic setting. To do that we study anisotropic analogues of the two central objects in time-frequency analysis, the Carleson operator and the bilinear Hilbert transform.

To state the main results of Chapters 1 and 2, let us fix a vector of positive integers

$\alpha = (\alpha_1, \dots, \alpha_n)$. We equip $\mathbb{R}^n$ with the anisotropic dilations given by

$$\delta_\lambda(x) = (\lambda^{\alpha_1} x_1, \dots, \lambda^{\alpha_n} x_n).$$

Let m be a bounded function on $\mathbb{R}^n$ that is smooth outside the origin and satisfies

$$m(\delta_\lambda \xi) = m(\xi)$$

for all $\lambda > 0$ and $\xi \neq 0$. Define

$$\mathscr{C}_m f(x) = \sup_{\xi \in \mathbb{R}^n} |\mathrm{M}_{-\xi} T_m \mathrm{M}_\xi f(x)|,$$

where T_m denotes the Fourier multiplier associated with m, i.e. $\widehat{T_m f} = \widehat{f} m$. If $n = 1$ and $m = \mathbf{1}_{(-\infty,0)}$, then $\mathscr{C}_m$ coincides with the Carleson operator.

Theorem 2. *Let m be a bounded function on $\mathbb{R}^n$ that is smooth outside the origin and satisfies $m(\xi) = m(\delta_\lambda(\xi))$ for all $\lambda > 0$ and $\xi \in \mathbb{R}^n$. Then we have*

$$\|\mathscr{C}_m f\|_{2,\infty} \leq C \|f\|_2,$$

with $C \in (0, \infty)$ depending on α and m.

The proof of this theorem is the main goal of Chapter 1 (see Theorem 1.1.1). It is based on Lacey and Thiele's time-frequency approach [LT00] and Pramanik and Terwilleger's generalization [PT03] thereof. The main difficulties arise in the tree estimate and reduction to the dyadic model operator. This provides a first step towards approaching Open Problem 1. A fundamental obstacle in solving Open Problem 1 is the low regularity of the multiplier of $\mathscr{C}^{(d)}$. In an attempt to approach Open Problem 1 via Theorem 2 we define a certain one-parameter family of toy model operators in Section 1.2 with the endpoint essentially being $\mathscr{C}^{(d)}$. While we cannot reach that endpoint, we can at least bound a certain range of the toy model operators.

For the special case of the *parabolic Carleson operator* $\mathscr{C}^{(2)}$ there are additional obstructions due to additional modulation symmetries. In particular, an L^2 bound for $\mathscr{C}^{(2)}$ would immediately imply an L^2 bound for Lie's quadratic Carleson operator [Lie09] (see Proposition 3.6.1).

We move on to discuss the main result of Chapter 2. Let B be a real and diagonal $n \times n$ matrix. We are interested in the bilinear operator

$$(f_1, f_2) \longmapsto \int_{\mathbb{R}^n} f_1(x+y) f_2(x+By) K(y) dy.$$

Here K is the kernel associated with the anisotropic multiplier m (that is, $\widehat{K} = m$). In the case $n = \alpha = 1$, $K(y) = p.v.\frac{1}{y}$ this operator coincides with the classical bilinear Hilbert transform studied by Lacey and Thiele [LT97b], [LT99]. The diagonal constraint on B is due to the fact that a non-diagonal B would break the dilation symmetry of the operator in the non-isotropic case. By duality it is equivalent to study the trilinear form

$$\Lambda_m(f_1, f_2, f_3) = \int_{\mathbb{R}^{2n}} f_1(x+y) f_2(x+By) f_3(x) K(y) d(x,y).$$

The main theorem of Chapter 2 reads as follows.

Theorem 3. *Let m be a bounded function on $\mathbb{R}^n$ that is smooth outside the origin and satisfies $m(\xi) = m(\delta_\lambda(\xi))$ for all $\lambda > 0$ and $\xi \in \mathbb{R}^n$. Further, suppose that* $\det B(1-B) \neq 0$. *Let $2 < p_1, p_2, p_3 < \infty$ with $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1$. Then there exists a constant $C \in (0, \infty)$ depending on $m, \alpha, B, p_1, p_2, p_3$ such that*

$$|\Lambda_m(f_1, f_2, f_3)| \leq C\|f_1\|_{p_1}\|f_2\|_{p_2}\|f_3\|_{p_3}.$$

In the proof we use the argument of Lacey and Thiele [LT97b] in reducing our trilinear form to a suitable model form. A difference is that our model form is continuous, rather than discrete. To prove boundedness of the model form we employ the framework of outer measure L^p spaces that was recently developed by Do and Thiele [DT15]. For that purpose we prove an extension of the generalized Carleson embedding theorem from [DT15] to the multidimensional anisotropic setting. An obstacle in proving that result is that the one-dimensional proof crucially depends on the linear ordering of $\mathbb{R}$.

In Chapter 3 we approach Open Problem 1 from a different angle. The content of this chapter is a joint work with Shaoming Guo, Lillian Pierce and Po-Lam Yung [GPRY16]. Let m, d be positive integers and f a Schwartz function in the plane. For $N \in \mathbb{R}$ let

$$H_N^{m,d} f(x, y) = p.v. \int_{\mathbb{R}} f(x - t, y - t^m) e^{iNt^d} \frac{dt}{t}.$$

The results from Chapter 3 can be summarized as follows.

Theorem 4. *Let m, d be positive integers and $p \in (1, \infty)$. Then there exists a constant $C_p \in (0, \infty)$ such that*

1. *if $m \neq d$ and $(m, d) \neq (2, 1)$, then*

$$\| \sup_{N\in\mathbb{R}} \|H_N^{m,d} f(x, y)\|_{L^p_y} \|_{L^p_x} \leq C_p \|f\|_p,$$

2. *if $d > 1$, then*

$$\| \sup_{N\in\mathbb{R}} \|H_N^{m,d} f(x, y)\|_{L^p_x} \|_{L^p_y} \leq C_p \|f\|_p.$$

To see the relation to Open Problem 1, consider the cases $d = m$ and $d = 1$. Observe that $H_N^{m,m} = M_{(0,N)} H_m M_{-(0,N)}$ and $H_N^{m,1} = M_{(N,0)} H_m M_{(-N,0)}$. Thus we recognize that, if $d \in \{1, m\}$, it is a special case of Open Problem 1 to show

$$\| \sup_{N\in\mathbb{R}} |H_N^{m,d} f| \|_p \lesssim \|f\|_p. \tag{4}$$

We cannot prove this estimate at the moment. However, note that if we interchange the order of supremum and L^p norm, we obtain a much weaker inequality that follows immediately from well-known singular Radon transform theory:

$$\sup_{N\in\mathbb{R}} \|H_N^{m,d} f\|_p \lesssim \|f\|_p.$$

In Theorem 4 we claim an improvement of this trivial inequality by interchanging "half" of the L^p norm with the supremum.
As mentioned above, Pierce and Yung have studied polynomial Carleson operators along

the paraboloid (3) in $\mathbb{R}^{n+1}$ with $n \geq 2$ in an earlier work [PY15]. The methods in that paper did not work in the planar case, which is the subject of this result. Another important aspect separating this work from [PY15] is that Theorem 4 treats some of the modulation invariant cases. Namely, if $d = m$ or $d = 1$, the operators under consideration feature a linear modulation symmetry. In fact, one can see via a partial Fourier transform and an application of Plancherel's theorem that our bounds for these cases imply Carleson's theorem. This idea goes back to Coifman and El Kohen, who used it in the context of Hilbert transforms along vector fields (also see the discussion in [BT13]) and will also play a role in our upcoming discussion of Chapter 4. Our analysis of the modulation invariant cases relies on a vector-valued estimate for the maximally truncated Carleson operator (see Theorem 3.3.1), which is a standard consequence of the scalar Carleson-Hunt theorem. If $d \neq m$ and $d > 1$, we do not depend on Carleson's theorem. A new ingredient in that case is a certain refinement of the method of Stein and Wainger [SW01]. Our core estimate depends on having good estimates from below for the determinant of a certain 4×4 matrix (see Lemma 3.4.5).

In Chapter 4 we prove L^p bounds for certain maximal operators and Hilbert transforms along variable monomial. The content of this chapter consists of a joint work with Shaoming Guo, Jonathan Hickman and Victor Lie [GHLR16].

To explain the motivation for this project let us recall a basic theorem in harmonic analysis. The Lebesgue differentiation theorem states that, for a locally integrable function f on $\mathbb{R}^n$, the limit

$$\lim_{r \to 0+} \frac{1}{|B_r|} \int_{B_r} f(x-y)dy$$

exists for almost every $x \in \mathbb{R}^n$ and is equal to $f(x)$. Here B_r denotes the ball of radius r around the origin. The key to proving this theorem is providing a weak $(1,1)$ bound for the Hardy-Littlewood maximal operator,

$$Mf(x) = \sup_{r>0} \frac{1}{|B_r|} \int_{B_r} |f(x-y)|dy.$$

There has been a lot of interest in the literature in replacing Euclidean balls by lower dimensional geometric objects such as spheres (resp. circles) or curves.

The following problem, known as Zygmund's conjecture (see [LL10] for further details), addresses the corresponding question for Lipschitz vector fields.

Open Problem 5. *Let f be locally square-integrable and u Lipschitz continuous. Do we have*

$$\lim_{\varepsilon \to 0+} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} f(x-t, y-u(x,y)t)dt = f(x)$$

for almost every $(x,y) \in \mathbb{R}^2$?

A counterexample based on Besicovitch-Kakeya sets shows that this is false for u being merely Hölder continuous of exponent strictly lower than one (see [LL10]).

For positive real numbers α, ε_0 and a function $u : \mathbb{R}^2 \to \mathbb{R}$ we define the maximal operator

$$\mathcal{M}_{u,\varepsilon_0}^{(\alpha)} f(x,y) = \sup_{0<\varepsilon<\varepsilon_0} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} |f(x-t, y-u(x,y)[t]^\alpha)|dt.$$

Here $[t]^\alpha$ may either stand for $|t|^\alpha$ or $\mathrm{sgn}(t)|t|^\alpha$. To solve Open Problem 5 it would suffice to give a weak $(2,2)$ bound for $\mathcal{M}^{(1)}_{u,\varepsilon_0}$ for Lipschitz u and some parameter $\varepsilon_0 > 0$. While we are not able to do that, we can answer the corresponding question for $\alpha \neq 1$.

Theorem 6. *Let α be a positive real number with $\alpha \neq 1$ and $u : \mathbb{R}^2 \to \mathbb{R}$ a function. Then the following hold:*

1. *if $2 < p \leq \infty$ and u is measurable, then we have*

$$\|\mathcal{M}^{(\alpha)}_{u,\infty} f\|_p \leq C_{p,\alpha} \|f\|_p,$$

2. *if u is Lipschitz continuous, then there exists $\varepsilon_0 = \varepsilon_0(\|u\|_{\mathrm{Lip}}) > 0$ such that for every $1 < p \leq 2$ we have*

$$\|\mathcal{M}^{(\alpha)}_{u,\varepsilon_0} f\|_p \leq C_{p,\alpha} \|f\|_p\,.$$

Here, $C_{p,\alpha} \in (0,\infty)$ is a constant that depends only on p and α.

As a corollary we obtain that for f locally in L^p for any $p > 1$ and u Lipschitz we have that

$$\lim_{\varepsilon\to 0+} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} f(x-t, y-u(x,y)[t]^{\alpha})dt = f(x)$$

holds for almost every $(x,y) \in \mathbb{R}^2$.

After obtaining this result we learned that the first item (the case $p > 2$) was already proven earlier by Marletta and Ricci in [MR98]. Their method of proof involves reduction to Bourgain's circular maximal operator from [Bou86] and uses that as a black box. We use an alternative approach based on local smoothing estimates in the spirit of Mockenhaupt, Seeger and Sogge [MSS92]. Another important ingredient is a certain interpolation scheme of Nagel, Stein and Wainger [NSW78]. Key components of our proof break down in the case $\alpha = 1$ because of the lack of curvature. However, we anticipate that other aspects of our proof might still shed some light on the case $\alpha = 1$.

It is also natural to consider the singular integral version of this problem. Let

$$\mathcal{H}^{(\alpha)}_u f(x,y) = p.v. \int_{\mathbb{R}} f(x-t, y-u(x,y)[t]^{\alpha}) \frac{dt}{t}.$$

We cannot prove the analogue of Theorem 6 for $\mathcal{H}^{(\alpha)}$. Instead, we need a one-variable assumption on the function u.

Theorem 7. *Let α be a positive real number with $\alpha \neq 1$ and $u : \mathbb{R}^2 \to \mathbb{R}$ a measurable function satisfying*

$$u(x,y) = u(x,0)$$

for almost every $(x,y) \in \mathbb{R}^2$. Then we have that for all $1 < p < \infty$ the following holds:

$$\|\mathcal{H}^{(\alpha)}_u f\|_p \leq C_{p,\alpha} \|f\|_p.$$

The constant $0 < C_{p,\alpha} < \infty$ depends only on p and α.

The corresponding result for the case $\alpha = 1$ was proven by Bateman and Thiele in [Bat13], [BT13] based on crucial earlier work by Lacey and Li [LL06], [LL10]. A key idea in their proof, that our proof also depends on, is to exploit the one-variable assumption by noticing that it causes our operator to commute with Littlewood-Paley projections in the second variable. Thus the argument naturally splits into a single annulus estimate and a square-function estimate. Their method of completing these two steps is very different from ours. In fact, the case $\alpha = 1$ implies Carleson's theorem and is thus best attacked using time-frequency methods. In contrast to that, we exploit curvature and use oscillatory estimates in the spirit of Stein and Wainger [SW01].

In order to achieve the L^p bounds for p other than two, we rely on a new pointwise estimate for comparing averages along curves to averages over rectangles using the shifted maximal function. By plugging in $u \equiv 1$ we thereby also obtain an alternative proof for the L^p boundedness of the singular Radon transform along the curve $(t, [t]^\alpha)$ when $p \neq 2$.

The case of Lipschitz u and $\alpha = 1$ corresponds to a well-known open problem (see [LL10]). S. Guo [Guo15], [Guo17a] was able to make some partial progress towards this open problem by extending the result of Bateman and Thiele to suitable measurable vector fields that are constant on Lipschitz curves.

We proceed to describe one more result from Chapter 4. Let $P_k^{(2)}$ denote a Littlewood-Paley projection in the second variable. That is, $\widehat{P_k^{(2)} f}(\xi, \eta)$ vanishes unless $|\eta| \approx 2^k$.

Theorem 8. *For every positive real α with $\alpha \neq 2$ and every $1 < p < \infty$, there exists $0 < C_{\alpha,p} < \infty$ such that for every measurable function $u = (u_1, u_2) : \mathbb{R} \to \mathbb{R}^2$ we have*

$$\Big\| p.v. \int_{\mathbb{R}} (P_k^{(2)} f)(x - t, y - u_1(x)t - u_2(x)[t]^\alpha) \frac{dt}{t} \Big\|_p \leq C_{\alpha,p} \|f\|_p,$$

uniformly in $k \in \mathbb{Z}$.

This should be understood as an extension of Bateman's single annulus estimate [Bat13]. The proof is by oscillatory integral estimates and uses Bateman's result as a black box to deal with the low-frequency component.

To conclude this introduction we return to Carleson's theorem by recording a simple consequence of Theorem 8 that can be proven using the partial Fourier transform trick.

Corollary 9. *For every positive real number α with $\alpha \neq 2$, we have*

$$\Big\| \sup_{u_1, u_2 \in \mathbb{R}} \Big| p.v. \int_{\mathbb{R}} f(x - t) e^{iu_1 t + iu_2 [t]^\alpha} \frac{dt}{t} \Big| \Big\|_2 \leq C_\alpha \|f\|_2,$$

with a constant C_α depending only on α.

The case $\alpha = 0$ is Carleson's theorem. Note that the case $\alpha = 2$ corresponds to Lie's quadratic Carleson operator $\mathscr{C}_2$ from (2) (see [Lie09]), which this result does not encompass due to the quadratic modulation symmetries present in that case.

Notation. Throughout this thesis we adopt the convention that C is a positive and finite constant that is only allowed to depend on certain fixed parameters depending on context. We do not care about the precise value of C and it may change from line to line. We also write $A \lesssim B$ to denote $A \leq C \cdot B$ and $A \approx B$ to denote $A \lesssim B$ and $B \lesssim A$. For convenience we choose to define the Fourier transform as

$$\mathcal{F} f(\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{ix \cdot \xi} f(x) dx.$$

We will mostly ignore the resulting factors involving π that pop up when we use Fourier inversion, Plancherel and other identities of Fourier analysis by including them into the constant. Moreover, all functions (in particular the function f) are test functions (say, Schwartz functions) unless specified otherwise. As a consequence, the estimates we obtain should be regarded as *a priori* estimates, which lend themselves to extension via standard density arguments which we will not comment on any further. All these are commonly used conventions in harmonic analysis.

Chapter 1

Anisotropic time-frequency analysis: The Carleson operator

1.1 Introduction

Let us consider $\mathbb{R}^n$ equipped with the anisotropic dilations given by

$$\delta_\lambda(x) = (\lambda^{\alpha_1} x_1, \ldots, \lambda^{\alpha_n} x_n). \tag{1.1.1}$$

Here $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a fixed vector of natural numbers. We write $|\alpha| = \sum_{i=1}^n \alpha_i$. For an integer ν, we say that a function m on $\mathbb{R}^n$ is in the class $\mathscr{M}^\nu$ if

1. m is bounded and contained in $C^\nu(\mathbb{R}^n \backslash \{0\})$, i.e. ν times continuously differentiable outside of the origin,

2. $m(\delta_\lambda(\xi)) = m(\xi)$ for all $\xi \neq 0$ and $\lambda > 0$.

Let us denote

$$\|m\|_{\mathscr{M}^\nu} = \sup_{0 \le \beta_1 + \cdots + \beta_n \le \nu} \sup_{\rho(\xi)=1} |\partial_1^{\beta_1} \cdots \partial_n^{\beta_n} m(\xi)|.$$

We define the Carleson operator associated with the multiplier m as

$$\mathscr{C}_m f(x) = \sup_{N \in \mathbb{R}^n} \left| \int_{\mathbb{R}^n} \widehat{f}(\xi) e^{ix\xi} m(\xi - N) d\xi \right|.$$

Then we can state our main result as follows.

Theorem 1.1.1. *Let ν_0 be a sufficiently large integer (depending only on α). There exists $C > 0$ depending only on α such that for all $m \in \mathscr{M}^{\nu_0}$ we have*

$$\|\mathscr{C}_m f\|_{2,\infty} \le C \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2. \tag{1.1.2}$$

Here $\|\cdot\|_{2,\infty}$ denotes the quasinorm of the Lorentz space $L^{2,\infty}(\mathbb{R}^n)$ (also called weak L^2) which is defined as

$$\|f\|_{2,\infty} = \sup_{\lambda > 0} \lambda |\{x \in \mathbb{R}^n : |f(x)| > \lambda\}|^{1/2}.$$

Remark 1.1.1. An inspection of the proof shows that we can choose ν_0 to be the smallest multiple of $\overline{\alpha}$ such that $\nu_0 > 3|\alpha| + 2$, where $\overline{\alpha}$ is the least common multiple of $\alpha_1, \ldots, \alpha_n$.

We do not claim that this is optimal. However, any improvement to a lower bound below $|\alpha|$ would require a different proof, since integrability of the integral in (1.3.2) is vital at every major step.

The proof of this theorem is based on the time-frequency techniques of Lacey and Thiele [LT00]. In the one-dimensional case $n = 1, \alpha = 1$ we recover the weak $(2, 2)$ bound for Carleson's operator, which immediately imply Carleson's theorem on the almost everywhere convergence of Fourier series [Car66] (up to transference). In the isotropic case $\alpha = (1, \dots, 1)$ the theorem follows from a result of Sjölin [Sjö71]. In [PT03], weak $(2, 2)$ bounds for the isotropic case are studied using the method of Lacey and Thiele. This is extended to strong (p, p) for $1 < p < \infty$ in [GTT04]. We speculate that Theorem 1.1.1 could also be extended to strong (p, p) for $1 < p < \infty$ using the methods from [GTT04].

The method of Lacey and Thiele involves several ingredients. The first step is a reduction to a discrete dyadic model operator that involves summation over certain regions in phase space which are called tiles. This is detailed in Section 1.3. In this step we encounter a small complication which is caused by the absence of rotation invariance in the anisotropic case. This is resolved using an anisotropic cone decomposition.

The next step is a certain procedure of combinatorial nature the purpose of which is to organize the tiles into certain collections (which are called trees) each of which is associated with a component of the operator that behaves more like a classical singular integral operator (see Lemma 1.4.4). The combinatorial part of the argument requires only little modification compared to the original procedure in [LT00] (see Sections 1.4, 1.5, 1.6). The technically most demanding part and the part where our proof differs most from the corresponding sections in [LT00] and [PT03] is the tree estimate (see Section 1.7).

Our main motivation for considering this anisotropic variant of Carleson's theorem roots in Open Problem 1 from Chapter 0. This will be further discussed in Section 1.2 along with an application of Theorem 1.1.1 to a certain family of rough multipliers that can be seen as a toy model for Open Problem 1. We will also discuss the special case of the parabolic Carleson operator in which there are further obstructions caused by additional modulation symmetries.

In Section 1.8 we collect the proofs of several standard estimates that are used throughout the main body of the argument.

Acknowledgments. The author thanks Christoph Thiele for his encouragement and countless valuable comments and conversations on this project. He is also indebted to Po-Lam Yung for numerous detailed and fruitful discussions, in particular (but not only) regarding certain aspects of the paper [PT03].

1.2 Motivation and an application

For a positive integer $d \geq 2$, let us consider the multiplier of the Hilbert transform along the curve (t, t^d) in the form

$$m_d(\xi, \eta) = p.v. \int_{\mathbb{R}} e^{i\xi t - i\eta t^d} \frac{dt}{t}, \ (\xi, \eta) \in \mathbb{R}^2. \tag{1.2.1}$$

Open Problem 1 asks for bounds for the Carleson operator along the curve (t, t^d). The multiplier m_d satisfies the anisotropic dilation symmetry

$$m_d(\lambda\xi, \lambda^d\eta) = m_d(\xi, \eta)$$

for $\lambda > 0$ and $(\xi, \eta) \neq 0$. However, Theorem 1.1.1 does not apply because m_d is too rough to be in the class $\mathscr{M}^\nu$ for any positive integer ν.

In an attempt to approach Open Problem 1 we introduce a family of toy model operators. In the following discussion we focus on the intersection of the quadrant $\{\xi \geq 0, \eta \geq 0\}$ with the region $\eta^{\frac{1}{d}} \leq 2\xi$. The other quadrants can be treated similarly (though depending on the parity of d the phase might not have a critical point in each quadrant; this is an inconsequential subtlety that we will ignore). Our restriction to the region $\eta^{\frac{1}{d}} \leq 2\xi$ is natural because stationary phase considerations show that m_d is smooth away from the axis $\eta = 0$. For $\xi > 0$ and $\eta > 0$ we define

$$m_{d,1}(\xi, \eta) = \left(\eta^{\frac{1}{d}}\xi^{-1}\right)^{\frac{d'}{2}} e^{i\left(\eta^{\frac{1}{d}}\xi^{-1}\right)^{-d'}} \psi(\eta^{\frac{1}{d}}\xi^{-1}), \tag{1.2.2}$$

where $\frac{1}{d} + \frac{1}{d'} = 1$ and ψ is a smooth cutoff function supported in $[-2, 2]$ and equal to one on a slightly smaller interval. We extend $m_{d,1}$ continuously by setting it equal to zero on the remainder of $\mathbb{R}^2$.

From a standard computation using the stationary phase principle (see [Ste93, Ch. VIII.1, Prop. 3]) we can see that, up to a negligible constant rescaling, this term constitutes the main contribution to the oscillatory integral in (1.2.1). The remainder term from stationary phase is smoother in the variables ξ and η and is therefore simpler to handle. We will ignore it for the purpose of this discussion.

From the definition we see directly that $m_{d,1}$ (and therefore also m_d) is only Hölder continuous of class $C^{\frac{1}{2(d-1)}}$ along the axis $\eta = 0$ while it is infinitely differentiable away from that axis.

Let us from here on denote

$$\zeta = \zeta(\xi, \eta) = \eta^{\frac{1}{d}}\xi^{-1}$$

for $\xi > 0, \eta > 0$. Since we do not know how to handle $m_{d,1}$ we introduce a family of modified, less oscillatory multipliers which is defined on the quadrant $\{\xi > 0, \eta > 0\}$ by

$$m_{d,\delta}(\xi, \eta) = \zeta^{\frac{d'}{2}} e^{i\zeta^{-d'\delta}} \psi(\zeta), \tag{1.2.3}$$

where $0 \leq \delta \leq 1$ is a parameter (and we again extend $m_{d,\delta}$ to the rest of $\mathbb{R}^2$ by zero).

The multiplier $m_{d,\delta}$ still fails to be in $\mathscr{M}^\nu$ for every positive integer ν. Thus, it may surprise that we can nevertheless apply Theorem 1.1.1 to bound $\mathscr{C}_{m_{d,\delta}}$ for small enough δ. For this purpose we express ψ in the form

$$\psi = \sum_{j \leq 0} \varphi_j,$$

where $\varphi_j(x) = \varphi(2^{-j}x)$ and φ is a smooth bump function supported in $[1/2, 2]$ and satisfying $\sum_{j\in\mathbb{Z}} \varphi_j(x) = 1$ for all $x \neq 0$. Then we have the following consequence of Theorem 1.1.1.

Corollary 1.2.1. *There exists $\delta_0 > 0$ such that for all $0 \le \delta < \delta_0$ we have*

$$\|\mathscr{C}_{m_{d,\delta}} f\|_{2,\infty} \lesssim \|f\|_2.$$

Proof. Let us write

$$m_{d,\delta,j}(\xi,\eta) = \zeta^{\frac{d'}{2}} e^{i\zeta^{-d'\delta}} \varphi_j(\zeta) \quad (\text{for } \xi > 0, \eta > 0),$$

$$T_j f(x,y) = \int_{\mathbb{R}^2} \widehat{f}(\xi,\eta) m_{d,\delta,j}(\xi,\eta) e^{ix\xi+iy\eta} d(\xi,\eta).$$

By the change of variables $\eta \mapsto 2^{jd}\eta$, we see that

$$T_j f(x,y) = 2^{j\frac{d'}{2}} 2^{jd} \int_{\mathbb{R}^2} \widehat{f}(\xi, 2^{jd}\eta) \widetilde{m}_{d,\delta,j}(\xi,\eta) e^{ix\xi+i2^{jd}y\eta} d(\xi,\eta) = 2^{j\frac{d'}{2}} \mathrm{D}_{2^{jd}} \widetilde{T}_j \mathrm{D}_{2^{-jd}} f(x,y),$$

where $\mathrm{D}_\lambda f(x,y) = f(x,\lambda y)$ and

$$\widetilde{m}_{d,\delta,j}(\xi,\eta) = \zeta^{\frac{d'}{2}} e^{i2^{-jd'\delta}\zeta^{-d'\delta}} \varphi(\zeta) \quad (\text{for } \xi > 0, \eta > 0),$$

$$\widetilde{T}_j f(x,y) = \int_{\mathbb{R}^2} \widehat{f}(\xi,\eta) \widetilde{m}_{d,\delta,j}(\xi,\eta) e^{ix\xi+iy\eta} d(\xi,\eta).$$

We have

$$\|\widetilde{m}_{d,\delta,j}\|_{\mathscr{M}^\nu} \lesssim 2^{-jd'\delta\nu}$$

for every integer $\nu \ge 0$ (where the implied constant depends on ν, d, δ and φ). Using Theorem 1.1.1 we therefore obtain

$$\|\mathscr{C}_{m_{d,\delta}}\|_{L^2 \to L^{2,\infty}} \lesssim \sum_{j \le 0} 2^{j\frac{d'}{2}} \|\mathscr{C}_{\widetilde{m}_{d,\delta,j}}\|_{L^2 \to L^{2,\infty}} \lesssim \sum_{j \le 0} 2^{jd'(\frac{1}{2} - \delta\nu_0)}.$$

Thus, setting $\delta_0 = \frac{1}{2}\nu_0^{-1}$ yields the claim. □

It is possible to increase δ_0 by lowering the number ν_0 in Theorem 1.1.1. However, we do not expect to be able to reach $\mathscr{C}_{m_{d,1}}$ (and therefore bound $\mathscr{C}_{m_d}$) in this way. On the other hand, note that we can bound $\mathscr{C}_{m_{d,0}}$ by this method even without the knowledge that $\nu_0 < \infty$.

For the case of the parabola, $d = 2$, there are some additional obstructions. Let us write the parabolic Carleson operator as

$$\mathscr{C}^{\mathrm{par}} f(x,y) = \sup_{N \in \mathbb{R}^2} \left| p.v. \int_{\mathbb{R}} f(x-t, y-t^2) e^{iN_1 t + iN_2 t^2} \frac{dt}{t} \right|.$$

Apart from the linear modulation symmetries given by

$$\mathscr{C}^{\mathrm{par}} f = \mathscr{C}^{\mathrm{par}} \mathrm{M}_N f$$

for $N \in \mathbb{R}^2$, there are some additional modulation symmetries. For a polynomial in two variables, $P = P(x,y)$, we write the corresponding polynomial modulation as

$$\mathrm{M}_P f(x,y) = e^{iP(x,y)} f(x,y).$$

Then we have that

$$\mathscr{C}^{\text{par}} \mathrm{M}_{Nx^2} f = \mathscr{C}^{\text{par}} f, \tag{1.2.4}$$
$$\mathscr{C}^{\text{par}} \mathrm{M}_{Nx(y+x^2)} f = \mathscr{C}^{\text{par}} f,$$
$$\mathscr{C}^{\text{par}} \mathrm{M}_{N(y+x^2)^2} f = \mathscr{C}^{\text{par}} f$$

hold for all $N \in \mathbb{R}$. We will see that these are all the polynomial modulation symmetries of the operator $\mathscr{C}^{\text{par}}$ (up to linear combination).

The quadratic modulation symmetry (1.2.4) suggests a connection to Lie's quadratic Carleson operator [Lie09]. Indeed, even a certain *partial* L^2 bound for $\mathscr{C}^{\text{par}}$ would immediately imply an L^2 bound for the quadratic Carleson operator (see Proposition 3.6.1 in Chapter 3 for the details).

In light of these symmetries it is natural to perform the change of variables $y \mapsto y+x^2$. Let $\tau(x,y) = (x, y+x^2)$. This is a measure-preserving map that can be understood as a kind of nonlinear shear. Denoting

$$Af(x,y) = p.v. \int_{\mathbb{R}} f(x-t, y-t^2) \frac{dt}{t}$$

and

$$Bf(x,y) = p.v. \int_{\mathbb{R}} f(x-t, y-2xt) \frac{dt}{t}$$

we observe that

$$Af = B(f \circ \tau^{-1}) \circ \tau.$$

It is a simple and curious fact that this observation allows us to conclude L^p bounds for A from L^p bounds for B, which are known to hold in the range $p > \frac{3}{2}$ by a more general result of Bateman and Thiele on the Hilbert transform along one-variable vector fields [BT13] (see Theorem 4.1.5) which is proven using techniques from time-frequency analysis. On the other hand of course, A is long since known to be bounded in the full range $p > 1$ by entirely different techniques from the theory of singular Radon transforms (this follows for example as a special case of Theorem 4.1.2 in Chapter 4).

Similarly, any L^p bound for $\mathscr{C}^{\text{par}}$ would be equivalent to an L^p bound for

$$\mathscr{C}^{\text{sh}} f(x,y) = \sup_{N \in \mathbb{R}^2} \left| p.v. \int_{\mathbb{R}} f(x-t, y-2xt) e^{iN_1 t + iN_2 t^2} \frac{dt}{t} \right|.$$

Let us return to our original discussion of modulation symmetries of the parabolic Carleson operator. For the operator $\mathscr{C}^{\text{sh}}$ these symmetries take a very simple form. Indeed, it is easy to check that we have

$$\mathscr{C}^{\text{sh}} \mathrm{M}_P f = \mathscr{C}^{\text{sh}} f$$

for a polynomial P if and only if P is of degree at most two. Changing variables back we see that the list of polynomial modulation symmetries that we gave above for $\mathscr{C}^{\text{par}}$ is complete.

1.3 Reduction to a model operator

Before we begin we need to introduce some more notation and definitions. We fix the anisotropic norm

$$\rho(x) = \max\{|x_i|^{\frac{1}{\alpha_i}} \,:\, i = 1, \dots, n\}.$$

Observe that ρ satisfies the triangle inequality with constant ≤ 1 (depending on α) and moreover,

$$\min(|x|_\infty, |x|_\infty^{1/|\alpha|_\infty}) \leq \rho(x) \leq \max(|x|_\infty, |x|_\infty^{1/|\alpha|_\infty}), \tag{1.3.1}$$

where $|x|_\infty = \max_i |x_i|$ for $x \in \mathbb{R}^n$. It is easy to verify that

$$\int_{\mathbb{R}^n} (1+\rho(x))^{-\nu} dx \lesssim 1 \text{ and } \sum_{u \in \mathbb{Z}^n} (1+\rho(u))^{-\nu} \lesssim 1 \tag{1.3.2}$$

hold for all $\nu > |\alpha|$ with constants depending only on ν and $|\alpha|$. We will also use the notation

$$\operatorname{dist}_\alpha(A, B) = \inf_{x \in A, y \in B} \rho(x-y)$$

and $\operatorname{dist}_\alpha(A, x) = \operatorname{dist}_\alpha(A, \{x\})$. For $a, b \in \mathbb{R}^n$ we write

$$[a, b] = \prod_{i=1}^n [a_i, b_i]$$

and similarly $(a, b), [a, b)$. We will refer to all such sets as *rectangles*. For a rectangle $I \subset \mathbb{R}^n$ we define $c(I)$ to be its center. By an *anisotropic cube* we mean a rectangle $[a, b]$ such that $b_i - a_i = \lambda^{\alpha_i}$ holds for all $i = 1, \dots, n$ and some $\lambda > 0$. We define the collection of *anisotropic dyadic cubes* by

$$\mathcal{D}^\alpha = \{[\delta_{2^k}(\ell), \delta_{2^k}(\ell+1)) \,:\, \ell \in \mathbb{Z}^n,\, k \in \mathbb{Z}\}.$$

Every two anisotropic dyadic cubes have the property that they are either disjoint or contained in one another. Moreover, for every $I \in \mathcal{D}^\alpha$ there exists a unique dyadic cube $I^+ \in \mathcal{D}^\alpha$ such that $|I^+| = 2^{|\alpha|}|I|$ and $I \subset I^+$. We call I^+ the *parent* of I and say that I is a *child* of I^+.

Definition 1.3.1. A *tile* P is a rectangle in $\mathbb{R}^n \times \mathbb{R}^n$ of the form

$$P = I_P \times \omega_P,$$

where $I_P, \omega_P \in \mathcal{D}^\alpha$ and $|I_P| \cdot |\omega_P| = 1$.

The set of tiles is denoted by $\overline{\mathcal{P}}$. Given a tile P we denote its *scale* by $k_P = |I_P|^{1/|\alpha|}$. For $r \in \{0,1\}^n$ and a tile P with $\omega_P = [\delta_{2^{-k_P}}(\ell), \delta_{2^{-k_P}}(\ell+1)]$ we define the *semi-tile* $P(r)$ by

$$P(r) = I_P \times \omega_{P(r)}, \text{ where } \omega_{P(r)} = \Big[\delta_{2^{-k_P}}(\ell + \frac{1}{2}r), \delta_{2^{-k_P}}(\ell + \frac{1}{2}(r+1))\Big].$$

Note that $P(r)$ is not a tile and also $\omega_{P(r)} \notin \mathcal{D}^\alpha$. The model operator is built up using a large family of wave packets adapted to tiles. It is convenient to generate this family by letting the symmetry group of our operator act on a single bump function. For this purpose, let ϕ be a Schwartz function on $\mathbb{R}^n$ such that $0 \le \widehat{\phi} \le 1$ with $\widehat{\phi}$ being supported in $[-\frac{b_0}{2}, \frac{b_0}{2}]^n$ and equal to 1 on $[-\frac{b_1}{2}, \frac{b_1}{2}]^n$, where $0 < b_1 < b_0 \ll 1$ are some fixed, small numbers whose ratio is not too large (it becomes clear what precisely is required in Section 1.7). For example, we may set $b_0 = \frac{1}{10}$, $b_1 = \frac{9}{100}$. We denote translation, modulation and dilation of a function f by

$$\mathrm{T}_y f(x) = f(x-y), \qquad (y \in \mathbb{R}^n)$$

$$\mathrm{M}_\xi f(x) = e^{ix\xi} f(x), \qquad (\xi \in \mathbb{R}^n)$$

$$\mathrm{D}^p_\lambda f(x) = \lambda^{-\frac{|\alpha|}{p}} f\left(\delta_{\lambda^{-1}}(x)\right), \qquad (\lambda, p > 0),$$

where $|\alpha| = \sum_{i=1}^n \alpha_i$.

Given a tile P and $N \in \mathbb{R}^n$ we define the wave packets ϕ_P, ψ^N_P on $\mathbb{R}^n$ by

$$\phi_P(x) = \mathrm{M}_{c(\omega_{P(0)})} \mathrm{T}_{c(I_P)} \mathrm{D}^2_{2^{k(P)}} \phi(x) \tag{1.3.3}$$

$$\widehat{\psi^N_P}(\xi) = \mathrm{T}_N m(\xi) \cdot \widehat{\phi_P}(\xi) \tag{1.3.4}$$

We would like to think of ϕ_P as being time-frequency supported in the semi-tile $P(0)$. However, as an instance of the Heisenberg uncertainty principle a non-zero function can only be compactly supported either in frequency or in time (indeed, by the Paley-Wiener theorem the Fourier transform of a compactly supported function is analytic). Here we have that $\widehat{\phi_P}$ is compactly supported in (a small cube centrally contained in) $\omega_{P(0)}$ and $|\phi_P|$ decays rapidly outside of I_P.

For $N \in \mathbb{R}^n$ and $r \neq 0$ we introduce the dyadic model sum operator

$$A^{r,m}_N f(x) = \sum_{P \in \overline{\mathcal{P}}} \langle f, \phi_P \rangle \psi^N_P(x) \mathbf{1}_{\omega_{P(r)}}(N). \tag{1.3.5}$$

Theorem 1.3.2. *For every large enough integer ν_0 there exists $C > 0$ depending only on ν_0, α and the choice of ϕ such that for all multipliers $m \in \mathscr{M}^{\nu_0}$ we have*

$$\| \sup_{N \in \mathbb{R}^n} |A^{r,m}_N f| \|_{2,\infty} \le C \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2. \tag{1.3.6}$$

In the isotropic case $\alpha = (1, \dots, 1)$ this theorem was proved by Pramanik and Terwilleger [PT03]. An extension to strong L^p for $p \in (1, \infty)$ is contained in [GTT04]. We speculate that one can also extend (1.3.6) to L^p for $p \in (1, \infty)$ bounds using the approach of [GTT04].

The proof of the theorem is contained in Sections 1.4, 1.5, 1.6, 1.7. We conclude this section by showing that Theorem 1.3.2 implies Theorem 1.1.1. For this purpose we employ the averaging procedure of Lacey and Thiele [LT00] combined with an anisotropic cone decomposition of the multiplier m. The term *(anisotropic) cone* will always refer to a subset $\Theta \subsetneq \mathbb{R}^n$ of the form

$$\Theta = \{\delta_t(\xi) \,:\, t > 0,\, \xi \in Q\}$$

for some cube $Q \subset \mathbb{R}^n$. Let us denote $\mathcal{B}_s = \{x : \rho(x) \le s\}$. Let

$$\mathbf{A}^{r,m} f(x) = \lim_{R\to\infty} \frac{1}{R^{2|\alpha|}} \int_{\mathcal{B}_R} \int_{\mathcal{B}_R} \int_0^1 \mathrm{M}_{-\eta} \mathrm{T}_{-y} \mathrm{D}^2_{2^{-s}} A^{r,m}_{2^{-s}\eta} \mathrm{D}^2_{2^s} \mathrm{T}_y \mathrm{M}_\eta f(x) ds dy d\eta. \quad (1.3.7)$$

Lemma 1.3.3. *For every $r \in \{0,1\}^n$ and every test function f, the function $\mathbf{A}^{r,m} f(x)$ is well-defined and also a test function. We have*

$$\widehat{\mathbf{A}^{r,m}} f(\xi) = \theta_r(\xi) m(\xi) \widehat{f}(\xi)$$

for some smooth function θ_r that is independent of m. Moreover, there exists a constant $\varepsilon_0 > 0$ and an anisotropic cone Θ_r such that

$$\theta_r(\xi) > \varepsilon_0 \quad \textit{for all } \xi \in \Theta_r.$$

and

$$(-\infty, \varepsilon_0]^n \subset \bigcup_{r\in\{0,1\}^n\setminus\{0\}} \Theta_r. \quad (1.3.8)$$

Proof. Well-definedness (that is, existence of the limit) follows by pondering the following proof with that issue in mind. By expanding definitions we see that

$$(\mathrm{M}_{-\eta} \mathrm{T}_{-y} \mathrm{D}^2_{2^{-s}} A^{r,m}_{2^{-s}\eta} \mathrm{D}^2_{2^s} \mathrm{T}_y \mathrm{M}_\eta f)^\wedge(\xi)$$

is equal to (up to a universal constant)

$$m(\xi) \sum_{P\in\overline{\mathcal{P}}} \langle \widehat{f}, \mathrm{T}_{-\eta+\delta_{2^s}(c(\omega_{P(0)}))} \mathrm{M}_{y-\delta_{2^{-s}}(c(I_P))} \mathrm{D}^2_{2^{s-k_P}} \widehat{\phi} \rangle$$
$$\times \mathrm{T}_{-\eta+\delta_{2^s}(c(\omega_{P(0)}))} \mathrm{M}_{y-\delta_{2^{-s}}(c(I_P))} \mathrm{D}^2_{2^{s-k_P}} \widehat{\phi}(\xi) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)),$$

where we have used that $m(\delta_{2^{-s}}(\xi)) = m(\xi)$. The previous display equals

$$m(\xi) \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}^n} \sum_{u\in\mathbb{Z}^n} 2^{-|\alpha|(s-k)} \int_{\mathbb{R}^n} \widehat{f}(\zeta) e^{i(y-\delta_{2^{-s+k}}(u+\frac{1}{2}))(\xi-\zeta)} \overline{\widehat{\phi}}\Big(\delta_{2^{-s+k}}(\zeta+\eta) - \Big(\ell + \frac{1}{4}\Big)\Big) d\zeta$$
$$\times \widehat{\phi}\Big(\delta_{2^{-s+k}}(\xi+\eta) - \Big(\ell + \frac{1}{4}\Big)\Big) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)).$$

Applying the Poisson summation formula to the summation in u and using the Fourier support information of the function ϕ we see that the previous display equals (up to a universal constant)

$$m(\xi) \widehat{f}(\xi) \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}^n} |\widehat{\phi}|^2 \Big(\delta_{2^{-s+k}}(\xi+\eta) - \Big(\ell + \frac{1}{4}\Big)\Big) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)).$$

Observe that the expression no longer depends on the variable y. It remains to compute the function $\theta_r(\xi) = c \cdot \lim_{R\to\infty} I_R(\xi)$, where c is a universal constant and

$$I_R(\xi) = \frac{1}{R^{|\alpha|}} \int_{\mathcal{B}_R} \int_0^1 \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}^n} |\widehat{\phi}|^2 \Big(\delta_{2^{-s+k}}(\xi+\eta) - \Big(\ell + \frac{1}{4}\Big)\Big) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)) ds d\eta.$$

Note the formula

$$\int_0^1 \sum_{k\in\mathbb{Z}} F(2^{k-s})ds = \frac{1}{\log 2}\int_0^\infty F(t)\frac{dt}{t},$$

which follows from a change of variables $2^{k-s} \to t$. Using this we have

$$I_R(\xi) = \frac{c}{R^{|\alpha|}}\int_{\mathcal{B}_R}\int_0^\infty \sum_{\ell\in\mathbb{Z}^n} |\widehat{\phi}|^2\Big(\delta_t(\xi+\eta) - \Big(\ell+\frac{1}{4}\Big)\Big)\mathbf{1}_{Q_r}(\delta_t(\eta)-\ell)\frac{dt}{t}d\eta,$$

where $Q_r = \left[\frac{1}{2}r, \frac{1}{2}(r+1)\right] = \prod_{i=1}^n \left[\frac{1}{2}r_i, \frac{1}{2}(r_i+1)\right]$ and $c = (\log 2)^{-1}$ (c may change from line to line in this proof). To simplify our expression further we perform the change of variables

$$\delta_t(\xi+\eta) - \ell \to \zeta$$

in the integration in η. This yields

$$I_R(\xi) = c\int_{\mathbb{R}^n}\int_0^\infty \chi(\zeta)\mathbf{1}_{Q_r}(\zeta - \delta_t(\xi))\left(\sum_{\ell\in\mathbb{Z}^n}\frac{\mathbf{1}_{\rho(\zeta+\ell-\delta_t(\xi))\le tR}}{(tR)^{|\alpha|}}\right)\frac{dt}{t}d\zeta \tag{1.3.9}$$

where we have set

$$\chi(\zeta) = |\widehat{\phi}|^2\Big(\zeta - \frac{1}{4}\Big).$$

Observe that the integrand in (1.3.9) is supported in a compact subset of $\mathbb{R}^n \times (0,\infty)$ (which depends on ξ). By counting the ℓ for which the summand is non-zero we see that for every fixed $\zeta, \xi \in \mathbb{R}^n$ and $t > 0$ the sum

$$\sum_{\ell\in\mathbb{Z}^n}\frac{\mathbf{1}_{\rho(\zeta+\ell-\delta_t(\xi))\le tR}}{(tR)^{|\alpha|}}$$

converges to a universal constant as $R \to \infty$. Thus, from Lebesgue's dominated convergence theorem we conclude that

$$\theta_r(\xi) = c\int_{\mathbb{R}^n}\int_0^\infty \chi(\zeta)\mathbf{1}_{Q_r}(\zeta-\delta_t(\xi))\frac{dt}{t}d\zeta. \tag{1.3.10}$$

Evidently we have $\theta_r(\delta_t(\xi)) = \theta_r(\xi)$ for every $t > 0$ and $\xi \in \mathbb{R}^n$. From our choice of ϕ we get that χ is supported on $Q^{(0)}$ and equal to one on $Q^{(1)}$, where

$$Q^{(j)} = \Big[\frac{1}{4} - \frac{b_j}{2}, \frac{1}{4} + \frac{b_j}{2}\Big]$$

for $j = 0, 1$. Let us set

$$\Theta_r = \{\delta_t(\xi) \,:\, \xi \in Q^{(1)} - Q_r\}.$$

Then we can read off (1.3.10) that θ_r is greater than some positive constant on $\Theta_r^{(1)}$. Note that

$$Q^{(j)} - Q_r = \Big[-\frac{1}{2}r - \Big(\frac{1}{4}+\frac{b_j}{2}\Big), -\frac{1}{2}r + \Big(\frac{1}{4}+\frac{b_j}{2}\Big)\Big].$$

Looking at the anisotropic cone generated by each of the regions $Q^{(1)} - Q_r$ we see that (1.3.8) is satisfied for sufficiently small ε_0. □

In the isotropic case $\alpha = (1, \dots, 1)$ we can assume without loss of generality that the multiplier m is supported in some arbitrarily chosen cone (see [PT03]). Then we could finish the proof of Theorem 1.1.1 by applying Theorem 1.3.2 to a fixed r and the multiplier $\widetilde{m}(\xi) = \theta_r^{-1}(\xi)m(\xi)$. However, due to the lack of rotation invariance this assumption becomes invalid in the anisotropic setting.

Proof of Theorem 1.1.1. Let $m \in \mathscr{M}^{\nu_0}$. Without loss of generality we may assume that m is supported in the "quadrant" $(-\infty, 0]^n$ (this is all that is left of rotation invariance in the anisotropic setting). By (1.3.8) we can choose smooth functions $(\varrho_r)_r$ such that ϱ_r is supported in Θ_r and

$$\sum_{r \in \{0,1\}^n \setminus \{0\}} \varrho_r(\xi) = 1$$

for $\xi \in (-\infty, 0]^n$. By the triangle inequality and Lemma 1.3.3, we have

$$\|\mathscr{C}_m f\|_{2,\infty} \leq \sum_{r \in \{0,1\}^n \setminus \{0\}} \| \sup_{N \in \mathbb{R}^n} |\mathbf{A}^{r, \theta_r^{-1} \varrho_r m} \mathrm{M}_N f| \|_{2,\infty}.$$

Here θ_r^{-1} refers to the function $\xi \mapsto (\theta_r(\xi))^{-1}$, which is bounded on Θ_r. By (1.3.7) and Minkowski's integral inequality, the previous is no greater than

$$\sum_{r \in \{0,1\}^n \setminus \{0\}} \limsup_{R \to \infty} \frac{1}{R^{2|\alpha|}} \int_{\mathcal{B}_R} \int_{\mathcal{B}_R} \int_0^1 \| \sup_{N \in \mathbb{R}^n} |A_N^{r, \theta_r^{-1} \varrho_r m} \mathrm{D}_{2^s}^2 \mathrm{T}_y \mathrm{M}_\eta f| \|_{2,\infty} ds dy d\eta,$$

which by Theorem 1.3.2 is bounded by

$$C \sum_{r \in \{0,1\}^n \setminus \{0\}} \|\theta_r^{-1} \varrho_r m\|_{\mathscr{M}^{\nu_0}} \|f\|_2 \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2.$$

□

1.4 Boundedness of the model operator

In this section we describe the proof of Theorem 1.3.2. We follow [LT00]. First, we perform some preliminary reductions. Given a measurable function $N : \mathbb{R}^n \to \mathbb{R}^n$ we define

$$Tf(x) = A_{N(x)}^r f(x).$$

Note that the estimate (1.3.6) is equivalent to showing

$$\|Tf\|_{2,\infty} \leq C \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2$$

with C not depending on the choice of the measurable function N. This has the benefit that we are now concerned with a linear operator, rather than a sublinear operator. This linearization was already used by Kolmogorov and Seliverstov [KS24] to prove a certain (very) weak predecessor of Carleson's theorem. By duality, it is equivalent to show

$$|\langle Tf, \mathbf{1}_E \rangle| \lesssim \|m\|_{\mathscr{M}^{\nu_0}} |E|^{\frac{1}{2}} \|f\|_2,$$

where E is an arbitrary measurable set. By scaling, we may assume without loss of generality that $\|f\|_2 = 1$ and $|E| \leq 1$. Thus, by the triangle inequality, it suffices to show that

$$\sum_{P\in\mathcal{P}} |\langle f, \phi_P\rangle \langle \mathbf{1}_{E\cap N^{-1}(\omega_{P(r)})}, \psi_P^{N(\cdot)}\rangle| \lesssim \|m\|_{\mathscr{M}^{\nu_0}}, \tag{1.4.1}$$

for all finite sets of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$, with the implied constant being independent of $f, E, N, \mathcal{P}$. Throughout this and the following sections we fix $r \in \{0,1\}^n \setminus \{0\}$. Before we continue we need to introduce certain collections of tiles called trees. There is a partial order on tiles defined by

$$P \leq P' \qquad \text{if} \qquad I_P \subset I_{P'} \quad \text{and} \quad c(\omega_{P'}) \in \omega_P.$$

Observe that two tiles are comparable with respect to $\leq$ if and only if they have a non-empty intersection.

Definition 1.4.1. A finite collection $\mathbf{T} \subset \overline{\mathcal{P}}$ of tiles is called a *tree* if there exists $P \in \mathbf{T}$ such that $P' \leq P$ for every $P' \in \mathbf{T}$. In that case, P is uniquely determined and referred to as the *top* of the tree $\mathbf{T}$. We denote the top of a tree $\mathbf{T}$ by $P_{\mathbf{T}} = I_{\mathbf{T}} \times \omega_{\mathbf{T}}$ and write $k_{\mathbf{T}} = |I_{\mathbf{T}}|^{1/|\alpha|}$.
A tree $\mathbf{T}$ is called a 1*–tree* if $c(\omega_{\mathbf{T}}) \notin \omega_{P(r)}$ for all $P \in \mathbf{T}$ and it is called a 2*–tree* if $c(\omega_{\mathbf{T}}) \in \omega_{P(r)}$ for all $P \in \mathbf{T}$. These names are due to historical reasons (see [LT00]).

The notion of a tree was first introduced by C. Fefferman [Fef73]. For a tile $P \in \overline{\mathcal{P}}$ we write

$$E_P = E \cap N^{-1}(\omega_P) \text{ and } E_{P(r)} = E \cap N^{-1}(\omega_{P(r)}).$$

The *mass* of a single tile P is defined as

$$\mathcal{M}(P) = \sup_{P'\geq P} \int_{E_{P'}} w_{P'}^{\nu_1}(x)dx, \tag{1.4.2}$$

where ν_1 is a fixed large integer depending only on $|\alpha|$ that is to be determined later and

$$w_P^{\nu}(x) = \mathrm{T}_{c(I_P)}\mathrm{D}^1_{2^{k(P)}} w^{\nu}(x),$$

where the weight w^ν takes the form

$$w^\nu(x) = (1+\rho(x))^{-\nu}.$$

For convenience we also write $w_P = w_P^{\nu_1}$. For a collection of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$ we define their mass as

$$\mathcal{M}(\mathcal{P}) = \sup_{P\in\mathcal{P}} \mathcal{M}(P) = \sup_{P\in\mathcal{P}} \sup_{P'\geq P} \int_{E_{P'}} w_{P'}(x)dx. \tag{1.4.3}$$

The *energy* of a collection of tiles $\mathcal{P}$ is defined as

$$\mathcal{E}(\mathcal{P}) = \sup_{\mathbf{T}\subset\mathcal{P}\ 2-\text{tree}} \left(\frac{1}{|I_T|} \sum_{P\in\mathbf{T}} |\langle f, \phi_P\rangle|^2 \right)^{1/2}. \tag{1.4.4}$$

These quantities and the following lemmas originate in [LT00].

Lemma 1.4.2 (Mass lemma)**.** *There exists $C > 0$ depending only on α such that for every finite set of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$ there is a decomposition $\mathcal{P} = \mathcal{P}_{\text{light}} \cup \mathcal{P}_{\text{heavy}}$ such that*

$$\mathcal{M}(\mathcal{P}_{\text{light}}) \leq 2^{-2}\,\mathcal{M}(\mathcal{P}) \tag{1.4.5}$$

and $\mathcal{P}_{\text{heavy}}$ is a union of a set $\mathcal{T}$ of trees such that

$$\sum_{\mathbf{T}\in\mathcal{T}} |I_{\mathbf{T}}| \leq \frac{C}{\mathcal{M}(\mathcal{P})}. \tag{1.4.6}$$

Lemma 1.4.3 (Energy lemma)**.** *There exists $C > 0$ depending only on α such that for every finite set of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$ there is a decomposition $\mathcal{P} = \mathcal{P}_{\text{low}} \cup \mathcal{P}_{\text{high}}$ such that*

$$\mathcal{E}(\mathcal{P}_{\text{low}}) \leq 2^{-1}\,\mathcal{E}(\mathcal{P}) \tag{1.4.7}$$

and $\mathcal{P}_{\text{high}}$ is a union of a set $\mathcal{T}$ of trees such that

$$\sum_{\mathbf{T}\in\mathcal{T}} |I_{\mathbf{T}}| \leq \frac{C}{\mathcal{E}(\mathcal{P})^2}. \tag{1.4.8}$$

Lemma 1.4.4 (Tree estimate)**.** *There exists $C > 0$ depending only on α such that if $m \in \mathscr{M}^{\nu_0}$, then the following inequality holds for every tree $\mathbf{T}$:*

$$\sum_{P\in\mathbf{T}} |\langle f, \phi_P\rangle\langle \psi_P^{N(\cdot)}, \mathbf{1}_{E_{P(r)}}\rangle| \leq C\|m\|_{\mathscr{M}^{\nu_0}} |I_{\mathbf{T}}|\mathcal{E}(\mathbf{T})\mathcal{M}(\mathbf{T}) \tag{1.4.9}$$

The proofs of these lemmas are contained in Sections 1.5, 1.6 and 1.7. By iterated application of these lemmas we obtain a proof of (1.4.1). This argument is literally the same as in [LT00], but we include it here for convenience of the reader. Let $\mathcal{P}$ be a finite collection of tiles. We will decompose $\mathcal{P}$ into disjoint sets $(\mathcal{P}_\ell)_{\ell\in\mathcal{N}}$ (where $\mathcal{N}$ is some finite set of integers) such that each $\mathcal{P}_\ell$ satisfies

$$\mathcal{M}(\mathcal{P}_\ell) \leq 2^{2\ell} \quad \text{and} \quad \mathcal{E}(\mathcal{P}_\ell) \leq 2^{\ell} \tag{1.4.10}$$

and is equal to the union of a set of trees $\mathcal{T}_\ell$ such that

$$\sum_{\mathbf{T}\in\mathcal{T}_\ell} |I_{\mathbf{T}}| \leq C2^{-2\ell}. \tag{1.4.11}$$

This is achieved by the following procedure:

(1) Initialize $\mathcal{P}^{\text{stock}} := \mathcal{P}$ and choose an initial ℓ that is large enough such that

$$\mathcal{M}(\mathcal{P}^{\text{stock}}) \leq 2^{2\ell} \quad \text{and} \quad \mathcal{E}(\mathcal{P}^{\text{stock}}) \leq 2^{\ell}. \tag{1.4.12}$$

(2) If $\mathcal{M}(\mathcal{P}^{\text{stock}}) > 2^{2(\ell-1)}$, then apply Lemma 1.4.2 to decompose $\mathcal{P}^{\text{stock}}$ into $\mathcal{P}_{\text{light}}$ and $\mathcal{P}_{\text{heavy}}$. We add[1] $\mathcal{P}_{\text{heavy}}$ to $\mathcal{P}_\ell$ and update $\mathcal{P}^{\text{stock}} := \mathcal{P}_{\text{light}}$ (thus, we now have $\mathcal{M}(\mathcal{P}^{\text{stock}}) \leq 2^{2(\ell-1)}$).

(3) If $\mathcal{E}(\mathcal{P}^{\text{stock}}) > 2^{\ell-1}$, then apply Lemma 1.4.3 to decompose $\mathcal{P}^{\text{stock}}$ into $\mathcal{P}_{\text{low}}$

[1] We can think of all the $\mathcal{P}_\ell$ as being initialized by the empty set.

and $\mathcal{P}_{\text{high}}$. We add $\mathcal{P}_{\text{high}}$ to $\mathcal{P}_\ell$ and update $\mathcal{P}^{\text{stock}} := \mathcal{P}_{\text{low}}$ (thus, we now have $\mathcal{E}(\mathcal{P}^{\text{stock}}) \leq 2^{\ell-1}$).

(4) If $\mathcal{P}^{\text{stock}}$ is not empty, then replace ℓ by $\ell - 1$ and go to Step (2).

Then we can finish the proof of (1.4.1) by using (1.4.10), (1.4.11), (1.4.9) and keeping in mind that we always have $\mathcal{M}(\mathcal{P}) \leq \|w^{\nu_1}\|_1$:

$$\sum_{P\in\mathcal{P}} |\langle f, \phi_P\rangle \langle \mathbf{1}_{E\cap N^{-1}(\omega_{P(r)})}, \psi_P^{N(\cdot)}\rangle| = \sum_{\ell\in\mathcal{N}} \sum_{\mathbf{T}\in\mathcal{T}_\ell} \sum_{P\in\mathbf{T}} |\langle f, \phi_P\rangle \langle \mathbf{1}_{E\cap N^{-1}(\omega_{P(r)})}, \psi_P^{N(\cdot)}\rangle|$$
$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} \sum_{\ell\in\mathcal{N}} 2^\ell \min(1, 2^{2\ell}) \sum_{\mathbf{T}\in\mathcal{T}_\ell} |I_{\mathbf{T}}| \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \sum_{\ell\in\mathbb{Z}} 2^{-\ell} \min(1, 2^{2\ell}) \lesssim \|m\|_{\mathscr{M}^{\nu_0}}.$$

To conclude this section we list several basic estimates for m, K, ϕ_P, ψ_P^N which are used during the remainder of the proof. All of these are very standard, but for the sake of completeness we provide proofs in Section 1.8. First, from the definition of $\mathscr{M}^\nu$ we have the symbol estimate

$$|\partial_i^\nu m(\xi)| \leq \|m\|_{\mathscr{M}^\nu} \rho(\xi)^{-\nu\alpha_i} \tag{1.4.13}$$

for every integer $\nu \leq \nu_0$ and $i = 1, \dots, n$. If we let K denote the corresponding kernel (that is, $\widehat{K} = m$), we have

$$|K(x)| \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} \rho(x)^{-|\alpha|} \tag{1.4.14}$$

for $x \neq 0$. This is a consequence of an anisotropic variant of the well-known Hörmander-Mikhlin multiplier theorem (we prove this estimate in Section 1.8).

Recall that $\overline{\alpha}$ denotes the least common multiple of $\alpha_1, \dots, \alpha_n$. For every integer $\nu \geq 0$ which is a multiple of $\overline{\alpha}$ we have that ψ_P^N satisfies the following decay estimate provided that $N \notin \omega_{P(0)}$:

$$|\psi_P^N(x)| \lesssim \|m\|_{\mathscr{M}^\nu} |I_P|^{1/2} w_P^\nu(x), \tag{1.4.15}$$

where the implicit constant depends only on ν, α and the choice of ϕ.

The next estimates concern the interaction of two wave packets associated with distinct tiles. Let $P, P' \in \overline{\mathcal{P}}$ be tiles. The idea is that if P, P' are disjoint (or equivalently, incomparable with respect to $\leq$) then their associated wave packets are almost orthogonal, i.e. $\langle \phi_P, \phi_{P'}\rangle$ is negligibly small. Indeed, if ω_P and $\omega_{P'}$ are disjoint, then we even have $\langle \phi_P, \phi_{P'}\rangle = 0$. However, as an artifact of the Heisenberg uncertainty principle, in the case that only I_P and $I_{P'}$ are disjoint, we need to deal with tails. The precise estimate we need is as follows. Assume that $|I_P| \geq |I_{P'}|$. Then for every integer $\nu \geq |\alpha| + 1$ we have that

$$|\langle \phi_P, \phi_{P'}\rangle| \lesssim |I_P|^{-\frac{1}{2}} |I_{P'}|^{\frac{1}{2}} (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu}, \tag{1.4.16}$$

where the implicit constant depends only on ν and ϕ. See [Thi06, Lemma 2.1] for the version of this estimate for one-dimensional wave packets. Similarly, we have

$$|\langle \psi_P^N, \psi_{P'}^N\rangle| \lesssim \|m\|_{\mathscr{M}^\nu}^2 |I_P|^{-\frac{1}{2}} |I_{P'}|^{\frac{1}{2}} (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu}, \tag{1.4.17}$$

for every integer $\nu \geq |\alpha| + 1$ which is a multiple of $\overline{\alpha}$, provided that $N \notin \omega_{P(0)} \cup \omega_{P'(0)}$.

1.5 Proof of the mass lemma

In this section we prove Lemma 1.4.2. The proof is in essence the same as in [LT00, Prop. 3.1]. Let $\mathcal{P}$ be a finite set of tiles and set $\mu = \mathcal{M}(\mathcal{P})$. We define the set of heavy tiles by

$$\mathcal{P}_{\text{heavy}} = \Big\{P \in \mathcal{P} \,:\, \mathcal{M}(P) > \frac{\mu}{4}\Big\}$$

and accordingly $\mathcal{P}_{\text{light}} = \mathcal{P} \backslash \mathcal{P}_{\text{heavy}}$. Then (1.4.5) is automatically satisfied. It remains to show (1.4.6). By the definition of mass (1.4.2) we know that for every $P \in \mathcal{P}_{\text{heavy}}$ there exists a $P' = P'(P) \in \overline{\mathcal{P}}$ with $P' \geq P$ such that

$$\int_{E_{P'}} w_{P'}(x)dx > \frac{\mu}{4} \tag{1.5.1}$$

Note that P' need not be in $\mathcal{P}$. Let $\mathcal{P}'$ be the maximal elements in

$$\{P'(P) \,:\, P \in \mathcal{P}_{\text{heavy}}\}$$

with respect to the partial order $\leq$ of tiles. Then $\mathcal{P}_{\text{heavy}}$ is a union of trees with tops in $\mathcal{P}'$. Therefore it suffices to show

$$\sum_{P \in \mathcal{P}'} |I_P| \leq \frac{C}{\mu} \tag{1.5.2}$$

First we rewrite (1.5.1) as

$$\sum_{j=0}^{\infty} \int_{E_P \cap (\delta_{2^j}(I_P) \backslash \delta_{2^{j-1}}(I_P))} w_P(x)dx > C\mu \sum_{j=0}^{\infty} 2^{-j}. \tag{1.5.3}$$

where we adopt the temporary convention that $\delta_{2^{-1}}(I_P) = \emptyset$ and for $j \geq 0$,

$$\delta_{2^j}(I_P) = \prod_{i=1}^{n} \Big[c(I_P)_i - 2^{(k_P+j)\alpha_i - 1}, c(I_P)_i + 2^{(k_P+j)\alpha_i - 1}\Big).$$

Thus, for every $P \in \mathcal{P}'$ there exists a $j \geq 0$ such that

$$\int_{E_P \cap (\delta_{2^j}(I_P) \backslash \delta_{2^{j-1}}(I_P))} \frac{dx}{(1 + 2^{-k_P}\rho(x - c(I_P)))^{\nu_1}} > C|I_P|\mu 2^{-j}. \tag{1.5.4}$$

Note that for $x \in \delta_{2^j}(I_P) \backslash \delta_{2^{j-1}}(I_P)$ we have

$$1 + 2^{-k_P}\rho(x - c(I_P)) \geq C2^j.$$

Using this we obtain from (1.5.4),

$$|I_P| < C\mu^{-1}|E_P \cap \delta_{2^j}(I_P)|2^{-(\nu_1 - 1)j}. \tag{1.5.5}$$

Summarizing, we have shown that for every $P \in \mathcal{P}'$ there exists $j \geq 0$ such that (1.5.5) holds. This leads us to define for every $j \geq 0$, a set of tiles $\mathcal{P}_j$ by

$$\mathcal{P}_j = \{P \in \mathcal{P}' \,:\, |I_P| < C\mu^{-1}|E_P \cap \delta_{2^j}(I_P)|2^{-j(\nu_1-1)}\}.$$

The estimate (1.5.2) will follow by summing over j if we can show that

$$\sum_{P\in\mathcal{P}_j} |I_P| \leq C2^{-j}\mu^{-1} \tag{1.5.6}$$

for all $j \geq 0$. To show (1.5.6) we use a covering argument reminiscent of Vitali's covering lemma. Fix $j \geq 0$. For every tile $P = I_P \times \omega_P$ we have an enlarged tile $\delta_{2^j}(I_P) \times \omega_P$ (this is not a tile anymore). We inductively choose $P_i \in \mathcal{P}_j$ such that $|I_{P_i}|$ is maximal among the $P \in \mathcal{P}_j \backslash \{P_0, \ldots, P_{i-1}\}$ and the enlarged tile of P_i is disjoint from the enlarged tiles of $P_0, \ldots, P_{i-1}$. Since $\mathcal{P}_j$ is finite, this process terminates after finitely many steps, so that we have selected a subset $\mathcal{P}'_j = \{P_0, P_1, \ldots\} \subset \mathcal{P}_j$ of tiles whose enlarged tiles are pairwise disjoint. By construction, for every $P \in \mathcal{P}_j$ there exists a unique $P' \in \mathcal{P}'_j$ such that $|I_P| \leq |I_{P'}|$ and the enlarged tiles of P and P' intersect. We call P *associated* with P'.

Now the claim is that if two tiles $P, Q \in \mathcal{P}_j$ are associated with the same $P' \in \mathcal{P}'_j$, then I_P and I_Q are disjoint. To see this note that ω_P intersects $\omega_{P'}$ by definition. Thus, since $|I_P| \leq |I_{P'}|$, we have $\omega_{P'} \subset \omega_P$. The same holds for Q. Therefore we have $\omega_{P'} \subset \omega_P \cap \omega_Q$. But $P, Q \in \mathcal{P}_j \subset \mathcal{P}'$ are disjoint tiles, so we must have $I_P \cap I_Q = \emptyset$. Moreover, all tiles P associated with P' satisfy $I_P \subset \delta_{2^{j+2}}(I_{P'})$. Therefore we get

$$\begin{aligned}
\sum_{P\in\mathcal{P}_j} |I_P| &= \sum_{P'\in\mathcal{P}'_j} \sum_{\substack{P\in\mathcal{P}_j \\ \text{assoc. with } P'}} |I_P| = \sum_{P'\in\mathcal{P}'_j} \Bigg| \bigcup_{\substack{P\in\mathcal{P}_j \\ \text{assoc. with } P'}} I_P \Bigg| \\
&\leq \sum_{P'\in\mathcal{P}'_j} 2^{(j+2)|\alpha|}|I_{P'}| \leq C\mu^{-1}2^{-j(\nu_1-|\alpha|-1)} \sum_{P'\in\mathcal{P}'_j} |E \cap N^{-1}(\omega_{P'}) \cap \delta_{2^j}(I_{P'})| \\
&\leq C2^{-j}\mu^{-1},
\end{aligned}$$

provided that $\nu_1 > |\alpha| + 1$. The penultimate inequality is a consequence of (1.5.5) and the last inequality follows, because the sets $N^{-1}(\omega_{P'}) \cap \delta_{2^j}(I_{P'})$ are disjoint and $|E| \leq 1$.

1.6 Proof of the energy lemma

In this section we prove Lemma 1.4.3. We adapt the argument of Lacey and Thiele [LT00, Prop. 3.2]. The tree selection algorithm of Lacey and Thiele relies on the natural ordering of real numbers. In our situation this can be replaced by any functional on $\mathbb{R}^n$ that separates $\omega_{P(0)}$ from $\omega_{P(r)}$ for every tile $P \in \overline{\mathcal{P}}$ (this was already observed in [PT03]). Let i_0 be such that $r_{i_0} = 1$ (exists because $r \neq 0$). Let us introduce the projection to the i_0th coordinate: $\pi_0 : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto x_{i_0}$. Then we have that

$$\pi_0(\xi) < \pi_0(\eta) \tag{1.6.1}$$

holds for every $\xi \in \omega_{P(0)}, \eta \in \omega_{P(r)}, P \in \overline{\mathcal{P}}$.
Let $\varepsilon = \mathcal{E}(\mathcal{P})$. For a 2–tree $\mathbf{T}_2$ we define

$$\Delta(\mathbf{T}_2) = \left(\frac{1}{|I_{\mathbf{T}_2}|} \sum_{P \in \mathbf{T}_2} |\langle f, \phi_P \rangle|^2 \right)^{1/2}.$$

We will now describe an algorithm to choose the desired collection of trees $\mathcal{T}$ and also an auxiliary collection of 2–trees $\mathcal{T}_2$:

(1) Initialize $\mathcal{T} := \mathcal{T}_2 := \emptyset$ and $\mathcal{P}^{\text{stock}} := \mathcal{P}$.

(2) Choose a 2–tree $\mathbf{T}_2 \subset \mathcal{P}^{\text{stock}}$ such that

   (a) $\Delta(\mathbf{T}_2) \geq \varepsilon/2$, and
   (b) $\pi_0(c(\omega_{\mathbf{T}_2}))$ is minimal among all the 2–trees in $\mathcal{P}^{\text{stock}}$ satisfying (a).

   If no such $\mathbf{T}_2$ exists, then terminate.

(3) Let $\mathbf{T}$ be the maximal tree in $\mathcal{P}^{\text{stock}}$ with top $P_{\mathbf{T}_2}$ (with respect to set inclusion).

(4) Add $\mathbf{T}$ to $\mathcal{T}$ and $\mathbf{T}_2$ to $\mathcal{T}_2$. Also, remove all the elements of $\mathbf{T}$ from $\mathcal{P}^{\text{stock}}$. Then continue again with Step (2).

Since $\mathcal{P}$ is finite it is clear that the algorithm terminates after finitely many steps. Also note for every $\mathbf{T} \in \mathcal{T}$ there exists a unique $\mathbf{T}_2 \in \mathcal{T}_2$ with $\mathbf{T}_2 \subset \mathbf{T}$, and vice versa. After the algorithm terminates we set $\mathcal{P}_{\text{low}} = \mathcal{P}^{\text{stock}}$ and $\mathcal{P}_{\text{high}}$ to be the union of the trees in $\mathcal{T}$. Then, (1.4.7) is automatically satisfied and it only remains to show

$$\sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}| \lesssim \varepsilon^{-2}. \tag{1.6.2}$$

Before we do that we establish a geometric property of the selected trees that will be crucial in the following. This has been referred to as *strong disjointness* (see [Lac04]).

Lemma 1.6.1. *Let $\mathbf{T}_2 \neq \mathbf{T}'_2 \in \mathcal{T}_2$ and $P \in \mathbf{T}_2, P' \in \mathbf{T}'_2$. If $\omega_P \subset \omega_{1P'}$, then $I_{P'} \cap I_{\mathbf{T}_2} = \emptyset$.*

Proof. Note that $c(\omega_{\mathbf{T}_2}) \in \omega_P \subset \omega_{P'(0)}$ while $c(\omega_{\mathbf{T}'_2}) \in \omega_{P'(r)}$. By (1.6.1) and condition (b) in Step (2) we therefore conclude that $\mathbf{T}_2$ was chosen before $\mathbf{T}'_2$ during the above algorithm. Let $\mathbf{T}$ be the tree in $\mathcal{T}$ such that $\mathbf{T}_2 \subset \mathbf{T}$. Thus, if $I_{P'}$ was not disjoint from $I_{\mathbf{T}_2} = I_{\mathbf{T}}$, then it would be contained in $I_{\mathbf{T}}$ and therefore $P' \leq P_{\mathbf{T}}$ which means it would have been included into $\mathbf{T}$ during Step (3). That is a contradiction. □

The sum in (1.6.2) equals

$$\sum_{T_2 \in \mathcal{T}_2} \Delta(\mathbf{T}_2)^{-2} \sum_{P \in \mathbf{T}_2} |\langle f, \phi_P \rangle| \leq 4\varepsilon^{-2} \sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2,$$

where $\bigcup \mathcal{T}_2 = \bigcup_{\mathbf{T}_2 \in \mathcal{T}_2} \mathbf{T}_2$. Let us write

$$\sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2 = \Big\langle \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P, f \Big\rangle \tag{1.6.3}$$

and use the Cauchy-Schwarz inequality to estimate this by

$$\Big\| \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P \Big\|_2, \tag{1.6.4}$$

where we used that $\|f\|_2 = 1$. So far we have shown that

$$\varepsilon^2 \sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}| \lesssim \Big\| \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P \Big\|_2. \tag{1.6.5}$$

Thus if we can show that

$$\Big\| \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P \Big\|_2^2 \lesssim \varepsilon^2 \sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}|, \tag{1.6.6}$$

then (1.6.2) follows. Expanding the L^2 norm in (1.6.6) we get that the left hand side is bounded by

$$\sum_{\substack{P,P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} |\langle f, \phi_P \rangle \langle f, \phi_{P'} \rangle \langle \phi_P, \phi_{P'} \rangle| + 2 \sum_{\substack{P,P' \in \bigcup \mathcal{T}_2, \\ \omega_P \subset \omega_{P'(0)}}} |\langle f, \phi_P \rangle \langle f, \phi_{P'} \rangle \langle \phi_P, \phi_{P'} \rangle|. \tag{1.6.7}$$

Here we have used that $\langle \phi_P, \phi'_P \rangle = 0$ if $\omega_{P(0)} \cap \omega_{P'(0)} = \emptyset$ and therefore either $\omega_P = \omega_{P'}$, $\omega_P \subset \omega_{P'(0)}$, or $\omega_{P'} \subset \omega_{P(0)}$ (the last two cases are symmetric). We treat both sums in this term separately. Estimating the smaller one of $|\langle f, \phi_P \rangle|$ and $|\langle f, \phi_{P'} \rangle$ by the larger one, we obtain that the first sum in (1.6.7) is

$$\lesssim \sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2 \sum_{\substack{P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} |\langle \phi_P, \phi_{P'} \rangle|.$$

Using (1.4.16) we estimate this by

$$\sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2 \sum_{\substack{P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu}. \tag{1.6.8}$$

Notice that $I_P \cap I_{P'} = \emptyset$ for $P \neq P'$ in the inner sum. This implies

$$\sum_{\substack{P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu} \lesssim \int_{\mathbb{R}^n} (1 + \rho(x))^{-\nu} dx \lesssim 1,$$

provided that $\nu > |\alpha|$. Therefore (1.6.8) is

$$\lesssim \sum_{\mathbf{T}_2 \in \mathcal{T}_2} \sum_{P \in \mathbf{T}_2} |\langle f, \phi_P \rangle|^2 \leq \varepsilon^2 \sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}|, \tag{1.6.9}$$

as desired. It remains to estimate the second sum in (1.6.7). To that end it suffices to show that

$$\sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} |\langle f, \phi_P \rangle \langle f, \phi_{P'} \rangle \langle \phi_P, \phi_{P'} \rangle| \lesssim \varepsilon^2 |I_{\mathbf{T}_2}|, \tag{1.6.10}$$

for every $\mathbf{T}_2 \in \mathcal{T}_2$, where

$$\mathcal{S}_P = \Big\{ P' \in \bigcup \mathcal{T}_2 \,:\, \omega_P \subset \omega_{P'(0)} \Big\}.$$

Here we follow the argument given in [Lac04]. Observe that if $P \in \mathbf{T}_2$, then $\mathcal{S}_P \cap \mathbf{T}_2 = \emptyset$. Interpreting the singleton $\{P\}$ as a 2–tree we obtain

$$|\langle f, \phi_P \rangle| \le \varepsilon |I_P|^{1/2} \tag{1.6.11}$$

for all $P \in \mathcal{P}$. Combining this with (1.4.16) we can estimate the left hand side of (1.6.10) by

$$\varepsilon^2 \sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} |I_{P'}| (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu}. \tag{1.6.12}$$

This is the point where we make use of the strong disjointness property. Indeed, Lemma 1.6.1 implies that $I_{\mathbf{T}_2} \cap I_{P'} = \emptyset$ for every $P \in \mathbf{T}_2, P' \in \mathcal{S}_P$. Moreover, it also implies that for $P' \ne P'' \in \mathcal{S}_P$ we have $I_{P'} \cap I_{P''} = \emptyset$. These facts facilitate the following estimate:

$$\sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} |I_{P'}| (1+2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu} \lesssim \sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} \int_{I_{P'}} (1+2^{-k_P} \rho(c(I_P) - x))^{-\nu} dx$$

$$\lesssim \sum_{P \in \mathbf{T}_2} \int_{(I_{\mathbf{T}_2})^c} (1 + 2^{-k_P} \rho(c(I_P) - x))^{-\nu}.$$

Since $\mathbf{T}_2$ is a tree, the last quantity can be estimated by

$$\sum_{k \le k_{\mathbf{T}_2}} \sum_{u \in Q_k \cap (\mathbb{Z}^n + \frac{1}{2})} \int_{(I_{\mathbf{T}_2})^c} (1 + \rho(u - \delta_{2^{-k}}(x)))^{-\nu} dx,$$

where $Q_k \in \mathcal{D}^\alpha$ is an anisotropic dyadic rectangle of scale $k_{\mathbf{T}_2} - k$ that is given by a rescaling of $I_{\mathbf{T}_2}$. The previous display is no greater than a constant times

$$\sum_{k \le k_{\mathbf{T}_2}} 2^{k|\alpha|} \left(\sum_{u \in Q_k \cap (\mathbb{Z}^n + \frac{1}{2})} (1 + \mathrm{dist}_\alpha((Q_k)^c, u))^{-|\alpha| - \gamma} \right) \left(\int_{\mathbb{R}^n} (1 + \rho(x))^{-(\nu - |\alpha| - \gamma)} dx \right), \tag{1.6.13}$$

where $\nu > 2|\alpha|$ and γ is a fixed and sufficiently small positive constant. The integral over x in the previous display is bounded by a constant depending on $\nu - |\alpha| - \gamma > |\alpha|$. To estimate the sum over u we note that for every u in the indicated range there exists a lattice point $v \in \partial Q_k \cap \mathbb{Z}^n$ such that $\mathrm{dist}_\alpha((Q_k)^c, u) \ge \frac{1}{2} \rho(v - u)$. Thus we may bound the sum over u by

$$\sum_{v \in \partial Q_k \cap \mathbb{Z}^n} \sum_{u \in \mathbb{Z}^n + \frac{1}{2}} (1 + \rho(v - u))^{-|\alpha| - \gamma} \lesssim |\partial Q_k \cap \mathbb{Z}^n| \lesssim 2^{(k_{\mathbf{T}_2} - k)|\alpha|_\infty}.$$

Thus, (1.6.13) is bounded by a constant times

$$2^{k_{\mathbf{T}_2} |\alpha|_\infty} \sum_{k \le k_{\mathbf{T}_2}} 2^{k(|\alpha| - |\alpha|_\infty)} \lesssim 2^{k_{\mathbf{T}_2} |\alpha|} = |I_{\mathbf{T}_2}|.$$

This proves (1.6.10).

1.7 Proof of the tree estimate

In this section we prove Lemma 1.4.4. The proof follows the method set forth in [LT00], but diverges from it in some technical points. For a rectangle $I = \prod_{i=1}^n I_i \in \mathcal{D}^\alpha$ we denote by $\widetilde{I}$ the enlarged rectangle defined by

$$\widetilde{I} = \prod_{i=1}^{n} (2^{\alpha_i+1} - 1) I_i.$$

Here λI_i is the interval of length $\lambda|I_i|$ with the same center as I_i. Let $\mathcal{J}$ be the partition of $\mathbb{R}^n$ that is given by the collection of maximal anisotropic dyadic rectangles $J \in \mathcal{D}^\alpha$ such that $\widetilde{J}$ does not contain any I_P with $P \in \mathbf{T}$ (maximal with respect to inclusion). Set $\varepsilon = \mathcal{E}(\mathbf{T})$ and $\mu = \mathcal{M}(\mathbf{T})$. Choose phase factors $(\epsilon_P)_P$ of modulus 1 such that

$$\sum_{P\in\mathbf{T}} |\langle f, \phi_P\rangle \langle \psi_P^{N(\cdot)}, \mathbf{1}_{E_{P(r)}}\rangle| = \int_{\mathbb{R}^n} \sum_{P\in\mathbf{T}} \epsilon_P \langle f, \phi_P\rangle \psi_P^{N(x)}(x) \mathbf{1}_{E_{P(r)}}(x) dx$$

$$\leq \left\| \sum_{P\in\mathbf{T}} \epsilon_P \langle f, \phi_P\rangle \psi_P^{N} \mathbf{1}_{E_{P(r)}} \right\|_1 \leq \mathcal{K}_1 + \mathcal{K}_2,$$

where

$$\mathcal{K}_1 = \sum_{J\in\mathcal{J}} \sum_{P\in\mathbf{T}, |I_P|\leq|J^+|} \|\langle f, \phi_P\rangle \psi_P^{N(\cdot)} \mathbf{1}_{E_{P(r)}}\|_{L^1(J)},$$

$$\mathcal{K}_2 = \sum_{J\in\mathcal{J}} \left\| \sum_{P\in\mathbf{T}, |I_P|>|J^+|} \epsilon_P \langle f, \phi_P\rangle \psi_P^{N(\cdot)} \mathbf{1}_{E_{P(r)}} \right\|_{L^1(J)}.$$

We first estimate $\mathcal{K}_1$. This is the easy part, since in the sum defining $\mathcal{K}_1$ we have that I_P is disjoint from $\widetilde{J}$. Again, interpreting the singleton $\{P\}$ as a 2–tree we see that (1.6.11) holds for all $P \in \mathbf{T}$. This gives

$$\mathcal{K}_1 \leq \varepsilon \sum_{J\in\mathcal{J}} \sum_{\substack{P\in\mathbf{T}\\ |I_P|\leq|J^+|}} 2^{|\alpha|k_P/2} \|\psi_P^{N(\cdot)} \mathbf{1}_{E_{P(r)}}\|_{L^1(J)}.$$

Using (1.4.15) the previous display is seen to be no larger than a constant times

$$\|m\|_{\mathscr{M}^{\nu_1+\nu_2}} \varepsilon \sum_{J\in\mathcal{J}} \sum_{\substack{P\in\mathbf{T}\\ |I_P|\leq|J^+|}} \int_{J\cap E_{P(r)}} w^{\nu_1+\nu_2}(2^{-k_P}(x - c(I_P)))dx$$

$$\leq \|m\|_{\mathscr{M}^{\nu_1+\nu_2}} \varepsilon\mu \sum_{J\in\mathcal{J}} \sum_{\substack{P\in\mathbf{T}\\ |I_P|\leq|J^+|}} 2^{|\alpha|k_P} \sup_{x\in J} w^{\nu_2}(2^{-k_P}(x - c(I_P))), \tag{1.7.1}$$

where ν_1 is as in (1.4.3) and ν_2 is to be determined later. Since I_P is disjoint from $\widetilde{J}$ we have

$$w(2^{-k_P}(x - c(I_P))) \lesssim w(2^{-k_P}\mathrm{dist}_\alpha(J, I_P))$$

for $x \in J$. Thus (1.7.1) is

$$\lesssim \|m\|_{\mathscr{M}^{\nu_1+\nu_2}} \varepsilon\mu \sum_{J\in\mathcal{J}} \sum_{\substack{k\in\mathbb{Z},\\ 2^{k|\alpha|}\le|J^+|}} 2^{|\alpha|k} \sum_{\substack{P\in\mathbf{T},\\ k_P=k}} w^{\nu_2}(2^{-k}\mathrm{dist}_\alpha(J, I_P)). \tag{1.7.2}$$

Before we proceed, we claim that for every $\nu > |\alpha|$, $k \in \mathbb{Z}$ and fixed $J \in \mathcal{J}$ with $2^{k|\alpha|} \le |J^+|$ we have

$$\sum_{\substack{P\in\mathbf{T},\\ k_P=k}} w^{\nu}(2^{-k}\mathrm{dist}_\alpha(J, I_P)) \lesssim 1, \tag{1.7.3}$$

where the implicit constant blows up as ν approaches $|\alpha|$. To verify the claim, let us assume for simpler notation that J is centered at the origin. Then by disjointness of I_P and $\widetilde{J}$ we have

$$\mathrm{dist}_\alpha(J, I_P) \gtrsim \mathrm{dist}_\alpha(0, I_P) \gtrsim 2^k\rho(m),$$

where $m = (m_1, \ldots, m_n) \in \mathbb{Z}^n$ is such that $I_P = \prod_{i=1}^n [2^{k\alpha_i}m_i, 2^{k\alpha_i}(m_i + 1))$. Thus the sum in (1.7.3) is

$$\lesssim \sum_{m\in\mathbb{Z}^n} (1 + \rho(m))^{-\nu},$$

which implies the claim.

Now let us write $\nu_2 = \nu_3 + \nu_4$ with ν_3, ν_4 large enough and estimate (1.7.2) by

$$\lesssim \|m\|_{\mathscr{M}^{\nu_1+\nu_2}} \varepsilon\mu \sum_{J\in\mathcal{J}} w^{\nu_3}(2^{-k_\mathbf{T}}\mathrm{dist}_\alpha(J, I_T)) \sum_{\substack{k\in\mathbb{Z},\\ 2^{k|\alpha|}\le|J^+|}} 2^{|\alpha|k}. \tag{1.7.4}$$

Here $k_\mathbf{T}$ is the scale of $I_\mathbf{T}$ and we have used (1.7.3) and

$$2^{-k}\mathrm{dist}_\alpha(J, I_P) \ge 2^{-k_\mathbf{T}}\mathrm{dist}_\alpha(J, I_\mathbf{T}).$$

Summing the geometric series, (1.7.4) is

$$\lesssim \|m\|_{\mathscr{M}^{\nu_1+\nu_2}} \varepsilon\mu \sum_{J\in\mathcal{J}} w^{\nu_3}(2^{-k_\mathbf{T}}\mathrm{dist}_\alpha(J, I_P))|J|.$$

The sum in that expression can be estimated as follows:

$$\sum_{J\in\mathcal{J}} w^{\nu_3}(2^{-k_\mathbf{T}}\mathrm{dist}_\alpha(J, I_P))|J| \lesssim \sum_{J\in\mathcal{J}} \int_J (1 + 2^{-k_\mathbf{T}}\rho(x - c(I_\mathbf{T})))^{-\nu_2}dx.$$

By disjointness of the J we can bound this by

$$\int_{\mathbb{R}^n} (1 + 2^{-k_\mathbf{T}}\rho(x - c(I_\mathbf{T})))^{-\nu_3}dx = |I_\mathbf{T}| \int_{\mathbb{R}^n} (1 + \rho(x))^{-\nu_3}dx \lesssim |I_\mathbf{T}|,$$

where the last inequality requires ν_2 to be strictly larger than $|\alpha|$. To summarize, we showed that

$$\mathcal{K}_1 \lesssim \|m\|_{\mathscr{M}^{\nu_1+\nu_2}} \varepsilon\mu|I_{\mathbf{T}}|,$$

provided that $\nu_2 > 2|\alpha|$.

Let us proceed to estimating $\mathcal{K}_2$. This is considerably more difficult. We may assume that the sum runs only over those J for which there is a $P \in \mathbf{T}$ such that $|I_P| > |J^+|$. Then $|I_{\mathbf{T}}| > |J^+|$ and $J \subset \widetilde{I_{\mathbf{T}}}$. From now on let such a J be fixed. Define

$$G_J = J \cap \bigcup_{P\in\mathbf{T}, |I_P|>|J^+|} E_{P(r)} \tag{1.7.5}$$

Before proceeding we prove the following.

Lemma 1.7.1. *There exists a constant $C > 0$ independent of J such that*

$$|G_J| \leq C\mu|J| \tag{1.7.6}$$

Proof. By definition of J, there exists $P_0 \in \mathbf{T}$ such that I_{P_0} is contained in $\widetilde{J^+}$. We claim that there exists a tile $P_0 < P' < P_{\mathbf{T}}$ such that $|I_{P'}| = |J^{++}|$. Indeed, note $|I_{P_0}| \leq |J^{++}|$. If there is equality, we simply take $P' = P_0$. Otherwise we take $I_{P'} \in \mathcal{D}^\alpha$ to be the unique dyadic ancestor of I_{P_0} such that $|I_{P'}| = |J^{++}|$ and choose $\omega_{P'}$ accordingly such that it contains $c(\omega_{\mathbf{T}})$. Now we have

$$|\omega_P| = |I_P|^{-1} \leq |J^{++}|^{-1} = |I_{P'}|^{-1} = |\omega_{P'}|$$

for every tile $P \in \mathbf{T}$ with $|I_P| > |J^+|$. This implies $\omega_P \subset \omega_{P'}$ and thus

$$G_J \subset J \cap E_{P'}.$$

As a consequence,

$$|G_J| \leq \int_{E_{P'}} \mathbf{1}_J(x)dx \lesssim |I_{P'}| \int_{E_{P'}} w_{P'}(x)dx \lesssim \mu|J|.$$

□

Let us define

$$F_J = \sum_{\substack{P\in\mathbf{T},\\ |I_P|>|J^+|}} \epsilon_P \langle f, \phi_P\rangle \psi_P^{N(\cdot)} \mathbf{1}_{E_{P(r)}}. \tag{1.7.7}$$

Since every tree can be written as the union of a 1–tree and a 2–tree, we may treat each of these cases separately.

1.7.1 The case of 1–trees

Assume that $\mathbf{T}$ is a 1–tree. This is the easier case. The reason is that for every $P, P' \in \mathbf{T}$, $\omega_P \neq \omega_{P'}$ we have that $\omega_{P(r)}$ and $\omega_{P'(r)}$ are disjoint and thus we have good orthogonality of the summands in (1.7.7). Using (1.6.11) and (1.4.15) we see that

$$|F_J(x)| \leq c_\nu \varepsilon \sum_{\substack{P\in\mathbf{T},\\ |I_P|>|J^+|}} (1+2^{-k_P}\rho(x-c(I_P)))^{-\nu}\mathbf{1}_{E_{P(r)}}(x).$$

Using disjointness of the $E_{P(r)}$ this can be estimated by

$$c_\nu\varepsilon \cdot \sup_{k\in\mathbb{Z}} \sum_{m\in\mathbb{Z}^n+\frac{1}{2}} (1+2^{-k}\rho(x-\delta_{2^k}(m)))^{-\nu}.$$

By an index shift we see that

$$\sum_{m\in\mathbb{Z}^n+\frac{1}{2}} (1+2^{-k}\rho(x-\delta_{2^k}(m)))^{-\nu} = \sum_{m\in\mathbb{Z}^n+\frac{1}{2}} (1+\rho(m+\gamma))^{-\nu},$$

where $\gamma \in [0,1]^n$ depends on k and x. The last sum is $\lesssim 1$ independently of γ provided that $\nu > |\alpha|$. Thus we proved the pointwise estimate

$$|F_J(x)| \lesssim c_\nu\varepsilon. \tag{1.7.8}$$

Combining this with the support estimate (1.7.6) we obtain

$$\|F_J\|_{L^1(J)} \lesssim c_\nu\varepsilon\mu|J|. \tag{1.7.9}$$

Summing over the pairwise disjoint $J \subset \widetilde{I_{\mathbf{T}}}$ we obtain

$$\mathcal{K}_2 \lesssim c_\nu\varepsilon\mu|I_{\mathbf{T}}|$$

as desired.

1.7.2 The case of 2–trees

Here we assume that $\mathbf{T}$ is a 2–tree. This is the core of the proof. The additional x-dependence present in the wave packets $\psi_P^{N(x)}$ makes this part more technically demanding than the congruent argument in [LT00]. This is also a main difficulty that Pramanik and Terwilleger had to overcome in [PT03]. Our argument is perhaps a bit simpler than that given in [PT03]. The anisotropic setting requires a few technical modifications.

The goal is again to obtain a pointwise estimate for F_J. In the following we fix $x \in J$ such that $F_J(x) \neq 0$. Observe that the $\omega_{P(r)}$, $P \in \mathbf{T}$ are nested. Let us denote the smallest (resp. largest) $\omega_{P(r)}$ (resp. ω_P) such that $x \in N^{-1}(\omega_{P(r)}) \cap E$ by ω_- (resp. ω_+). Let $k_+ \in \mathbb{Z}$ be such that $|\omega_+| = 2^{k_+|\alpha|}$ and $k_- \in \mathbb{Z}$ such that $|\omega_-| = 2^{-k_-|\alpha|-n}$ (note from the definition that $\omega_- \notin \mathcal{D}^\alpha$ if $\alpha \neq (1,\dots,1)$). Then the nestedness property implies

$$F_J(x) = \sum_{\substack{P\in\mathbf{T},\\ k_+\leq k_P\leq k_-}} \epsilon_P\langle f, \phi_P\rangle\psi_P^{N(x)}(x)$$

Define

$$h_x = \mathrm{M}_{c(\omega_+)}\mathrm{D}^1_{2^{k_+}}\phi_+ - M_{c(\omega_-)}D^1_{2^{k_-}}\phi_-,$$

where $\phi_+(x) = b_1^{-n}\phi(b_1^{-1}x)$ and ϕ_- is a Schwartz function satisfying $0 \le \widehat{\phi_-} \le 1$ such that $\widehat{\phi_-}$ is supported on $[-\frac{b_2}{2}, \frac{b_2}{2}]$ and equals to one on $[-\frac{b_3}{2}, \frac{b_3}{2}]$, where $b_{j+2} = \frac{1}{2} + b_j$ for $j = 0, 1$. From the definition we see that $\widehat{h_x}$ is supported on $b_0 b_1^{-1}\omega_+ \cap (2b_3\omega_-)^c$ and equal to one on $\omega_+ \cap (2b_2\omega_-)^c$. In particular, $\widehat{h_x}(\xi)$ equals to one if $\xi \in \operatorname{supp}\phi_P$ and $k_+ \le k_P \le k_-$ and vanishes if k_P is outside this range. For technical reasons that become clear further below we need the support of $\widehat{h_x}$ to keep a certain distance to ω_-. We obtain

$$F_J(x) = \sum_{P\in\mathbf{T}} \epsilon_P \langle f, \phi_P\rangle (\psi_P^{N(x)} * h_x)(x).$$

Fix $\xi_0 \in \omega_{\mathbf{T}}$. We decompose

$$F_J(x) = \sum_{P\in\mathbf{T}} \epsilon_P\langle f, \phi_P\rangle(\psi_P^{\xi_0} * h_x)(x) + \sum_{P\in\mathbf{T}} \epsilon_P\langle f, \phi_P\rangle(\psi_P^{N(x)} - \psi_P^{\xi_0}) * h_x)(x) \quad (1.7.10)$$

$$= G * \mathrm{M}_{\xi_0}K * h_x(x) + G * (\mathrm{M}_{N(x)}K - \mathrm{M}_{\xi_0}K) * h_x(x) \quad (1.7.11)$$

where

$$G = \sum_{P\in\mathbf{T}} \epsilon_P\langle f, \phi_P\rangle\phi_P. \quad (1.7.12)$$

Before proceeding with the proof we record the following simple variant of a standard fact about maximal functions (see [Duo01]).

Lemma 1.7.2. *Let $\lambda > 0$ and w be an integrable function on $\mathbb{R}^n$ which is constant on $\{\rho(y) \le \lambda\}$ and radial and decreasing with respect to ρ, i.e.*

$$w(x) \le w(y)$$

if $\rho(x) \ge \rho(y)$, with equality if $\rho(x) = \rho(y)$. Let $x \in \mathbb{R}^n$ and $J \subset \mathbb{R}^n$ be such that $J \subset \{y : \rho(x-y) \le \lambda\}$. Then we have

$$|F * w|(x) \le \|w\|_1 \sup_{J\subset I} \frac{1}{|I|}\int_I |F(y)|dy,$$

where the supremum is taken over all anisotropic cubes $I \subset \mathbb{R}^n$.

Proof. First we assume that w is a step function. That is,

$$w(y) = \sum_{j=1}^{\infty} c_j \mathbf{1}_{\rho(y)\le r_j}$$

with $\lambda \le r_1 < r_2 < \cdots$. Then we have

$$F * w(x) = \sum_j r_j^{|\alpha|} c_j \frac{1}{r_j^{|\alpha|}} \int_{\rho(x-y)\le r_j} |F(y)|dy \le \|w\|_1 \sup_{J\subset I}\frac{1}{|I|}\int_I |F(y)|dy.$$

The general case follows by approximation of w by step functions and an application of Lebesgue's dominated convergence theorem. ☐

Since

$$|h_x(y)| \lesssim 2^{-k_+|\alpha|}|\phi_+|(\delta_{2^{-k_+}}(y)) + 2^{-k_-|\alpha|}|\phi_-|(\delta_{2^{-k_-}}(y)) \quad (1.7.13)$$

and $x \in J$, $|J| \leq 2^{k_+|\alpha|} \leq 2^{k_-|\alpha|}$ we have from Lemma 1.7.2 that

$$|G * \mathrm{M}_{\xi_0} K * h_x(x)| \lesssim \sup_{J \subset I} \frac{1}{|I|} \int_I |G * \mathrm{M}_{\xi_0} K(y)| dy. \tag{1.7.14}$$

Let us assume for the moment that we also have the estimate

$$|G * (\mathrm{M}_{N(x)} K - \mathrm{M}_{\xi_0} K) * h_x(x)| \lesssim \|m\|_{\mathscr{M}^\nu} \sup_{J \subset I} \frac{1}{|I|} \int_I |G(y)| dy \tag{1.7.15}$$

for some large enough integer ν. We will first show how to finish the proof from here. At the end of the section we will then show that (1.7.15) indeed holds.
From (1.7.11), (1.7.14), (1.7.15) and Lemma 1.7.1 we see that

$$\sum_{\substack{J \in \mathcal{J}, \\ J \subset \widetilde{I_{\mathbf{T}}}}} \|F_J\|_{L^1(J)} \lesssim \mu \sum_{\substack{J \in \mathcal{J}, \\ J \subset \widetilde{I_{\mathbf{T}}}}} |J| \left(\sup_{J \subset I} \frac{1}{|I|} \int_I |G * \mathrm{M}_{\xi_0} K(y)| dy + \|m\|_{\mathscr{M}^\nu} \sup_{J \subset I} \frac{1}{|I|} \int_I |G(y)| dy \right)$$

By disjointness of the $J \in \mathcal{J}$ this is no greater than

$$\mu \left(\|\mathcal{M}(G * \mathrm{M}_{\xi_0} K)\|_{L^1(\widetilde{I_{\mathbf{T}}})} + \|m\|_{\mathscr{M}^\nu} \|\mathcal{M}(G)\|_{L^1(\widetilde{I_{\mathbf{T}}})} \right), \tag{1.7.16}$$

where $\mathcal{M}$ denotes the maximal function defined by

$$\mathcal{M}F(y) = \sup_{y \in I} \frac{1}{|I|} \int_I |F|,$$

where the supremum runs over all anisotropic cubes $I \subset \mathbb{R}^n$. Clearly, $\mathcal{M}$ is a bounded operator $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. To see this, note that it is bounded pointwise by a composition of one-dimensional Hardy-Littlewood maximal functions applied in each component.

Applying the Cauchy-Schwarz inequality and the L^2 boundedness of $\mathcal{M}$ we see that (1.7.16) is

$$\lesssim \mu |I_T|^{\frac{1}{2}} \left(\|G * \mathrm{M}_{\xi_0} K\|_2 + \|m\|_{\mathscr{M}^\nu} \|G\|_2 \right).$$

By repeating the arguments that lead to the proof of (1.6.6), using (1.4.16) or (1.4.17), respectively, we obtain that

$$\|G * \mathrm{M}_{\xi_0} K\|_2 + \|m\|_{\mathscr{M}^\nu} \|G\|_2 \lesssim \|m\|_{\mathscr{M}^\nu} \varepsilon |I_{\mathbf{T}}|^{\frac{1}{2}}$$

for large enough ν (it should be $> 2|\alpha|$ and a multiple of $\overline{\alpha}$). This concludes the proof.

It remains to prove (1.7.15). Let us write

$$R(y) = (\mathrm{M}_{N(x)} K - \mathrm{M}_{\xi_0} K) * h_x(y).$$

The goal is to estimate R in a way that allows us to apply Lemma 1.7.2. We will give two different estimates for R. The first one is only effective if $\rho(y)$ is large and the second one if $\rho(y)$ is small. Let us start with the first estimate. By Fourier inversion, we can write $R(y)$ (up to a constant) as

$$\int_{\mathbb{R}^n} (m(\xi - N(x)) - m(\xi - \xi_0)) \widehat{h_x}(\xi) e^{i\xi y} d\xi. \tag{1.7.17}$$

Fix y and let i be such that $\rho(y) = |y_i|^{1/\alpha_i}$. Then we integrate by parts in the ith component to see that (1.7.17) is bounded by

$$\lesssim \rho(y)^{-\nu'\alpha_i} \int_{\mathbb{R}^n} \left| \partial_{\xi_i}^{\nu'} \Big[(m(\xi - N(x)) - m(\xi - \xi_0)) \widehat{h_x}(\xi) \Big] \right| d\xi \tag{1.7.18}$$

for every integer $\nu' \geq 0$, where we have used that $\rho(y) \geq 2^{k_-}$ to estimate $|\delta_{2^{-k_-}}(y)| \geq 2^{-k_-}\rho(y)$. Let ℓ be a non-negative integer. Using (1.4.13) we obtain

$$\left| \partial_{\xi_i}^{\ell} \Big[m(\xi - N(x)) - m(\xi - \xi_0) \Big] \right| \leq \|m\|_{\mathscr{M}^\ell} (\rho(\xi - N(x))^{-\ell\alpha_i} + \rho(\xi - \xi_0)^{-\ell\alpha_i}). \tag{1.7.19}$$

Recall that ξ_0 and $N(x)$ are contained in ω_- and the integrand of (1.7.18) is supported on $b_0 b_1^{-1}\omega_+ \cap (2b_3\omega_-)^c$. Also, there exist $\omega_1, \dots, \omega_M \in \mathcal{D}^\alpha$ such that

$$\omega_- \subsetneq \omega_1 \subsetneq \cdots \subsetneq \omega_M = \omega_+$$

and $|\omega_j| = 2^{-k_j|\alpha|}$ with $k_1 = k_-$ and $k_{j+1} = k_j - 1$. If $\xi \in (2b_3\omega_-)^c$ we have

$$\min(\rho(\xi - N(x)), \rho(\xi - \xi_0)) \gtrsim 2^{-k_-}. \tag{1.7.20}$$

On the other hand, if $\xi \in (b_0 b_1^{-1}\omega_j) \cap \omega_{j-1}^c$ for $j = 2, \dots, M$, then

$$\min(\rho(\xi - N(x)), \rho(\xi - \xi_0)) \gtrsim 2^{-k_j}. \tag{1.7.21}$$

Combining (1.7.19) and (1.7.20), (1.7.21) we get

$$|\partial_{\xi_i}^{\ell} \Big[m(\xi - N(x)) - m(\xi - \xi_0) \Big]| \lesssim \|m\|_{\mathscr{M}^\ell} \sum_{j=1}^{M} 2^{k_j \ell \alpha_i} \mathbf{1}_{b_0 b_1^{-1}\omega_j}(\xi). \tag{1.7.22}$$

We also have

$$|\partial_{\xi_i}^{\ell} \widehat{h_x}(\xi)| \lesssim 2^{k_+\ell\alpha_i} \mathbf{1}_{b_0 b_1^{-1}\omega_+}(\xi) + 2^{k_-\ell\alpha_i} \mathbf{1}_{2b_3\omega_-}(\xi). \tag{1.7.23}$$

Thus we see from (1.7.22) and (1.7.23) that for all $i = 1, \dots, n$ and $0 \leq \ell \leq \nu'$ we obtain

$$\int_{\mathbb{R}^n} \left| \partial_{\xi_i}^{\ell} \Big[m(\xi - N(x)) - m(\xi - \xi_0) \Big] \partial_{\xi_i}^{\nu' - \ell} \widehat{h_x}(\xi) \right| d\xi \lesssim \|m\|_{\mathscr{M}^{\nu'}} 2^{k_-(\nu'\alpha_i - |\alpha|)},$$

provided that $\nu'\alpha_i \geq |\alpha|$. Setting $\nu = \nu'\alpha_i \geq \nu'$, we have shown that

$$|R(y)| \lesssim \|m\|_{\mathscr{M}^\nu} 2^{-k_-|\alpha|} (2^{-k_-}\rho(y))^{-\nu}, \tag{1.7.24}$$

where ν is a multiple of $\overline{\alpha}$ satisfying $\nu \geq |\alpha|$. It remains to find a good estimate for $R(y)$ when $\rho(y)$ is small. Let us estimate

$$|R(y)| \leq R_+(y) + R_-(y),$$

where

$$R_\pm = |(\mathrm{M}_{N(x)}K - \mathrm{M}_{\xi_0}K) * \mathrm{D}^1_{2^{k_\pm}} \phi_\pm|.$$

The first claim is that if $\rho(y) \le 2^{k_\pm+1}$, then

$$R_\pm(y) \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} 2^{-k_\pm|\alpha|}. \tag{1.7.25}$$

(Here and throughout the proof of this claim $\pm$ always stands for a fixed choice of sign, either $+$ or $-$.) To see this, we first estimate $R_\pm(y)$ by

$$2^{-k_\pm|\alpha|} \int_{\mathbb{R}^n} |(e^{i(N(x)-\xi_0)z} - 1)K(z)\phi_\pm(2^{-k_\pm}(y-z))|dz \lesssim 2^{-k_\pm|\alpha|}(\mathbf{I} + \mathbf{II}),$$

where

$$\mathbf{I} = \int_{\rho(z)\le 2^{k_\pm+2}} |(e^{i(N(x)-\xi_0)z} - 1)K(z)\phi_\pm(2^{-k_\pm}(y-z))|dz, \quad \text{and}$$

$$\mathbf{II} = \sum_{j=2}^{\infty} \int_{2^{k_\pm+j}\le\rho(z)\le 2^{k_\pm+j+1}} |K(z)\phi_\pm(2^{-k_\pm}(y-z))|dz.$$

We first estimate $\mathbf{I}$. Changing variables $z \mapsto \delta_{2^{k_\pm+2}}(z)$ we see that

$$\mathbf{I} \lesssim \int_{\rho(z)\le 1} |(e^{i\delta_{2^{k_\pm+2}}(N(x)-\xi_0)z} - 1)K(z)|dz.$$

Since $|K(z)| \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}}\rho(z)^{-|\alpha|}$, the previous display is

$$\lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} |\delta_{2^{k_\pm+2}}(N(x)-\xi_0)| \int_{\rho(z)\le 1} |z|\rho(z)^{-|\alpha|}dz.$$

Using (1.3.1) and $\rho(\delta_{2^{k_\pm+2}}(N(x)-\xi_0)) = 2^{k_\pm+2}\rho(N(x)-\xi_0) \lesssim 1$ we can bound this further as

$$\lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} \int_{\rho(z)\le 1} \rho(z)^{1-|\alpha|}dz \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}}.$$

This proves that $\mathbf{I} \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}}$. It remains to treat $\mathbf{II}$. Here we make use of the fact that we have $\rho(y-z) \ge 2^{k_\pm+j-1}$ in the integrand of $\mathbf{II}$, because of our assumption $\rho(y) \le 2^{k_\pm}$. Using the decay of $\phi_\pm$ we obtain

$$\mathbf{II} \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} \sum_{j=2}^{\infty} 2^{-j} \int_{2^{k_\pm+j}\le\rho(z)\le 2^{k_\pm+j+1}} \rho(z)^{-|\alpha|}dz \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}}.$$

Thus we have proved (1.7.25). The only further ingredient which we need in order to verify (1.7.15) is a good estimate for $R_+(y)$ when $2^{k_++1} \le \rho(y) \le 2^{k_-}$. In order to do this we need to do a slightly more careful decomposition. Let us write

$$Q_\ell = [\delta_{2^{k_+}}(\ell), \delta_{2^{k_+}}(\ell+1)) = \prod_{i=1}^{n} [2^{k_+\alpha_i}\ell_i, 2^{k_+\alpha_i}(\ell_i+1))$$

for $\ell \in \mathbb{Z}^n$. Assume that $y \in Q_\ell$ with $1 \le |\ell|_\infty < 2^{k_- - k_+}$. We have

$$R_+(y) \le 2^{-k_+|\alpha|} \sum_{s\in\mathbb{Z}^n} \int_{Q_s} |(e^{i(N(x)-\xi_0)z} - 1)K(z)\phi_+(2^{-k_+}(y-z))|dz. \tag{1.7.26}$$

Moreover, the same estimates that were used to prove (1.7.25) yield

$$\int_{Q_s} |(e^{i(N(x)-\xi_0)z} - 1)K(z)\phi_+(2^{-k_+}(y-z))|dz$$
$$\lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} 2^{k_+-k_-}(1+\rho(s-\ell))^{-\nu}(1+\rho(s))^{1-|\alpha|}$$

Plugging this inequality into (1.7.26) we obtain

$$\begin{aligned} R_+(y) &\lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} 2^{-k_-} 2^{k_+(1-|\alpha|)} \sum_{s\in\mathbb{Z}^n} (1+\rho(s-\ell))^{-\nu}(1+\rho(s))^{1-|\alpha|} \\ &\lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} 2^{-k_-}(2^{k_+}\rho(\ell))^{1-|\alpha|}, \end{aligned}$$

where the last inequality requires ν to be large enough. Therefore,

$$R_+(y) \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}} 2^{-k_-}\rho(y)^{1-|\alpha|}. \tag{1.7.27}$$

Finally, summarizing (1.7.24), (1.7.25) and (1.7.27) we have shown that

$$|R(y)| \lesssim \|m\|_{\mathscr{M}^{\nu}}(w_0(y) + w_+(y) + w_-(y) + w_1(y)),$$

where

$$w_0(y) = 2^{-k_-|\alpha|}(2^{-k_-}\rho(y))^{-\nu}\mathbf{1}_{\rho(y)\geq 2^{k_-}},$$
$$w_\pm(y) = 2^{-k_\pm|\alpha|}\mathbf{1}_{\rho(y)\leq 2^{k_\pm+1}},$$
$$w_1(y) = 2^{-k_-}\rho(y)^{1-|\alpha|}\mathbf{1}_{2^{k_++1}\leq\rho(y)\leq 2^{k_-}},$$

and $\nu \geq |\alpha|+1$ is an integer multiple of $\overline{\alpha}$. Each of these functions is integrable with an $L^1(\mathbb{R}^n)$ norm not depending on k_-, k_+, radial and decreasing with respect to ρ in the sense of Lemma 1.7.2 and constant on $\{\rho(y) \leq 2^{k_+}\}$ or $\{\rho(y) \leq 2^{k_-}\}$. Thus, applying Lemma 1.7.2 to each of these functions yields (1.7.15).

1.8 Proofs of auxiliary estimates

In this section we prove (1.4.14), (1.4.15), (1.4.16) and (1.4.17).

Proof of (1.4.14). The proof is a straight-forward adaptation of the proof in the isotropic case (see [Ste93, Ch. VI.4, Prop. 2, p. 245]). Let φ be a smooth function supported in $1/2 \leq \rho(\xi) \leq 2$ such that

$$\sum_{j\in\mathbb{Z}} \varphi_j(\xi) = 1$$

for all $\xi \neq 0$, where $\varphi_j(\xi) = \varphi(\delta_{2^{-j}}(\xi)$. Let us set

$$K_j(x) = \int_{\mathbb{R}^n} e^{ix\xi} m(\xi)\varphi_j(\xi)d\xi.$$

Then it suffices to show that

$$\sum_{j\in\mathbb{Z}} |K_j(x)| \lesssim \|m\|_{\mathscr{M}^{|\alpha|+1}}\rho(x)^{-|\alpha|}.$$

We will show this by estimating $|K_j(x)|$ in two different ways. First, by the triangle inequality we have

$$|K_j(x)| \leq \|m\|_\infty 2^{j|\alpha|}. \tag{1.8.1}$$

To obtain the second estimate let us think of x as being fixed and let i be such that $\rho(x) = |x_i|^{1/\alpha_i}$. From integration by parts and the triangle inequality we see that

$$|K_j(x)| \leq |x_i|^{-\nu} \int_{\mathbb{R}^n} |\partial_{\xi_i}^\nu (m(\xi)\varphi_j(\xi))| d\xi \lesssim \|m\|_{\mathscr{M}^\nu} \rho(x)^{-\nu\alpha_i} 2^{j(|\alpha|-\nu\alpha_i)} \tag{1.8.2}$$

for every integer $\nu \geq 0$. From (1.8.1) and (1.8.2) we obtain

$$\sum_{j\in\mathbb{Z}} |K_j(x)| \lesssim \|m\|_{\mathscr{M}^\nu} \left(\sum_{2^j \leq \rho(x)^{-1}} 2^{j|\alpha|} + \rho(x)^{-\nu\alpha_i} \sum_{2^j > \rho(x)^{-1}} 2^{j(|\alpha|-\nu\alpha_i)} \right).$$

Choosing ν to be the smallest integer $> \frac{|\alpha|}{\alpha_i}$ (which is $\leq |\alpha|+1$) we obtain the desired conclusion. Here we remark that the lower bound required on ν could be improved by following the proof of the corresponding improved version of the Hörmander-Mikhlin theorem. □

Proof of (1.4.15). Expanding definitions and using Fourier inversion we see that, up to a universal constant, $|\psi_P^N(x)|$ is equal to

$$2^{k_P|\alpha|/2} \left| \int_{\mathbb{R}^n} e^{i\xi(x-c(I_P))} m(\xi - N) \widehat{\phi}(\delta_{2^{k_P}}(\xi - c(\omega_{P(0)}))) d\xi \right|.$$

Via a change of variables $\delta_{2^{k_P}}(\xi - c(\omega_{P(0)})) \to \zeta$ and using that $m(\xi) = m(\delta_{2^{k_P}}(\xi))$ this becomes

$$2^{-k_P|\alpha|/2} \left| \int_{\mathbb{R}^n} e^{i\zeta\delta_{2^{-k_P}}(x-c(I_P))} m(\zeta + \delta_{2^{k_P}}(c(\omega_{P(0)}) - N)) \widehat{\phi}(\zeta) d\zeta \right|. \tag{1.8.3}$$

Let us fix x and P and take i to be such that $\rho(x - c(I_P)) = |x_i - c(I_P)_i|^{1/\alpha_i}$. From repeated integration by parts we see that (1.8.3) is bounded by

$$2^{-k_P|\alpha|/2} (2^{-k_P} \rho(x - c(I_P)))^{-\nu'\alpha_i} \int_{\mathbb{R}^n} \left| \partial_{\zeta_i}^{\nu'} (m(\zeta + \delta_{2^{k_P}}(c(\omega_{P(0)}) - N)) \widehat{\phi}(\zeta)) \right| d\zeta,$$

for every integer $\nu' \geq 0$. We set $\nu' = \nu/\alpha_i$. Since $N \notin \omega_{P(0)}$ we have $|\zeta + \delta_{2^{k_P}}(c(\omega_{P(0)}) - N)| \gtrsim 1$. Therefore,

$$\int_{\mathbb{R}^n} \left| \partial_{\zeta_i}^{\nu'} (m(\zeta + \delta_{2^{k_P}}(c(\omega_{P(0)}) - N)) \widehat{\phi}(\zeta)) \right| d\zeta \lesssim \|m\|_{\mathscr{M}^{\nu'}} \leq \|m\|_{\mathscr{M}^\nu}.$$

This concludes the proof of (1.4.15) in the case that $\rho(x - c(I_P)) \geq 1$. In the case $\rho(x - c(I_P)) \leq 1$ we simply use the triangle inequality on (1.8.3). □

Proof of (1.4.16) *and* (1.4.17). If $c(I_P) = c(I_{P'})$, the estimates are trivial. Thus we may assume $c(I_P) \neq c(I_{P'})$. We have

$$|\langle \phi_P, \phi_{P'} \rangle| \leq |I_P|^{-\frac{1}{2}} |I_{P'}|^{-\frac{1}{2}} \int_{\mathbb{R}^n} |\phi(\delta_{2^{k_P}}(x - c(I_P))) \phi(\delta_{2^{k_{P'}}}(x - c(I_{P'})))| dx. \tag{1.8.4}$$

Since

$$\rho(c(I_P) - c(I_{P'})) \le \rho(x - c(I_P)) + \rho(x - c(I_{P'})),$$

at least one of $\rho(x - c(I_P)), \rho(x - c(I_{P'}))$ is $\ge \frac{1}{2}\rho(c(I_P) - c(I_{P'}))$. Thus, splitting the integral over x accordingly, using rapid decay of ϕ and the fact that $\|\phi_P\|_1 = |I_P|^{\frac{1}{2}}\|\phi\|_1$, the right hand side of (1.8.4) is no greater than a constant times

$$|I_P|^{-\frac{1}{2}}|I_{P'}|^{\frac{1}{2}}(1+2^{-k_P}\rho(c(I_P) - c(I_{P'})))^{-\nu} + |I_P|^{\frac{1}{2}}|I_{P'}|^{-\frac{1}{2}}(1+2^{-k_{P'}}\rho(c(I_P) - c(I_{P'})))^{-\nu}.$$

Recalling that we assumed $|I_P| \ge |I_{P'}|$ we see that the previous display is bounded by a constant times

$$|I_P|^{-\frac{1}{2}}|I_{P'}|^{\frac{1}{2}}(1 + 2^{-k_P}\rho(c(I_P) - c(I_{P'})))^{-\nu},$$

provided that $\nu \ge |\alpha| + 1$. This proves (1.4.16). The estimate (1.4.17) can be proven in the same way, by using the decay estimate (1.4.15). □

Chapter 2

Anisotropic time-frequency analysis: The bilinear Hilbert transform

2.1 Introduction

In this chapter we study an anisotropic variant of the bilinear Hilbert transform. We continue to use the notations $\alpha, \delta_\lambda(x), \rho(x), \mathrm{T}_y f, \mathrm{D}_\lambda^p f, \mathrm{M}_\eta f, \mathscr{M}^\nu$ which were introduced in Chapter 1. Let m be a bounded function that is sufficiently smooth outside of the origin and satisfies $m(\delta_\lambda(\xi)) = m(\xi)$ for all $\xi \neq 0, \lambda > 0$. By K we again denote the associated kernel, that is $\widehat{K} = m$. Let B be a real and diagonal $n \times n$ matrix. Consider the bilinear operator

$$(f_1, f_2) \longmapsto \int_{\mathbb{R}^n} f_1(x+y) f_2(x+By) K(y) dy.$$

The case $n = \alpha = 1$, $K(y) = p.v.\frac{1}{y}$ is the classical bilinear Hilbert transform studied by Lacey and Thiele [LT97a], [LT97b], [LT98], [LT99]. The diagonal constraint on B is due to the fact that a non-diagonal B would break the anisotropic dilation symmetry of our operator if $\alpha \neq (1, \ldots, 1)$. By duality, it is equivalent to study the trilinear form

$$\Lambda_m(f_1, f_2, f_3) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f_1(x+y) f_2(x+By) f_3(x) K(y) dx dy.$$

Our main theorem is the following.

Theorem 2.1.1. *Let $m \in \mathscr{M}^{\nu_0}$ for a sufficiently large integer ν_0 (depending only on α). Suppose that* $\det B(1-B) \neq 0$. *Let $2 < p_1, p_2, p_3 < \infty$ with $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1$. Then there exists a constant $C > 0$ such that*

$$|\Lambda_m(f_1, f_2, f_3)| \leq C \|f_1\|_{p_1} \|f_2\|_{p_2} \|f_3\|_{p_3}.$$

Remark 2.1.1. We will not make the number ν_0 explicit in the proof in order to simplify notation.

To prove this result we employ the framework of outer measure L^p spaces that was recently developed by Do and Thiele [DT15]. This framework can be seen as an alternative to the original formulation of the Lacey-Thiele approach. Indeed, it is also possible to formulate the proof of Carleson's theorem using the outer measure language (see the recent work by Uraltsev [Ura16], where the focus lies on variation norm bounds for the one-dimensional Carleson operator).

The outer measure language nicely modularizes the bulk of the proof into two separate parts. The first part is a reduction of the original operator to a certain model form that involves an integration over the operator's symmetry space and is more akin to the language of outer measure spaces. Do and Thiele [DT15] give a short and elegant argument for performing this reduction in the case of the classical bilinear Hilbert transform. This approach no longer works in our setting. Instead, we adapt the argument of Lacey and Thiele [LT97b]. This is the content of Section 2.2.

The second part involves the application of a certain generalized Carleson embedding theorem (see Section 2.3). The generalized Carleson embedding contains the main technical components of the proof. Our proof is based on the proof of the one-dimensional version given in [DT15]. The multidimensional setting causes certain technicalities in the tent selection procedure. The problem to overcome is basically that the one-dimensional proof uses the natural ordering on $\mathbb{R}$ when identifying upper and lower parts of tents. Roughly speaking, we address this issue by doing the selection componentwise. This seems to be quite arbitrary and other approaches are possible. See also Section 1.6 and the comments on ordering in the introduction and Section 5 of Pramanik and Terwilleger's work [PT03].

Note that in contrast to Demeter and Thiele [DT10], we study a fully non-degenerate case, whereas their analysis is focused on the more problematic degenerate cases.

Acknowledgments. The author thanks Christoph Thiele for his guidance and many useful discussions.

2.2 Reduction to a model form

We first introduce some notation. Let $0 < \epsilon_0 < 1$ be a small constant to be chosen later. A Schwartz function φ on $\mathbb{R}^n$ will be called *admissible* if $\widehat{\varphi}(\xi) = 0$ for $\rho(\xi) > \epsilon_0$. For $(y, \eta, t) \in \mathbb{R}^n \times \mathbb{R}^n \times (0, \infty)$ we denote

$$\varphi_{y,\eta,t}(x) = \mathrm{T}_y \mathrm{M}_\eta \mathrm{D}_t^2 \varphi(x) = e^{i\eta(x-y)} t^{-|\alpha|/2} \varphi(\delta_{t^{-1}}(x-y)).$$

For short we write $\varphi_{y,t} = \varphi_{y,0,t}$. Slightly abusing notation we also write

$$\varphi_{t,y,\eta}(x) = \mathrm{D}_t^2 \mathrm{T}_y \mathrm{M}_\eta \varphi(x) = e^{i\eta(\delta_{t^{-1}}(x)-y)} t^{-|\alpha|/2} \varphi(\delta_{t^{-1}}(x) - y).$$

We can use the relation $\varphi_{y,\eta,t} = \varphi_{t,\delta_{t^{-1}}(y),\delta_t(\eta)}$ to switch between the two whenever it seems convenient. For vectors $x_1, \ldots, x_m \in \mathbb{R}^n$ we write

$$\operatorname{diam}\{x_1, \ldots, x_m\} = \max_{i,j} |x_i - x_j|.$$

The reduction presented here resembles the approach of Lacey and Thiele in [LT97b], [LT99] to reduce the classical bilinear Hilbert transform. We need two lemmas.

Lemma 2.2.1. *Let $L \in \mathbb{R}^n$ with $|L| > 1$ sufficiently large. We have*

$$K(y) = \int_0^\infty K_t(y) \frac{dt}{t}$$

with $K_t(y) = \mathrm{D}_t^1 K_1(y) = t^{-|\alpha|} K_1(\delta_{t^{-1}}(y))$, where K_1 is a Schwartz function such that $\widehat{K_1}$ is supported on $\rho(y - L) \leq 1$.

Proof. Pick a Schwartz function a_0 on $\mathbb{R}$ with $\int_0^\infty a_0(t) \frac{dt}{t} = 1$. Let $a(\xi) = a_0(\rho(\xi))$ for $\xi \in \mathbb{R}^n$. Then

$$m(\xi) = \int_0^\infty m(\xi) a_0(t) \frac{dt}{t} = \int_0^\infty m(\delta_t(\xi)) a(\delta_t(\xi)) \frac{dt}{t}.$$

The claim follows with $K_1 = (m \cdot a)^\vee$ by applying the Fourier transform to both sides. □

Lemma 2.2.2 (Gabor reproducing formula)**.** *Let f be a Schwartz function in $\mathbb{R}^n$ and $\|\varphi\|_2 = 1$. Then for all $t > 0$ and $x \in \mathbb{R}^n$ we have*

$$f(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \langle f, \varphi_{t,y,\eta} \rangle \varphi_{t,y,\eta}(x) dy d\eta. \tag{2.2.1}$$

Proof. Note that, by a change of variables, the right hand side of (2.2.1) is invariant under replacing $\varphi_{t,y,\eta}$ by $\varphi_{y,\eta,t}$. From Fourier inversion, we have

$$\int_{\mathbb{R}^n} \langle f, \varphi_{y,\eta,t} \rangle \varphi_{y,\eta,t}(x) dy = c_n \int_{\mathbb{R}^n} \widehat{f}(\xi) |\widehat{\varphi_{\eta,t}}(\xi)|^2 e^{ix\xi} d\xi$$

for some absolute constant c_n. This implies the claim:

$$\begin{aligned} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \langle f, \varphi_{y,\eta,t} \rangle \varphi_{y,\eta,t}(x) dy d\eta &= c_n \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \widehat{f}(\xi) |\widehat{\varphi_{\eta,t}}(\xi)|^2 e^{ix\xi} d\xi d\eta \\ &= c_n \int_{\mathbb{R}^n} \widehat{f}(\xi) e^{ix\xi} \int_{\mathbb{R}^n} |\widehat{\varphi}(\eta)|^2 d\eta d\xi \\ &= \int_{\mathbb{R}^n} \widehat{f}(\xi) e^{ix\xi} d\xi = f(x). \end{aligned}$$

Here we have used that $c_n \|\widehat{\varphi}\|_2^2 = \|\varphi\|_2^2 = 1$. □

Now we begin by expanding each of the functions f_1, f_2, f_3 into a Gabor representation and splitting the kernel:

$$\begin{aligned}\Lambda_m(f_1,f_2,f_3) &= \int_0^\infty \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f_1(x+y)f_2(x+By)f_3(x)K_t(y)dxdy\frac{dt}{t}\\ &= \int_0^\infty \int\limits_{\mathbb{R}^{6n}} \langle f_1,\varphi_{t,u_1,\eta_1}\rangle\langle f_2,\varphi_{t,u_2,\eta_2}\rangle\langle f_3,\varphi_{t,u_3,\eta_3}\rangle\\ &\qquad \times\Big(\int\limits_{\mathbb{R}^{2n}} \varphi_{t,u_1,\eta_2}(x+y)\varphi_{t,u_2,\eta_2}(x+By)\varphi_{t,u_3,\eta_3}(x)K_t(y)d(x,y)\Big)d(u,\eta)\frac{dt}{t}\\ &= \int_0^\infty \int\limits_{\mathbb{R}^{6n}} t^{-|\alpha|/2}\langle f_1,\varphi_{t,u_1,\eta_1}\rangle\langle f_2,\varphi_{t,u_2,\eta_2}\rangle\langle f_3,\varphi_{t,u_3,\eta_3}\rangle G(u,\eta,t)d(u,\eta)\frac{dt}{t},\end{aligned} \tag{2.2.2}$$

where we have set

$$G(u,\eta,t) = t^{|\alpha|/2}\int_{\mathbb{R}^{2n}} \varphi_{t,u_1,\eta_2}(x+y)\varphi_{t,u_2,\eta_2}(x+By)\varphi_{t,u_3,\eta_3}(x)K_t(y)d(x,y).$$

and $u=(u_1,u_2,u_3), \eta=(\eta_1,\eta_2,\eta_3)$ with $u_i,\eta_i\in\mathbb{R}^n$. We proceed to estimate G in two different ways. By the triangle inequality we obtain that $|G(u,\eta,t)|$ is bounded by

$$t^{-2|\alpha|}\int_{\mathbb{R}^{2n}} |\varphi(\delta_{t^{-1}}(x+y)-u_1)\varphi(\delta_{t^{-1}}(x+By)-u_2)\varphi(\delta_{t^{-1}}(x)-u_3)K_1([y]_{t^{-1}})|dxdy \tag{2.2.3}$$

$$= \int_{\mathbb{R}^{2n}} |\varphi(x+y-u_1)\varphi(x+By-u_2)\varphi(x-u_3)K_1(y)|dxdy.$$

At least one of the parameters $x+y-u_1, x+By-u_2, x-u_3, y$ is larger than $C\mathrm{diam}\{u_1,u_2,u_3\}$ where $C>0$ is a constant only depending on B. This is because we can write u_i-u_j for $i,j=1,2,3$ as linear combinations of the parameters above. Namely,

$$u_1-u_2 = -(x+y-u_1)+(x+By-u_2)-(B-I)y$$

and therefore

$$|u_1-u_2| \le (3+\|B\|)\max\{|x+y-u_1|, |x+By-u_2, |x-u_3|, |y|\},$$

similarly for u_2-u_3 and u_1-u_3. Since φ, K_1 are Schwartz functions we get

$$|G(u,\eta,t)| \le C_m(1+\mathrm{diam}\{u_1,u_2,u_3\})^{-m} \tag{2.2.4}$$

for every $m\in\mathbb{N}$. On the other hand we can use Plancherel's theorem in $\mathbb{R}^{2n}$ to see that $G(u,\eta,t)$ equals (up to a universal multiplicative constant)

$$t^{|\alpha|/2}\int_{\mathbb{R}^{2n}} \mathcal{F}[\varphi_{t,u_1,\eta_1}(x+y)\varphi_{t,u_2,\eta_2}(x+By)](\xi,\tau)\mathcal{F}[\varphi_{t,u_3,\eta_3}(x)K_t(y)](\xi,\tau)d(\xi,\tau). \tag{2.2.5}$$

We have

$$\mathcal{F}[\varphi_{t,u_1,\eta_1}(x+y)\varphi_{t,u_2,\eta_2}(x+By)](\xi,\tau) = \int_{\mathbb{R}^{2n}} e^{-i(\xi\cdot x+\eta\cdot y)}\varphi_{t,u_1,\eta_1}(x+y)\varphi_{t,u_2,\eta_2}(x+By)d(x,y).$$

Substituting $v = x+y$ and $w = x+By$ and using Fubini we find the last display to be a constant multiple of

$$\widehat{\varphi_{t,u_1,\eta_1}}(-AB\xi + A\tau)\widehat{\varphi_{t,u_2,\eta_2}}(A\xi - A\tau+),$$

where $A = (1-B)^{-1}$. By Fubini's theorem we also see

$$\mathcal{F}[\varphi_{t,u_3,\eta_3}(x)K_t(y)](\xi,\tau) = \widehat{\varphi_{t,u_3,\eta_3}}(\xi)\widehat{K_t}(\tau).$$

Plugging this back into (2.2.5) gives

$$G(u,\eta,t) = Ct^{|\alpha|/2}\int_{\mathbb{R}^{2n}} \widehat{\varphi_{t,u_1,\eta_1}}(-AB\xi + A\tau)\widehat{\varphi_{t,u_2,\eta_2}}(A\xi - A\tau)\widehat{\varphi_{t,u_3,\eta_3}}(\xi)\widehat{K_t}(\tau)d(\xi,\tau).$$

The triangle inequality yields

$$\begin{aligned}|G(u,\eta,t)| &\leq Ct^{2|\alpha|}\int_{\mathbb{R}^{2n}} |\widehat{\varphi}(\delta_t(-AB\xi + A\tau)-\eta_1)\widehat{\varphi}(\delta_t(A\xi - A\tau)-\eta_2)\widehat{\varphi}(\delta_t(\xi)-\eta_3)\widehat{K_1}(\delta_t(\tau))|d\xi d\tau \\ &= C\int_{\mathbb{R}^{2n}} |\widehat{\varphi}(-AB\xi + A\tau-\eta_1)\widehat{\varphi}(A\xi - A\tau-\eta_2)\widehat{\varphi}(\xi-\eta_3)\widehat{K_1}(\tau)|d\xi d\tau\end{aligned}$$

Suppose that $G(u,\eta,t) \neq 0$. Then there exist $\tau,\xi \in \mathbb{R}^n$ such that the integrand is non-zero. Therefore $\rho(\tau - L) \leq 1$ and $\rho(\xi - \eta_3) \leq b$. We deduce,

$$\begin{aligned}\rho(\eta_3 - (B^{-1}L - A^{-1}B^{-1}\eta_1)) &\leq Cb, \\ \rho(\eta_3 - (L + A^{-1}\eta_2)) &\leq Cb,\end{aligned}$$

where C depends only on α, B. We change the variables in (2.2.2) introducing new variables $\eta, \lambda_1, \lambda_2, u, v_1, v_2$ as follows:

$$\begin{cases}\lambda_1 &= \eta_3 - B^{-1}L + A^{-1}B^{-1}\eta_1, \\ \lambda_2 &= \eta_3 - L - A^{-1}\eta_2, \\ \eta &= \eta_3,\end{cases} \qquad \begin{cases} v_1 &= u_3 - u_1, \\ v_2 &= u_3 - u_2, \\ u &= u_3.\end{cases}$$

Dropping the integration in λ_1, λ_2 over the compact set $K = \{\rho(\lambda_1) \leq Cb, \rho(\lambda_2) \leq Cb\}$, applying the triangle inequality and the estimate (2.2.3) leaves us with

$$|\Lambda_m(f_1,f_2,f_3)| \leq C_\nu \sup_{\lambda\in K}\int_{\mathbb{R}^{2n}} F(\lambda,v)(1+\max\{|v_1|,|v_2|\})^{-\nu}dv,$$

where $\lambda = (\lambda_1,\lambda_2), v = (v_1,v_2)$, ν a large integer that is to be determined later and

$$F(\lambda,v) = \int_0^\infty\int_{\mathbb{R}^{2n}} t^{-\frac{|\alpha|}{2}}|\langle f_1, \varphi_{t,u+v_1,-AB\eta+AL+AB\lambda_1}\rangle\langle f_2,\varphi_{t,u+v_2,A\eta-AL-A\lambda_2}\rangle\langle f_3,\varphi_{t,u,\eta}\rangle|d(u,\eta)\frac{dt}{t}.$$

Theorem 2.1.1 follows if there is some integer ν' such that

$$|F(\lambda, v)| \le C(1 + \max\{|v_1|, |v_2|\})^{\nu'} \|f_1\|_{p_1} \|f_2\|_{p_2} \|f_3\|_{p_3} \tag{2.2.6}$$

for all $\lambda \in K$, $v_1, v_2 \in \mathbb{R}^n$.

2.3 Boundedness of the model form

In this section we prove the following lemma, which implies estimate (2.2.6) and therefore the claim of Theorem 2.1.1. We make use of the L^p theory for outer measures introduced by Do and Thiele in [DT15]. This lemma should be seen as a generalized variant of Lemma 6.2 in [DT15].

Lemma 2.3.1. *Let $2 < p_1, p_2, p_3 < \infty$ such that $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1$, $\varphi^{(j)}$ admissible, M_j real, invertible, diagonal $n \times n$ matrices and $a_j \in \mathbb{R}^n$ pairwise different. There exist $C > 0$ such that*

$$\left| \int_0^\infty \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} t^{-|\alpha|/2} \prod_{j=1}^3 \langle f_j, \varphi^{(j)}_{u, M_j \eta + \delta_{t^{-1}}(a_j), t} \rangle du d\eta \frac{dt}{t} \right| \le C \|f_1\|_{p_1} \|f_2\|_{p_2} \|f_3\|_{p_3}. \tag{2.3.1}$$

The constant C takes the form

$$C = C_{n, M_j, a_j, p_j} \max_{j=1,2,3} \sup_{\beta \in \mathbb{N}_0^n, x \in \mathbb{R}^n} |D^\beta \varphi^{(j)}(x)| (1 + |x|)^\nu \tag{2.3.2}$$

for some sufficiently large integer ν.

For the sake of simpler notation we will not track the dependence of the constant on the decay of the bump functions $\varphi^{(j)}$ during the proof. It is clear from the proof that the constant takes the form (2.3.2).

Proof. The plan is to apply an outer Hölder inequality, Proposition 3.4 in [DT15], and then generalized Carleson embedding. Let $X = \mathbb{R}^n \times \mathbb{R}^n \times (0, \infty)$. We consider the outer measure induced by the premeasure $\sigma(\mathbf{T}(x, \xi, s)) = s^{|\alpha|}$ on the collection $\mathcal{T}$ of all tents

$$\mathbf{T} = \mathbf{T}(x, \xi, s) = \{(y, \eta, t) \in X : \rho(y - x) \le s - t, \rho(\eta - \xi) \le Lt^{-1}\}$$

with $(x, \xi, s) \in X$, where $L > 0$ is a large number to be chosen later. For all Borel measurable functions F on X and all $\mathbf{T} \in \mathcal{T}$ we define the size

$$S(F)(\mathbf{T}) = \frac{1}{\sigma(\mathbf{T})} \int_{\mathbf{T}} |F(y, \eta, t)| dy d\eta \frac{dt}{t}.$$

We set

$$F_j(y, \eta, t) = \langle f_j, \varphi^{(j)}_{y, M_j \eta + \delta_{t^{-1}}(a_j), t} \rangle$$

for $j = 1, 2, 3$, $(y, \eta, t) \in X$. By an application of Proposition 3.6 in [DT15] we see that the left hand side of (2.3.1) is bounded by $C\|t^{|\alpha|/2} F_1 F_2 F_3\|_{L^1(X, \sigma, S)}$. In order to apply the outer Hölder inequality we need to define three more sizes. First define

$$\mathbf{T}^{(j)} = \mathbf{T}^{(j)}(x, \xi, s) = \{(y, \eta, t) \in X : \rho(y - x) \le s - t, \rho(M_j^{-1}(\eta - \xi) - M_j^{-1} \delta_{t^{-1}}(a_j)) \le bt^{-1}\}$$

for $j=1,2,3$, where b is given by

$$b=\frac{\min_{i\neq j}\{\rho(a_i-a_j)\}}{4\max_i\{\rho(M_i)\}}>0.$$

Here $\rho(M)$ is defined as $\rho((\mu_1,\dots,\mu_n))$ for a diagonal matrix $M=\operatorname{diag}(\mu_1,\dots,\mu_n)$. The choice of b is such that the $\mathbf{T}^{(j)}$ are pairwise disjoint. Indeed, assume that $(y,\eta,t)\in\mathbf{T}^{(i)}\cap\mathbf{T}^{(j)}$ for $i\neq j$. By symmetry we may assume that $i=1,j=2$. Then we have

$$\begin{cases}\rho(M_1^{-1}(\eta-\xi)-M_1^{-1}\delta_{t^{-1}}(a_1)) & \leq bt^{-1},\\ \rho(M_2^{-1}(\eta-\xi)-M_2^{-1}\delta_{t^{-1}}(a_2)) & \leq bt^{-1}\end{cases}$$

Multiplication by M_1,M_2 in the respective inequalities gives

$$\rho((\eta-\xi)-\delta_{t^{-1}}(a_j))\leq\max\{\rho(M_1),\rho(M_2)\}bt^{-1}$$

for $j=1,2$ and thus $\rho(a_1-a_2)\leq 2\max\{\rho(M_1),\rho(M_2)\}b$, which is a contradiction. We define the sizes

$$S_j(F)(\mathbf{T})=\Big(\frac{1}{\sigma(\mathbf{T})}\int_{\mathbf{T}\setminus\mathbf{T}^{(j)}}|F(y,\eta,t)|^2dyd\eta\frac{dt}{t}\Big)^{\frac12}+\sup_{(y,\eta,t)\in\mathbf{T}}t^{-|\alpha|/2}|F(y,\eta,t)|.$$

It remains to show compatibility of the sizes in order to apply the outer Hölder inequality. Setting $F(y,\eta,t)=t^{-|\alpha|/2}F_1(y,\eta,t)F_2(y,\eta,t)F_3(y,\eta,t)$, we have

$$\begin{aligned}\sigma(\mathbf{T})S(F)(\mathbf{T})&=\int_{\mathbf{T}}|F(y,\eta,t)|dyd\eta\frac{dt}{t}\\&=\int_{\mathbf{T}\setminus(\mathbf{T}^{(1)}\cup\mathbf{T}^{(2)}\cup\mathbf{T}^{(3)})}|F(y,\eta,t)|dyd\eta\frac{dt}{t}+\sum_{j=1}^{3}\int_{\mathbf{T}\cap\mathbf{T}^{(j)}}|F(y,\eta,t)|dyd\eta\frac{dt}{t}\\&\leq\Big(\sup_{(y,\eta,t)\in\mathbf{T}}t^{-|\alpha|/2}|F_1(y,\eta,t)|\Big)\prod_{j=2}^{3}\Big(\int_{\mathbf{T}\setminus\mathbf{T}^{(j)}}|F_j(y,\eta,t)|^2dyd\eta\frac{dt}{t}\Big)^{\frac12}+\\&+\sum_{j=1}^{3}\sup_{(y,\eta,t)\in\mathbf{T}}t^{-|\alpha|/2}|F_j(y,\eta,t)|\prod_{j\neq k}\Big(\int_{\mathbf{T}\setminus\mathbf{T}^{(k)}}|F_k(y,\eta,t)|^2dyd\eta\frac{dt}{t}\Big)^{\frac12}.\end{aligned}$$

Here we have used the Cauchy-Schwarz inequality, as well as disjointness of the tents $\mathbf{T}^{(j)}$ in the second summand. As a consequence we obtain

$$S(F)(\mathbf{T})\leq 4\prod_{j=1}^{3}S_j(F_j)(\mathbf{T}).$$

Combined with the previous, the outer Hölder inequality implies that the left hand side of (2.3.1) is no greater than

$$C\prod_{j=1}^{3}\|F_j\|_{L^{p_j}(X,\sigma,S_j)}.$$

The generalized Carleson embedding theorem (Theorem 2.4.1) now gives

$$\|F_j\|_{L^{p_j}(X,\sigma,S_j)} \le C\|f_j\|_{p_j}.$$

This concludes the proof. □

2.4 A generalized Carleson embedding theorem

In this section we provide a more general variant of the generalized Carleson embedding theorem proved in [DT15] (see Theorem 5.1). To formulate the embedding theorem we first need to set up some definitions. Let $X = \mathbb{R}^n \times \mathbb{R}^n \times (0,\infty)$. Fix a real, invertible, diagonal $n \times n$ matrix $M = \operatorname{diag}\{M_1, \dots, M_n\}$ and a vector $a \in \mathbb{R}^n$ such that $\rho(M) \le 1$ and $\rho(a) \le h < 1$ with $0 < h < 1$ being a sufficiently small number to be chosen later. By $\mathcal{T}$ we denote the collection of tents

$$\mathbf{T} = \mathbf{T}(x,\xi,s) = \{(y,\eta,t) \in X \,:\, \rho(y-x) \le s-t, \rho(M(\eta-\xi) + \delta_{t^{-1}}(a)) \le t^{-1}\}.$$

A tent is the outer measure equivalent of a tree from the Lacey-Thiele method (see Definition 1.4.1). We equip X with the outer measure ω generated by the premeasure on tents given by $\sigma(\mathbf{T}(x,\xi,s)) = s^{|\alpha|}$. Additionally, we need the inner tents

$$\mathbf{T}^b = \mathbf{T}^b(x,\xi,s) = \{(y,\eta,t) \in X \,:\, \rho(y-x) \le s-t, \rho(\eta-\xi) \le bt^{-1}\},$$

where $0 < b < \epsilon$ is a parameter and $0 < \epsilon < 1$ is a sufficiently small constant to be chosen later. To achieve shorter formulas we will sometimes denote points in the outer measure space by $u = (x,\xi,s) \in X$ and we define the measure μ on X by $d\mu(u) = dxd\xi\frac{dt}{t}$. Accordingly, $\varphi_u = \varphi_{y,\eta,t}$. For any Borel measurable function f on X and $\mathbf{T} \in \mathcal{T}$ we define the sizes

$$S_2(F)(\mathbf{T}) = \Big(\frac{1}{\sigma(\mathbf{T})}\int_{\mathbf{T}\setminus\mathbf{T}^b} |F(v)|^2 d\mu(v)\Big)^{\frac{1}{2}}, S_\infty(F)(\mathbf{T}) = \sup_{v=(y,\eta,t)\in\mathbf{T}} t^{-|\alpha|/2}|F(v)|,$$

$$S(F)(\mathbf{T}) = S_2(F)(\mathbf{T}) + S_\infty(F)(\mathbf{T})$$

Theorem 2.4.1. *Let M, a, b be as above and φ an admissible Schwartz function. For $u \in X$ we define $F(u) = \langle f, \varphi_u\rangle$. Then for $2 < p \le \infty$ there exists $0 < C < \infty$ such that*

$$\|F\|_{L^p(X,\sigma,S)} \le C\|f\|_p \tag{2.4.1}$$

and

$$\|F\|_{L^{2,\infty}(X,\sigma,S)} \le C\|f\|_2. \tag{2.4.2}$$

The proof is based on the proof of the one-dimensional embedding theorem, found in [DT15]. Our multidimensional setting causes a number of minor technical complications. Before we turn to the proof let us mention a few auxiliary statements which we will need at different places.

Lemma 2.4.2. *Let $\varphi, \widetilde{\varphi}$ be Schwartz functions on $\mathbb{R}^n$. If $0 < t < s$ and $x, y \in \mathbb{R}^n$, then* *for all sufficiently large integers m we have*

$$|\langle \varphi_{x,t}, \widetilde{\varphi}_{y,s}\rangle| \leq Ct^{|\alpha|/2} s^{-|\alpha|/2}(1+|\delta_{s^{-1}}(x-y)|)^{-m}.$$

The proof for the case $n = \alpha = 1$ can be found in [Thi06, Lemma 2.1]. For the anisotropic version see the estimate (1.4.16) and its proof in Section 1.8 in Chapter 1. If the Fourier transform of a Schwartz function in one variable is supported away from the origin, then its primitive decays rapidly.

Lemma 2.4.3. *Let φ be a Schwartz function on $\mathbb{R}$ with $\operatorname{supp}\widehat{\varphi} \subset [-\delta,\delta]^c$ and $0 < \delta < 1$. Then for all $m \in \mathbb{N}_0$ we have*

$$\Big|\int_{-\infty}^{x} \varphi(y)dy\Big| \leq C_m \delta^{-m} \|\varphi\|_1 (1+|x|)^{-m}$$

where C_m is independent of φ, x, δ.

We turn to the proof of Theorem 2.4.1. The claim will follow by Marcinkiewicz interpolation between the endpoints $p = 2$ and $p = \infty$. See [DT15] for further details on Marcinkiewicz interpolation for L^p spaces of outer measures.

2.4.1 The endpoint $p = \infty$

To prove the bound (2.4.1) for $p = \infty$ we need to show

$$S_2(F)(\mathbf{T}(u)) + S_\infty(F)(\mathbf{T}(u)) \leq C\|f\|_\infty$$

for all $u \in X$. We handle the S_2 and S_∞ separately. First,

$$S_\infty(F)(\mathbf{T}(u)) \leq \sup_{(y,\eta,t)\in\mathbf{T}(u)} t^{-|\alpha|/2}\int_{\mathbb{R}^n} |f(x)\varphi_{y,\eta,t}(x)|dx \leq \|\varphi\|_1\|f\|_\infty.$$

To bound the L^2 size we first prove a lemma.

Lemma 2.4.4. *For all $u \in X$ and $f \in L^2(\mathbb{R}^n)$ we have*

$$\int_{\mathbf{T}(u)\setminus\mathbf{T}^b(u)} |F(v)|^2 d\mu(v) \leq C\|f\|_2^2. \tag{2.4.3}$$

Proof. Fix $u = (x,\xi,s)$. Then

$$\mathbf{T}(u)\setminus\mathbf{T}^b(u) \subset \{(y,\eta,t) \in X \,:\, bt^{-1} \leq \rho(\eta-\xi) \leq ct^{-1}\},$$

where $c = 2\rho(M^{-1}) > b$. Therefore the left hand side of (2.4.3) is bounded by

$$\int_0^\infty \int_{\mathbb{R}^n}\int_{bt^{-1}\leq\rho(\eta-\xi)\leq ct^{-1}} |F(y,\eta,t)|^2 d\eta dy \frac{dt}{t}.$$

By a substitution $\gamma = \delta_t(\eta-\xi)$ in the variable η and Fubini's theorem this becomes

$$\int_{b\leq\rho(\gamma)\leq c}\int_0^\infty\int_{\mathbb{R}^n} s|t^{-|\alpha|/2}F(y,\delta_{t^{-1}}(\gamma)+\xi,t)|^2 dy\frac{dt}{t}d\gamma.$$

Dropping the outer integral over a compact set we see that it suffices to show

$$\int_0^\infty \int_{\mathbb{R}^n} |\langle f, \psi_{y,\gamma,t}\rangle|^2 dy \frac{dt}{t} \leq C\|f\|_2^2, \tag{2.4.4}$$

for all $\gamma \in \mathbb{R}^n$ with $b \leq \rho(\gamma) \leq c$, where

$$\psi_{y,\gamma,t}(y') = e^{i(\xi+\delta_{t^{-1}}(\gamma))(y'-y)} t^{-|\alpha|} \varphi(\delta_{t^{-1}}(y'-y)) = \mathrm{T}_y \mathrm{M}_{\xi+\delta_{t^{-1}}(\gamma)} \mathrm{D}_t^1 \varphi(y').$$

Note that the Fourier transform of the function $\mathrm{M}_{-\xi}\psi_{y,\gamma,t}$ is supported away from zero by the support assumption on φ. It is tempting to invoke Calderón's reproducing formula at this point. However this is not possible, because the parameter γ causes translations on the Fourier side making it impossible to account for the normalization of the bump functions needed in Calderón's reproducing formula. Instead we must use an almost orthogonality argument and careful estimates of the bump functions. We claim that it suffices to show

$$\int_0^\infty \int_{\mathbb{R}^n} |\langle \psi_{y,\gamma,t}, \psi_{z,\gamma,r}\rangle| dz \frac{dr}{r} = \int_0^\infty \int_{\mathbb{R}^n} |\langle \psi_{y,\gamma,t}, \psi_{z,\gamma,r}\rangle| dz \frac{dr}{r} \leq C. \tag{2.4.5}$$

Indeed, let A denote the left hand side of (2.4.4). Then we have by the Cauchy-Schwarz inequality that

$$A^2 \leq \left\| \int_0^\infty \int_{\mathbb{R}^n} \langle f, \psi_{y,\gamma,t}\rangle \psi_{y,\gamma,t} dy \frac{dt}{t} \right\|_2^2 \|f\|_2^2$$

From expanding the L^2 norm we see that this is no greater than

$$\|f\|_2^2 \int_0^\infty \int_{\mathbb{R}^n} \int_0^\infty \int_{\mathbb{R}^n} |\langle f, \psi_{y,\gamma,t}\rangle \langle \psi_{y,\gamma,t}, \psi_{z,\gamma,r}\rangle \langle f, \psi_{z,\gamma,r}\rangle| dy \frac{dt}{t} dz \frac{dr}{r}.$$

By estimating the smaller one of $|\langle f, \psi_{y,\gamma,t}\rangle|$ and $|\langle f, \psi_{z,\gamma,r}\rangle|$ by the larger one and using symmetry we estimate this by

$$2\|f\|_2^2 \int_0^\infty \int_{\mathbb{R}^n} |\langle f, \psi_{y,\gamma,t}\rangle|^2 \Big(\int_0^\infty \int_{\mathbb{R}^n} |\langle \psi_{y,\gamma,t}, \psi_{z,\gamma,r}\rangle| dz \frac{dr}{r} \Big) dy \frac{dt}{t},$$

which by (2.4.5) is at most $2C\|f\|_2^2 A$. Dividing both sides by A we obtain (2.4.4). It remains to show (2.4.5). We first estimate the region $t \leq r$. Writing $\langle \psi_{y,\gamma,t}, \psi_{z,\gamma,r}\rangle = \langle \mathrm{M}_{-\xi}\psi_{y,\gamma,t}, \mathrm{M}_{-\xi}\psi_{z,\gamma,r}\rangle$, doing partial integration in every component, each time integrating $\mathrm{M}_{-\xi}\psi_{y,\gamma,t}$ while differentiating $\mathrm{M}_{-\xi}\psi_{z,\gamma,r}$ and using Lemma 2.4.3 we obtain

$$|\langle \psi_{y,\gamma,t}, \psi_{z,\gamma,r}\rangle| \leq C \int_{\mathbb{R}^n} r^{-2|\alpha|} (1 + |\delta_{t^{-1}}(y'-y)|)^{-(n+1)} (1 + |\delta_{r^{-1}}(y'-z)|)^{-(n+1)} dx$$

Therefore,

$$\begin{aligned}
&\int_t^\infty \int_{\mathbb{R}^n} |\langle \psi_{y,\gamma,t}, \psi_{z,\gamma,r}\rangle| dz \frac{dr}{r} \\
&\leq C \int_t^\infty \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} r^{-2|\alpha|}(1+|\delta_{t^{-1}}(y'-y)|)^{-(n+1)}(1+|\delta_{r^{-1}}(y'-z)|)^{-(n+1)} dy' dz \frac{dr}{r} \\
&\leq C \int_t^\infty \int_{\mathbb{R}^n} r^{-|\alpha|-1}(1+|\delta_{t^{-1}}(y'-y)|)^{-(n+1)} dy' dr \\
&\leq C \int_{\mathbb{R}^n} t^{-|\alpha|}(1+|\delta_{t^{-1}}(y'-y)|)^{-(n+1)} dy' \leq C
\end{aligned}$$

The region $t \geq r$ is estimated analogously. □

We return to finishing the proof for the endpoint $p = \infty$. Let $f \in L^\infty(\mathbb{R}^n)$. To make use of the lemma we apply a cut-off to f to obtain an L^2 function and estimate the remaining part by a rough tail estimate. Namely, let

$$D = \{y \in \mathbb{R}^n \,:\, \rho(x-y) \leq 2s\} \text{ and } f = f_1 + f_2,$$

where $f_1 = f\mathbf{1}_D$ and $f_2 = f - f_1$. Accordingly we let $F_1(v) = \langle f_1, \varphi_v\rangle$ and $F_2(v) = \langle f_2, \varphi_v\rangle$. We estimate $S_2(F_1)(\mathbf{T}(u))$ and $S_2(F_2)(\mathbf{T}(u))$ separately. By definition we have $f_1 \in L^2(\mathbb{R}^n)$ with $\|f_1\|_2 \leq s^{|\alpha|/2}\|f\|_\infty$. This and Lemma 2.4.4 give the required bound on $S_2(F_1)(\mathbf{T}(u))$. It remains to estimate the L^2 size on F_2. Let $(y, \eta, t) \in \mathbf{T}(u)$. Then we have $t < s$ and $\rho(y-x) \leq s - t < s$. Thus,

$$\begin{aligned}
|F(y,\eta,t)| &\leq \int_{\rho(w-x)\geq 2s} |f(w) t^{-|\alpha|/2}\varphi(\delta_{t^{-1}}(w-y))| dw \\
&\leq \|f\|_\infty t^{|\alpha|/2} \int_{\rho(w)\geq s} |t^{-|\alpha|}\varphi(\delta_{t^{-1}}(w))| dw \\
&\leq t^{|\alpha|/2}\|f\|_\infty \int_{\rho(w)\geq s/t} |\varphi(w)| dw \\
&\leq C t^{|\alpha|/2}\|f\|_\infty \int_{\rho(w)\geq s/t} \rho(w)^{-|\alpha|}(1+|w|)^{-(n+1)} dw \\
&\leq C t^{3|\alpha|/2} s^{-|\alpha|}\|f\|_\infty
\end{aligned}$$

Hence $S_2(F_2)(\mathbf{T}(u))$ is no greater than

$$C\|f\|_\infty \Big(\frac{1}{s^{|\alpha|}} \int_0^s \int_{\rho(\xi-\eta)\leq ct^{-1}} \int_{\rho(x-y)\leq s} t^{3|\alpha|} s^{-2|\alpha|} dy d\eta \frac{dt}{t}\Big)^{1/2} \leq C\|f\|_\infty.$$

2.4.2 The endpoint $p = 2$

Proving the weak L^2 bound involves selecting a countable collection of tents covering the "bad" points of X in an good way by maximizing certain parameters. For this procedure to make sense we require the space of tents to be discrete. That is, the choice of tops will be restricted to a discrete subset of X. This motivates the following. Let $0 < \epsilon_1, \epsilon_2 < 1$ be small constants to be determined later. Then X_Δ is defined to be the set of $(x,\xi,s) \in X$ such that there exist $k \in \mathbb{Z}, m, l \in \mathbb{Z}^n$ with

$$x = \delta_{\epsilon_1 2^k}(m), \xi = \delta_{\epsilon_2 2^{-k}}(l), s = 2^k.$$

For $(y,\eta,t),(x,\xi,s) \in X$ we say that (y,η,t) *is centrally contained in* $\mathbf{T}(x,\xi,s)$, if

$$\begin{aligned} \delta s < t \leq 2\delta s, \\ \rho(x-y) \leq \epsilon_1 s, \\ \rho(\xi-\eta) \leq \epsilon_2 s^{-1}. \end{aligned}$$

where $0 < \delta < 1$ is a constant to be determined later. By $\mathcal{T}_\Delta$ we denote the set of tents $\mathbf{T}(u)$ such that $u \in X_\Delta$. Let ω_Δ be the outer measure on X generated by the premeasure $\sigma_\Delta = \sigma|_{\mathcal{T}_\Delta}$. We claim that the original outer measure ω and its discretized version ω_Δ are equivalent. To show this we need to be able to cover tents by tents with tops in X_Δ.

Lemma 2.4.5. *Let $(x',\xi',s') \in X$. Then there exists $(x,\xi,s) \in X_\Delta$ such that (x',ξ',s') is centrally contained in $\mathbf{T}(x,\xi,s)$. Moreover, there exist 2^n points $(x,\xi^{(1)},s),\dots,(x,\xi^{(2^n)},s) \in X_\Delta$ such that*

$$\mathbf{T}(x',\xi',s') \subset \bigcup_{i=1}^{2^n} \mathbf{T}(x,\xi^{(i)},s), \text{ and} \tag{2.4.6}$$

$$\bigcap_{i=1}^{2^n} \mathbf{T}^b(x,\xi^{(i)},s) \cap \mathbf{T}(x',\xi',s') \subset \mathbf{T}^b(x',\xi',s'). \tag{2.4.7}$$

Proof. There exists exactly one $k \in \mathbb{Z}$ such that $2^k \in \left[\frac{s'}{2\delta}, \frac{s'}{\delta}\right)$. Choose $s = 2^k$. Then there exist $m, l \in \mathbb{Z}^n$ such that for $x = \delta_{\epsilon_1 s}(m)$ and $\xi = \delta_{\epsilon_2 s^{-1}}(l)$ we have $\rho(x'-x) \leq \epsilon_1 s$ and $\rho(\xi'-\xi) \leq \epsilon_2 s^{-1}$. In every component j we can decide whether to choose ξ_j to the left of ξ'_j or to the right of ξ'_j. For every j with $1 \leq j \leq n$ choose $\xi_{1,j} \in \mathbb{R}$ with $\xi_{1,j} \geq \xi'_j$ and $\xi_{-1,j} \in \mathbb{R}$ with $\xi_{-1,j} \leq \xi'_j$ such that $|\xi_{\pm 1,j} - \xi_j|^{1/\alpha_j} \leq \epsilon_2 s^{-1}$ and $(\xi_{1,j})_{j=1,\dots,n}, (\xi_{-1,j})_{j=1,\dots,n} \in X_\Delta$. Pick a bijection $\iota : \{1,\dots,2^n\} \to \{1,-1\}^n$, $i \mapsto (\iota_1(i),\dots,\iota_n(i))$. For every $i \in \{1,\dots,2^n\}$ set $\xi^{(i)} = (\xi_{\iota_j(i),j})_{j=1,\dots,n}$. By construction, $\xi^{(i)} \in X_\Delta$ and (x',ξ',s') is centrally contained in $\mathbf{T}(x,\xi^{(i)},s)$ for all i. It remains to show (2.4.6), (2.4.7). For this purpose we assume without loss of generality that $M_j \geq 0$ for all $1 \leq j \leq n$. Let $(y,\eta,t) \in \mathbf{T}(x',\xi',s')$ and $1 \leq j \leq n$. If $M_j(\eta_j - \xi'_j) + a_j t^{-\alpha_j} \geq 0$ then

$$\begin{aligned} -t^{-\alpha_j} &\leq -\left(\frac{s'}{2\delta}\right)^{-\alpha_j} \leq -(\epsilon_2 (s')^{-1})^{\alpha_j} \leq M_j(\xi'_j - \xi_{1,j}) \\ &\leq M_j(\xi'_j - \xi_{1,j}) + M_j(\eta_j - \xi'_j) + a_j t^{-\alpha_j} \\ &= M_j(\eta_j - \xi_{1,j}) + a_j t^{-\alpha_j} \\ &\leq M_j(\eta_j - \xi'_j) + a_j t^{-\alpha_j} \leq t^{-\alpha_j} \end{aligned}$$

Likewise, in the case $M_j(\eta_j - \xi'_j) + a_j t^{-\alpha_j} \leq 0$ we have

$$-t^{-\alpha_j} \leq M_j(\eta_j - \xi_{-1,j}) + a_j t^{-\alpha_j} \leq t^{-\alpha_j}.$$

Thus, by applying ι^{-1} to the vector of signs of $M_j(\eta_j - \xi'_j) + a_j t^{-\alpha_j}$ we obtain an i such that $(y, \eta, t) \in \mathbf{T}(x, \xi^{(i)}, s)$. This proves (2.4.6). Assume that $(y, \eta, t) \in \mathbf{T}(x', \xi', s')$ and $(y, \eta, t) \in \mathbf{T}^b(x, \xi^{(i)}, s)$ for all $1 \leq i \leq 2^n$. If $\eta_j \geq \xi'_j$ then

$$-bt^{-\alpha_j} \leq 0 \leq \eta_j - \xi'_j \leq \eta_j - \xi_{-1,j} \leq bt^{-\alpha_j},$$

whereas if $\eta_j \leq \xi'_j$ then

$$-bt^{-\alpha_j} \leq \eta_j - \xi_{1,j} \leq \eta_j - \xi'_j \leq 0 \leq bt^{-\alpha_j}.$$

This shows (2.4.7). □

The lemma implies $\omega \leq C\omega_\Delta$ and therefore the equivalence of ω and ω_Δ. We also define the discrete sizes $S_{2,\Delta}$, $S_{\infty,\Delta}$, S_Δ by restriction of the respective original sizes. As a consequence of the above we have that the norms $\|\cdot\|_{L^{2,\infty}(X,\sigma,S)}$ and $\|\cdot\|_{L^{2,\infty}(X,\sigma_\Delta,S_\Delta)}$ are equivalent. To prove the endpoint estimate (2.4.2) it therefore suffices to show its discrete variant

$$\|F\|_{L^{2,\infty}(X,\sigma_\Delta,S_\Delta)} \leq C\|f\|_2. \tag{2.4.8}$$

To demonstrate this it suffices to find for each $\lambda > 0$ a collection $\mathcal{E} \subset \mathcal{T}_\Delta$ of tents such that

$$\sum_{\mathbf{T}\in\mathcal{E}} \sigma(\mathbf{T}) \leq C\lambda^{-2}\|f\|_2^2 \quad \text{and} \tag{2.4.9}$$

$$S(F\mathbf{1}_{X\setminus E})(\mathbf{T}') \leq \lambda \tag{2.4.10}$$

for all $\mathbf{T}' \in \mathcal{T}_\Delta$, where $E = \bigcup_{\mathbf{T}\in\mathcal{E}} \mathbf{T}$. As before we will proceed by treating the L^∞ and L^2 portions of the size separately.

Lemma 2.4.6 (L^∞ size)**.** *For every $\lambda > 0$ there exists a collection $\mathcal{E}^0 \subset \mathcal{T}_\Delta$ of tents such that*

$$\sum_{\mathbf{T}\in\mathcal{E}^0} \sigma(\mathbf{T}) \leq C\lambda^{-2}\|f\|_2^2 \quad \textit{and} \tag{2.4.11}$$

and $t^{-|\alpha|/2}|F(v)| \leq \frac{\lambda}{2}$ for all $v = (y, \eta, t) \in X\setminus E^0$, where $E^0 = \bigcup_{\mathbf{T}\in\mathcal{E}^0} \mathbf{T}$.

In order to formulate the L^2 size lemma we need a few more definitions. The L^2 tent selection algorithm requires us to sort tents by frequency. We overcome the problem of the absence of a natural ordering on $\mathbb{R}^n$ by performing the selection in each component. This requires us to consider the following auxiliary tents. For $1 \leq j \leq n$ and $u = (x, \xi, s) \in X$ we set

$$\mathbf{T}^{b,j}(u) = \{(y, \eta, t) \in X \,:\, \rho(y - x) \leq s - t, |\eta_j - \xi_j| \leq (bt^{-1})^{\alpha_j}\}.$$

Then we have

$$\mathbf{T}^b(u) = \bigcap_{j=1}^{n} \mathbf{T}^{b,j}(u). \tag{2.4.12}$$

For a fixed frequency $\xi \in \mathbb{R}^n$ we also set

$$X_\xi^{+,j} = \{(y,\eta,t) \in X \,:\, \eta_j \geq \xi_j\}, X_\xi^{-,j} = X \backslash X_\xi^{+,j}$$

and we define the auxiliary sizes

$$S_2^{\pm,j}(F)(\mathbf{T}(u)) = \Big(\frac{1}{s^{|\alpha|}} \int_{X_\xi^{\pm,j} \cap \mathbf{T}(u) \cap (\mathbf{T}^{b,j}(u))^c} |F(v)|^2 d\mu(v)\Big)^{1/2},$$

$$S_2^j = S_2^{+,j} + S_2^{-,j},$$

for $u = (x,\xi,s) \in X_\Delta$.

Lemma 2.4.7 (L^2 size)**.** *Let $1 \leq j \leq n$. For every $\lambda > 0$ there exist $\mathcal{E}^{+,j}, \mathcal{E}^{-,j} \subset \mathcal{T}_\Delta$ such that*

$$\sum_{\mathbf{T} \in \mathcal{E}^{+,j} \cup \mathcal{E}^{-,j}} \sigma(\mathbf{T}) \leq C\lambda^{-2} \|f\|_2^2 \quad \textit{and} \tag{2.4.13}$$

$$S_2^j(F\mathbf{1}_{X \backslash (E^0 \cup E^{+,j} \cup E^{-,j})})(T') \leq \frac{\lambda}{2n} \tag{2.4.14}$$

for all $T' \in \mathcal{T}_\Delta$, where $E^{\pm,j} = \bigcup_{\mathbf{T} \in \mathcal{E}^{\pm,j}} \mathbf{T}$.

Before we turn to the proof of these claims, we show how they imply the desired conclusion. Fix $\lambda > 0$. Define

$$\mathcal{E} = \mathcal{E}^0 \cup \bigcup_{j=1}^n \mathcal{E}^{+,j} \cup \mathcal{E}^{-,j}.$$

Then (2.4.9) holds. It remains to show that (2.4.10) is true. Set $E = \bigcup_{\mathbf{T} \in \mathcal{E}} \mathbf{T}$. Then (2.4.12) and Lemmas 2.4.6 and 2.4.7 imply

$$S_2(F\mathbf{1}_{X\backslash E})(\mathbf{T}) \leq \sum_{j=1}^n S_2^j(F\mathbf{1}_{X\backslash E}) \leq \frac{\lambda}{2},$$

$$S(F\mathbf{1}_{X\backslash E})(\mathbf{T}) = S_\infty(F\mathbf{1}_{X\backslash E})(\mathbf{T}) + S_2(F\mathbf{1}_{X\backslash E})(\mathbf{T}) \leq \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda.$$

Now it only remains to prove Lemma 2.4.6 and 2.4.7. Both proofs consist of three main components. First an algorithm to select the proper tents is established. Then one appeals to an almost orthogonality argument using a disjointness property of the selected tents. Finally one has to make sure that all bad points are really covered.

Proof of Lemma 2.4.6. Fix $\lambda > 0$. By renormalizing we may assume $\|f\|_2 = 1$. We want to cover the set of bad points

$$\Omega = \{(y,\eta,t) \in X : t^{-|\alpha|/2} |F(y,\eta,t)| > \lambda\}$$

with tents. By the Cauchy-Schwarz inequality we have $|F(y,\eta,t)| \leq \|\varphi\|_2$. Thus, if

$(y, \eta, t) \in \Omega$ then we have an a priori bound on t given by

$$t \leq C\lambda^{-\frac{2}{|\alpha|}}.$$

We set up the following iterative procedure to generate an increasing chain of collections of tents $\mathcal{E}_0 \subset \mathcal{E}_1 \subset \mathcal{E}_2 \subset \cdots$. Initially define $\mathcal{E}_0 = \emptyset$. Assume that we have already constructed $\mathcal{E}_0, \ldots, \mathcal{E}_{k-1}$. If $\Omega \backslash \mathcal{E}_{k-1}$ is empty then the algorithm terminates. Otherwise, among the non-empty set

$$\{(v, u) \in X \times X_\Delta \,:\, v \in \Omega \backslash \mathcal{E}_{k-1},\, v \,\text{centrally contained in}\, \mathbf{T}(u)\}$$

we choose a pair (v_k, u_k), $v_k = (y_k, \eta_k, t_k), u_k = (x_k, \xi_k, s_k)$ such that s_k is maximal. This is possible by the a priori bound on t_k and because s_k is chosen from a discrete set. By N we denote the number of steps after which the algorithm terminates or $N = \infty$ if it never terminates. We claim that

$$\sum_{k=1}^{N} s_k^{|\alpha|} \leq C\lambda^{-2}.$$

For every $m \in \mathbb{N}_0$ let K_m be the set of k with $1 \leq k \leq N$ such that

$$2^m \lambda \leq t^{-|\alpha|/2} |F(y_k, \eta_k, t_k)| \leq 2^{m+1}\lambda.$$

Then we have

$$\sum_{k=1}^{N} s_k^{|\alpha|} \leq C \sum_{m=0}^{\infty} \sum_{k \in K_m} t_k^{|\alpha|} \leq C \sum_{m=0}^{\infty} 2^{-2m} \lambda^{-2} \sum_{k \in K_m} |F(y_k, \eta_k, t_k)|^2.$$

As a consequence it is enough to show

$$\sum_{k \in K_m} |F(y_k, \eta_k, t_k)|^2 \leq C$$

For convenience we will write $\varphi_k = \varphi_{y_k, \eta_k, t_k}$. Setting $A = \sum_{k \in K_m} |\langle f, \varphi_k \rangle|^2$ we have by Cauchy-Schwarz

$$\begin{aligned} A^2 &\leq \Big\| \sum_{k \in K_m} \langle f, \varphi_k \rangle \varphi_k \Big\|_2^2 \leq \sum_{k,l \in K_m} \langle f, \varphi_k \rangle \langle \varphi_k, \varphi_l \rangle \langle \varphi_l, f \rangle \\ &\leq 2 \sum_{k,l \in K_m, s_l \leq s_k} |\langle f, \varphi_k \rangle \langle \varphi_k, \varphi_l \rangle \langle \varphi_l, f \rangle| \end{aligned}$$

Since $k, l \in K_m$ they are at most by a factor 2 apart. So the previous display is no greater than

$$4 \sum_{k,l \in K_m, s_l \leq s_k} t_l^{|\alpha|/2} t_k^{-|\alpha|/2} |\langle f, \varphi_k \rangle^2 | \langle \varphi_k, \varphi_l \rangle|.$$

Estimating $|\langle \varphi_k, \varphi_l \rangle|$ by Lemma 2.4.2 we see that it is now enough to show that we have for all $k \in K_m$

$$\sum_{l \in K_m, s_l \leq s_k, \langle \varphi_k, \varphi_l \rangle \neq 0} t_l^{|\alpha|}(1 + |\delta_{t_k^{-1}}(y_k - y_l)|)^{-(n+1)} \leq C t_k^{|\alpha|} \tag{2.4.15}$$

To show this we need to exploit a disjointness property of the selected tents.

Claim. If $l', l \in K_m$ are such that $l' \leq l$, $s_{l'}, s_l \leq s_k$ and $\langle \varphi_k, \varphi_l \rangle \neq 0$, $\langle \varphi_k, \varphi_{l'} \rangle \neq 0$, then

$$\{\rho(\delta_{s_l^{-1}}(x - x_l)) \leq \epsilon\} \cap \{\rho(\delta_{s_{l'}^{-1}}(x - x_{l'})) \leq \epsilon\} = \emptyset.$$

Proof of the claim. Suppose not. Then we claim that $(y_{l'}, \eta_{l'}, t_{l'})$ is an element of the tent $\mathbf{T}(x_l, \xi_l, s_l)$. This would yield a contradiction since $l' \leq l$ means that $T_{l'}$ was chosen prior to $\mathbf{T}_l$. Indeed we have

$$\begin{aligned} \rho(y_{l'} - x_l) &\leq \rho(y_{l'} - x_{l'}) + \rho(x_{l'} - x_l) \\ &\leq \epsilon_1 s_{l'} + 2\epsilon s_l < s_l - t_{l'} \end{aligned}$$

if $\delta < \frac{1}{2}(1 - 2\epsilon - \epsilon_1)$ which we can certainly arrange. Also by looking at the support of $\widehat{\varphi}$ we see,

$$\begin{aligned} &\rho(M(\eta_{l'} - \xi_l) + \delta_{t_{l'}^{-1}}(a)) \\ &\leq \rho(\eta_{l'} - \eta_l) + \rho(\eta_l - \xi_l) + h t_{l'}^{-1} \\ &\leq (7\epsilon_0 + 2\epsilon_2 \delta + h) t_{l'}^{-1} \leq t_{l'}^{-1} \end{aligned}$$

provided $\epsilon_0, \epsilon, \delta, h$ are appropriately chosen, which is certainly possible. □

We shall now prove (2.4.15). By central containment the left hand side is no greater than

$$\begin{aligned} &C \sum_{l \in K_m, s_l \leq s_k, \langle \varphi_k, \varphi_l \rangle \neq 0} s_l^{|\alpha|}(1 + \delta_{s_k^{-1}}(x_k - x_l))^{-(n+1)} \\ &\leq C \sum_{l \in K_m, s_l \leq s_k, \langle \varphi_k, \varphi_l \rangle \neq 0} \int_{\{\rho(x - x_l) \leq s_l \epsilon\}} (1 + \delta_{s_k^{-1}}(x - x_k))^{-(n+1)} dx \end{aligned}$$

The above claim shows that this can be estimated by

$$C \int_{\mathbb{R}^n} (1 + \delta_{s_k^{-1}}(x - x_k))^{-(n+1)} \leq C t_k^{|\alpha|}$$

Set $\mathcal{E}^0 = \bigcup_{k=1}^N \mathcal{E}_k$. It remains to be shown that $E_0 = \bigcup_{\mathbf{T} \in \mathcal{E}^0} \mathbf{T}$ really covers Ω. This is clear if $N < \infty$. If on the other hand $N = \infty$, then s_k must converge to 0 as $k \to \infty$. Thus, for any $(y, \eta, t) \in X$ we must have $s_k < t$ for large enough k. If in addition (y, η, t) is in Ω then by maximality of the s_k it had to be chosen before the kth step, so $\Omega \subset E^0$ as required. □

Proof of Lemma 2.4.7. Fix $\lambda > 0$. We may assume without loss of generality that $\widehat{f}$ has compact support. The proof of this translates verbatim from [DT15, Section 5.2].

We also renormalize so that $\|f\|_2 = 1$. Fix $1 \le j \le n$. We will write $\pi(\xi) = \xi_j$ for $\xi \in \mathbb{R}^n$ to reduce indices. Also, we will omit the index j from the indices in $E^{\pm,j}, S^{\pm,j}, X^{\pm}_{\xi}$.

We construct a sequence of tents $(\mathbf{T}(u_k))_k$ with $u_k \in X_\Delta$. Suppose $u_1, \dots, u_{k-1}$ have been selected already. A point $u = (x, \xi, s) \in X_\Delta$ is called *bad* if

$$S_2^+(F\mathbf{1}_{X\setminus E^0 \cup E^+_{k-1}})(\mathbf{T}(u)) \ge \epsilon\lambda^2$$

If there is no bad point, then the algorithm terminates. Otherwise, we choose $u_k = (x_k, \xi_k, s_k)$ to be the bad points such that $\pi(\xi_k)$ and s_k are maximized in lexicographical order, where the frequency component takes precedence. This is possible because we have a priori bounds on s by Lemma 2.4.3 and on $\pi(\xi)$ by Lemma 2.4.3 and the compact support assumption.

Define $E^+_k = \bigcup_{i=1}^k \mathbf{T}(u_i)$, $\mathbf{T}_k = \mathbf{T}(u_k)$, $\mathbf{T}^{b,j}_k = \mathbf{T}^{b,j}(u_k)$, $X^+_k = X^+_{\xi_k}$ and

$$\mathbf{T}^*_k = \mathbf{T}_k \cap X^+_k \cap (\mathbf{T}^{b,j}_k \cup E^0 \cup E^+_{k-1})^c.$$

By N we denote the number of steps after which the algorithm terminates or we set $N = \infty$ in case the algorithm does not terminate.

We need to show that

$$\sum_{k=1}^N s_k^{|\alpha|} \le C\lambda^{-2} \tag{2.4.16}$$

By definition we have

$$\sum_{k=1}^N s_k^{|\alpha|} \le \epsilon^{-1}\lambda^{-2} \sum_{k=1}^N \int_{\mathbf{T}^*_k} |F(v)|^2 d\mu(v). \tag{2.4.17}$$

Therefore it is enough to show

$$\sum_{k=1}^N \int_{\mathbf{T}^*_k} |\langle f, \varphi_v\rangle|^2 d\mu(v) \le C \tag{2.4.18}$$

Denote the left hand side of the last display by A. Then by the Cauchy-Schwarz inequality we have

$$\begin{aligned}
A^2 &\le \Big\| \sum_{k=1}^N \int_{\mathbf{T}^*_k} \langle f, \varphi_v\rangle \varphi_v d\mu(v) \Big\|_2^2 \\
&= \sum_{k,l=1}^N \int_{\mathbf{T}^*_k} \int_{\mathbf{T}^*_l} \langle f, \varphi_v\rangle \langle \varphi_v, \varphi_{v'}\rangle \langle \varphi_{v'}, f\rangle d\mu(v') d\mu(v) \\
&\le \sum_{k,l=1}^N \int_{\mathbf{T}^*_k} \int_{\mathbf{T}^*_l, L^{-1}t \le t' \le Lt} \langle f, \varphi_v\rangle \langle \varphi_v, \varphi_{v'}\rangle \langle \varphi_{v'}, f\rangle d\mu(v') d\mu(v) \\
&\qquad + 2 \sum_{k,l=1}^N \int_{\mathbf{T}^*_k} \int_{\mathbf{T}^*_l, Lt' \le t} \langle f, \varphi_v\rangle \langle \varphi_v, \varphi_{v'}\rangle \langle \varphi_{v'}, f\rangle
\end{aligned}$$

where $L > 1$ is a large number to be determined later.

The diagonal part can be easily estimated to be bounded by CA . To bound the off-diagonal part we need to exploit the tent selection procedure to get a suitable disjointness property.

Claim. Set $\Omega_{y,\eta,t} = \{(z,\zeta,r) \in \mathbf{T}_l^* \text{ for some } l \text{ s.t. } Lr \le t, \langle \varphi_{y,\eta,t}, \varphi_{z,\zeta,r}\rangle \neq 0\}$. There exists a function $Q : \mathbb{R}^n \to (0,\infty)$ such that

$$\Omega_{y,\eta,t} \subset \{\rho(z - x_k) > s_k - t, r \in [Q(z), LQ(z))\}.$$

Proof of the claim. For the purpose of clarity we assume $M_j \ge 0$. Suppose that $(z,\zeta,r) \in \Omega_{y,\eta,t}$. Then

$$\rho(\eta - \zeta) \le \epsilon_0(L^{-1} + 1)r^{-1}$$

by support considerations of $\widehat{\varphi}$. Also, $(y,\eta,t) \in \mathbf{T}_k \cap X_k^+$ implies

$$\pi(M(\eta - \xi_k) + \delta_{t^{-1}}(a)) \le t^{-\alpha_j},$$

so that we have

$$\pi(\eta - \xi_k) \le M_j^{-1}(1 - a_j)t^{-\alpha_j} \le M_j^{-1}t^{-\alpha_j}.$$

Moreover, $(z,\zeta,r) \in X_l^+ \cap (\mathbf{T}_l^{b,j})^c$ implies

$$\pi(\zeta - \xi_l) > b^{\alpha_j} r^{-\alpha_j}.$$

Therefore we obtain

$$\begin{aligned}\pi(\xi_k - \xi_l) &= \pi(\xi_k - \eta) + \pi(\eta - \zeta) + \pi(\zeta - \xi_l)\\ &\ge -M_j^{-1}t^{-\alpha_j} - (\epsilon_0(L^{-1}+1))^{\alpha_j} r^{-\alpha_j} + b^{\alpha_j} r^{-\alpha_j}\\ &\ge (b^{\alpha_j} - M_j^{-1}L^{-\alpha_j} - (\epsilon_0(L^{-1}+1))^{\alpha_j})r^{-\alpha_j} \ge 0\end{aligned}$$

provided that L, ϵ_0 are chosen properly. From this we can conclude that $\mathbf{T}_k$ has been selected prior to $\mathbf{T}_l$. Now we prove $\rho(z - x_k) \ge s_k - t$. We have

$$\begin{aligned}&\rho(M(\zeta - \xi_k) + \delta_{r^{-1}}(a)) \le \rho(M(\zeta - \eta)) + \rho(M(\zeta - \eta) + \delta_{t^{-1}}(a)) + \rho(\delta_{r^{-1}}(a)) + \rho(\delta_{t^{-1}}(a))\\ &\le ((\epsilon_0 + h)(L^{-1} + 1) + L^{-1})r^{-1} \le r^{-1}\end{aligned}$$

provided that ϵ_0, h, L are chosen properly, which is possible. Since $\mathbf{T}_k$ was chosen before $\mathbf{T}_l$ we have $(z,\zeta,r) \notin \mathbf{T}_k$ we must have $|z - x_k| \ge s_k - r \ge s_k - t$.

To show the other half of the claim pick another point $(z',\zeta',r') \in \Omega_{y,\eta,t}$. If we assume $Lr' \le r$, then by an analogous argument as above we get that $\mathbf{T}_l$ was chosen prior to $T_{l'}$ which in particular implies $z \neq z'$. Therefore if two points in $\Omega_{y,\eta,t}$ have the same spatial component, then their scales can be at most by a factor of L apart. In other words, the set of all points in $\Omega_{y,\eta,t}$ with a given fixed spatial component has their scale contained in an interval of the form $[Q, L \cdot Q]$ for some $Q \in (0,\infty)$. □

The claim enables us to estimate the off-diagonal part by CA . We have shown (2.4.16). If $N < \infty$ then there exist no more bad points after the algorithm has terminated. Suppose that $N = \infty$. If $\pi(\xi_k) \to -\infty$, then we are done. Otherwise $\pi(\xi_k)$ converges to some value ν_1 in a discrete set. We restart the iteration continuing to choose bad points, which are not covered yet, giving a sequence of points with frequency components $\xi_{(1),k}$. We must have a strict inequality $\pi(\xi_{(1),k}) < \nu_1$. If this new sequence also fails to terminate or diverge towards $-\infty$ it must converge towards some $\nu_2 < \nu_1$. We rerun the argument again and iterate. In the worst case we end up with an infinite sequence $\nu_1 > \nu_2 > \nu_3 > \cdots$. Since the ν_i are restricted to be within a discrete set they must converge towards $-\infty$, which means that if we unite all the tents chosen and call this collection $\mathcal{E}^{+,j}$, we are done. One treats the size S_2^- similarly by reversing appropriate inequality signs in the above argument. □

Chapter 3

Polynomial Carleson operators along monomial curves in the plane

This chapter consists of a joint publication with Shaoming Guo, Lillian B. Pierce and Po-Lam Yung [GPRY16], which will appear in The Journal of Geometric Analysis. The copyright is held by Mathematica Josephina, Inc. and the article appears as part of this thesis in accordance with their copyright regulations.

3.1 Introduction

3.1.1 Historical background.

In 1966, Carleson [Car66] proved an L^2 bound for the Carleson operator

$$f(x) \longmapsto \sup_{N\in\mathbb{R}} \Big| p.v. \int_{\mathbb{R}} f(x-t) e^{iNt} \frac{dt}{t} \Big|. \tag{3.1.1}$$

This provided the key step in proving almost everywhere convergence of Fourier series of L^2 functions and thereby resolved a conjecture of Luzin. The L^p boundedness of the Carleson operator for $1 < p < \infty$ was then shown by Hunt [Hun68], and further proofs of Carleson's theorem were later given by Fefferman [Fef73] and Lacey and Thiele [LT00].

E. M. Stein suggested the following generalization: fix a natural number d and consider the operator given by

$$f(x) \longmapsto \sup_{P} \left| \int_{\mathbb{R}^n} f(x-y) e^{iP(y)} K(y) dy \right|, \tag{3.1.2}$$

where K is an appropriately chosen Calderón-Zygmund kernel and the supremum runs over all real-valued polynomials P of degree at most d in n variables. Stein asked whether this *polynomial Carleson operator* is bounded from L^p to L^p for $1 < p < \infty$. Stein and Wainger [SW01] used a TT^* argument and certain oscillatory integral estimates of van der Corput type to obtain L^p bounds for a variant of the operator (3.1.2), where the polynomial P is restricted to the set of polynomials of degree at most d that vanish to at least second order at the origin (so in particular, have no linear term; of course constant terms may be disregarded). In dimension $n = 1$, a positive answer to Stein's full question was provided by Lie [Lie09], [Lie11], who developed a sophisticated time-frequency approach. In higher dimensions $n > 1$, boundedness of the full polynomial Carleson operator remains an open problem.

Pierce and Yung [PY15] have introduced a new aspect to the study of polynomial Carleson operators, by considering an operator that also features Radon-type behavior in the sense of integration along an appropriate hypersurface. More precisely, they considered the operator

$$f(x,y) \longmapsto \sup_P \left| \int_{\mathbb{R}^n} f(x-t, y-|t|^2) e^{iP(t)} K(t) dt \right|, \tag{3.1.3}$$

acting on functions f on $\mathbb{R}^n \times \mathbb{R}$ where $n \geq 2$, K is a Calderón-Zygmund kernel, and the supremum runs over a suitable vector subspace of the space of all real-valued polynomials P of degree at most d in n variables. In particular, this allowable subspace requires that the polynomials considered should omit linear as well as certain types of quadratic terms. The key result of [PY15] then proves L^p, $1 < p < \infty$, bounds for this operator, via a method of proof based on square functions, TT^* techniques in the spirit of Stein and Wainger [SW01], and certain refined van der Corput estimates. Notably, the method of [PY15] does not work in the planar case $n = 1$, which is the main subject of the present paper. Our goal here is to prove bounds for a new class of polynomial Carleson operators along curves in the plane, and to demonstrate the curious feature that even *partial* results for these new operators along curves are in some sense as strong as Carleson's original theorem (and its variants) in the purely one-dimensional setting.

3.1.2 Statements of main results.

Let m, d be positive integers and f a Schwartz function on $\mathbb{R}^2$. For $N \in \mathbb{R}$ let

$$H_N f(x,y) = H_N^{m,d} f(x,y) = p.v. \int_{\mathbb{R}} f(x-t, y-t^m) e^{iNt^d} \frac{dt}{t}.$$

The natural goal, in the spirit of Carleson operators, is to prove that for all $1 < p < \infty$,

$$\left\| \sup_{N\in\mathbb{R}} |H_N f| \right\|_{L^p(dxdy)} \leq C \|f\|_{L^p(dxdy)}. \tag{3.1.4}$$

This would be analogous to the results of Stein and Wainger [SW01] in the Radon-type context of (3.1.2). We recall the useful strategy of linearization via a linearizing stopping time function: we define for an arbitrary measurable function $N(x,y) : \mathbb{R}^2 \mapsto \mathbb{R}$ the operator $f \mapsto H_{N(x,y)} f(x,y)$. Then proving

$$\|H_{N(x,y)} f(x,y)\|_{L^p(dxdy)} \leq C \|f\|_{L^p(dxdy)}$$

with a constant C independent of the choice of the function N is equivalent to proving (3.1.4).

To prove this inequality appears to be out of reach of our current methods. Recalling instead that a special case of [SW01] already shows that for any integer $d > 1$ the operator

$$f(x) \longmapsto p.v. \int_{\mathbb{R}} f(x-t) e^{iN(x)t^d} \frac{dt}{t} \tag{3.1.5}$$

is bounded on $L^p(\mathbb{R})$ for $1 < p < \infty$, we are motivated to consider the case when we twist the operator (3.1.5) with an additional Radon transform, while preserving the dependence of the linearizing function N on one variable only.

Thus for an arbitrary measurable function $N : \mathbb{R} \to \mathbb{R}$, we define our main operators of interest:

$$A_N^{m,d} f(x,y) := H_{N(x)}^{m,d} f(x,y) \tag{3.1.6}$$

and

$$B_N^{m,d} f(x,y) := H_{N(y)}^{m,d} f(x,y). \tag{3.1.7}$$

Before turning to our main results, we briefly note that certain special cases of these operators may be treated immediately: namely, for $d \geq 1$ the operators $A_N^{1,d}$ and $B_N^{1,d}$ are bounded on $L^p(\mathbb{R}^2)$ for $1 < p < \infty$. Indeed even the operator $\sup_{N \in \mathbb{R}} |H_N^{1,d} f(x,y)|$ is bounded on $L^p(\mathbb{R}^2)$ for $1 < p < \infty$. This follows immediately by integrating Carleson's theorem for (3.1.1) (in the case $d = 1$), or the result of Stein and Wainger [SW01] for (3.1.2) (in the case $d > 1$), along the straight lines of slope 1 in $\mathbb{R}^2$, using Fubini's theorem.

The remaining cases, with $m > 1$, are highly nontrivial. We formulate our main results as two theorems, which despite superficial similarities have quite different flavors, due to the differing symmetry groups of the involved operators (see Section 3.2.1). Our first main result can be stated as follows.

Theorem 3.1.1. *Let $N : \mathbb{R} \to \mathbb{R}$ be a measurable function and $d, m > 1, d \neq m$ integers. Then for $1 < p < \infty$,*

$$\left\| A_N^{m,d} f \right\|_p \leq C \|f\|_p, \tag{3.1.8}$$

$$\left\| B_N^{m,d} f \right\|_p \leq C \|f\|_p, \tag{3.1.9}$$

with the constant $0 < C < \infty$ depending only on d, m, p and not on N, f.

Note that uniformity of (3.1.8) in N is tantamount to the estimate

$$\left\| \sup_{N \in \mathbb{R}} \|H_N f(x,y)\|_{L^p(dy)} \right\|_{L^p(dx)} \leq C \|f\|_p. \tag{3.1.10}$$

Similarly, (3.1.9) corresponds to

$$\left\| \sup_{N \in \mathbb{R}} \|H_N f(x,y)\|_{L^p(dx)} \right\|_{L^p(dy)} \leq C \|f\|_p. \tag{3.1.11}$$

Our proof of Theorem 3.1.1 proceeds via van der Corput estimates, and does not depend on Carleson's theorem; this is in contrast to our second result, which we state as follows.

Theorem 3.1.2. *Let $N:\mathbb{R}\to\mathbb{R}$ be a measurable function. Then for $1<p<\infty$,*

$$\left\|A_N^{m,1}f\right\|_p \le C\|f\|_p, \quad \text{for any integer } m\ge 3, \tag{3.1.12}$$

$$\left\|B_N^{m,m}f\right\|_p \le C\|f\|_p, \quad \text{for any integer } m\ge 2, \tag{3.1.13}$$

with the constant C depending only on m,p and not on N,f.

A novel feature of our proof of Theorem 3.1.2 is that we combine Carleson's theorem with TT^* estimates in the spirit of Stein and Wainger. One surprising feature of our proof, compared to the original work [SW01] is that these TT^* estimates can handle certain cases of phase polynomials with a linear term (c.f. estimates (3.5.16)–(3.5.18)).

Remark 3.1.1. One is led to ask what happens to the remaining nontrivial ($m>1$) cases that are not covered by Theorems 3.1.1 and 3.1.2, namely $A_N^{2,1}$, $A_N^{m,m}$ and $B_N^{m,1}$ where $m>1$ is an integer. The key again lies in the symmetries of these operators: they are different from the symmetries of the operators in Theorems 3.1.1 and 3.1.2, and this points to why our current proofs do not apply in these situations. Despite these difficulties, at least the L^2 bounds for all these problematic cases still follow from known Carleson theorems via partial Fourier transform and Plancherel's theorem; see Section 3.6.2. The full L^p bounds remain an open problem in these cases.

3.1.3 Consequences of bounding partial Carleson operators

We now turn to the surprising feature that L^2 bounds for partial operators along curves imply L^2 bounds for Carleson-type operators acting on functions of one variable. Here we summarize several deductions of this kind; proofs are given in Section 3.6.

First, L^2 bounds for certain operators $A^{m,1}$ and $B^{m,m}$ are in some sense equivalent to an L^2 bound for Carleson's operator. More precisely, for any integer $m\ge 1$, the L^2 boundedness of $A_N^{m,1}$ implies the L^2 boundedness for the one-dimensional Carleson operator (3.1.1), by a Plancherel argument (see Section 6.1). In the other direction, we use the boundedness of the maximal truncated Carleson operator (3.3.2) (itself dominated by the Carleson operator according to the inequality (3.7.1)) to prove the L^2 boundedness of $A^{m,1}$ for $m\ge 3$ in Theorem 3.1.2, while the L^2 bound for $A^{2,1}$ may be deduced from Carleson's theorem (see Section 3.6.2).

Similarly, for any *odd* integer $m\ge 1$, the L^2 boundedness of $B_N^{m,m}$ implies an L^2 bound for the one-dimensional Carleson operator (see Section 3.6.1), while in the other direction we use the maximal truncated Carleson operator to prove Theorem 3.1.2.

Of course, the most natural challenge in the setting of Carleson operators along curves in the plane is the quadratic Carleson operator along the parabola defined by

$$\mathscr{C}^{\text{par}}f(x,y)=\sup_{N\in\mathbb{R}^2}\left|H_N^{\text{par}}f(x,y)\right|, \tag{3.1.14}$$

where for f a Schwartz function on $\mathbb{R}^2$,

$$H_N^{\text{par}}f(x,y)=p.v.\int_{\mathbb{R}} f(x-t,y-t^2)e^{iN_1t+iN_2t^2}\frac{dt}{t}.$$

This operator combines all the features that have proved troublesome in the study of (3.1.3) in [PY15]: apart from acting on functions in the plane, the phase consists entirely of the problematic linear and quadratic terms. Assuming that $N_1, N_2 : \mathbb{R} \to \mathbb{R}$ are arbitrary measurable functions depending only on x, observe that for $N_1 = 0$ this gives our operator $A^{2,2}$ (which our present arguments cannot treat) and for $N_2 = 0$ it gives our problematic operator $A^{2,1}$ (which again our present arguments cannot treat). So we are quite far from knowing how to bound (3.1.14).

But in the spirit of studying partial versions of Carleson operators, we point out that even a partial estimate for H_N^{par} of the form

$$\left\| \sup_{N \in \mathbb{R}^2} \|H_N^{\mathrm{par}} f\|_{L^2(dy)} \right\|_{L^2(dx)} \leq C\|f\|_2, \tag{3.1.15}$$

would suffice to imply an analogue over $\mathbb{R}$ of Lie's L^2 result on the quadratic Carleson operator [Lie09]; see Section 3.6.1 for details. These considerations indicate the interest in pursuing the partial Carleson operators we consider.

3.2 Overview of the methods

3.2.1 Symmetries of our operators

To make precise the differences between Theorems 3.1.1 and 3.1.2, we now characterize symmetries of the operators $A_N^{m,d}$ and $B_N^{m,d}$ as m and d vary. First there is an anisotropic dilation symmetry. If we denote

$$D_\lambda f(x, y) = f(\lambda x, \lambda^m y) \tag{3.2.1}$$

for $\lambda > 0$, then

$$D_\lambda^{-1} H_N^{m,d} D_\lambda = H_{\lambda^{-d} N}^{m,d}.$$

Second, due to the convolution structure, $H_N^{m,d}$ commutes with translations of the plane, for any m, d.

Third, the operators in Theorem 3.1.2 additionally have certain modulation symmetries. Let

$$M_{\xi,\zeta} f(x, y) = e^{ix\xi + iy\zeta} f(x, y) \tag{3.2.2}$$

for $\xi, \zeta \in \mathbb{R}$. Then if $d = 1$, we have

$$M_{\xi,0}^{-1} A_N^{m,1} M_{\xi,0} = A_{N-\xi}^{m,1} \tag{3.2.3}$$

for all $\xi \in \mathbb{R}$. Similarly if $d = m$, we have

$$M_{0,\zeta}^{-1} B_N^{m,m} M_{0,\zeta} = B_{N-\zeta}^{m,m} \tag{3.2.4}$$

for all $\zeta \in \mathbb{R}$. Simultaneous translation and modulation invariance is a characteristic property of the Carleson operator. Hence we are led to use Carleson's theorem in parts of the proof of Theorem 3.1.2.

Finally, we remark briefly that for $A_N^{2,1}$, the modulation symmetries are more involved. The problem is that in addition to the modulation symmetry (3.2.3), it also has a certain quadratic modulation symmetry. Let

$$Q_b f(x, y) = e^{ibx^2} f(x, y). \tag{3.2.5}$$

Then

$$Q_b^{-1} M_{0,b}^{-1} A_N^{2,1} M_{0,b} Q_b = A_{N-2bx}^{2,1}. \tag{3.2.6}$$

Recall that for the operator $A_N^{2,1}$, the linearizing function N depends on the variable x. Thus, by $N - 2bx$ we mean the function $x \mapsto N(x) - 2bx$, also only depending on x. Moreover, notice that in (3.2.6), the linear modulation acts on the y variable, while the quadratic modulation acts on the x variable. Hence there is a certain "twist" in this modulation symmetry.

3.2.2 Method of proof: Theorem 3.1.1

We now sketch the proof of Theorem 3.1.1. The strategy follows broadly that of Stein and Wainger, but the means of obtaining the key estimates is necessarily different. More precisely, we proceed by splitting the integral defining $A_N^{m,d}$ or $B_N^{m,d}$ into two parts, according to the size of the phase Nt^d: for Nt^d sufficiently small, we compare the resulting operator to a maximal truncated Hilbert transform along a curve, and for Nt^d large, we use TT^* and van der Corput estimates to handle the operator that arises. It is in the treatment of this latter operator where we must assume that the stopping time depends on one variable only, so that we may perform a Fourier transform in the free variable, along which the linearizing function is constant. This idea goes back to Coifman and El Kohen, who used it in the context of Hilbert transforms along vector fields (see the discussion in Bateman and Thiele [BT13]).

Another important ingredient is a certain refinement of Theorem 1 of [SW01]. The main novelty is that our core estimate, which we now record, allows us to consider phases with monomials of fractional exponents.

Lemma 3.2.1. *Fix real numbers $\alpha, \beta > 0$, $\alpha \neq \beta$, $\alpha, \beta \neq 1$. Let ψ be smooth and supported on $[1, 2]$. For $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$ and $t > 0$, let*

$$\Phi^\lambda(t) = e^{i\lambda_1 t^\alpha + i\lambda_2 t^\beta} \psi(t)/t, \tag{3.2.7}$$

and set $\Phi^\lambda(-t) = 0$. For $a > 0$, let

$$\Phi_a^\lambda(t) = a^{-1} \Phi^\lambda(t/a). \tag{3.2.8}$$

Let $|\lambda| = |\lambda_1| + |\lambda_2|$. Then there exists $\gamma_0 > 0$ such that for all $r \geq 1$ and all $F \in L^2(\mathbb{R})$,

$$\left\| \sup_{a>0, |\lambda| \geq r} |F * \Phi_a^\lambda| \right\|_{L^2(dx)} \lesssim r^{-\gamma_0} \|F\|_{L^2(dx)}.$$

Remark 3.2.1. For $\alpha, \beta \in \mathbb{N}$ this is merely a special case of Stein and Wainger's Theorem 1 in [SW01], but to prove Lemma 3.2.1 in full generality requires estimates of a very different flavor. See also the work of the first author [Guo16] for a similar result regarding a phase composed of a single fractional monomial. Fractional exponents appear naturally during the analysis of the operators $B^{m,d}$ via a change of variables $t^m \to t$ (for instance, see (3.4.6) and (3.5.14)). (In addition, Theorems 3.1.1 and 3.1.2 could be somewhat generalized to non-integral m, d, but we have chosen the integer setting for our main results, to avoid unnecessary complications.)

The key contrast of our proof of Lemma 3.2.1 with the corresponding result in Stein and Wainger [SW01] appears in the proof of Lemma 3.4.1. The strategy is to linearize the operator $F \mapsto \sup_{a>0,|\lambda|\geq r} |F * \Phi_a^\lambda|$ using stopping times for a, λ, and to bound an oscillatory integral by showing that for all but a small exceptional region of the integral, the phase has a large derivative of some order. Our proof enables us to make the exceptional region independent of the precise stopping time λ, thus obviating the need for the small set maximal functions that appear in [SW01]; at the cost of restricting our attention to phases with only two monomials, we are also able to handle fractional powers.

3.2.3 Method of proof: Theorem 3.1.2

Next, we sketch the proof of Theorem 3.1.2. To analyze $A_N^{m,1}$, where $m \geq 3$ is an integer, we first decompose the operator as

$$A_N^{m,1} = \sum_{k\in\mathbb{Z}} A_N^{m,1} \circ P_k,$$

where P_k is a Littlewood-Paley projection onto frequency $\sim 2^k$ in the y-variable. In view of the modulation invariance (3.2.3) in the x-variable, this is the only viable Littlewood-Paley decomposition we can use for the operator $A^{m,1}$; a Littlewood-Paley decomposition in the x-variable is doomed to fail. We also note that the Littlewood-Paley projection in the y-variable commutes with $A_N^{m,1}$, since the stopping time N in the operator $A_N^{m,1}$ depends only on x but not on y.

Now to analyze each Littlewood-Paley piece of $A_N^{m,1}$, we decompose the integral

$$A_N^{m,1} \circ P_k f(x,y) = \int_{\mathbb{R}} (P_k f)(x-t, y-t^m) e^{iN(x)t} \frac{dt}{t}$$

into two parts, where t is small or large compared to the frequency 2^k. For t small, we compare the resulting integral to a maximally truncated Carleson operator in the x-variable; this is natural in view of the remarks in Section 3.1.3. The error will be given by a strong maximal function, since $P_k f$ is localized in frequency in the y-variable. For t large, we need to use a van der Corput estimate: again we take advantage of the fact that the stopping time N of $A_N^{m,1}$ depends only on x, to take a partial Fourier transform in the y-variable.

In order to reassemble the various Littlewood-Paley pieces, the main ingredient is a vector-valued estimate for the maximally truncated Carleson operator (Theorem 3.3.1).

A similar strategy works for $B_N^{m,m}$ for $m \geq 2$ an integer. There is, however, an interesting distinction depending on whether m is odd or even: when m is odd, we need to use the maximally truncated Carleson operator in the y-variable, whereas when m is even, the component of the operator that would correspond to the maximally truncated Carleson operator magically vanishes. (See equation (3.5.5), and the discussion immediately thereafter.)

Roughly speaking, our proof of Theorem 3.1.2 works because the linearizing function depends on the same variable in which the modulation invariance occurs, so the other variable is at our disposal to use Plancherel's theorem and localize in frequency via Littlewood-Paley decomposition. Essential parts of this proof fail in the remaining cases $A^{2,1}$, $A^{m,m}$ and $B^{m,1}$, where $m > 1$. For $A^{m,m}$, the linearizing function varies with x, so we would like to use Plancherel's theorem in y and localize in the y frequency. However, the modulation invariance in (3.2.4) causes translation invariance in the y frequency so that any attempt at doing a Littlewood-Paley decomposition is doomed from the start. Similar behavior occurs for $B^{m,1}$.

3.3 Preliminaries

3.3.1 Notation

The notation $A \lesssim B$ always means $A \le C \cdot B$ with $0 < C < \infty$ depending only on m, d and the function ψ chosen below (and within the proof of Lemma 3.2.1, on α, β). Similarly, $A \approx B$ means $C_1 A \le B \le C_2 A$ with $0 < C_1 \le C_2 < \infty$ and the same dependence. We use the Fourier transform $\hat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-i\xi x}dx$ with inverse $\check{g}(x) = (2\pi)^{-1}\int_{\mathbb{R}} g(\xi)e^{ix\xi}d\xi$ and Plancherel identity $\|f\|_2 = (2\pi)^{-1/2}\|\hat{f}\|_2$.

3.3.2 Littlewood-Paley decomposition

Once and for all we fix a smooth function $\psi : \mathbb{R} \to \mathbb{R}$ supported on $\{t : 1/2 \le |t| \le 2\}$ such that $0 \le \psi(t) \le 1$ and $\sum_{k\in\mathbb{Z}} \psi_k(t) = 1$ for all $t \neq 0$, where $\psi_k(t) = \psi(2^{-k}t)$. Define the associated Littlewood-Paley projection of a function F on $\mathbb{R}$ by

$$F_k(w) = P_k F(w) = \int_{\mathbb{R}} F(u)\check{\psi}_k(w-u)du, \tag{3.3.1}$$

where $\check{\psi}_k$ denotes the inverse Fourier transform of the function ψ_k. The standard Littlewood-Paley estimates apply, in the form

$$\|F\|_p \lesssim \Big\| \Big(\sum_k |P_k F|^2\Big)^{1/2} \Big\|_p \lesssim \|F\|_p.$$

We will apply this in the x-variable or y-variable of $f(x, y)$, depending which is free.

3.3.3 Vector-valued inequalities

In this section we collect several vector-valued estimates that will play important roles in our work.

Define the maximally truncated Carleson operator by

$$\mathscr{C}^* F(x) = \sup_{N\in\mathbb{R}, \varepsilon>0} \Big| p.v. \int_{|t|\le\varepsilon} F(x-t)e^{iNt}\frac{dt}{t} \Big|. \tag{3.3.2}$$

Note that this operator is usually studied with the inequality $|t| \le \varepsilon$ being reversed; we may of course reduce to that case by subtracting the Carleson operator from $\mathscr{C}^*$.

Theorem 3.3.1. *For $1 < p < \infty$,*

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} |\mathscr{C}^* F_k|^2\Big)^{1/2} \Big\|_p \lesssim \Big\| \Big(\sum_{k\in\mathbb{Z}} |F_k|^2\Big)^{1/2} \Big\|_p, \tag{3.3.3}$$

with a constant depending only on p.

We assemble the necessary results to verify Theorem 3.3.1 in Section 3.7.1.

Next, let $\mathscr{M}$ be the maximal operator of Radon-type along the curve (t, t^m):

$$\mathscr{M} f(x,y) = \sup_{r>0} \frac{1}{2r} \int_{-r}^{r} |f(x-t, y-t^m)| dt. \tag{3.3.4}$$

This is known to be a bounded operator of L^p for $1 < p \leq \infty$ (e.g. by a small modification of the proof in the case of the parabola (t, t^2), [Ste93, Chapter XI §1.2, §2]). We require a vector-valued inequality for $f_k := P_k f$, with P_k acting on either the x-variable or y-variable (to be specified later):

Theorem 3.3.2. *For $1 < p < \infty$ we have*

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} |\mathscr{M} f_k|^2 \Big)^{1/2} \Big\|_p \lesssim \Big\| \Big(\sum_{k\in\mathbb{Z}} |f_k|^2 \Big)^{1/2} \Big\|_p, \tag{3.3.5}$$

with a constant depending only on p.

This result is stated in [RdFRT86, Theorem 2.5], as a consequence obtainable from a more general theory. For completeness, we offer a brief, self-contained proof for our special case in Section 3.7.2; we thank E. M. Stein for sharing with us this method of proof, which appears in a significantly more general form in the preprint [MST15, Appendix A, Theorem A.1].

Finally, we will need two one-variable Hardy-Littlewood maximal functions in the plane, denoted by M_1 and M_2, respectively. Indeed, they will act on the first and second variables respectively:

$$M_1 f(x,y) = \sup_{r>0} \frac{1}{2r} \int_{-r}^{r} |f(x-u, y)| du \tag{3.3.6}$$

$$M_2 f(x,y) = \sup_{r>0} \frac{1}{2r} \int_{-r}^{r} |f(x, y-t)| dt. \tag{3.3.7}$$

They are bounded on $L^p(\mathbb{R}^2)$ for all $1 < p < \infty$, and satisfy the following vector-valued inequality, which follows easily by integrating a corresponding result of Fefferman and Stein:

Theorem 3.3.3. *For $1 < p < \infty$ and $i = 1, 2$, we have*

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} |M_i f_k|^2 \Big)^{1/2} \Big\|_p \lesssim \Big\| \Big(\sum_{k\in\mathbb{Z}} |f_k|^2 \Big)^{1/2} \Big\|_p, \tag{3.3.8}$$

with a constant depending only on p.

See e.g. [Ste93, Chapter II §1.1] for further details.

3.4 The asymmetric case: Theorem 3.1.1

First we prove Theorem 3.1.1, assuming Lemma 3.2.1; then in Section 3.4.3 we prove the lemma.

For convenience, we define the auxiliary variable $z = z(x, y)$ to be understood as indicating either $z(x, y) = x$ or $z(x, y) = y$, so that $N(z)$ can mean either $N(x)$ or $N(y)$.

To simplify notations, we also define

$$Tf(x,y) = H^{m,d}_{N(z)} f(x,y), \tag{3.4.1}$$

with m, d satisfying the conditions of Theorem 3.1.1. With ψ_ℓ as defined in Section 3.3.2, define for each $\ell \in \mathbb{Z}$

$$T_\ell f(x,y) = \int_{\mathbb{R}} f(x-t, y-t^m) e^{iN(z)t^d} \psi_\ell(t) \frac{dt}{t}.$$

Let $n : \mathbb{R} \to \mathbb{Z}$ be such that for all $z \in \mathbb{R}$,

$$2^{-n(z)d} \leq |N(z)| < 2^{-(n(z)-1)d}. \tag{3.4.2}$$

Then we decompose $T = T^{(1)} + T^{(2)}$ with

$$T^{(1)} f(x,y) = \sum_{\ell \leq n(z)} T_\ell f(x,y)$$

and

$$T^{(2)} f(x,y) = \sum_{\ell > 0} T_{n(z)+\ell} f(x,y).$$

The motivation for this decomposition is that when $\ell \leq n(z)$, $\psi_\ell(t)$ localizes to $|t| \leq 2^{\ell+1} \leq 2^{n(z)+1}$ and the exponential factor $e^{iN(z)t^d}$ is well approximated by 1. Consequently we write $T^{(1)} f(x,y)$ as

$$\sum_{\ell \leq n(z)} \int_{\mathbb{R}} f(x-t, y-t^m)(e^{iN(z)t^d} - 1)\psi_\ell(t) \frac{dt}{t} + \sum_{\ell \leq n(z)} \int_{\mathbb{R}} f(x-t, y-t^m) \psi_\ell(t) \frac{dt}{t}. \tag{3.4.3}$$

We may estimate the absolute value of the first summand brutally by applying (3.4.2):

$$\lesssim \sum_{\ell \leq n(z)} \int_{\mathbb{R}} |f(x-t, y-t^m)| \cdot |N(z) t^{d-1} \psi_\ell(t)| dt \lesssim \frac{1}{2^{n(z)+2}} \int_{-2^{n(z)+1}}^{2^{n(z)+1}} |f(x-t, y-t^m)| dt.$$

The right hand side is bounded by $\mathscr{M} f(x,y)$, where $\mathscr{M}$ denotes the maximal operator along (t, t^m) defined in (3.3.4).

The second summand in (3.4.3) is bounded in absolute value by the maximal truncated Hilbert transform along the curve (t, t^m), defined by

$$\mathcal{H}^* f(x,y) = \sup_{\varepsilon, R > 0} \Big| \int_{\varepsilon < |t| < R} f(x-t, y-t^m) \frac{dt}{t} \Big|, \tag{3.4.4}$$

plus an error term bounded by $\mathscr{M} f(x,y)$ (which arises at the endpoint when passing from smooth bump functions to a sharp truncation). Thus in total we have obtained the pointwise estimate

$$|T^{(1)} f| \lesssim \mathscr{M} f + \mathcal{H}^* f.$$

Since both $\mathcal{H}^*$, $\mathscr{M}$ are known to be bounded in L^p, $1 < p < \infty$ (for example, by slight modifications of Stein and Wainger's work for (t, t^2) in [SW78]), we may conclude that

$$\|T^{(1)}f\|_p \lesssim \|f\|_p$$

for all $1 < p < \infty$.

It remains to show the same for $T^{(2)}$. Let

$$S_\ell f(x,y) = T_{n(z)+\ell} f(x,y);$$

we claim that it suffices to prove that there exists some $\gamma_0 > 0$ such that for all $\ell > 0$,

$$\|S_\ell f\|_2 \lesssim 2^{-\gamma_0 \ell}\|f\|_2. \tag{3.4.5}$$

Indeed, the triangle inequality implies the pointwise estimate $|S_\ell f| \lesssim \mathscr{M}f$, so that we immediately obtain $\|S_\ell f\|_p \lesssim \|f\|_p$ for all $1 < p < \infty$; by interpolation with (3.4.5) we then obtain for any $1 < p < \infty$ there exists some $\gamma_p > 0$ such that

$$\|S_\ell f\|_p \lesssim 2^{-\gamma_p \ell}\|f\|_p.$$

Finally, summing over $\ell \geq 0$ gives

$$\|T^{(2)}f\|_p \lesssim \|f\|_p.$$

All that remains is to prove (3.4.5); we proceed by distinguishing two cases.

3.4.1 The $B_N^{m,d}$ operators.

Here we consider the case $z(x,y) = y$. Applying Plancherel's theorem in the free x-variable, we obtain

$$\|S_\ell f(x,y)\|_{L^2(dx)} = (2\pi)^{-1/2}\Big\| \int_{\mathbb{R}} g_\xi(y - t^m) e^{iN(y)t^d - i\xi t} \psi_{n(y)+\ell}(t) \frac{dt}{t} \Big\|_{L^2(d\xi)},$$

where

$$g_\xi(y) = \int_{\mathbb{R}} e^{-i\xi x} f(x,y)dx.$$

Therefore to prove (3.4.5) it will suffice to prove a bound of the form

$$\Big\| \int_{\mathbb{R}} F(y - t^m) e^{iN(y)t^d - i\xi t} \psi_{n(y)+\ell}(t) \frac{dt}{t} \Big\|_{L^2(dy)} \lesssim 2^{-\gamma_0 \ell}\|F\|_2,$$

uniformly in $\xi \in \mathbb{R}$, for all single variable functions F. Recall that the cutoff function $\psi_{n(y)+\ell}$ has supports both in the positive half line and in the negative half line. Accordingly let us split the integration over t into a positive and a negative part. We consider the positive part; the negative component is treated in an entirely analogous way. Changing variables $t^m \mapsto t$, we see that it suffices to show there exists some $\gamma_0 > 0$ such that for all $\ell > 0$ and all $F \in L^2(\mathbb{R})$,

$$\Big\| \int_0^\infty F(y - t) e^{iN(y)t^{d/m} - i\xi t^{1/m}} \psi_{n(y)+\ell}(t^{1/m}) \frac{dt}{t} \Big\|_{L^2(dy)} \lesssim 2^{-\gamma_0 \ell}\|F\|_2, \tag{3.4.6}$$

uniformly in ξ. In fact (3.4.6) is an immediate consequence of the key Lemma 3.2.1, with $\alpha = d/m$, $\beta = 1/m$. To see this, we first rewrite $N(y) = 2^{-n(y)d+r(y)d}$ with $0 < r(y) < 1$ for all y. Then for $a \in \mathbb{R}$ and $\lambda \in \mathbb{R}^2$, we define $\Phi_a^\lambda := a^{-1}\Phi^\lambda(t/a)$, where

$$\Phi^\lambda(t) := e^{i\lambda_1 t^\alpha + i\lambda_2 t^\beta}\psi(t^{1/m})t^{-1} \quad \text{for } t > 0,$$

and $\Phi^\lambda(t) = 0$ for $t \leq 0$. One then observes that the integral on the left hand side of (3.4.6) is equal to $F * \Phi_a^\lambda(y)$, with parameters

$$a = 2^{(n(y)+\ell)m}, \qquad \lambda_1 = 2^{\ell d + r(y)d}, \qquad \lambda_2 = -\xi 2^{n(y)+\ell}.$$

Then (recalling $\ell > 0$, $0 < r(y) < 1$), we have

$$|\lambda| = |\lambda_1| + |\lambda_2| \geq 2^{\ell d + r(y)d} \geq 2^{\ell d},$$

and we see from Lemma 3.2.1 that for any fixed $\ell > 0$,

$$\Big\| \int_0^\infty F(y-t)e^{iN(y)t^{d/m} - i\xi t^{1/m}}\psi_{n(y)+\ell}(t^{1/m})\frac{dt}{t}\Big\|_{L^2(dy)} \lesssim \| \sup_{a>0, |\lambda| \geq 2^{\ell d}} |F * \Phi_a^\lambda| \|_2 \lesssim 2^{-\gamma_0 \ell}\|F\|_2,$$

as desired. This proves (3.4.6) and hence (3.4.5) in this case.

3.4.2 The $A_N^{m,d}$ operators.

Here we treat the case $z(x, y) = x$. Applying Plancherel's theorem in the free y-variable, we obtain

$$\Big\|S_\ell f(x, y)\Big\|_{L^2(dy)} = (2\pi)^{-1/2}\Big\| \int_{\mathbb{R}} g_\eta(x - t)e^{iN(x)t^d - i\eta t^m}\psi_{n(x)+\ell}(t)\frac{dt}{t}\Big\|_{L^2(d\eta)}$$

with

$$g_\eta(x) = \int_{\mathbb{R}} e^{-i\eta y} f(x, y)dy.$$

By Plancherel's theorem it suffices to show that there exists $\gamma_0 > 0$ such that for each $\ell > 0$,

$$\Big\| \int_{\mathbb{R}} F(x - t)e^{iN(x)t^d - i\eta t^m}\psi_{n(x)+\ell}(t)\frac{dt}{t}\Big\|_{L^2(dx)} \lesssim 2^{-\gamma_0 \ell}\|F\|_2$$

uniformly in η. Now it is clear that we may proceed similarly to (3.4.6), and deduce this bound from Lemma 3.2.1 with $\alpha = d$, $\beta = m$.

3.4.3 Proof of Lemma 3.2.1

In order to complete the proof of Theorem 3.1.1, it remains to prove Lemma 3.2.1. Due to a minor technical issue we will assume the pair $\{\alpha, \beta\} \neq \{2, 3\}$ in the proof. However, this case is of course already covered by Stein and Wainger's work [SW01, Theorem 1].

In fact it suffices to prove there exists γ_0 such that for all $r \geq 1$,

$$\| \sup_{a>0,\, r \leq |\lambda| \leq 2r} |F * \Phi_a^\lambda| \|_2 \lesssim r^{-\gamma_0}\|F\|_2. \tag{3.4.7}$$

With this result in hand, we immediately obtain the desired result,

$$\Big\| \sup_{a>0,\, |\lambda|\geq r} |F * \Phi_a^\lambda| \Big\|_2 \leq \sum_{k=0}^{\infty} \Big\| \sup_{a>0,\, 2^k r\leq|\lambda|\leq 2^{k+1}r} |F * \Phi_a^\lambda| \Big\|_2 \lesssim r^{-\gamma_0} \|F\|_2.$$

We proceed by linearizing the supremum. For measurable functions $a : \mathbb{R} \to (0, \infty)$, $\lambda : \mathbb{R} \to \mathbb{R}^2$ with $r \leq |\lambda(u)| \leq 2r$ for all $u \in \mathbb{R}$, we define an operator $\Lambda : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by

$$\Lambda F(u) = F * \Phi_{a(u)}^{\lambda(u)}(u) = \int_{\mathbb{R}} F(t)\Phi_{a(u)}^{\lambda(u)}(u-t)dt.$$

The bound (3.4.7) will follow from proving $\|\Lambda\|_{2\to 2} \lesssim r^{-\gamma_0}$ for some $\gamma_0 > 0$ with the implicit constant independent of a, λ. Since $\|\Lambda\|_{2\to 2} = \|\Lambda\Lambda^*\|_{2\to 2}^{1/2}$, we will in fact prove

$$\|\Lambda\Lambda^*\|_{2\to 2} \lesssim r^{-2\gamma_0}. \tag{3.4.8}$$

We calculate

$$\Lambda\Lambda^* F(u) = \int_{\mathbb{R}} F(s)(\Phi_{a_1}^\nu * \tilde{\Phi}_{a_2}^\mu)(u-s)ds, \tag{3.4.9}$$

with $\tilde{\Phi}(u) := \overline{\Phi(-u)}$ and $\nu = \lambda(u), \mu = \lambda(s), a_1 = a(u), a_2 = a(s)$. Note that by rescaling we may write

$$(\Phi_{a_1}^\nu * \tilde{\Phi}_{a_2}^\mu)(s) = a_2^{-1}(\Phi_{a_1/a_2}^\nu * \tilde{\Phi}_1^\mu)(a_2^{-1}s) = a_1^{-1}(\Phi_1^\nu * \tilde{\Phi}_{a_2/a_1}^\mu)(a_1^{-1}s).$$

Thus we will deduce (3.4.8) from applying the following bounds, which are the heart of the proof:

Lemma 3.4.1. *There exists $\gamma_1 > 0$ such that for $0 < h \leq 1$, $r \leq |\nu|, |\lambda| \leq 2r$ we have*

$$|(\Phi_h^\nu * \tilde{\Phi}_1^\mu)(s)| \lesssim r^{-\gamma_1}\mathbf{1}_{\{|s|\leq 4\}}(s) + \mathbf{1}_{\{|s|\leq r^{-\gamma_1}\}}(s), \tag{3.4.10}$$

$$|(\Phi_1^\nu * \tilde{\Phi}_h^\mu)(s)| \lesssim r^{-\gamma_1}\mathbf{1}_{\{|s|\leq 4\}}(s) + \mathbf{1}_{\{|s|\leq r^{-\gamma_1}\}}(s). \tag{3.4.11}$$

Remark 3.4.1. Note that the exceptional sets in (3.4.10), (3.4.11) do not depend on ν, μ. This is in contrast to [SW01, Lemma 4.1]. As a consequence we do not require Stein and Wainger's small set maximal function [SW01, Proposition 3.1].

We first proceed with the proof of Lemma 3.2.1, and then prove Lemma 3.4.1 in Section 3.4.4. Applying (3.4.10) and (3.4.11) appropriately (depending on whether $a_1 \geq a_2$ or $a_1 \leq a_2$), we deduce

$$|(\Phi_{a_1}^\nu * \tilde{\Phi}_{a_2}^\mu)(s)| \lesssim r^{-\gamma_1} \sum_{k=1}^{2} \left(a_k^{-1}\mathbf{1}_{\{|s|\leq 4a_k\}}(s) + (a_k r^{-\gamma_1})^{-1}\mathbf{1}_{\{|s|\leq r^{-\gamma_1}a_k\}}(s) \right).$$

Thus for any $G \in L^2$ we may compute

$$\begin{aligned} |\langle \Lambda\Lambda^* F, G\rangle| &= \Big| \int_{\mathbb{R}}\int_{\mathbb{R}} (\Phi_{a_1}^\nu * \tilde{\Phi}_{a_2}^\mu)(u-s)F(s)\overline{G(u)}dsdu \Big| \\ &\lesssim r^{-\gamma_1} \left(\int_{\mathbb{R}} MF(u)|G(u)|du + \int_{\mathbb{R}} |F(s)|MG(s)ds \right), \end{aligned}$$

where M denotes the standard one-variable Hardy-Littlewood maximal function. (Here the important point is that we have integrated first in whichever variable was independent of the stopping time a_k, for the two terms $k = 1, 2$.) Via the Cauchy-Schwarz inequality and the boundedness of M on L^2, we obtain

$$|\langle \Lambda\Lambda^* F, G\rangle| \lesssim r^{-\gamma_1} \|F\|_2 \|G\|_2.$$

This completes the proof of (3.4.7) with $\gamma_0 = \gamma_1/2$.

3.4.4 Proof of Lemma 3.4.1

We will only prove (3.4.10), as (3.4.11) follows by symmetry. By definition,

$$(\Phi_h^\nu * \tilde{\Phi}_1^\mu)(s) = \int_{\mathbb{R}} e^{i\nu_1 t^\alpha + i\nu_2 t^\beta - i\mu_1 (ht-s)^\alpha - i\mu_2 (ht-s)^\beta} \frac{\psi(t)}{t} \frac{\overline{\psi(ht-s)}}{ht-s} dt. \tag{3.4.12}$$

First notice that the support of $\Phi_h^\nu * \tilde{\Phi}_1^\mu(s)$ is contained in $\{s : |s| \leq 4\}$.

In order to apply van der Corput estimates, we need to analyze when the phase function

$$Q(t,s) = \nu_1 t^\alpha + \nu_2 t^\beta - \mu_1 (ht-s)^\alpha - \mu_2 (ht-s)^\beta$$

has a large derivative of some order. Here we recall that $\nu_i = \lambda_i(u)$ are fixed with respect to t, s (the relevant variables of integration in (3.4.9)), and that $r \leq |\nu_1| + |\nu_2| \leq 2r$. On the other hand, $\mu_i = \lambda_i(u-s)$ depends on s (in an unknown way), and thus our strategy is to make our argument independent of μ_1, μ_2.

Case 1: Suppose that $0 < h \leq h_0$, where $0 < h_0 < 1$ is to be determined later, depending on r, α, β; this is the easier case.

Let $0 < \varepsilon_1 < 1$ be small and fixed. Within the support of $\psi(t)\overline{\psi(ht-s)}$, we estimate

$$|\partial_t Q(t,s)| \geq |\alpha\nu_1 t^{\alpha-1} + \beta\nu_2 t^{\beta-1}| - hCr,$$

where C is a positive constant only depending on the exponents α, β. Let us define the function

$$F(t) = \alpha\nu_1 t^{\alpha-1} + \beta\nu_2 t^{\beta-1},$$

and its associated exceptional set

$$E = \{t \in [1/2, 2] : |F(t)| \leq \tau r^{1-\varepsilon_1}\},$$

where τ is a positive constant that depends only on α, β and is to be determined later. Our strategy will be to choose τ so that $|E|$ is small and then apply van der Corput's lemma outside of E.

We will prove (at the end of the considerations for Case 1):

Lemma 3.4.2. *There exists a choice of τ (depending only on α, β) such that*

$$|E| \lesssim r^{-\varepsilon_2} \tag{3.4.13}$$

with $\varepsilon_2 = \varepsilon_1/|\beta - \alpha|$.

Assuming τ is chosen as in the lemma, we now specify h_0 to be such that

$$h_0 C = \frac{1}{2}\tau r^{-\varepsilon_1}. \tag{3.4.14}$$

Then whenever $h \le h_0$, for all $t \in [1/2, 2] \setminus E$,

$$|\partial_t Q(t,s)| \gtrsim r^{1-\varepsilon_1}. \tag{3.4.15}$$

We now split the integral in (3.4.12) according to whether $t \in [1/2, 2] \setminus E$ or $t \in E$. We estimate the portion of the integral over E trivially by the measure of E, which is small $\lesssim r^{-\varepsilon_2}$ by Lemma 3.4.2.

We will estimate the portion of the integral over $[1/2, 2] \setminus E$ by applying van der Corput's lemma combined with the lower bound (3.4.15).

Here we encounter a delicate point: as stated in [Ste93, Chapter VIII §1.2] van der Corput's lemma for a first derivative assumes monotonicity. We circumvent this assumption as follows. We first note that E (and thus also $[1/2, 2]\setminus E$) is a finite union of intervals, with the number of intervals being bounded by a small absolute constant. To see this note that the equation

$$\alpha\nu_1 t^{\alpha-1} + \beta\nu_2 t^{\beta-1} \pm \tau r^{1-\varepsilon_1} = 0$$

has at most 3 solutions in $t > 0$ (see for example [SW70, Lemma 3]).

Thus we may apply the following slight variant of van der Corput's lemma (proved at the end of Case 1) to each such interval:

Lemma 3.4.3. *Suppose ϕ is real-valued and smooth in (a, b) and that both $|\phi'(x)| \ge \sigma_1$ and $|\phi''(x)| \le \sigma_2$ for all $t \in (a, b)$. Then*

$$\left|\int_a^b e^{i\lambda\phi(t)} dt\right| \le (a-b)\left(\frac{\sigma_2}{\sigma_1^2}\right)\lambda^{-1}.$$

Here we note that for s fixed, we have that $Q(t,s)$ is C^∞ with respect to t for all t in the support of $\psi(t)\overline{\psi(ht-s)}$; in particular note that both $t, ht-s$ are bounded away from the origin. We also verify trivially that for all such t,

$$|\partial_t^2 Q(t,s)| \lesssim r, \tag{3.4.16}$$

with a constant depending only on α, β. Hence applying Lemma 3.4.3 with the bounds (3.4.15) and (3.4.16) to each of the finitely many finite-length intervals in $[1/2, 2] \setminus E$, we obtain for each such portion of the integral a bound of size $\lesssim r(r^{1-\varepsilon_1})^{-2} = r^{-(1-2\varepsilon_1)}$. In total, combining this with our trivial estimate for the portion of the integral over E, we have proved

$$|(\Phi_h^\nu * \tilde{\Phi}_1^\mu)(s)| \lesssim r^{-(1-2\varepsilon_1)} + r^{-\varepsilon_2} \lesssim r^{-\varepsilon_3},$$

for all $|s| \le 4$, for a suitable $\varepsilon_3 > 0$, which suffices for (3.4.10) in this case. All that remains is to verify Lemmas 3.4.2 and 3.4.3.

Proof of Lemma 3.4.2. We observe that if one of $|\nu_1|, |\nu_2|$ dominates the other then $|F(t)|$ is large, that is $|F(t)| \gtrsim r$. More precisely, recall that $|\nu_1| + |\nu_2| \approx r$, and suppose that $|\nu_2|/|\nu_1| \leq c_0$ for some small constant c_0 (so in particular $|\nu_1| \gtrsim r$). Then

$$|F(t)| = |\nu_1| \left| \alpha t^{\alpha-1} + \beta \frac{\nu_2}{\nu_1} t^{\beta-1} \right|;$$

if c_0 is chosen sufficiently small (with respect to α, β) we may guarantee that for all $t \in [1/2, 2]$,

$$\alpha t^{\alpha-1} \geq 2 \left| \beta \frac{\nu_2}{\nu_1} t^{\beta-1} \right|$$

and hence

$$|F(t)| \geq |\nu_1| \frac{\alpha}{2} t^{\alpha-1} \geq c_1 r,$$

say. We may argue similarly to obtain $|F(t)| \geq c_1' r$ if $|\nu_1|/|\nu_2| \leq c_0'$ where c_0' depends only on α, β.

By choosing $\tau < \min\{c_0, c_0'\}$ (hence depending only on α, β) we then see that if E is to be non-empty, we must be in the regime where $(c_0')^{-1} \leq |\nu_1|/|\nu_2| \leq c_0$, that is, $|\nu_1| \approx |\nu_2|$. In this case, we will deduce that (3.4.13) holds. Suppose that $\alpha < \beta$; write $c := \nu_2/\nu_1$ so that $|c| \in [(c_0')^{-1}, c_0]$. Then

$$F(t) = \alpha \nu_1 t^{\alpha-1}(1 + c(\beta/\alpha) t^{\beta-\alpha}),$$

so that for all $t \in E$ we must have

$$r|1 + c(\beta/\alpha)t^{\beta-\alpha}| \lesssim |F(t)| \leq \tau r^{1-\varepsilon_1},$$

that is, t must satisfy

$$|1 + c(\beta/\alpha)t^{\beta-\alpha}| \lesssim r^{-\varepsilon_1}.$$

The measure of such t is $\lesssim r^{-\varepsilon_1/(\beta-\alpha)}$, with an implicit constant dependent on α, β. For the case $\alpha > \beta$ we argue in an entirely analogous way. This proves Lemma 3.4.2. □

Proof of Lemma 3.4.3. We recall the proof of the original van der Corput lemma in the case of a first derivative [Ste93, Ch VIII, Proposition 2], which bounds the integral in question by

$$\lambda^{-1} \int_a^b \left| \frac{d}{dt} \left(\frac{1}{\phi'} \right) \right| dt = \lambda^{-1} \int_a^b \left| \frac{\phi''(t)}{\phi'(t)^2} \right| dt,$$

where we have evaluated the derivative rather than invoking monotonicity of ϕ' to bring the absolute values outside the integral. The inequality claimed in Lemma 3.4.3 then clearly follows. □

We have now concluded the proof of Lemma 3.4.1 in Case 1.

Case 2. In the remaining case, $h_0 \leq h \leq 1$. Fix any small $0 < \varepsilon_4 < 1$; if $|s| \leq r^{-\varepsilon_4}$ we use the triangle inequality to bound (3.4.12) trivially by 1, which is sufficient for the second term in (3.4.10). Thus from now on we assume that

$$|s| \geq r^{-\varepsilon_4} \tag{3.4.17}$$

and work to obtain a small bound for the integral. Note that as a vector,

$$\begin{pmatrix} \partial_t Q(t,s) \\ \partial_t^2 Q(t,s) \\ \partial_t^3 Q(t,s) \\ \partial_t^4 Q(t,s) \end{pmatrix} = M_{t,s} \begin{pmatrix} \alpha\nu_1 t^{\alpha-1} \\ -\alpha\mu_1 (ht-s)^{\alpha-1} \\ \beta\nu_2 t^{\beta-1} \\ -\beta\mu_2(ht-s)^{\beta-1} \end{pmatrix}, \tag{3.4.18}$$

where $M_{t,s}$ is the 4×4 matrix

$$M_{t,s} = \begin{pmatrix} 1 & h & 1 & h \\ a_1 t^{-1} & a_1 h^2 (ht-s)^{-1} & b_1 t^{-1} & b_1 h^2 (ht-s)^{-1} \\ a_2 t^{-2} & a_2 h^3 (ht-s)^{-2} & b_2 t^{-2} & b_2 h^3 (ht-s)^{-2} \\ a_3 t^{-3} & a_3 h^4 (ht-s)^{-3} & b_3 t^{-3} & b_3 h^4 (ht-s)^{-3} \end{pmatrix}, \tag{3.4.19}$$

and

$$\begin{aligned} a_1 &= \alpha - 1, \quad a_2 = (\alpha-1)(\alpha-2), \quad a_3 = (\alpha-1)(\alpha-2)(\alpha-3), \\ b_1 &= \beta - 1, \quad b_2 = (\beta-1)(\beta-2), \quad b_3 = (\beta-1)(\beta-2)(\beta-3). \end{aligned}$$

If we can show that $|\det M_{t,s}|$ is sufficiently large, that is

$$|\det M_{t,s}| \gtrsim r^{-\kappa} \tag{3.4.20}$$

for some $\kappa > 0$, then we will apply the following lemma (whose proof we defer to the end of the section):

Lemma 3.4.4. *Let A be an invertible $n \times n$ matrix and $x \in \mathbb{R}^n$. Then*

$$|Ax| \geq |\det A| \|A\|^{1-n} |x|,$$

where $\|A\|$ denotes the matrix norm $\sup_{|x|=1} |Ax|$.

Note that $\|M_{t,s}\| \lesssim 1$ (since we only consider t in the support of $\psi(t)\overline{\psi(ht-s)}$, so that both $t, ht-s$ are bounded away from the origin). If we have shown (3.4.20) for t in a certain interval, then applying Lemma 3.4.4 to (3.4.18), we see that throughout that interval,

$$\left(\sum_{k=1}^{4} |\partial_t^k Q(t,s)|^2 \right)^{1/2} \gtrsim r^{-\kappa} |(\nu_1, \nu_2, \mu_1, \mu_2)^T| \gtrsim r^{1-\kappa}. \tag{3.4.21}$$

Then applying the van der Corput lemma to that portion of the integral (3.4.12) shows that portion is bounded by $r^{-(1-\kappa)/4}$. (Note: to be precise, if only the first order term $|\partial_t Q(t,s)|$ dominates in (3.4.21), then we must apply the variant Lemma 3.4.3 of the van der Corput lemma, using the trivial upper bound $|\partial_t^2 Q(t,s)| \lesssim r$, similar to our argument in Case 1. This will result in a bound for the portion of the integral over that interval of size $\lesssim r^{-(1-2\kappa)}$, which is sufficient.)

In fact, we will show that $|\det M_{t,s}|$ is sufficiently large in this manner for all but a small exceptional set E of t, with measure $\lesssim r^{-\kappa'}$ for some small $\kappa' > 0$. (As in our argument in Case 1, we will also note that this exceptional set is a union of a finite number of intervals, dependent only on α, β, so that we may apply the above argument to each individual component of $[1/2, 2] \setminus E$.) Thus this strategy is sufficient to complete the proof of (3.4.10).

We require the following purely algebraic identity.

Lemma 3.4.5. *Let $a_0, \ldots, a_3, b_0, \ldots, b_3, x, y$ be arbitrary real numbers. Then*

$$\begin{vmatrix} a_0 & a_0 & b_0 & b_0 \\ a_1 x & a_1 y & b_1 x & b_1 y \\ a_2 x^2 & a_2 y^2 & b_2 x^2 & b_2 y^2 \\ a_3 x^3 & a_3 y^3 & b_3 x^3 & b_3 y^3 \end{vmatrix} = (c(x^2+y^2) + dxy)(x-y)^2 xy,$$

where c, d are given by

$$c = -\begin{vmatrix} a_0 & b_0 \\ a_1 & b_1 \end{vmatrix} \cdot \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}, \qquad d = c + \begin{vmatrix} a_0 & b_0 \\ a_3 & b_3 \end{vmatrix} \cdot \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}. \tag{3.4.22}$$

Proof. Expand the determinant along the first row, combine the terms corresponding to the first and second columns, and those corresponding to the third and fourth columns, respectively. Then again expand each of the two resulting 3×3 determinants along the first row. □

Let $\tilde{M}$ denote the matrix in Lemma 3.4.5. Rescaling the individual rows and columns of $M_{t,s}$ appropriately to clear denominators, we see that

$$\det M_{t,s} = h^2 t^{-6} (ht - s)^{-6} \det \tilde{M}$$

where within $\tilde{M}$ we set $a_0 = b_0 = 1$, $x = ht - s$ and $y = ht$. Then we may apply Lemma 3.4.5 to compute

$$\det M_{t,s} = t^{-5} (ht - s)^{-5} s^2 h^3 S(t), \tag{3.4.23}$$

with

$$S(t) = h^2 (2c + d) t^2 - h(2c + d) st + cs^2,$$

where c, d are as in (3.4.22). Note that with $a_0 = b_0 = 1$ and the other a_i, b_i as specified above, then $c \neq 0$ is equivalent to $\alpha \neq \beta$, $\alpha, \beta \neq 1, 2$.

Now, in order to verify that $|\det M_{t,s}|$ is sufficiently large, as in (3.4.20), we distinguish between two cases.

Case 2A. Suppose first that $2c + d = 0$, so that $S(t) = cs^2$. We must then verify that $c \neq 0$. Since $2c + d = 0$, clearly if $c = 0$ then $d = 0$. But recall from above that $c = 0$ implies that either $\alpha = 2$ or $\beta = 2$ (since the hypotheses of Lemma 3.2.1 already ruled out $\alpha = \beta$, $\alpha = 1$ or $\beta = 1$). Recall also that we assume in this stage of the proof that the pair (α, β) is not $(2, 3)$.

Suppose that $\alpha = 2$. Then we would have $d = (\beta - 1)^2 (\beta - 2)^2 (\beta - 3)(\alpha - 1)$, which is clearly non-zero (since $\beta \neq 3$). Analogously we see that $\beta = 2$ leads to a contradiction. Thus we may conclude that $c \neq 0$, and recalling $|s| \geq r^{-\varepsilon_4}$ from (3.4.17) and $h \geq h_0 \gtrsim r^{-\varepsilon_1}$ from (3.4.14), we may compute immediately from (3.4.23) that

$$|\det M_{t,s}| \gtrsim r^{-4\varepsilon_4 - 3\varepsilon_1},$$

holds for all $t \in [1/2, 2]$. This verifies (3.4.20) and allows us to apply the van der Corput lemma to bound the full integral (3.4.12) by $r^{-\kappa}$ for some $\kappa > 0$, completing the proof of (3.4.10) in this case.

Case 2B. The final case we must consider is when $2c+d \neq 0$. Fix any small $0 < \varepsilon_5 < 1$ and define

$$E = \{t \in [1/2, 2] \,:\, |S(t)| \leq r^{-2\varepsilon_5 - 2\varepsilon_1}\}.$$

Note first that E is a union of at most two intervals, since S is a quadratic polynomial. Then for $t \in [1/2, 2] \setminus E$, (3.4.23) in combination with $|s| \geq r^{-\varepsilon_4}$, $h \gtrsim r^{-\varepsilon_1}$ implies

$$|\det M_{t,s}| \gtrsim r^{-2\varepsilon_4 - 5\varepsilon_1 - 2\varepsilon_5},$$

verifying (3.4.20) so that we may apply the van der Corput lemma to bound the portion of the integral over $[1/2, 2] \setminus E$ by $\lesssim r^{-\kappa}$ for some $\kappa > 0$, which suffices for the first term in (3.4.10).

We will bound the portion of the integral over E trivially, so all that remains is to verify that E has small measure, for which we call upon the following lemma (see Christ [Chr85, Lemma 3.3]):

Lemma 3.4.6. *Let $I \subset \mathbb{R}$ be an interval, $k \in \mathbb{N}$, $f \in C^k(I)$, and suppose that for some $\sigma > 0$, $|f^{(k)}(x)| \geq \sigma$ for all $x \in I$. Then there exists a constant $0 < C < \infty$ depending only on k such that for every $\rho > 0$,*

$$|\{x \in I \,:\, |f(x)| \leq \rho\}| \leq C\Big(\frac{\rho}{\sigma}\Big)^{1/k}. \tag{3.4.24}$$

By the choice of h_0 in (3.4.14) we have

$$|S''(t)| \gtrsim h^2 \gtrsim h_0^2 \gtrsim r^{-2\varepsilon_1}.$$

Thus by Lemma 3.4.6, we have $|E| \lesssim r^{-\varepsilon_5}$, which suffices for the second term in (3.4.10).

All that remains to complete the proof of Lemma 3.4.1, and hence of the main Lemma 3.2.1, is to verify Lemma 3.4.4.

Proof of Lemma 3.4.4. First we show $\|A^{-1}\| \leq \|A\|^{n-1}/|\det A|$. By homogeneity we can assume $\|A\| = 1$. Then all the eigenvalues of AA^* are between 0 and 1. Let λ be the smallest eigenvalue of AA^*. Then $\|A^{-1}\| = \lambda^{-1/2} \leq \det(AA^*)^{-1/2} = |\det A|^{-1}$. Therefore in general,

$$|x| = |A^{-1}Ax| \leq \|A^{-1}\| \cdot |Ax| \leq \|A\|^{n-1}|\det A|^{-1}|Ax|,$$

as desired. □

3.5 The symmetric case: Theorem 3.1.2

Here we prove Theorem 3.1.2. We present the proof in detail only for $B^{m,m}$; thus in the following we write $T = B^{m,m}$. The proof for $A^{m,1}$ is, mutatis mutandis, analogous, and we merely sketch the necessary changes in Section 3.5.3.

Recalling the function ψ_k fixed in Section 3.3.2, we define the Littlewood-Paley projection in the free x-variable by

$$f_k(x, y) = P_k f(x, y) = \int_{\mathbb{R}} f(u, y)\check{\psi}_k(x - u)du, \tag{3.5.1}$$

where $\check{\psi}_k$ denotes the inverse Fourier transform of the function ψ_k. In particular, note that $TP_k = P_kT$.

3.5.1 Single annulus estimate.

We fix $k_0 \in \mathbb{Z}$ and split the operator as $T = T^{(1)}_{k_0} + T^{(2)}_{k_0}$, where for any fixed k, $T^{(1)}_k$ is defined as

$$T^{(1)}_k f(x,y) := p.v. \int_{|t|\le 2^{-k}} f(x-t, y-t^m) e^{iN(y)t^m} \frac{dt}{t}$$

and $T^{(2)}_k := T - T^{(1)}_k$ accordingly.

Our key estimate for $T^{(1)}_{k_0}$ is the pointwise bound:

$$|T^{(1)}_{k_0} P_{k_0} f(x,y)| \lesssim \mathscr{C}^* P_{k_0} f(x,y) + M_1 M_2 P_{k_0} f(x,y), \tag{3.5.2}$$

where the maximally truncated one-variable Carleson operator $\mathscr{C}^*$ is defined as in (3.3.2); here our understanding is that $\mathscr{C}^*$ acts only on the second variable. Also, M_1 and M_2 refer to the Hardy-Littlewood maximal function in the first and second variables respectively, as defined in (3.3.6), (3.3.7).

We will prove (3.5.2) by using the fact that under the single annulus assumption

$$f = P_{k_0} f, \tag{3.5.3}$$

the function $f(x-t, y-t^m)$ is well approximated by $f(x, y-t^m)$. Precisely, we assume (3.5.3) and estimate

$$|T^{(1)}_{k_0} f(x,y)| \le \mathbf{I} + \mathbf{II},$$

where

$$\mathbf{I} = \int_{|t|\le 2^{-k_0}} |f(x-t, y-t^m) - f(x, y-t^m)| \frac{dt}{|t|}, \tag{3.5.4}$$

$$\mathbf{II} = \Big| p.v. \int_{|t|\le 2^{-k_0}} f(x, y-t^m) e^{iN(y)t^m} \frac{dt}{t} \Big|. \tag{3.5.5}$$

At this point there is a striking dichotomy in our treatment, depending on the parity of m: if m is even, the term $\mathbf{II}$ vanishes identically due to the integrand being an odd function. On the other hand, if m is odd, we can change variables $t^m \mapsto t$ (appropriately in the cases $t > 0, t < 0$) to see

$$\mathbf{II} \lesssim \sup_{\varepsilon>0} \Big| p.v. \int_{|t|\le\varepsilon} f(x, y-t) e^{iN(y)t} \frac{dt}{t} \Big| \le \mathscr{C}^* f(x,y), \tag{3.5.6}$$

where the maximally truncated Carleson operator acts only on the second variable. This contributes the first term in (3.5.2).

Next, we note that the first term $\mathbf{I}$ can be estimated by a maximal function due to the single annulus assumption (3.5.3). We write

$$f(x-t,y-t^m)-f(x,y-t^m)=\int_{\mathbb{R}} f(x-u,y-t^m)(\check{\psi}_{k_0}(u-t)-\check{\psi}_{k_0}(u))du. \quad (3.5.7)$$

By the rapid decay of the first derivative of $\check{\psi}$ we certainly have

$$|\frac{d}{d\xi}(\psi_{k_0})\check{}\,|=|\frac{d}{d\xi}(2^{k_0}\check{\psi}(2^{k_0}\xi))|\leq 2^{2k_0}(1+|2^{k_0}\xi|)^{-2}. \quad (3.5.8)$$

Now suppose that for some $j\geq 0$, u is in the annulus

$$2^{-k_0+j-1}\leq |u|\leq 2^{-k_0+j}, \quad (3.5.9)$$

so that for $|t|\leq 2^{-k_0}$ we have both $2^{-k_0+j-2}\leq |u|,|u-t|\leq 2^{-k_0+j+1}$. Thus applying the mean value theorem and the decay (3.5.8), for u in the annulus (3.5.9) we have

$$|\check{\psi}_{k_0}(u-t)-\check{\psi}_{k_0}(u)|\lesssim |t|\cdot 2^{2(k_0-j)},$$

where the implicit constant depends only on the choice of ψ.

Therefore (3.5.7) can be estimated in absolute value by

$$\lesssim |t|2^{2k_0}\sum_{j=0}^{\infty}2^{-2j}\int_{|u|\leq 2^{-k_0+j}}|f(x-u,y-t^m)|du. \quad (3.5.10)$$

This allows us to bound the term **I** by

$$\lesssim \sum_{j=0}^{\infty}2^{-j}\frac{1}{2^{-k_0}2^{-k_0+j}}\int_{|u|\leq 2^{-k_0+j}}\int_{|t|\leq 2^{-k_0}}|f(x-u,y-t^m)|dtdu. \quad (3.5.11)$$

We may dominate this by the maximal functions M_1 and M_2 as follows. Indeed, we focus temporarily on the inner integration in t in (3.5.11):

$$\frac{1}{2^{-k_0}}\int_{|t|\leq 2^{-k_0}}|f(x-u,y-t^m)|dt\leq \quad \frac{C}{2^{-k_0}}\int_{|s|\leq 2^{-k_0m}}|f(x-u,y-s)||s|^{\frac{1}{m}-1}ds$$

Since $|s|^{\frac{1}{m}-1}\mathbf{1}_{|s|\leq 2^{-k_0m}}$ is radially decreasing and integrable in s, with integral equal to $C2^{-k_0}$, by [Ste70, Chapter III Theorem 2], we can bound the above by $\lesssim M_2f(x-u,y)$. Thus (3.5.11) is bounded by $M_1M_2f(x,y)$. This completes the proof of the inequality (3.5.2) for $T^{(1)}_{k_0}$.

We now turn to estimating $T^{(2)}_{k_0}f$, still under the single annulus assumption (3.5.3). Let us define for any integer ℓ

$$T_\ell f(x,y)=\int_{\mathbb{R}} f(x-t,y-t^m)e^{iN(y)t^m}\psi_\ell(t)\frac{dt}{t}.$$

Then certainly

$$|T^{(2)}_{k_0}f|\lesssim \mathscr{M}f+\sum_{\ell=0}^{\infty}|T_{-k_0+\ell}f|; \quad (3.5.12)$$

here we need merely observe that the maximal operator $\mathscr{M}$ along (t, t^m) (defined in (3.3.4)) arises in (3.5.12) due to the transition to smooth cutoffs.

Next, we claim that (still under the assumption (3.5.3)) there exists a constant $\gamma > 0$ such that for all $\ell \geq 0$,

$$\|T_{-k_0+\ell} f\|_2 \lesssim 2^{-\gamma\ell} \|f\|_2. \tag{3.5.13}$$

To prove (3.5.13) we proceed similarly to Section 3.4.1. First we apply Plancherel's theorem in the free x-variable, so that it is equivalent to prove that for $g_\xi(y) = \int_{\mathbb{R}} e^{-i\xi x} f(x, y) dx$,

$$\Big\| \int_{\mathbb{R}} g_\xi(y - t^m) e^{iN(y)t^m - i\xi t} \psi_{-k_0+\ell}(t) \frac{dt}{t} \Big\|_{L^2(d\xi, dy)} \lesssim 2^{-\gamma\ell} \|g_\xi\|_{L^2(d\xi, dy)}.$$

In particular, we note that due to the assumption (3.5.3), g_ξ is nonzero only in the frequency annulus $2^{k_0-1} \leq |\xi| \leq 2^{k_0+1}$.

We then split the integral into a positive and a negative part, which are dealt with analogously. We focus here on the positive portion of the integral; by a change of variables $t^m \mapsto t$ the claim (3.5.13) is reduced to showing

$$\Big\| \int_0^\infty F(y - t) e^{iN(y)t - i\xi t^{1/m}} \psi_{-k_0+\ell}(t^{1/m}) \frac{dt}{t} \Big\|_{L^2(dy)} \lesssim 2^{-\gamma\ell} \|F\|_{L^2(dy)} \tag{3.5.14}$$

for all single variable functions F, uniformly in $2^{k_0-1} \leq |\xi| \leq 2^{k_0+1}$.

As in the proof of Lemma 3.2.1 we proceed by the TT^* method. For convenience we write

$$\tilde{\psi}_k(t) = \psi_k(t^{1/m}) \mathbf{1}_{(0,\infty)}(t),$$

and denote the operator on the left hand side of (3.5.14) by $\tilde{T}$. Then $\|\tilde{T}\|_{2\to 2} = \|\tilde{T}\tilde{T}^*\|_{2\to 2}^{1/2}$, where

$$\tilde{T}\tilde{T}^* F(y) = \int_{\mathbb{R}} F(y - s) K_{N(y), N(y-s)}(s) ds \tag{3.5.15}$$

and for any $\lambda_1, \lambda_2 \in \mathbb{R}$ the kernel K_{λ_1, λ_2} is given by

$$K_{\lambda_1,\lambda_2}(s) = \int_{\mathbb{R}} e^{i\lambda_1 t - i\lambda_2(t-s) - i\xi(t^{1/m} - (t-s)^{1/m})} \frac{\tilde{\psi}_{-k_0+\ell}(t-s)}{t-s} \frac{\tilde{\psi}_{-k_0+\ell}(t)}{t} dt.$$

Via the substitution $t \mapsto \rho t$ with $\rho = 2^{m(-k_0+\ell)}$ we obtain

$$\rho K_{\lambda_1,\lambda_2}(\rho s) = \int_{\mathbb{R}} e^{i\lambda_1\rho t - i\lambda_2\rho(t-s) - i\xi 2^{-k_0+\ell}(t^{1/m} - (t-s)^{1/m})} \frac{\tilde{\psi}_0(t-s)}{t-s} \frac{\tilde{\psi}_0(t)}{t} dt.$$

We need to analyze the phase function

$$Q(t, s) = \lambda_1 \rho t - \lambda_2 \rho(t - s) + \eta(t^{1/m} - (t - s)^{1/m}), \tag{3.5.16}$$

where $\eta = -\xi 2^{-k_0+\ell}$, so in particular

$$2^{\ell-1} \leq |\eta| \leq 2^{\ell+1}. \tag{3.5.17}$$

On first sight this may not look promising, because the phase function includes linear terms which tend to cause trouble (compare Stein and Wainger [SW01]). However, it turns out that we are allowed to take derivatives to isolate the non-linear term (recall $m \geq 2$) because we know by (3.5.17) that its coefficient η is large. Taking two derivatives with respect to t, we obtain

$$\partial_t^2 Q(t,s) = c\eta(t^\alpha - (t-s)^\alpha)$$

with $\alpha = \frac{1}{m} - 2$ and $c = \frac{1}{m}(\frac{1}{m} - 1)$. Suppose that $|s| \geq 2^{-\ell/2}$. Then by the mean value theorem and (3.5.17), $|\partial_t^2 Q(t,s)| \gtrsim 2^{\ell/2}$ throughout the region of t and $t-s$ considered (depending only on the support of $\tilde{\psi}_0$). In this case, an application of the second derivative test shows that

$$|\rho K_{\lambda_1,\lambda_2}(\rho s)| \lesssim 2^{-\ell/4}.$$

On the other hand, if $|s| \leq 2^{-\ell/2}$ we merely use the triangle inequality for the trivial estimate

$$|\rho K_{\lambda_1,\lambda_2}(\rho s)| \lesssim 1.$$

Altogether we have proved, with $\rho = 2^{m(-k_0+\ell)}$,

$$|K_{\lambda_1,\lambda_2}(s)| \lesssim 2^{-\ell/4}\rho^{-1}\mathbf{1}_{\{|s|\leq 4\rho\}}(s) + \rho^{-1}\mathbf{1}_{\{|s|\leq 2^{-\ell/2}\rho\}}(s),$$

uniformly in λ_1, λ_2. Applying this in (3.5.15) allows us to deduce that

$$|\tilde{T}\tilde{T}^* F(y)| \lesssim 2^{-\ell/4} MF(y), \tag{3.5.18}$$

where MF denotes the standard one-variable Hardy-Littlewood maximal function. An application of the L^2 estimate for M now implies our claim (3.5.14) with $\gamma = 1/8$; by Plancherel we then finally obtain (3.5.13).

3.5.2 Square function estimate

In this section we assemble the single annulus estimates of the previous section to derive the L^p boundedness of our operator $T = B^{m,m}$. This application of the Littlewood-Paley theory is in the spirit of Bateman and Thiele [BT13].

In view of the relation $TP_k = P_k T$ and the standard Littlewood-Paley inequalities, we have

$$\|Tf\|_p \lesssim \Big\| \Big(\sum_{k\in\mathbb{Z}} |TP_k f|^2 \Big)^{1/2} \Big\|_p \lesssim \Big\| \Big(\sum_{k\in\mathbb{Z}} |T_k^{(1)} P_k f|^2 \Big)^{1/2} \Big\|_p + \Big\| \Big(\sum_{k\in\mathbb{Z}} |T_k^{(2)} P_k f|^2 \Big)^{1/2} \Big\|_p$$

In the term for $T_k^{(1)}$ on the right hand side we apply the estimate (3.5.2); then by applying the vector-valued estimates of Theorems 3.3.1 and 3.3.3 to the maximally truncated Carleson operator and the one-variable maximal function, we obtain

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} |T_k^{(1)} P_k f|^2 \Big)^{1/2} \Big\|_p \lesssim \Big\| \Big(\sum_{k\in\mathbb{Z}} |P_k f|^2 \Big)^{1/2} \Big\|_p \lesssim \|f\|_p.$$

For $T_k^{(2)}$, we recall that by (3.5.12)

$$|T_k^{(2)} P_k f| \lesssim \mathscr{M} P_k f + \sum_{\ell=0}^{\infty} |T_{-k+\ell} P_k f|,$$

so that by Minkowski's inequality for integrals,

$$\Big\| \Big(\sum_{k \in \mathbb{Z}} |T_k^{(2)} P_k f|^2 \Big)^{1/2} \Big\|_p \lesssim \Big\| \Big(\sum_{k \in \mathbb{Z}} |\mathscr{M} P_k f|^2 \Big)^{1/2} \Big\|_p + \sum_{\ell=0}^{\infty} \Big\| \Big(\sum_{k \in \mathbb{Z}} |T_{-k+\ell} P_k f|^2 \Big)^{1/2} \Big\|_p.$$

We may apply Theorem 3.3.2 to obtain a bound $\lesssim \|f\|_p$ for the first term; for the second term it would suffice to show that for each $1 < p < \infty$ there exists $\gamma > 0$ such that for every $\ell \geq 0$,

$$\Big\| \Big(\sum_{k \in \mathbb{Z}} |T_{-k+\ell} P_k f|^2 \Big)^{1/2} \Big\|_p \lesssim 2^{-\gamma \ell} \Big\| \Big(\sum_{k \in \mathbb{Z}} |P_k f|^2 \Big)^{1/2} \Big\|_p \lesssim 2^{-\gamma \ell} \|f\|_p. \qquad (3.5.19)$$

For $p = 2$ this follows from (3.5.13). But recall that we always have the simple estimate $|T_\ell P_k f| \lesssim \mathscr{M} P_k f$ for all $k, \ell \in \mathbb{Z}$ by the triangle inequality. Therefore Theorem 3.3.2 implies a bound without decay, namely

$$\Big\| \Big(\sum_{k \in \mathbb{Z}} |T_{-k+\ell} P_k f|^2 \Big)^{1/2} \Big\|_r \lesssim \|f\|_r, \qquad (3.5.20)$$

valid for all $1 < r < \infty$. Now in general, (3.5.19) follows for all $1 < p < \infty$ by interpolating between the $L^2 \to L^2(\ell^2)$ case of (3.5.19) and the $L^r \to L^r(\ell^2)$ bound (3.5.20) for the vector-valued map $f \mapsto \{T_{-k+\ell} P_k f\}_{k \in \mathbb{Z}}$. The proof of Theorem 3.1.2 is now complete.

3.5.3 Remarks on the proof for $A^{m,1}$

As mentioned above, we will not repeat the proof explicitly for $A^{m,1}$, but merely complement the sketch already provided in Section 3.2.3 by pointing out two key modifications. Of course one interchanges the roles of the x and y variables. In addition:

(1) The cancellation miracle for even m in the term **II** of (3.5.5) does not occur for $A^{m,1}$. Instead one always needs to invoke Carleson's theorem in the form of Theorem 3.3.1, analogous to the computation already carried out for odd m in (3.5.6).

(2) In the treatment of $A^{m,1}$, the restriction $m \neq 2$ originates because the relevant phase function analogous to (3.5.16) is

$$Q(t,s) = \lambda_1 \rho t - \lambda_2 \rho (t-s) + \eta(t^m - (t-s)^m).$$

Visibly, when $m = 2$, the phase function $Q(t,s)$ is now *linear* in t, so that its second derivative vanishes, and consequently we fail in this case to obtain a good bound for the kernel.

3.6 Deductions for partial Carleson operators

3.6.1 L^2 consequences of partial Carleson bounds

As stated in Section 3.1.3, the L^2 boundedness of $A_N^{m,1}$, for any fixed integer $m \geq 1$, implies the L^2 boundedness of Carleson's operator (3.1.1). Similarly, the L^2 boundedness of $B_N^{m,m}$, when $m \geq 1$ is an odd integer, implies the L^2 boundedness of Carleson's operator. (Of course, in the work of this paper, our logic is actually the other way round: in proving Theorem 3.1.2, we used Carleson's theorem as a black box.)

We will see how to carry out these deductions from a more general argument we now give in the context of the quadratic Carleson operator $\mathscr{C}^{\text{par}}$ along the parabola (defined in equation (3.1.14)). We prove that an inequality of the form (3.1.15) would imply the analogue over $\mathbb{R}$ of Lie's result [Lie09] on the one-variable quadratic Carleson operator $\mathscr{C}_Q$:

Proposition 3.6.1. *Assume the veracity of the estimate*

$$\Big\| \int_{\epsilon \leq |t| \leq R} f(x-t, y-t^2) e^{iN_1(x)t + iN_2(x)t^2} \frac{dt}{t} \Big\|_{L^2(dxdy)} \leq C\|f\|_2, \tag{3.6.1}$$

for all Schwartz functions f, where $N_1, N_2 : \mathbb{R} \mapsto \mathbb{R}$ are measurable functions, $0 < \epsilon < R$ are real parameters, and the constant C is independent of f, N_1, N_2, ϵ, R. Then the operator

$$f \mapsto \mathscr{C}_Q f(x) := \sup_{N \in \mathbb{R}^2} \Big| p.v. \int_{\mathbb{R}} f(x-t) e^{iN_1 t + iN_2 t^2} \frac{dt}{t} \Big|$$

is bounded on $L^2(\mathbb{R})$.

Note that in our assumed bound (3.6.1), the linearizing functions N_1, N_2 are independent of y, so this is a far weaker assumption than the conjectured L^2 bound for $\mathscr{C}^{\text{par}}$ in (3.1.14). In the argument that we will now give for (3.6.1), if we replace the curve (t, t^2) by (t, t^m) and the phase by $N_1(x)t + N_2(x)t^m$, and furthermore specify that N_2 is identically zero, we may deduce Carleson's original theorem from the partial bound for $A_N^{m,1}$ for any integer $m \geq 1$; or, if we specify N_1 is identically zero, we may deduce Carleson's original theorem from the partial bound for $B_N^{m,m}$ for m an odd integer. (When m is even, under the specification $N_1 \equiv 0$, the operator in (3.6.1) would vanish, due to the integrand being an odd function.)

In general, to prove Proposition 3.6.1, we use an elementary tensor $f(x,y) = h(x)g(y)$, where h, g are real Schwartz functions, in which case (3.6.1) implies

$$\Big\| \int_{\epsilon \leq |t| \leq R} h(x-t) e^{iN_1(x)t + iN_2(x)t^2} g(y-t^2) \frac{dt}{t} \Big\|_{L^2(dxdy)} \leq C\|h\|_2 \|g\|_2.$$

Applying Plancherel's theorem in the y variable we obtain

$$\Big\| \int_{\epsilon \leq |t| \leq R} h(x-t) e^{iN_1(x)t + iN_2(x)t^2} \widehat{g}(\eta) e^{-i\eta t^2} \frac{dt}{t} \Big\|_{L^2(dxd\eta)} \leq C\|h\|_2 \|g\|_2. \tag{3.6.2}$$

Suppose for the time being that we have chosen g such that we have an estimate of the form

$$\Big\| \int_{\epsilon \le |t| \le R} h(x-t) e^{iN_1(x)t + iN_2(x)t^2} \widehat{g}(\eta)(e^{-i\eta t^2} - 1) \frac{dt}{t} \Big\|_{L^2(dx d\eta)} \le C \|h\|_2 \|g\|_2. \quad (3.6.3)$$

We would deduce from (3.6.2) and (3.6.3) that

$$\Big\| \int_{\epsilon \le |t| \le R} h(x-t) e^{iN_1(x)t + iN_2(x)t^2} \widehat{g}(\eta) \frac{dt}{t} \Big\|_{L^2(dx d\eta)} \le C \|h\|_2 \|g\|_2,$$

so that by Plancherel and Fubini,

$$\Big\| \int_{\epsilon \le |t| \le R} h(x-t) e^{iN_1(x)t + iN_2(x)t^2} \frac{dt}{t} \Big\|_2 \le C \|h\|_2.$$

Via Fatou's lemma this gives the L^2 boundedness of the quadratic Carleson operator $h \mapsto \mathscr{C}_Q h$, as claimed in Proposition 3.6.1.

To obtain the estimate (3.6.3), we choose δ with $0 < \delta < 1/R^2$ and specify that g be a Schwartz function on $\mathbb{R}$ such that $\widehat{g}$ is supported on $[-\delta, \delta]$ and $\|g\|_2 > 0$. Then by Minkowski's inequality and Fubini, the left hand side of (3.6.3) is bounded by

$$\|h\|_2 \int_{\epsilon \le |t| \le R} \|\widehat{g}(\eta)(e^{-i\eta t^2} - 1)\|_{L^2(d\eta)} \frac{dt}{|t|}. \quad (3.6.4)$$

The mean value theorem, followed by Plancherel, shows that for $|t| \le R$,

$$\int_{\mathbb{R}} \Big| \widehat{g}(\eta)(e^{-i\eta t^2} - 1) \Big|^2 d\eta \le \delta^2 t^4 \|g\|_2^2.$$

This implies that (3.6.4) is no greater than

$$\|h\|_2 \|g\|_2 \cdot \delta \int_{\epsilon \le |t| \le R} |t| dt \le \|h\|_2 \|g\|_2, \quad (3.6.5)$$

which completes the proof of (3.6.3), and hence Proposition 3.6.1.

3.6.2 L^2 deductions for partial Carleson operators

Remark 3.1.1 stated that L^2 bounds for $A_N^{2,1}$, $A_N^{m,m}$ and $B_N^{m,1}$ (with $m > 1$) follow from known Carleson theorems. We briefly indicate these deductions, which are along the lines of arguments in Sections 3.4.1 and 3.4.2. By Plancherel's theorem in the free y-variable,

$$\|A_N^{m,m} f\|_{L^2(dxdy)} = \Big\| \int_{\mathbb{R}} g_\eta(x-t) e^{i(N(x)-\eta)t^m} \frac{dt}{t} \Big\|_{L^2(dx d\eta)}$$

where $g_\eta(x) = \int_{\mathbb{R}} e^{-i\eta y} f(x,y) dy$. Then an L^2 bound of the form

$$\Big\| \int_{\mathbb{R}} g_\eta(x-t) e^{i(N(x)-\eta)t^m} \frac{dt}{t} \Big\|_{L^2(dx)} \lesssim \|g_\eta\|_{L^2(dx)},$$

uniform in η, follows from Stein and Wainger [SW01] (since $m > 1$), and this suffices.

In the next case,

$$\|A_N^{2,1} f\|_{L^2(dxdy)} = \Big\| \int_{\mathbb{R}} g_\eta(x-t) e^{iN(x)t - i\eta t^2} \frac{dt}{t} \Big\|_{L^2(dxd\eta)}.$$

Observe that

$$i\eta t^2 = i\eta(x-t)^2 - i\eta x^2 + 2i\eta xt.$$

Define $Q_\eta f(x) = e^{i\eta x^2} f(x)$ and set $\widetilde{N}(x) = N(x) - 2\eta x$. Then,

$$\int_{\mathbb{R}} g_\eta(x-t) e^{iN(x)t - i\eta t^2} \frac{dt}{t} = e^{i\eta x^2} \int_{\mathbb{R}} Q_{-\eta} g_\eta(x-t) e^{i\widetilde{N}(x)t} \frac{dt}{t} = Q_\eta H_{\widetilde{N}(x)} Q_{-\eta} g_\eta(x),$$

where $H_N f(x) = \int_{\mathbb{R}} f(x-t) e^{iNt} \frac{dt}{t}$. Since Q_η is an isometry in L^2, our claim follows from the L^2 bound for the Carleson operator.

In the final case, by Plancherel's theorem in the free x-variable,

$$\|B_N^{m,1} f\|_{L^2(dxdy)} = \Big\| \int_{\mathbb{R}} g_\xi(x - t^m) e^{i(N(x)-\eta)t} \frac{dt}{t} \Big\|_{L^2(d\xi dy)}$$

where $g_\xi(x) = \int_{\mathbb{R}} e^{-i\xi x} f(x,y) dx$. Thus the required L^2 bound follows from sending $t \mapsto t^{1/m}$ and applying Guo [Guo16] to the resulting operator, which has one fractional monomial in the phase.

3.7 Proof of vector-valued inequalities

3.7.1 Proof of Theorem 3.3.1

We assemble results from the Grafakos texts [Gra14a], [Gra14b]. By [Gra14b, Lemma 6.3.2], there is a positive constant $c > 0$ such that for any $1 \le p < \infty$, for all $f \in L^p(\mathbb{R})$ we have the pointwise inequality

$$\mathscr{C}^* f \le cMf + M(\mathscr{C} f), \tag{3.7.1}$$

where M is the standard one-dimensional Hardy-Littlewood maximal function. Since the vector-valued $L^p(\ell^2)$ inequality analogous to (3.3.3) is known to hold for the Hardy-Littlewood maximal function (see e.g. [Ste93, Chapter II §1.1]), the problem is then reduced to proving the analogue of (3.3.3) for the Carleson operator $\mathscr{C}$. In fact, this is a special case of [Gra14b, Exercise 6.3.4], which claims that for all $1 < p, r < \infty$ and all weights $w \in A_p$,

$$\Big\| \Big(\sum_k |\mathscr{C} f_k|^r \Big)^{1/r} \Big\|_{L^p(w)} \le C_{p,r}(w) \Big\| \Big(\sum_k |f_k|^r \Big)^{1/r} \Big\|_{L^p(w)} \tag{3.7.2}$$

for all sequences of functions $f_k \in L^p(w)$. This inequality may be verified, following Grafakos, by the method of extrapolation. We need only note that [Gra14b, Theorem 6.3.3] provides a weighted estimate

$$\|\mathscr{C} f\|_{L^p(w)} \le C(p, [w]_{A_p}) \|f\|_{L^p(w)}, \tag{3.7.3}$$

for every $1 < p < \infty$ and $w \in A_p$. This is sufficient to prove (3.7.2) for all the stated values of r, p by applying the vector-valued extrapolation result [Gra14a, Corollary 7.5.7]. (Here we remark on the detail that Corollary 7.5.7, to which we appeal, requires that $C(p, [w]_{A_p})$ be an increasing function in $[w]_{A_p}$. We can ensure that this is the case if we have the statement, slightly stronger than (3.7.3), that for every $B > 0$ there exists a constant $C_p(B)$ such that for all $w \in A_p$ with $[w]_{A_p} \leq B$ we have $\|\mathscr{C} f\|_{L^p(w)} \leq C_p(B)\|f\|_{L^p(w)}$; such a statement is verified by the explicit version of (3.7.3) given by Lerner and Di Plinio [DPL14, Theorem 1.1].)

Alternatively, once one has the pointwise inequality (3.7.1) and has consequently reduced matters to proving an $L^p(\ell^2)$ vector-valued inequality for $\mathscr{C}$, one can turn to the original result [RdFRT86] in the $L^p(\ell^2)$ case, or the recent streamlined proof [DS15, Theorem 7.1].

3.7.2 Proof of Theorem 3.3.2

We recall that the scalar-valued L^p-bound for $\mathscr{M}$ was obtained by comparing it to a square function [Ste93, Chapter XI §1.2]. Indeed, let $\chi(t)$ be a non-negative smooth function with compact support on the interval $[-2, 2]$, such that $\chi(t) \equiv 1$ on $[-1, 1]$. For $k \in \mathbb{Z}$, let $\chi_k(t) = 2^{-k}\chi(2^{-k}t)$, $d\mu_k(x, y) = \delta_{y=x^m}\chi_k(x)$, and

$$A_k f(x, y) = f * d\mu_k(x, y) = \int_{\mathbb{R}} f(x - t, y - t^m)\chi_k(t)dt.$$

Also let $\phi(x, y)$ be a smooth function with compact support on the unit ball in $\mathbb{R}^2$, normalized such that

$$\int_{\mathbb{R}^2} \phi(x, y)dxdy = \int_{\mathbb{R}} \chi(t)dt.$$

For $k \in \mathbb{Z}$, let $\phi_k(x, y) = 2^{-(m+1)k}\phi(2^{-k}x, 2^{-mk}y)$, and

$$B_k f(x, y) = f * \phi_k(x, y) = \int_{\mathbb{R}^2} f(x - u, y - v)\phi_k(u, v)dudv.$$

Then for non-negative functions f, we have the pointwise inequality

$$\mathscr{M} f \leq \sup_{k\in\mathbb{Z}} B_k f + Sf, \tag{3.7.4}$$

where S is the following square function:

$$Sf := \left(\sum_{k\in\mathbb{Z}} |A_k f - B_k f|^2\right)^{1/2}. \tag{3.7.5}$$

Now $\sup_{k\in\mathbb{Z}} B_k f$ is bounded by the standard maximal function associated to non-isotropic 'squares' of sizes $R \times R^m$ on $\mathbb{R}^2$. It is known that a vector-valued estimate holds for the maximal function associated to these non-isotropic squares; that is an analogue of Theorem 3.3.3. Thus the inequality (3.3.5) of Theorem 3.3.2 holds for $1 < p < \infty$ if we have $\sup_{k\in\mathbb{Z}} B_k$ in place of $\mathscr{M}$ on the left hand side. Hence to prove the desired form of (3.3.5), all we need to do is to establish

$$\left\| \left(\sum_{\ell \in \mathbb{Z}} |Sf_\ell|^2 \right)^{1/2} \right\|_{L^p} \lesssim_p \left\| \left(\sum_{\ell \in \mathbb{Z}} |f_\ell|^2 \right)^{1/2} \right\|_{L^p} \tag{3.7.6}$$

where S is defined by (3.7.5), and $1 < p < \infty$.

The following scalar-valued inequality for $1 < p < \infty$ is already known [Ste93, Section 4, Theorem 11]:

$$\|Sf\|_{L^p} \lesssim_p \|f\|_{L^p}. \tag{3.7.7}$$

But to deduce (3.7.6) we will instead use a related scalar-valued inequality for a signed operator. For ϵ_k a random sequence of signs ± 1, define

$$Tf := \sum_{k \in \mathbb{Z}} \epsilon_k (A_k f - B_k f).$$

It is known that

$$\|Tf\|_{L^p} \lesssim_p \|f\|_{L^p} \tag{3.7.8}$$

for all $1 < p < \infty$, independent of the signs ϵ_k. At the end of this section, we briefly recall a proof of this, for which one uses crucially the non-vanishing of the curvature of the curve (t, t^m), but we first deduce (3.7.6) from (3.7.8).

To do so, note that since T is linear, the Marcinkiewicz-Zygmund theorem implies that

$$\| \, |Tf_\ell|_{\ell^2} \, \|_{L^p} \lesssim_p \| \, |f_\ell|_{\ell^2} \, \|_{L^p}$$

for $1 < p < \infty$, i.e.

$$\left\| \left| \sum_{k \in \mathbb{Z}} \epsilon_k (A_k f_\ell - B_k f_\ell) \right|_{\ell^2(d\ell)} \right\|_{L^p} \lesssim_p \| \, |f_\ell|_{\ell^2} \, \|_{L^p}.$$

(We write $\ell^2(d\ell)$ to emphasize that the ℓ^2 norm is taken with respect to the variable ℓ.) Now we take the expectation, denoted $\mathbb{E}$, over all the possible choices of ϵ_k; by Khintchine's inequality,

$$\left(\sum_{k \in \mathbb{Z}} |A_k f_\ell - B_k f_\ell|^2 \right)^{1/2} \simeq \mathbb{E} \left| \sum_{k \in \mathbb{Z}} \epsilon_k (A_k f_\ell - B_k f_\ell) \right|.$$

Taking the $\ell^2(d\ell)$ and then L^p norms on both sides, we get

$$\begin{aligned}\left\| \left(\sum_{k,\ell\in\mathbb{Z}} |A_k f_\ell - B_k f_\ell|^2 \right)^{1/2} \right\|_{L^p} &\simeq \left\| \left(\mathbb{E} \left| \sum_{k\in\mathbb{Z}} \epsilon_k (A_k f_\ell - B_k f_\ell) \right| \right)_{\ell^2(d\ell)} \right\|_{L^p} \\ &\leq \mathbb{E} \left\| \left| \sum_{k\in\mathbb{Z}} \epsilon_k (A_k f_\ell - B_k f_\ell) \right|_{\ell^2(d\ell)} \right\|_{L^p} \\ &\lesssim_p \mathbb{E} \| \, |f_\ell|_{\ell^2} \, \|_{L^p} \\ &= \| \, |f_\ell|_{\ell^2} \, \|_{L^p}.\end{aligned}$$

(The first inequality is the Minkowski inequality.) The left hand side above is precisely $\| \, |Sf_\ell|_{\ell^2} \, \|_{L^p}$. This proves (3.7.6), and hence (3.3.5) of Theorem 3.3.2, for $1 < p < \infty$.

There are at least two ways of proving (3.7.8). One is by complex interpolation, along the lines of arguments in [SW78, Section 4], which we will not discuss here. Alternatively, we can deduce (3.7.8) from a result of Duoandikoetxea and Rubio de Francia [DRdF86] without using complex interpolation. To do so, let

$$d\sigma_k = \epsilon_k (d\mu_k - \phi_k dx dy).$$

Then $d\sigma_k$ has total mass $\|d\sigma_k\| \lesssim 1$, and its Fourier transform satisfies

$$|\widehat{d\sigma_k}(\xi,\eta)| \lesssim \min\{2^k \|(\xi,\eta)\|, (2^k \|(\xi,\eta)\|)^{-1/m}\}.$$

(Here we see the curvature of (t, t^m).) Furthermore, the operator $\sup_{k\in\mathbb{Z}} |f * |d\sigma_k||$ is bounded by the maximal Radon transform along the curve (t, t^m) plus the Hardy-Littlewood maximal operator adapted to certain non-isotropic balls in $\mathbb{R}^2$. It follows that $\sup_{k\in\mathbb{Z}} |f * |d\sigma_k||$ is bounded on $L^q(\mathbb{R}^2)$ for all $1 < q < \infty$. Thus Theorem B of Duoandikoetxea and Rubio de Francia [DRdF86] applies, and shows that $Tf = \sum_{k\in\mathbb{Z}} f * d\sigma_k$ is bounded on L^p for all $1 < p < \infty$. This completes our proof of (3.7.8).

Acknowledgments. We would like to thank C. Thiele and E. M. Stein for many helpful comments and discussions. Pierce is supported in part by NSF DMS-1402121. Roos is supported by the German National Academic Foundation. Yung is supported in part by the Hong Kong Research Grant Council Early Career Grant CUHK24300915. This collaboration was initiated at the Hausdorff Center for Mathematics and Oberwolfach, and continued at the joint AMS-EMS-SPM 2015 international meeting at Porto. The authors thank all institutions involved for gracious and productive work environments.

Chapter 4

Maximal functions and Hilbert transforms along variable non-flat homogeneous curves

This chapter consists of a joint publication with Shaoming Guo, Jonathan Hickman and Victor Lie [GHLR16], which will appear in the Proceedings of the London Mathematical Society. The copyright is held by the London Mathematical Society and the article appears as part of this thesis in accordance with their copyright regulations.

4.1 Introduction

This paper focuses on the study of certain maximal and singular integral operators that act by integration along *variable* homogeneous curves in the plane, of the form

$$\Gamma_u^\alpha(t) := (t, u \cdot [t]^\alpha)\,,$$

where here and throughout the paper, the exponent α is a fixed positive real number, the notation $[t]^\alpha$ stands for either $|t|^\alpha$ or $\operatorname{sgn}(t)|t|^\alpha$, while the "coefficient" $u(\cdot,\,\cdot)$ is allowed to change depending on the base point $(x, y) \in \mathbb{R}^2$.

More precisely, given $u : \mathbb{R}^2 \to \mathbb{R}$ a *measurable* function and $0 < \varepsilon_0 \leq \infty$ a parameter, we consider the following objects:

- the (ε_0-truncated) **maximal operator along** Γ_u^α, defined by

$$\mathcal{M}_{u,\varepsilon_0}^{(\alpha)} f(x, y) = \sup_{0<\varepsilon<\varepsilon_0} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} |f(x - t, y - u(x, y)[t]^\alpha)| dt. \tag{4.1.1}$$

- the (ε_0-truncated) **Hilbert transform along** Γ_u^α, given by

$$\mathcal{H}_{u,\varepsilon_0}^{(\alpha)} f(x, y) = \text{p.v.} \int_{-\varepsilon_0}^{\varepsilon_0} f(x - t, y - u(x, y)[t]^\alpha) \frac{dt}{t}. \tag{4.1.2}$$

For convenience, in what follows we will use the convention that $[t]^1 = t$. Moreover, when $\alpha = 1$, we will leave out the dependence on α and simply write $\mathcal{M}_{u,\epsilon_0}$ and $\mathcal{H}_{u,\varepsilon_0}$. The same principle applies to $\epsilon_0 = \infty$: in this case we will simply write $\mathcal{M}_u^{(\alpha)}$ and $\mathcal{H}_u^{(\alpha)}$, respectively.

A difficult problem in the area of harmonic analysis is to understand the weakest possible regularity assumptions on u that guarantee the L^p boundedness of $\mathcal{M}^{(\alpha)}_{u,\varepsilon_0}$ and $\mathcal{H}^{(\alpha)}_{u,\varepsilon_0}$. Our aim in this paper is to provide a partial solution to this problem when we impose a nontrivial curvature condition by requiring $\alpha \neq 1$.

We will now state our main results. The first result regards the boundedness of the maximal operator (4.1.1) and extends an earlier result of Marletta and Ricci [MR98].

Theorem 4.1.1. *Let $\alpha > 0$ and $\alpha \neq 1$. Then the following hold:*

1. [MR98] *If $u : \mathbb{R}^2 \to \mathbb{R}$ is measurable, then for every $2 < p \leq \infty$ we have*

$$\|\mathcal{M}^{(\alpha)}_u f\|_p \leq C_{p,\alpha} \|f\|_p \,. \tag{4.1.3}$$

2. *If $u : \mathbb{R}^2 \to \mathbb{R}$ is Lipschitz, then there exists $\varepsilon_0 = \varepsilon_0(\|u\|_{\mathrm{Lip}}) > 0$ such that for every $1 < p \leq 2$ we have*

$$\|\mathcal{M}^{(\alpha)}_{u,\varepsilon_0} f\|_p \leq C_{p,\alpha} \|f\|_p \,. \tag{4.1.4}$$

Here $C_{p,\alpha} \in (0, \infty)$ is a constant that depends only on p and α.

The second main result regards the boundedness of the Hilbert transform (4.1.2).

Theorem 4.1.2. *Let $\alpha > 0$ and $\alpha \neq 1$. Let $u : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function and assume that*

$$u(x, y) = u(x, 0) \text{ for every } x, y \in \mathbb{R}. \tag{4.1.5}$$

Then we have that for all $1 < p < \infty$ the following holds:

$$\|\mathcal{H}^{(\alpha)}_u f\|_p \leq C_{p,\alpha} \|f\|_p. \tag{4.1.6}$$

The constant $C_{p,\alpha} \in (0, \infty)$ depends only on p and α.

Remark 4.1.1. We would first like to stress that the analogue of estimate (4.1.3) for the Hilbert transform $\mathcal{H}^{(\alpha)}_u$ fails for every $p \in (1, \infty)$ if we only assume u to be measurable. This follows by a straight-forward modification of Karagulyan's [Kar07] construction of a counterexample in the case $\alpha = 1$. However, if we assume u to be Lipschitz it is possible that the analogue of (4.1.4) for $\mathcal{H}^{(\alpha)}_u$ holds. Unfortunately, as of now, we are not able to prove or disprove this.

Remark 4.1.2. Notice that unlike the situation described in Theorem 4.1.1, in Theorem 4.1.2 we do not require any regularity assumptions on the function $u(x, 0)$. Next, we remark that as opposed to (4.1.4), the ϵ_0-truncation is not present in statement (4.1.6). This is a direct consequence of a standard scaling argument that makes the truncation in $\mathcal{H}^{(\alpha)}_{u,\epsilon_0}$ from (4.1.2) become irrelevant. Nevertheless, the one-variable assumption (4.1.5) should be understood as being strictly stronger than the Lipschitz assumption imposed in Theorem 4.1.1. Indeed, we have the following corollary of Theorem 4.1.1.

Corollary 4.1.3. *Under the same assumptions as in Theorem 4.1.2, we have*

$$\|\mathcal{M}_u^{(\alpha)} f\|_p \leq C_{p,\alpha}\|f\|_p \tag{4.1.7}$$

for all $1 < p \leq \infty$.

The proof of Corollary 4.1.3 is via an anisotropic scaling argument which we sketch presently. First of all, by the scaling $x \to \lambda x, y \to \lambda^\alpha y$, and a density argument, it suffices to show that

$$\|\mathcal{M}_{u,\epsilon_0}^{(\alpha)} f\|_p \leq C_{p,\alpha}\|f\|_p, \tag{4.1.8}$$

for all compactly supported smooth functions f. Here ϵ_0 is the same as in Theorem 4.1.1. Now we approximate the chosen measurable function u pointwisely by a sequence of Lipschitz functions $\{u_n\}_{n\in\mathbb{N}}$ satisfying $u_n(x, y) = u_n(x, 0)$ for every $(x, y) \in \mathbb{R}^2$ and whose Lipschitz norms might grow to infinity. By changing variables $x \to x, y \to \lambda y$ in the L^p integration on the left hand side of (4.1.4), we obtain (4.1.8) with u replaced by u_n, and with a constant independent of $n \in \mathbb{N}$. In the end, we apply the dominated convergence theorem to conclude (4.1.8).

The curvature condition is fundamental in the proofs of both Theorem 4.1.1 and Theorem 4.1.2. Our approach relies on stationary phase methods, TT^*-arguments, local smoothing estimates and square function estimates. We speculate that the case $\alpha = 1$ will require a combination of time-frequency techniques and methods presented in the current paper, but this subject remains open for future investigation.

In the following we present the historical evolution of the subject that motivated our interest for this study. We then continue with a discussion of our main results, embedding them in the historical context and also state some further results.

4.1.1 Historical Background

The historical landmark that generated much of the literature discussed below is the so-called Zygmund conjecture. This long-standing open problem asks whether Lipschitz regularity of u suffices to guarantee any non-trivial L^p bounds for the maximal operator

$$\mathcal{M}_{u,\varepsilon_0} f(x, y) = \sup_{0<\varepsilon<\varepsilon_0} \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} |f(x - t, y - u(x, y)t)|dt, \tag{4.1.9}$$

provided ε_0 is small enough depending on $\|u\|_{\mathrm{Lip}}$. A counterexample based on a construction of the Besicovitch-Kakeya set shows that we cannot expect any L^p bounds other than the trivial L^∞ bound if u is only assumed to be Hölder continuous with some exponent strictly smaller than one. The analogous question for $\mathcal{H}_{u,\varepsilon_0}$ is also widely open. For a detailed discussion of these conjectures, the interested reader is invited to consult Lacey and Li [LL10].

Notice that these two fundamental problems address the boundedness properties of (4.1.1) and (4.1.2) along Γ_u^α in the following context:

- *no regularity* above Lipschitz class is assumed;
- *no curvature* in the t parameter is present, *i.e.* $\alpha = 1$, and thus these objects have rich classes of symmetries.

Partial progress towards understanding the above open problems developed relatively slowly in the past three decades; in direct relation with the itemization above it revolved around the nature of regularity and/or suitable curvature conditions imposed on the vector field u.

Bourgain [Bou89] proved that for every real analytic function u there exists $\epsilon_0 > 0$ such that the associated maximal operator $\mathcal{M}_{u,\epsilon_0}$ is bounded on L^2. His argument can be extended to L^p for all $p > 1$ by using a suitable interpolation argument. For smooth vector fields, Christ, Nagel, Stein and Wainger [CNSW99] proved, under some extra curvature conditions, that the associated maximal operator and singular integral operators are bounded on L^p for $p > 1$. The analogous result to that of Bourgain for singular integral operators was proved by Stein and Street [SS12]. Indeed, the result in [SS12] is far more general: in addition to the case of curves $(t, u \cdot t^\alpha)$ with u analytic and $\alpha \in \mathbb{N}$, they, in fact, consider all polynomials with analytic coefficients.

The first major breakthrough in terms of *regularity* came when Lacey and Li [LL06] brought tools from time-frequency analysis into the problem of Hilbert transforms along vector fields. To state these results, we introduce some notation.

Let $\psi_0 : \mathbb{R} \to \mathbb{R}$ be a non-negative smooth function supported on the set $[-2, -1/2] \cup [1/2, 2]$ such that for all $t \neq 0$,

$$\sum_{l\in\mathbb{Z}} \psi_l(t) = 1\,, \tag{4.1.10}$$

where $\psi_l(t) := \psi_0(2^{-l}t)$. For every $k \in \mathbb{Z}$, let $P_k^{(2)}$ denote the Littlewood-Paley projection in the y-variable corresponding to ψ_k. That is,

$$P_k^{(2)} f(x, y) := \int_{\mathbb{R}} f(x, y - \eta)\check{\psi}_k(\eta)d\eta. \tag{4.1.11}$$

Similarly, we define $P_k^{(1)}$. Now we are ready to state the main result of Lacey and Li.

Theorem 4.1.4 ([LL06]). *Let $u : \mathbb{R}^2 \to \mathbb{R}$ be an arbitrary measurable function. For every $p \geq 2$ there exists $0 < C_p < \infty$ such that the following hold:*

- *For all $k \in \mathbb{Z}$ we have*

$$\|\mathcal{H}_u P_k^{(2)} f\|_{2,\infty} \leq C_2 \|P_k^{(2)} f\|_2. \tag{4.1.12}$$

- *For all $p > 2$ and $k \in \mathbb{Z}$ we have*

$$\|\mathcal{H}_u P_k^{(2)} f\|_p \leq C_p \|P_k^{(2)} f\|_p. \tag{4.1.13}$$

A few years later, by further developing Lacey and Li's methods in [LL06] and [LL10], Bateman [Bat13] and Bateman and Thiele [BT13] proved the following result.

Theorem 4.1.5 ([Bat13], [BT13]). *Let $u : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function satisfying*

$$u(x, y) = u(x, 0) \quad a.e. \quad x, y \in \mathbb{R}. \tag{4.1.14}$$

Then for every $1 < p < \infty$ there exists $0 < C_p < \infty$ such that

$$\|\mathcal{H}_u P_k^{(2)} f\|_p \leq C_p \|P_k^{(2)} f\|_p \tag{4.1.15}$$

uniformly in $k \in \mathbb{Z}$*. Moreover, for all* $p > 3/2$*, we have*

$$\|\mathcal{H}_u f\|_p \leq C_p \|f\|_p . \tag{4.1.16}$$

An earlier result on maximal operators and Hilbert transforms along one-variable vector fields can be found in Carbery, Seeger, Wainger and Wright [CSWW99]. An especially interesting aspect of that work is that they gave an endpoint result on the product Hardy space $H^1_{\text{prod}}(\mathbb{R}\times\mathbb{R})$ under a certain curvature assumption on the function u.

To pass from (4.1.15) to (4.1.16), Bateman and Thiele [BT13] relied crucially on the commutation relation

$$\mathcal{H}_u P_k^{(2)} = P_k^{(2)} \mathcal{H}_u. \tag{4.1.17}$$

Unfortunately, this relation fails for the maximal operator $\mathcal{M}_u$ as it is only a sub-linear operator. For this reason the following problem still remains open.

Open Problem 1. *Let* $u : \mathbb{R}^2 \to \mathbb{R}$ *be a measurable function satisfying* (4.1.14)*. Is* $\mathcal{M}_u$ *bounded on* $L^p(\mathbb{R}^2)$ *for some* $p < \infty$*?*

This open problem served as our original motivation for the aim formulated at the beginning of our current paper. Further motivation for our study is provided by the rich literature addressing the boundedness of maximal and singular Radon transforms, focusing on curvature conditions. This corresponds to the case of Γ_u^α when $\alpha \neq 1$.

Apart from the work [MR98] that has been mentioned already, one more relevant result from this body of literature is due to Seeger and Wainger [SW03]. In this paper the variable curve $(t, u(x,y)\cdot[t]^\alpha)_{t\in\mathbb{R}}$ appears as a special case of the more general curve $\Gamma(x,y,t)$ which satisfies some convexity and doubling hypothesis uniformly in (x,y). For such curves $\Gamma(x,y,t)$, the authors proved that the associated maximal operator and singular integral operators are bounded on L^p for $p > 1$.

For more results in the same spirit, we refer to Nagel, Stein and Wainger [NSW79], Seeger [See94] and the references therein.

4.1.2 Comments on the main results

We will start our discussion with Theorem 4.1.1.

As mentioned earlier, item (1) is the restatement of the result of Marletta and Ricci in [MR98]. To prove this result they used Bourgain's result on the circular maximal operator [Bou86] as a black box. In contrast, we will provide an alternative approach that is more self-contained. In a certain sense we are unraveling the mechanism behind the boundedness of the circular maximal operator, by using the local smoothing estimates of Mockenhaupt, Seeger and Sogge [MSS92] and the l^2L^p decoupling inequalities for cones of Wolff [Wol00], Bourgain [Bou13] and Bourgain and Demeter [BD15]. We do not claim any originality in this approach: that local smoothing estimates can be used to prove the boundedness of the circular maximal operator has already been pointed out in [MSS92]. Moreover, the observation that decoupling inequalities for cones can provide certain progress toward the local smoothing conjecture is due to Wolff [Wol00].

The next comment regards both items (1) and (2): recall that one of the main obstacles in the analysis of the maximal operator is that the analogue of the commutation relation (4.1.17) fails due to sublinearity (see also (4.1.19) below). Thus, in order to prove estimates (4.1.3) and (4.1.4) our strategy is to work with all frequency annuli at the same time and take advantage of the non-trivial curvature provided by $\Gamma_u^\alpha(t) = (t, u(x,0)[t]^\alpha)$ in the situation $\alpha \neq 1$.

The method of proving Theorem 4.1.1 in the absence of the analogue of the commutation relation (4.1.17) might also provide some insight toward Open Problem 1.

We now focus our discussion on Theorem 4.1.2. This result can be regarded as the "curved" analogue of (4.1.16) from Bateman and Thiele's Theorem 4.1.5. In fact, in the next subsection we will state another result that includes the single annulus version of both Theorem 4.1.2 and Theorem 4.1.5, corresponding to (4.1.15) (see Theorem 4.1.6 below). Moreover, in a forthcoming paper of the first, third and fourth author we will be relying in part on the ideas developed by the third author in [Lie15a], [Lie15b] in order to extend Theorem 4.1.2 to the setting of general curves (not necessarily homogeneous) obeying some suitable smoothness and curvature conditions.

Regarding the proof of Theorem 4.1.2, we rely on several ingredients. Following the general scheme in [BT13], we first prove a single annulus estimate

$$\|\mathcal{H}_u^{(\alpha)} P_k^{(2)} f\|_p \lesssim \|P_k^{(2)} f\|_p \tag{4.1.18}$$

for all $p > 1$ and then we use the commutation relation

$$\mathcal{H}_u^{(\alpha)} P_k^{(2)} = P_k^{(2)} \mathcal{H}_u^{(\alpha)} \tag{4.1.19}$$

to pass to a square function estimate. However, it is worth stressing here that the methods through which we achieve (4.1.18) and then (4.1.6) are quite different from the ones in [BT13]: there the authors use time-frequency techniques while in our case we rely on almost-orthogonality, stationary-phase and TT^* methods derived from the presence of curvature.

This difference is also reflected in estimate (4.1.6) from Theorem 4.1.2 where we have an improved L^p range including bounds for all p close to 1. In contrast with this, the potential range for estimate (4.1.16) in Theorem 4.1.5 to hold is $p > 4/3$. This exponent is related to the exponents in the variation norm Carleson theorem [OST$^+$12]. We refer to Bateman and Thiele [BT13] for a more detailed discussion.

In order to achieve the L^p bounds for all $p > 1$ we develop a pointwise estimate for taking averages along variable curves, via the shifted (strong) Hardy-Littlewood maximal function[1]. This pointwise estimate has a natural geometric interpretation: roughly speaking, it says that the averages along a thickened segment[2] of the curve $(t, u \cdot |t|^\alpha)$ can be pointwisely controlled, up to a small logarithmic loss, by a sum of averages taken over a number of rectangles. We remark that in the case $u \equiv 1$, our proof of Theorem 4.1.2 reduces to an alternative proof for the L^p boundedness of the classical singular Radon transform along the curve $(t, [t]^\alpha)$ for $p \neq 2$, which does not seem to have appeared in the literature.

[1] See (4.3.19) below.
[2] That is, a small neighborhood of a segment.

4.1.3 Further results

Here we present two more results.

As already mentioned above, the first result encompasses the single annulus estimates corresponding to Theorem 4.1.2 and Theorem 4.1.5.

Theorem 4.1.6. *Let* $u : \mathbb{R}^2 \to \mathbb{R}^2$ *be a measurable function with* $u(x, y) = (u_1(x, y), u_2(x, y))$*. We then define the Hilbert transform along the variable polynomial curve*

$$(t, u_1(x, y)t + u_2(x, y)[t]^\alpha)_{t\in\mathbb{R}}$$

by

$$\mathcal{H}_u^{(\alpha)} f(x, y) = \text{p.v.} \int_{\mathbb{R}} f(x - t, y - u_1(x, y)t - u_2(x, y)[t]^\alpha)\frac{dt}{t}. \tag{4.1.20}$$

Suppose now that

$$u(x, y) = u(x, 0) \quad a.e.\ \ x, y \in \mathbb{R}\,. \tag{4.1.21}$$

Then, for each $\alpha > 0$ *with* $\alpha \neq 2$ *and each* $p > 1$*, there exists* $C_{\alpha,p} > 0$ *such that*

$$\|\mathcal{H}_u^{(\alpha)} P_k^{(2)} f\|_p \le C_{\alpha,p}\|P_k^{(2)} f\|_p, \tag{4.1.22}$$

uniformly in $k \in \mathbb{Z}$*, where here we recall that* $P_k^{(2)}$ *stands for the Littlewood-Paley projection operator in the* y*-variable.*

Notice that, by applying a partial Fourier transform in the y-variable and using Plancherel, our theorem implies the following.

Corollary 4.1.7. *For each* $\alpha > 0$ *with* $\alpha \neq 2$*, we have*

$$\left\| \sup_{u_1,u_2\in\mathbb{R}} \left| \text{p.v.} \int_{\mathbb{R}} f(x - t)e^{iu_1t+iu_2[t]^\alpha}\frac{dt}{t} \right| \right\|_2 \le C_\alpha\|f\|_2, \tag{4.1.23}$$

with a constant C_α *depending only on* α*.*

Remark that the case $\alpha = 2$ is a deep result due to the third author [Lie09] which Corollary 4.1.7 does not encompass due to the quadratic modulation symmetries present if $\alpha = 2$. The third author also proved bounds for the full polynomial Carleson operator [Lie11],

$$f \mapsto \sup_P \left| \text{p.v.} \int_{\mathbb{R}} f(x - t)e^{iP(t)}\frac{dt}{t} \right|, \tag{4.1.24}$$

where the supremum goes over all polynomials P with real coefficients of degree less than a fixed number. The proof uses a sophisticated time-frequency approach. If α is a positive integer, (4.1.23) is of course a corollary of the L^2 bounds for (4.1.24). It is interesting however, that we can prove (4.1.23) essentially using only Carleson's theorem as a black box.

Theorem 4.1.6 is about a single annulus estimate. It can be viewed as an extension of Bateman's result [Bat13]. The proof of this result is a combination of the stationary phase method with an application of Bateman's single annulus estimate [Bat13].

Next, we will state a second single annulus estimate. It is the counterpart of Lacey and Li's result Theorem 4.1.4, and will provide a key insight towards proving Theorem 4.1.1. However, it requires a completely different proof compared with Theorem 4.1.6.

Theorem 4.1.8. *Given $\alpha > 0$ and $\alpha \neq 1$. For each measurable function $u : \mathbb{R}^2 \to \mathbb{R}$, we have*

$$\|H_u^{(\alpha)} P_k^{(2)} f\|_p \leq C_{p,\alpha} \|P_k^{(2)} f\|_p, \tag{4.1.25}$$

for each $p > 2$. Here $C_{p,\alpha}$ does not depend on $k \in \mathbb{Z}$.

The range $p > 2$ is sharp in the sense that Theorem 4.1.8 fails for $p \leq 2$. This can be seen by testing the estimate against the characteristic function of the unit ball.

The proof relies on the local smoothing estimates by Mockenhaupt, Seeger and Sogge [MSS92]. The local smoothing estimates only work for $p > 2$ which is reflected in the constraint $p > 2$ in the above theorem. It is worth mentioning that we will not need the full strength of the local smoothing estimates, but only an "$\varepsilon-$amount" of them, and only for a single $p > 2$. For this reason, we are able to provide a simple and self-contained proof of the local smoothing estimates we need, via decoupling inequalities for cones in $\mathbb{R}^3$. This is the content of Section 4.6. We would like to emphasize again that the approach is due to Wolff [Wol00]. Moreover, in terms of the decoupling inequalities we use for cones in $\mathbb{R}^3$, we do not need the full range $2 \leq p \leq 6$ in Bourgain and Demeter [BD15], but only the range $2 \leq p \leq 4$. The decoupling inequalities for p in this range again have a simple proof. For the sake of completeness, we include it here, see Section 4.7. This argument is due to Bourgain [Bou13].

Structure of the paper.

- In Section 4.2 we prove Theorem 4.1.6. This is a single annulus version of the estimate in Theorem 4.1.2.
- In Section 4.3 we prove the full Theorem 4.1.2. The proof will rely on a vector-valued estimate for the shifted maximal operator.
- In Section 4.4 we show Theorem 4.1.8 whose proof serves as a preparation for the corresponding proof of Theorem 4.1.1.
- In Section 4.5 we provide the proof of Theorem 4.1.1. Theorems 4.1.8 and 4.1.1 rely on the local smoothing estimate in Theorem 4.6.1.
- In Section 4.6 we provide the proof of Theorem 4.6.1 via decoupling inequalities for cones in $\mathbb{R}^3$, following the approach of Wolff [Wol00].
- In Section 4.7 we provide a proof of the decoupling inequalities we need, following the approach of Bourgain [Bou13].

Acknowledgments. S.G. and J.R. would like to thank Christoph Thiele for his guidance on this project. They also thank Brian Street and Po-Lam Yung for useful discussions.

4.2 A single annulus estimate

In this section we prove Theorem 4.1.6. The special case $v \equiv 0$ will later be a key ingredient in the proof of Theorem 4.1.2.

Dropping the dependence on u, v and α in our notation, we now set[3]

$$\mathcal{H}f(x,y) := \int_{\mathbb{R}} f(x-t, y-v(x)t-u(x)[t]^{\alpha})\frac{dt}{t}. \tag{4.2.1}$$

Recall that throughout this section we always assume $\alpha \notin \{1,2\}$. We intend to show that

$$\|\mathcal{H}P_k^{(2)}f\|_p \lesssim \|P_k^{(2)}f\|_p, \tag{4.2.2}$$

for each $p>1$ and $k \in \mathbb{Z}$. The proof is a combination of the TT^* method in the spirit of Stein and Wainger [SW01], and the single annulus estimate for Hilbert transforms along one-variable vector fields by Bateman [Bat13]. Here we will need a maximally truncated version of Bateman's result.

We start the proof of (4.2.2). By an anisotropic scaling

$$x \to x, y \to \lambda y, \tag{4.2.3}$$

it suffices to prove (4.2.2) for $k=0$. In the rest of this section we will always assume that $f = P_0^{(2)}f$. Furthermore, we assume without loss of generality that $u(x)>0$ for almost every x. The case $u(x)<0$ can be handled similarly.

Observe that

$$\mathcal{H}f(x,y) = \sum_{l\in\mathbb{Z}} \int_{\mathbb{R}} f(x-t, y-v(x)t-u(x)[t]^{\alpha})\psi_l(u(x)^{1/\alpha}t)\frac{dt}{t}. \tag{4.2.4}$$

Here ψ_l is as defined in (4.1.10). Writing $\phi_0 = \sum_{l\le 0}\psi_l$, we split the operator $\mathcal{H}$ into two parts:

$$\begin{aligned}\mathcal{H}f(x,y) = &\int_{\mathbb{R}} f(x-t, y-v(x)t-u(x)[t]^{\alpha})\phi_0(u(x)^{1/\alpha}t)\frac{dt}{t}\\ &+\sum_{l\in\mathbb{N}} \int_{\mathbb{R}} f(x-t, y-v(x)t-u(x)[t]^{\alpha})\psi_l(u(x)^{1/\alpha}t)\frac{dt}{t}.\end{aligned} \tag{4.2.5}$$

We bound these two terms separately in the following two subsections.

4.2.1 Low frequency part

Here we treat the first summand on the right hand side of (4.2.5). The idea is to compare it with the (maximally truncated) Hilbert transform along the one-variable vector field $(t, v(x)t)_{t\in\mathbb{R}}$ given by

$$\tilde{\mathcal{H}}^* f(x,y) := \int_{\mathbb{R}} f(x-t, y-v(x)t)\phi_0(u(x)^{1/\alpha}t)\frac{dt}{t}. \tag{4.2.6}$$

[3] Here and throughout the remainder of this text we will omit the principal value notation.

We want the estimate

$$\|\widetilde{\mathcal{H}}^* f\|_p \lesssim \|f\|_p \tag{4.2.7}$$

to hold for all $p > 1$. In the case $v \equiv 0$ this follows from the boundedness of the maximally truncated Hilbert transform. For an arbitrary v it is a result essentially due to Bateman [Bat13]. For a stronger variation norm estimate, see [Guo17b] by the first author. Now we look at the difference, which is given by

$$\int_{\mathbb{R}} [f(x-t, y-v(x)t-u(x)[t]^\alpha) - f(x-t, y-v(x)t)]\, \phi_0(u(x)^{1/\alpha}t)\frac{dt}{t}. \tag{4.2.8}$$

Recall that

$$f(x,y) = \int_{\mathbb{R}} f(x, y-z)\check{\psi}_0(z)dz. \tag{4.2.9}$$

Substituting this identity into (4.2.8) we obtain

$$\int_{\mathbb{R}}\int_{\mathbb{R}} f(x-t, y-v(x)t-z)\left[\check{\psi}_0(z-u(x)[t]^\alpha) - \check{\psi}_0(z)\right]\phi_0(u(x)^{1/\alpha}t)\frac{dt}{t}dz. \tag{4.2.10}$$

Using the key restriction $|u(x)^{1/\alpha}t| \lesssim 1$ derived from (4.2.10), we apply the fundamental theorem of calculus to deduce

$$\left|\check{\psi}_0(z-u(x)[t]^\alpha) - \check{\psi}_0(z)\right| \lesssim \sum_{m\in\mathbb{Z}} \frac{1}{(|m|+1)^2}\mathbb{1}_{[m,m+1]}(z)\cdot u(x)|t|^\alpha. \tag{4.2.11}$$

Due to the sufficiently fast decay of $(|m|+1)^{-2}$, we will see that the summation in m does not cause any problems. For every $m \in \mathbb{Z}$ we consider the term

$$\int_m^{m+1}\int_{\mathbb{R}} |f(x-t, y-v(x)t-z)| u(x)|t|^\alpha \phi_0(u(x)^{1/\alpha}t)\frac{dt}{|t|}dz \tag{4.2.12}$$

arising from applying (4.2.11) to (4.2.10) in the range $m < z < m+1$. To bound this object in L^p we will make use of the following simple observation.

Lemma 4.2.1. *Let $\Gamma : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function and $1 \le p \le \infty$. If K is a non-negative measurable function of two variables such that*

$$f \longmapsto \int_{\mathbb{R}} f(x-t)K(x,t)dt \tag{4.2.13}$$

is bounded as an operator $L^p(\mathbb{R}) \to L^p(\mathbb{R})$ with constant C, then also

$$f \longmapsto \int_{\mathbb{R}} f(x-t, y-\Gamma(x,t))K(x,t)dt \tag{4.2.14}$$

is bounded as an operator $L^p(\mathbb{R}^2) \to L^p(\mathbb{R}^2)$ with constant C.

Proof. Take the L^p norm of the right hand side of (4.2.14). For fixed x consider the quantity

$$\left(\int_{\mathbb{R}}\left|\int_{\mathbb{R}} f(x-t, y-\Gamma(x,t))K(x,t)dt\right|^p dy\right)^{1/p}. \tag{4.2.15}$$

Applying Minkowski's integral inequality we bound this by

$$\int_{\mathbb{R}} \|f(x-t, y-\Gamma(x,t))\|_{L^p(dy)} K(x,t)dt. \tag{4.2.16}$$

Notice that $\Gamma(x,t)$ is independent of y, hence by a simple change of variable, (4.2.16) is equal to

$$\int_{\mathbb{R}} \|f(x-t, y)\|_{L^p(dy)} K(x,t)dt. \tag{4.2.17}$$

Now take the L^p norm in x and apply the hypothesis (4.2.13) to $g(x) = \|f(x,\cdot)\|_p$. This concludes the proof of Lemma 4.2.1. □

By Minkowski's inequality, the L^p norm of (4.2.12) is no greater than

$$\int_m^{m+1} \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} |f(x-t, y-v(x)t-z)| u(x)|t|^\alpha \phi_0(u(x)^{1/\alpha}t) \frac{dt}{|t|} \right|^p dydx \right)^{1/p} dz. \tag{4.2.18}$$

Defining

$$K(x,t) := u(x)|t|^\alpha \phi_0(u(x)^{1/\alpha}t) \frac{1}{|t|} \tag{4.2.19}$$

we see that the operator in (4.2.13) is dominated by the Hardy-Littlewood maximal operator, which is bounded on $L^p(\mathbb{R})$ for all $p > 1$. Hence (4.2.18) is bounded by (a constant multiple of) $\|f\|_p$. This completes the proof for the first summand on the right hand side of (4.2.5).

4.2.2 High frequency part

Here we handle the second summand on the right hand side of (4.2.5). By the triangle inequality it suffices to prove that there exists $\gamma > 0$ such that

$$\left\| \int_{\mathbb{R}} f(x-t, y-v(x)t-u(x)[t]^\alpha) \psi_l(u(x)^{1/\alpha}t) \frac{dt}{t} \right\|_p \lesssim 2^{-\gamma l} \|f\|_p. \tag{4.2.20}$$

By Lemma 4.2.1, (4.2.20) holds for all $p > 1$ without the exponentially decaying factor $2^{-\gamma l}$.

Hence by interpolation it suffices to prove (4.2.20) for $p = 2$. This will be the goal of the present subsection.

To proceed, we apply a partial Fourier transform to the left hand side of (4.2.20) in the y-variable. In view of Plancherel's theorem, (4.2.20) for $p = 2$ is equivalent to

$$\left\| \int_{\mathbb{R}} g_\eta(x-t) e^{iv(x)\eta t + iu(x)\eta [t]^\alpha} \psi_l(u(x)^{1/\alpha}t) \frac{dt}{t} \right\|_{L^2(dxd\eta)} \lesssim 2^{-\gamma l} \|f\|_2, \tag{4.2.21}$$

where

$$g_\eta(x) := \int_{\mathbb{R}} e^{-i\eta y} f(x,y) dy \tag{4.2.22}$$

denotes the Fourier transform of the function f in its second variable. By Fubini's theorem this reduces to proving the following one-dimensional estimate:

$$\left\| \int_{\mathbb{R}} g(x-t) e^{iv(x)t + iu(x)[t]^\alpha} \psi_l(u(x)^{1/\alpha} t) \frac{dt}{t} \right\|_2 \lesssim 2^{-\gamma l} \|g\|_2 \tag{4.2.23}$$

for all g in $L^2(\mathbb{R})$.

Observe that the phase function contains a linear term. This amounts to providing L^2 bounds for a modulation invariant operator with a special polynomial phase.

Given the shape of the operator one might guess that one has to use the time-frequency approach implemented by the third author in [Lie09], [Lie11]. However, using crucially that $\alpha \neq 1, 2$, one can rely exclusively on TT^* arguments in order to prove (4.2.23). A similar idea first appeared in a slightly simpler context in [GPRY16] by Pierce, Yung, the first author and the fourth author.

The first step to prove (4.2.23) is to decompose the integral inside the norm on the left hand side of (4.2.23) into regions where t is either positive or negative. Both parts are treated in the same way, so we only detail the estimate for the positive part. Accordingly, we denote

$$Tg(x) := \int_0^\infty g(x-t) e^{iv(x)t + iu(x)t^\alpha} \psi_l(u(x)^{1/\alpha} t) \frac{dt}{t}. \tag{4.2.24}$$

Then we have

$$TT^* g(y) = \int_{\mathbb{R}} (\Phi^l_{u(y),v(y)} * \tilde{\Phi}^l_{u(x),v(x)})(y-x) g(x) dx, \tag{4.2.25}$$

where here we set

$$\Phi^l_{u,v}(\xi) := e^{iv\xi + iu\xi^\alpha} \frac{\psi(2^{-l} u^{1/\alpha} \xi)}{\xi} \quad \text{and} \quad \tilde{\Phi}^l_{u,v}(\xi) := \overline{\Phi^l_{u,v}(-\xi)}, \tag{4.2.26}$$

with $\psi(\xi) := \psi_0(\xi)\chi_{(0,\infty)}(\xi)$. The kernel of TT^* satisfies

$$\begin{aligned} &|\Phi^l_{u(y),v(y)} * \tilde{\Phi}^l_{u(x),v(x)}|(\xi) \\ &= \left| \int_{\mathbb{R}} e^{i(v(y)-v(x))\eta + iu(y)\eta^\alpha - iu(x)(\eta-\xi)^\alpha} \frac{\psi(2^{-l}u(y)^{1/\alpha}\eta)}{\eta} \frac{\psi(2^{-l}u(x)^{1/\alpha}(\eta-\xi))}{\eta-\xi} d\eta \right|. \end{aligned} \tag{4.2.27}$$

Let us assume for the moment that $u(x) \le u(y)$. Denoting

$$h := \left(\frac{u(x)}{u(y)} \right)^{1/\alpha} \quad \text{and} \quad a := 2^{-l} u(x)^{1/\alpha}, \tag{4.2.28}$$

via the change of variables

$$2^{-l} u(y)^{1/\alpha} \eta \to \eta, \tag{4.2.29}$$

we see that (4.2.27) equals to

$$a \left| \int_{\mathbb{R}} e^{iw\eta + i2^{\alpha l}\eta^\alpha - i2^{\alpha l}(h\eta - a\xi)^\alpha} \frac{\psi(\eta)}{\eta} \frac{\psi(h\eta - a\xi)}{h\eta - a\xi} d\eta \right| \tag{4.2.30}$$

where w is some quantity depending on x, y and l, the value of which will be irrelevant to us. In the case $u(y) \le u(x)$ we interchange the roles of $u(x)$ and $u(y)$.

To finish this argument we use the following oscillatory integral estimate.

Lemma 4.2.2. *As above, assume $\alpha \notin \{1,2\}$ and let ψ be a smooth function supported in $[1/2,2]$. Then, there exists $\lambda > 0$ such that for all $\xi, w \in \mathbb{R}$, $0 < h \le 1$ and $l > 0$ we have*

$$\begin{aligned} &\left| \int_{\mathbb{R}} e^{iw\eta + i2^{\alpha l}\eta^\alpha - i2^{\alpha l}(h\eta-\xi)^\alpha} \frac{\psi(\eta)}{\eta} \frac{\psi(h\eta-\xi)}{h\eta-\xi} d\eta \right| \\ &\lesssim \mathbb{1}_{[-2^{-\lambda l}, 2^{-\lambda l}]}(\xi) + 2^{-\lambda l}\mathbb{1}_{[-2,2]}(\xi). \end{aligned} \tag{4.2.31}$$

We postpone the proof to the end of this section. Using the lemma we deduce that

$$|\Phi^l_{u(y),v(y)} * \tilde{\Phi}^l_{u(x),v(x)}|(\xi) \lesssim \sum_{i=1,2} (a_i \mathbb{1}_{[-2^{-\lambda l}, 2^{-\lambda l}]}(a_i\xi) + 2^{-\lambda l} a_i \mathbb{1}_{[-2,2]}(a_i \xi)),$$

where $a_1 := 2^{-l}u(y)^{1/\alpha}, a_2 := 2^{-l}u(x)^{1/\alpha}$. Therefore we have

$$\begin{aligned} |\langle TT^* g, h\rangle| &\le \int_{\mathbb{R}}\int_{\mathbb{R}} |(\Phi^l_{u(y),v(y)} * \tilde{\Phi}^l_{u(x),u(y)})(y-x) g(x) h(y)| dx dy \\ &\lesssim 2^{-\lambda l}\left(\int_{\mathbb{R}}\int_{\mathbb{R}} Mg(x)|h(y)| dx dy + \int_{\mathbb{R}}\int_{\mathbb{R}} |g(x)| Mh(y) dx dy \right) \\ &\lesssim 2^{-\lambda l} \|g\|_2 \|h\|_2. \end{aligned} \tag{4.2.32}$$

Here M denotes the Hardy-Littlewood maximal function and we have used its L^2 boundedness as well as the Cauchy-Schwarz inequality in the last step. This concludes the proof of (4.2.23).

Proof of Lemma 4.2.2. Denote the left hand side of (4.2.31) by I_ξ. First note that $I_\xi = 0$ if $|\xi| > 2$. Next, if $|\xi| \le 2^{-\lambda l}$, then the estimate follows from the triangle inequality and so we also assume that $|\xi| > 2^{-\lambda l}$. In the following we consider only η such that the integrand in the integral defining I_ξ is not zero. This implies that $\eta, h\eta - \xi \in [1/2, 2]$. We analyze the phase function

$$Q_\xi(\eta) := w\eta + 2^{\alpha l}(\eta^\alpha - (h\eta - \xi)^\alpha). \tag{4.2.33}$$

Note that

$$Q''_\xi(\eta) = \alpha(\alpha-1)2^{\alpha l}(\eta^{\alpha-2} - h^2(h\eta-\xi)^{\alpha-2}), \tag{4.2.34}$$

$$Q'''_\xi(\eta) = \alpha(\alpha-1)(\alpha-2)2^{\alpha l}(\eta^{\alpha-3} - h^3(h\eta-\xi)^{\alpha-3}), \tag{4.2.35}$$

and observe that the vector $X := \begin{pmatrix} Q''_\xi(\eta) \\ Q'''_\xi(\eta) \end{pmatrix}$ can be written as

$$\alpha(\alpha-1)2^{\alpha l} \begin{pmatrix} 1 & 1 \\ (\alpha-2)\eta^{-1} & (\alpha-2)h(h\eta-\xi)^{-1} \end{pmatrix} \begin{pmatrix} \eta^{\alpha-2} \\ -h^2(h\eta-\xi)^{\alpha-2} \end{pmatrix}. \tag{4.2.36}$$

This point of the argument crucially depends on the hypothesis $\alpha \neq 2$. Denoting the

2×2 matrix in the above expression by M we calculate

$$|\det(M)| = \frac{|(\alpha-2)\xi|}{(h\eta-\xi)\eta} \gtrsim |\xi| > 2^{-\lambda l}. \tag{4.2.37}$$

This allows us to estimate

$$|X| \gtrsim 2^{(\alpha-\lambda)l}. \tag{4.2.38}$$

Invoking van der Corput's lemma [Ste93, Chapter VIII.1] we conclude that

$$I_\xi \lesssim 2^{-(\alpha-\lambda)/3l} = 2^{-\lambda l}\,, \tag{4.2.39}$$

where we have set $\lambda := \frac{\alpha}{4}$. □

4.3 Proof of Theorem 4.1.2

Throughout this section we omit the dependence on u, α and simply refer to our operator as $\mathcal{H}$. Recalling the commutation relation[4]

$$\mathcal{H}P_k^{(2)} = P_k^{(2)}\mathcal{H}\,, \tag{4.3.1}$$

by Littlewood-Paley theory it suffices to prove that

$$\Big\|\Big(\sum_{k\in\mathbb{Z}} |\mathcal{H}P_k^{(2)} f|^2\Big)^{1/2}\Big\|_p \lesssim \|f\|_p\,. \tag{4.3.2}$$

We stress here that the one-variable assumption (4.1.5) is the key fact that guarantees the commutation relation (4.3.1). This is the only place in this section where the one-variable assumption (4.1.5) is explicitly used. An implicit appearance is in the estimate (4.3.10) for the case $p = 2$, which is the content of the previous section.

We return to the proof of (4.3.2). In Section 4.2 we already established that

$$\|\mathcal{H}P_k^{(2)} f\|_p \lesssim \|P_k^{(2)} f\|_p$$

holds (this is the case $v \equiv 0$). Here we should note that the proof in Section 4.2 needs a small modification in the case when $v \equiv 0$ and $\alpha = 2$. Namely, in that situation the exponential decay estimate (4.2.20) is essentially a special case of a well-known result due to Stein and Wainger (see [SW01, Theorem 1]).

Fix now $k \in \mathbb{Z}$. In view of the shape of the phase of our multiplier, we decompose our operator into a low and high frequency component, respectively:

$$\begin{aligned}\mathcal{H}P_k^{(2)} f(x,y) &= \sum_{l\in\mathbb{Z}} \int_{\mathbb{R}} (P_k^{(2)} f)(x-t, y-u(x)[t]^\alpha)\psi_l(u(x)^{1/\alpha}t)\frac{dt}{t} \\ &= \Big(\sum_{l\le -k/\alpha} + \sum_{l>-k/\alpha}\Big)\int_{\mathbb{R}} (P_k^{(2)} f)(x-t, y-u(x)[t]^\alpha)\psi_l(u(x)^{1/\alpha}t)\frac{dt}{t}.\end{aligned} \tag{4.3.3}$$

[4] Recall that $P_k^{(2)}$ denotes a Littlewood-Paley projection in the second variable.

We denote

$$T_{k,0}f(x,y) := \sum_{l \le -k/\alpha} \int_{\mathbb{R}} f(x-t, y-u(x)[t]^\alpha)\psi_l(u(x)^{1/\alpha}t)\frac{dt}{t} \tag{4.3.4}$$

and, for $j \geq 1$,

$$T_{k,j}f(x,y) := \int_{\mathbb{R}} f(x-t, y-u(x)[t]^\alpha)\psi_{-\frac{k}{\alpha}+j}(u(x)^{1/\alpha}t)\frac{dt}{t}. \tag{4.3.5}$$

Using the triangle inequality we obtain

$$\Big\|\Big(\sum_{k\in\mathbb{Z}} |\mathcal{H}P_k^{(2)}f|^2\Big)^{1/2}\Big\|_p \lesssim \Big\|\Big(\sum_{k\in\mathbb{Z}} |T_{k,0}P_k^{(2)}f|^2\Big)^{1/2}\Big\|_p + \sum_{j\in\mathbb{N}} \Big\|\Big(\sum_{k\in\mathbb{Z}} |T_{k,j}P_k^{(2)}f|^2\Big)^{1/2}\Big\|_p. \tag{4.3.6}$$

As in the previous section (see (4.2.8) – (4.2.18)) we treat $T_{k,0}P_k^{(2)}f$ as a perturbation of

$$\sum_{l \le -k/\alpha} \int_{\mathbb{R}} P_k^{(2)}f(x-t,y)\psi_l(u(x)^{1/\alpha}t)\frac{dt}{t}. \tag{4.3.7}$$

This yields

$$|T_{k,0}P_k^{(2)}f| \lesssim M_S(P_k^{(2)}f) + H^*(P_k^{(2)}f). \tag{4.3.8}$$

Here M_S denotes the strong maximal function and H^* a maximally truncated Hilbert transform applied in the first variable. Indeed, one may deduce (4.3.8) using the same arguments as in Section 4.2.1. The vector-valued estimates for M_S follow from the corresponding estimates for the one dimensional Hardy-Littlewood maximal function which are well-known (see Stein [Ste93, Chapter II.1]). Similarly, the vector-valued estimates for H^* follow from Cotlar's inequality and the vector-valued estimates for the Hilbert transform and the maximal function. Thus we have

$$\Big\|\Big(\sum_{k\in\mathbb{Z}} |T_{k,0}P_k^{(2)}f|^2\Big)^{1/2}\Big\|_p \lesssim \Big\|\Big(\sum_{k\in\mathbb{Z}} |P_k^{(2)}f|^2\Big)^{1/2}\Big\| \lesssim \|f\|_p \tag{4.3.9}$$

for all $p > 1$. This finishes the proof for the first term on the right hand side of (4.3.6).

To bound the second term in (4.3.6) we will prove that there exists $\gamma_p > 0$ such that

$$\Big\|\Big(\sum_{k\in\mathbb{Z}} |T_{k,j}P_k^{(2)}f|^2\Big)^{1/2}\Big\|_p \lesssim 2^{-\gamma_p j}\|f\|_p. \tag{4.3.10}$$

For $p = 2$ this follows from (4.2.20) with $v \equiv 0$. Note here again that in the case $v \equiv 0$, $\alpha = 2$ the estimate (4.2.20) is a consequence of [SW01, Theorem 1]. Hence, by interpolation it suffices to prove

$$\Big\|\Big(\sum_{k\in\mathbb{Z}} |T_{k,j}P_k^{(2)}f|^2\Big)^{1/2}\Big\|_p \lesssim j^4\|f\|_p. \tag{4.3.11}$$

Let us first note that if p is sufficiently close to 2, then (4.3.11) follows immediately from an interpolation argument. To carry out this interpolation, we observe the trivial pointwise bound

$$|T_{k,j}P_k^{(2)}f| \lesssim 2^{\alpha j} M_S(P_k^{(2)}f). \tag{4.3.12}$$

This implies that

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} |T_{k,j}P_k^{(2)}f|^2 \Big)^{1/2} \Big\|_p \lesssim 2^{\alpha j}\|f\|_p \tag{4.3.13}$$

for all $p > 1$. Interpolating with the bound for $p = 2$ in (4.3.10), we can find a positive constant $\varepsilon_0 > 0$ such that (4.3.10) holds true for all $p \in (2-\varepsilon_0, 2+\varepsilon_0)$.

Recall that our goal is to prove (4.3.11) for all $p > 1$. For convenience we choose to present only the case $[t]^\alpha = |t|^\alpha$. The other case $[t]^\alpha = \operatorname{sgn}(t)|t|^\alpha$ can be treated by the same arguments.

Our strategy is to derive a sufficiently fine-grained pointwise estimate of $|T_{k,j}P_k^{(2)}f|$ by appropriate shifted maximal functions and then apply vector-valued bounds to conclude (4.3.11). The use of the shifted maximal operator was inspired by the work of the third author [Lie15b] (see there Section 2.4., Lemma 2).

First let us consider the case $k = 0$. By definition we have

$$T_{0,j}P_0^{(2)}f(x,y) = \iint_{\mathbb{R}^2} f(x-t, y-s-u(x)|t|^\alpha)\frac{\psi_j(u(x)^{\frac{1}{\alpha}}t)}{t}\check{\psi}(s)dtds. \tag{4.3.14}$$

Up to Schwartz tails in s, this is essentially an average of f over a thickened segment of a translate of the curve $(t, u(x)|t|^\alpha)$. The idea is to cut up this thickened curve segment into pieces that are well approximated by rectangles.

Taking absolute values and using the triangle inequality we see that the previous display is

$$\lesssim \frac{1}{\lambda_{x,j}}\int_{|t|\approx\lambda_{x,j}}\int_{\mathbb{R}} |f(x-t, y-s-u(x)|t|^\alpha)\check{\psi}(s)|dsdt, \tag{4.3.15}$$

where $\lambda_{x,j} := 2^j u(x)^{-1/\alpha}$ and the notation $|t| \approx \lambda$ means $\frac{1}{2}\lambda \le |t| \le 2\lambda$. By the rapid decay of $\check{\psi}$ this is

$$\lesssim \sum_{\tau\in\mathbb{Z}} \frac{1}{(1+|\tau|)^{10}}\frac{1}{\lambda_{x,j}}\int_{|t|\approx\lambda_{x,j}}\int_\tau^{\tau+1} |f(x-t, y-s-u(x)|t|^\alpha)|dsdt. \tag{4.3.16}$$

Once at this point, the intuition is given by the following observation: the function $f = P_0^{(2)}f$ "sees" the $y-$universe in unit steps; that is, f is morally $y-$constant on segments of length one. Consequently, it is natural to further discretize the location of t in intervals on which the variation of the term $u(x)|t|^\alpha$ does not exceed the order of one. This invites us to consider the following construction.

Set $\delta_{x,j} := 2^{-(\alpha-1)j}u(x)^{-1/\alpha}$ and cover the region $\frac{1}{2}\lambda_{x,j} \le |t| \le 2\lambda_{x,j}$ by intervals $\{I_m\}_{m=0}^{N_j-1}$ where

$$I_m = \Big\{t : \frac{1}{2}\lambda_{x,j} + m\delta_{x,j} \le |t| \le \frac{1}{2}\lambda_{x,j} + (m+1)\delta_{x,j}\Big\}$$

and $N_j \in \mathbb{N}$ is such that $\frac{3}{2}\lambda_{x,j} \le N_j\delta_{x,j} \le 2\lambda_{x,j}$. Notice that $N_j \sim 2^{\alpha j}$ and moreover that N_j can be chosen independently of x.

With this we have

$$\frac{1}{\lambda_{x,j}}\int_{|t|\approx\lambda_{x,j}}\int_{\tau}^{\tau+1}|f(x-t,y-s-u(x)|t|^{\alpha})|dsdt$$

$$\lesssim \frac{1}{N_j}\sum_{m=0}^{N_j-1}\frac{1}{|I_m|}\int_{I_m}\int_0^1 |f(x-t,y-s-\tau-u(x)|t|^{\alpha})|dsdt. \tag{4.3.17}$$

We now set

$$\mathbf{C}_m := \Big\{(t, s+\tau+u(x)|t|^{\alpha}) \,:\, t\in I_m,\, s\in[0,1]\Big\} \subset I_m\times J_m,$$

where

$$J_m = \left[\tau+u(x)\Big(\frac{1}{2}\lambda_{x,j}+m\delta_{x,j}\Big)^{\alpha}, \tau+1+u(x)\Big(\frac{1}{2}\lambda_{x,j}+(m+1)\delta_{x,j}\Big)^{\alpha}\right].$$

Notice that, because of our choice of $\delta_{x,j}$, the thickened curve segment $\mathbf{C}_m$ is contained in a rectangular region of comparable area. Indeed, we have by the mean value theorem,

$$|J_m| \sim_{\alpha} 1+u(x)\delta_{x,j}\lambda_{x,j}^{\alpha-1} \sim_{\alpha} 1.$$

Thus we further have that (4.3.17) is bounded by a constant multiple of

$$\frac{1}{N_j}\sum_{m=0}^{N_j-1}\frac{1}{|I_m\times J_m|}\iint_{I_m\times J_m}|f(x-t,y-s)|dtds.$$

Given a non-negative parameter σ, we define the *shifted maximal operator* as

$$M^{(\sigma)}f(z) := \sup_{z\in I\subset\mathbb{R}}\frac{1}{|I|}\int_{I^{(\sigma)}}|f(\zeta)|d\zeta. \tag{4.3.18}$$

Here the supremum goes over all bounded intervals I containing z, and $I^{(\sigma)}$ denotes a shift of the interval $I=[a,b]$ given by

$$I^{(\sigma)} := [a-\sigma\cdot|I|, b-\sigma\cdot|I|]\cup[a+\sigma|I|, b+\sigma|I|].$$

Note that

$$\frac{1}{|I_m\times J_m|}\iint_{I_m\times J_m}|f(x-t,y-s)|dtds \le M_1^{(\sigma_m^{(1)})}M_2^{(\sigma_m^{(2)}+\tau)}f(x,y),$$

where

$$\begin{cases} \sigma_m^{(1)} & := 2^{\alpha j-1}+m, \\ \sigma_m^{(2)} & := c_{\alpha}(2^j+2^{-(\alpha-1)j}m)^{\alpha}, \end{cases}$$

and c_α is a constant only depending on α and $M_1^{(n)}$ (respectively, $M_2^{(n)}$) denotes the shifted maximal operator applied in the first (respectively, second) variable. Notice that since $N_j\sim 2^{\alpha j}$ and $m<N_j$ we have that $\sigma_m^{(i)}\lesssim 2^{\alpha j}$ for $i=1,2$.

Altogether we have now proved that

$$|T_{0,j}P_0^{(2)}f(x,y)| \lesssim \sum_{\tau\in\mathbb{Z}} \frac{1}{(1+|\tau|)^{10}} \frac{1}{N_j} \sum_{m=0}^{N_j-1} M_1^{(\sigma_m^{(1)})} M_2^{(\sigma_m^{(2)}+\tau)} f(x,y).$$

By a scaling argument
we have that for all $k \in \mathbb{Z}$ the following holds:

$$|T_{k,j}P_k^{(2)}f(x,y)| \lesssim \sum_{\tau\in\mathbb{Z}} \frac{1}{(1+|\tau|)^{10}} \frac{1}{N_j} \sum_{m=0}^{N_j-1} M_1^{(\sigma_m^{(1)})} M_2^{(\sigma_m^{(2)}+\tau)} P_k^{(2)} f(x,y). \tag{4.3.19}$$

Inserting these two bounds into the left hand side of (4.3.11) yields

$$\Big\|\Big(\sum_{k\in\mathbb{Z}}\Big(\sum_{\tau\in\mathbb{Z}} \frac{1}{(1+|\tau|)^{10}} \frac{1}{N_j} \sum_{m=0}^{N_j-1} M_1^{(\sigma_m^{(1)})} M_2^{(\sigma_m^{(2)}+\tau)} P_k^{(2)} f(x,y)\Big)^2\Big)^{1/2}\Big\|_p. \tag{4.3.20}$$

By the triangle inequality this is no greater than

$$\sum_{\tau\in\mathbb{Z}} \frac{1}{(1+|\tau|)^{10}} \frac{1}{N_j} \sum_{m=0}^{N_j-1} \Big\|\Big(\sum_{k\in\mathbb{Z}}\Big(M_1^{(\sigma_m^{(1)})} M_2^{(\sigma_m^{(2)}+\tau)} P_k^{(2)} f(x,y)\Big)^2\Big)^{1/2}\Big\|_p. \tag{4.3.21}$$

Thus, to show (4.3.11) it suffices to prove that

$$\Big\|\Big(\sum_{k\in\mathbb{Z}}\Big(M_1^{(\sigma_m^{(1)})} M_2^{(\sigma_m^{(2)}+\tau)} P_k^{(2)} f(x,y)\Big)^2\Big)^{1/2}\Big\|_p \lesssim j^4 \cdot \log^2(2+|\tau|)\|f\|_p. \tag{4.3.22}$$

Since $\sigma_m^{(i)} \lesssim 2^{\alpha j}$ for $i = 1, 2$, by Fubini's theorem we have that (4.3.22) is a consequence of the following vector-valued estimate for the one-dimensional shifted maximal operator:

$$\left\|\Big(\sum_{k\in\mathbb{Z}} |M^{(n)} f_k|^2\Big)^{1/2}\right\|_p \lesssim (\log\langle n\rangle)^2 \Big\|\Big(\sum_{k\in\mathbb{Z}} |f_k|^2\Big)^{1/2}\Big\|_p, \tag{4.3.23}$$

where we adopt the Japanese bracket notation $\langle n\rangle := 2 + |n|$. We give the proof of this last statement below.

Let $\mathcal{D}$ denote the set of dyadic intervals $I = [2^k m, 2^k(m+1))$ with $k, m \in \mathbb{Z}$. In accordance with the above definition we have,

$$I^{(n)} = [2^k(m-n), 2^k(m-n+1)) \in \mathcal{D}$$

for $n \in \mathbb{Z}$. For simplicity we discuss only the dyadic variant of $M^{(n)}$, defined as

$$M^{(n)} f(x) = \sup_{x\in I\in\mathcal{D}} \frac{1}{|I|} \int_{I^{(n)}} |f(y)|dy.$$

Everything here carries over to the non-dyadic version with the constants having the same dependence on n (see Muscalu [Mus14, p. 741]).

Theorem 4.3.1. *For $1 < p < \infty, 1 < q \le \infty$ we have*

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} |M^{(n)} f_k|^q \Big)^{1/q} \Big\|_p \lesssim (\log\langle n\rangle)^2 \Big\| \Big(\sum_{k\in\mathbb{Z}} |f_k|^q \Big)^{1/q} \Big\|_p. \tag{4.3.24}$$

We did not find a reference for this result in the literature, so we provide a short self-contained proof. The scalar version of this estimate involves standard Calderón-Zygmund techniques and can be found in [Mus14], where the author attributes it to Stein [Ste93].

Before starting the proof we make few more observations: firstly, notice that the endpoints $1 < p = q \le \infty$ and $q = \infty$ of Theorem 4.3.1 follow immediately from the scalar version, and thus interpolation establishes the result for $1 < p \le q < \infty$. Secondly, the exponent 2 for the log loss is only chosen for convenience; the proof actually gives a slightly better exponent.

Now the proof we present below relies on a weighted estimate in the spirit of Fefferman-Stein [FS71]:

Lemma 4.3.2. *Let $\omega \ge 0$ be a locally integrable function. For all $\lambda > 0$,*

$$\omega(\{x : M^{(n)} f(x) > \lambda\}) \lesssim \lambda^{-1} \Big(\log\langle n\rangle \|f\|_{L^1(M^{(-n)}\omega)} + \|f\|_{L^1(M\omega)} \Big), \tag{4.3.25}$$

where M denotes the Hardy-Littlewood maximal function.

Proof. Fix $\lambda > 0$. Let $\mathcal{I}$ be the collection of maximal dyadic intervals I such that

$$\frac{1}{|I|} \int_I |f| > \lambda. \tag{4.3.26}$$

Given $I \in \mathcal{I}$ we denote by $\mathcal{J}_i^I$ the collection of dyadic intervals J such that $|J| = 2^{-i}|I|$, $J^{(n)} \subset I$ and $\frac{1}{|J|}\int_{J^{(n)}} |f| > \lambda$. This is the i-th generation of shifted subintervals of I. We call i *large* if $2^{-i}|n| < 1$ and otherwise we call i *small*. It will be important that this depends only on n. Denote $\mathcal{J}^I = \bigcup_{i\ge 0} \mathcal{J}_i^I$. Observe that

$$\{x : M^{(n)} f(x) > \lambda\} \subset \bigcup_{I\in\mathcal{I}} \bigcup_{J\in\mathcal{J}^I} J. \tag{4.3.27}$$

Indeed, if x is such that $M^{(n)} f(x) > \lambda$ then there exists $J \in \mathcal{D}$ with $x \in J$ and $\frac{1}{|J|}\int_{J^{(n)}} |f| > \lambda$. By definition of $\mathcal{I}$ there is some $I \in \mathcal{I}$ with $J^{(n)} \subset I$ and therefore $x \in J \in \mathcal{J}^I$. The crucial observation is that if $J \in \mathcal{J}_i^I$ and i is large, then J is contained in $3I$.

Thus we can estimate

$$\omega(\{x : M^{(n)} f(x) > \lambda\}) \le \sum_{I\in\mathcal{I}} \Big(\sum_{\substack{J\in\mathcal{J}_i^I, \\ i \text{ small}}} \int_J \omega + \int_{3I} \omega \Big). \tag{4.3.28}$$

For $J \in \mathcal{J}_i^I$, since $(J^{(n)})^{(-n)} = J$ we have that

$$\int_J \omega \le \frac{1}{\lambda} \int_{J^{(n)}} |f(x)| \cdot \Big(\frac{1}{|J|} \int_J \omega(y) dy \Big) dx \le \frac{1}{\lambda} \int_{J^{(n)}} |f(x)| M^{(-n)} \omega(x) dx. \tag{4.3.29}$$

Similarly,

$$\int_{3I} \omega \lesssim \frac{1}{\lambda} \int_I |f(x)| M\omega(x) dx. \tag{4.3.30}$$

Thus we can estimate (4.3.28) by

$$\lambda^{-1}\Big(\sum_{I\in\mathcal{I}} \sum_{\substack{J\in\mathcal{J}_i^I, \\ i \text{ small}}} \int_{J^{(n)}} |f| M^{(-n)}\omega + \sum_{I\in\mathcal{I}} \int_I |f| M\omega \Big). \tag{4.3.31}$$

For fixed i the $J \in \mathcal{J}_i^I$ are disjoint and we have $\bigcup_{J\in\mathcal{J}_i^I} J^{(n)} \subset I$. Since there are about $\log\langle n\rangle$ small i, the previous display is bounded by

$$\lambda^{-1} \left(\log\langle n\rangle \|f\|_{L^1(M^{(-n)}\omega)} + \|f\|_{L^1(M\omega)} \right). \tag{4.3.32}$$

This finishes the proof of Lemma 4.3.2. □

Now let us treat the range $p \geq q$. Lemma 4.3.2 says that $M^{(n)}$ is a bounded operator $L^1(\widetilde{M}^{(-n)}\omega) \to L^{1,\infty}(\omega)$ with constant $\lesssim \log\langle n\rangle$ where $\widetilde{M}^{(n)} = M^{(n)} + M$. By interpolation with the trivial L^∞ bound we get

$$\int (M^{(n)} f_k)^q \omega \lesssim \log\langle n\rangle \int |f_k|^q \widetilde{M}^{(-n)}\omega \tag{4.3.33}$$

for all $1 < q < \infty$. Let r be the dual exponent of p/q. We have

$$\Big\| \Big(\sum_{k\in\mathbb{Z}} |M^{(n)} f_k|^q \Big)^{1/q} \Big\|_p^q = \Big\| \sum_{k\in\mathbb{Z}} |M^{(n)} f_k|^q \Big\|_{p/q} = \sup_{\|\omega\|_r \leq 1} \int \sum_{k\in\mathbb{Z}} (M^{(n)} f_k)^q \omega. \tag{4.3.34}$$

Using the previous estimate we bound this by

$$\log\langle n\rangle \int \sum_{k\in\mathbb{Z}} |f_k|^q \widetilde{M}^{(-n)}\omega. \tag{4.3.35}$$

By Hölder's inequality and the scalar logarithmic bound for $M^{(n)}$ we have

$$\int \sum_{k\in\mathbb{Z}} |f_k|^q \widetilde{M}^{(-n)}\omega \leq \Big\| \Big(\sum_{k\in\mathbb{Z}} |f_k|^q \Big)^{\frac{1}{q}} \Big\|_p^q \|\widetilde{M}^{(-n)}\omega\|_r \lesssim (\log\langle n\rangle)^{\frac{1}{r}} \Big\| \Big(\sum_{k\in\mathbb{Z}} |f_k|^q \Big)^{\frac{1}{q}} \Big\|_p^q. \tag{4.3.36}$$

This finishes the proof of our theorem.

4.4 Proof of Theorem 4.1.8

In this section we prove Theorem 4.1.8. The main tool we will be using is the local smoothing estimate from Theorem 4.6.1.

We start the proof by recalling that

$$\mathcal{H}_u^{(\alpha)} f(x,y) := \int_{\mathbb{R}} f(x-t, y-u(x,y)[t]^\alpha) \frac{dt}{t}. \tag{4.4.1}$$

We will only present the proof of the case $\alpha > 1$; the remaining case $0 < \alpha < 1$ can be treated using the same methods and is somewhat easier. As α and u will always be fixed, we will leave out the dependence on them in our notation and simply use T to denote $\mathcal{H}_u^{(\alpha)}$. In this section, we will prove

$$\|TP_k^{(2)}f\|_p \lesssim \|P_k^{(2)}f\|_p, \tag{4.4.2}$$

for all $p > 2$ and all measurable function $u : \mathbb{R}^2 \to \mathbb{R}$, with a bound independent of $k \in \mathbb{Z}$ and u. By the anisotropic scaling

$$x \to x, y \to \lambda y, \tag{4.4.3}$$

it suffices to prove (4.4.2) for $k = 0$.

In order to simplify our presentation, we introduce some notation. We let

$$z = (x, y) \quad \text{and} \quad u_z := u(x, y), \tag{4.4.4}$$

and set v_z to be the unique integer such that

$$2^{v_z} \leq u_z < 2^{v_z+1}. \tag{4.4.5}$$

For a given $k_0 \in \mathbb{Z}$, define

$$u_z^{(k_0)} := 2^{k_0 - v_z} u_z. \tag{4.4.6}$$

Observe that $u_z^{(v_z)} = u_z$, $u_z^{(k_0)} \in [2^{k_0}, 2^{k_0+1})$ and $u_z^{(k_0)} = 2^{k_0} u_z^{(0)}$.

Denote

$$T_{k_0} f(x, y) := \int_{\mathbb{R}} f(x - t, y - u_z^{(k_0)}[t]^\alpha) \frac{dt}{t}. \tag{4.4.7}$$

Remark 4.4.1. Roughly speaking, we will bound $TP_0^{(2)}f$ by the "square function"

$$\Big(\sum_{k_0 \in \mathbb{Z}} |T_{k_0} P_0^{(2)} f|^p \Big)^{1/p}. \tag{4.4.8}$$

Here we are using an l^p sum instead of an l^2 sum. This is because p is always larger than two. At first glance, it might be a bit surprising that the term (4.4.8) is still a bounded operator. The proof of this fact will be achieved by applying a finer decomposition for the function f, and then seeking for enough "off-diagonal" decay via a local smoothing estimate.

We now begin our analysis by performing a dyadic decomposition of the kernel $\frac{1}{t}$ around the singularity $t = 0$. In particular, let

$$T_{k_0, l} f(x, y) := \int_{\mathbb{R}} f(x - t, y - u_z^{(k_0)}[t]^\alpha) \psi_l((u_z^{(0)})^\beta t) \frac{dt}{t}, \tag{4.4.9}$$

where $\beta = \frac{1}{\alpha - 1}$. Recall that $\alpha > 1$ and so β is always positive. The motivation for using the factor $(u_z^{(0)})^\beta$ and the choice of β will become clear much later during the main argument (see the proof of Lemma 4.4.1, specifically (4.4.31)).

Next, we perform a Littlewood-Paley decomposition in the x-variable, and write

$$T_{k_0} P_0^{(2)} f = \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} T_{k_0,l} P_k^{(1)} P_0^{(2)} f. \tag{4.4.10}$$

We will split the sums in $l, k \in \mathbb{Z}$ into two major cases according to the behavior of the phase of the "multiplier" (see the discussion preceding (4.4.11) below) associated with $T_{k_0,l}$:

1. **low frequency case**: $l < \max\{-k, -k_0/\alpha\}$;

   In this situation the operator $T_{k_0,l}$ behaves like a one-dimensional convolution operator.

2. **high frequency case**: $l \geq \max\{-k, -k_0/\alpha\}$;

   This case splits into two subcases:

   (a) l sits below the critical point of the phase;

   (b) l sits above the critical point of the phase.

In what follows we explain the heuristic for the above partition of our analysis.

Fix $l, k \in \mathbb{Z}$ and focus on the function $T_{k_0,l} P_k^{(1)} f$. Imagine for the moment that the function $u_z^{(k_0)}$ is a constant $u^{(k_0)}$ on the interval $[2^{k_0}, 2^{k_0+1})$. Then $T_{k_0,l} P_k^{(1)}$ becomes a convolution type operator, and hence it makes sense to speak about its multiplier as

$$\int_{\mathbb{R}} e^{i 2^l t \xi + i u^{(k_0)} 2^{\alpha l} [t]^\alpha \eta} \psi_0([u_z^{(0)}]^\beta t) \frac{dt}{t}. \tag{4.4.11}$$

Recall that $\xi \sim 2^k$ and $\eta \sim 1$. In the situation described by item (1) either $2^l \xi \lesssim 1$ or $u^{(k_0)} 2^{\alpha l} \eta \lesssim 1$. When one of these two inequalities occurs, say $u^{(k_0)} 2^{\alpha l} \eta \lesssim 1$, then in the expression (4.4.9), we can view $f(x-t, y-u^{(k_0)}[t]^\alpha)$ as a perturbation of $f(x-t, y)$. This is what we meant when saying that the operator $T_{k_0,l}$ behaves like a one-dimensional operator.

Assume now that we are in the situation of item (2), that is $l \geq \max\{-k, -k_0/\alpha\}$. In this case we have two possibilities:

$$k < k_0/\alpha \text{ and } k \geq k_0/\alpha\,. \tag{4.4.12}$$

In the first instance, $k < k_0/\alpha$, the phase function in (4.4.11) does not admit any critical point, which makes this case much easier to handle. This is the reason for which we will only focus on the latter case $k \geq k_0/\alpha$.

Now, as explained above, the cutoff between case (2a) and (2b) is indicated by the stationary phase principle. Analyzing the stationary points we deduce the requirement

$$2^{l+k} \sim 2^{\alpha l + k_0} \implies |l - \beta \cdot (k - k_0)| \lesssim 1\,. \tag{4.4.13}$$

Once at this point, we let the situation in item (2a) be defined by the conditions:

$$k \geq k_0/\alpha \;\text{ and }\; l < \beta \cdot (k - k_0)\,. \tag{4.4.14}$$

while the situation in item (2b) be defined by the conditions:

$$k \geq k_0/\alpha \text{ and } l \geq \beta \cdot (k - k_0)\,. \tag{4.4.15}$$

With this heuristic, based on items (1) and (2) above we split (4.4.10) as

$$\begin{aligned} T_{k_0} P_0^{(2)} f &= \sum_{k\in\mathbb{Z}} \sum_{l\leq \max\{-k,-k_0/\alpha\}} T_{k_0,l} P_k^{(1)} P_0^{(2)} f + \sum_{k\in\mathbb{Z}} \sum_{l> \max\{-k,-k_0/\alpha\}} T_{k_0,l} P_k^{(1)} P_0^{(2)} f \\ &=: I_{k_0} + II_{k_0}. \end{aligned} \tag{4.4.16}$$

These two terms are treated separately in the following two subsections.

4.4.1 The high frequency case

In this subsection, we will treat the term II_{k_0} which is the main term in (4.4.16).

Recall from our heuristic that in the situation $k < k_0/\alpha$ the phase function in (4.4.11) does not admit a critical point. That makes this case easier to handle. The precise arguments are the same as for the main term, so we will not detail them here. That is, we will only treat the case $k \geq k_0/\alpha$. Accordingly we redefine

$$II_{k_0} := \sum_{k\geq k_0/\alpha} \sum_{l>-k_0/\alpha} T_{k_0,l} P_k^{(1)} P_0^{(2)} f. \tag{4.4.17}$$

As $\alpha > 1$, in this situation we always have

$$-k_0/\alpha \leq \beta \cdot (k - k_0).$$

Thus we split II_{k_0} into two parts

$$\begin{aligned} II_{k_0} &= \sum_{k\geq \frac{k_0}{\alpha}} \sum_{l>-k_0/\alpha}^{\beta\cdot(k-k_0)} T_{k_0,l} P_k^{(1)} P_0^{(2)} f + \sum_{k\geq \frac{k_0}{\alpha}} \sum_{l>\beta\cdot(k-k_0)} T_{k_0,l} P_k^{(1)} P_0^{(2)} f \\ &=: II_{k_0}^{(1)} + II_{k_0}^{(2)}\,. \end{aligned} \tag{4.4.18}$$

Our goal here will be to prove that

$$\Big\| \Big(\sum_{k_0\in\mathbb{Z}} |II_{k_0}^{(j)}|^p \Big)^{1/p} \Big\|_p \lesssim \|f\|_p \qquad \text{for } j \in \{1, 2\}\,. \tag{4.4.19}$$

For this, we first apply the change of variables[5] $l \to l + \beta(k - k_0)$, and use Fubini to deduce that

$$II_{k_0}^{(1)} = \sum_{l=-\infty}^{0} \sum_{k>\frac{k_0}{\alpha}-\frac{l}{\beta}} T_{k_0,\beta(k-k_0)+l} P_k^{(1)} P_0^{(2)} f\,, \tag{4.4.20}$$

[5]Note that $\beta(k - k_0)$ might not be an integer but this is irrelevant.

and

$$II^{(2)}_{k_0} = \sum_{l=1}^{\infty} \sum_{k \geq \frac{k_0}{\alpha}} T_{k_0,\beta(k-k_0)+l} P_k^{(1)} P_0^{(2)} f \, . \tag{4.4.21}$$

In order to prove (4.4.19) it will be sufficient to show that for all $p > 2$ we have

$$\Big\| \Big(\sum_{k_0 \in \mathbb{Z}} | \sum_{k > \frac{k_0}{\alpha} - \frac{l}{\beta}} T_{k_0,\beta\cdot(k-k_0)+l} P_k^{(1)} P_0^{(2)} f|^p \Big)^{1/p} \Big\|_p \lesssim 2^{-\gamma_p |l|} \|f\|_p, \tag{4.4.22}$$

and

$$\Big\| \Big(\sum_{k_0 \in \mathbb{Z}} | \sum_{k \geq \frac{k_0}{\alpha}} T_{k_0,\beta\cdot(k-k_0)+l} P_k^{(1)} P_0^{(2)} f|^p \Big)^{1/p} \Big\|_p \lesssim 2^{-\gamma_p l} \|f\|_p, \tag{4.4.23}$$

for some $\gamma_p > 0$. Here and throughout the paper γ_p is a positive constant that is allowed to change from line to line. We claim that for $p > 2$ there exists $\gamma_p > 0$ such that

- if $l \leq 0$ and $k > \frac{k_0}{\alpha} - \frac{l}{\beta}$, then

$$\|T_{k_0,\beta(k-k_0)+l} P_k^{(1)} P_0^{(2)} f\|_p \lesssim 2^{-\gamma_p\cdot(\alpha\beta k - \beta k_0 + l)} \|P_k^{(1)} P_0^{(2)} f\|_p, \text{ and} \tag{4.4.24}$$

- if $l > 0$ and $k \geq \frac{k_0}{\alpha}$, then

$$\|T_{k_0,\beta(k-k_0)+l} P_k^{(1)} P_0^{(2)} f\|_p \lesssim 2^{-\gamma_p\cdot(\alpha\beta k - \beta k_0 + \alpha l)} \|P_k^{(1)} P_0^{(2)} f\|_p. \tag{4.4.25}$$

Before we prove this claim, we will demonstrate how it is used to show (4.4.22) and (4.4.23). Here we only prove (4.4.22); the estimate (4.4.23) follows in essentially the same way.

For a fixed $l \leq 0$, we expand the L^p norm on the left hand side of (4.4.22), and then apply (4.4.24) to obtain

$$\Big(\sum_{k_0 \in \mathbb{Z}} \big(\sum_{k > \frac{k_0}{\alpha} - \frac{l}{\beta}} 2^{-\gamma_p\cdot(\alpha\beta k - \beta k_0 + l)} \|P_k^{(1)} P_0^{(2)} f\|_p \big)^p \Big)^{\frac{1}{p}}.$$

By applying Hölder's inequality to the summation in k, we obtain that the above expression is bounded by

$$\Big(\sum_{k_0 \in \mathbb{Z}} \sum_{k > \frac{k_0}{\alpha} - \frac{l}{\beta}} 2^{-\gamma_p\cdot(\alpha\beta k - \beta k_0 + l)} \|P_k^{(1)} P_0^{(2)} f\|_p^p \Big)^{\frac{1}{p}}.$$

Now we apply Fubini's theorem and exchange the order of summations in k and k_0 to further bound this by

$$2^{\gamma_p\cdot l(\alpha-1)} \big(\sum_{k \in \mathbb{Z}} \|P_k^{(1)} P_0^{(2)} f\|_p^p \big)^{\frac{1}{p}} \lesssim 2^{\gamma_p\cdot l(\alpha-1)} \|f\|_p.$$

This finishes the proof of the desired estimate (4.4.22).

It remains to prove (4.4.24) and (4.4.25). The idea is to reduce to the local smoothing estimate provided in Theorem 4.6.1. We will do this in a slightly more general setting in view of further applications in Section 4.5. To this end, given parameters $u > 0, w, l \in \mathbb{Z}$, we introduce

$$A_{u,w,l}f(z) := \int f(x-t, y-u[t]^\alpha)\psi_l(2^{\frac{w}{\alpha}}(2^{-v}u)^\beta t)\frac{dt}{|t|},$$

where $v \in \mathbb{Z}$ is such that $u \in [2^v, 2^{v+1})$. Note that $2^{-v}u \in [1,2)$. With this notation we have $T_{k_0,l} = A_{2^{k_0}u_z^{(0)},0,l}$. The extra parameter w will only be needed in the Section 4.5.

Lemma 4.4.1. *Let $r = r_z : \mathbb{R}^2 \to [1,2)$ be measurable and $m, k, v, l \in \mathbb{Z}$ such that*

$$M := \max\{k + l - \frac{w}{\alpha}, m + l\alpha - w + v\} \geq 0. \tag{4.4.26}$$

Then, for every $p > 2$ there exists $\gamma_p > 0$ such that

$$\|A_{2^v r_z,w,l}P_k^{(1)}P_m^{(2)}f(z)\|_{L_z^p} \lesssim_p 2^{-\gamma_p M}\|f\|_p, \tag{4.4.27}$$

where the implicit constant depends only on p and α.

Before we proceed to proving this statement let us first note that it directly implies (4.4.24) and (4.4.25) (here we use the identity $\beta + 1 = \alpha\beta$ and $\alpha > 1$).

Proof of Lemma 4.4.1. We have

$$A_{2^v r_z,w,l}P_k^{(1)}P_m^{(2)}f(z) = \int_{\mathbb{R}} P_k^{(1)}P_m^{(2)}f(x-t, y-2^v r_z[t]^\alpha)\psi_l(2^{\frac{w}{\alpha}}r_z^\beta t)\frac{dt}{|t|}. \tag{4.4.28}$$

By a change of variables $t \to 2^{l-\frac{w}{\alpha}}t$ we obtain

$$\int_{\mathbb{R}} P_k^{(1)}P_m^{(2)}f(x - 2^{l-\frac{w}{\alpha}}t, y - 2^{l\alpha-w+v}r_z[t]^\alpha)\psi_0(r_z^\beta t)\frac{dt}{|t|}. \tag{4.4.29}$$

Define $D_{a,b}f(x,y) := f(2^a x, 2^b y)$ and

$$Bf(z) := \int_{\mathbb{R}} f(x-t, y-r_z[t]^\alpha)\psi_0(r_z^\beta t)\frac{dt}{|t|}. \tag{4.4.30}$$

Changing variables $t \to r_z^{-\beta}t$ and using the identity $\beta = \alpha\beta - 1$ we see that B can be written in terms of the averaging operator from (4.6.1):

$$Bf(z) = \int_{\mathbb{R}} f(x - r_z^{-\beta}t, y - r_z^{-\beta}[t]^\alpha)\psi_0(t)\frac{dt}{|t|} = A_{r_z^{-\beta}}f(z). \tag{4.4.31}$$

From (4.4.28) – (4.4.30) we see that

$$A_{2^v r_z,w,l}P_k^{(1)}P_m^{(2)}f(z) = D_{-l+\frac{w}{\alpha},-l\alpha+w-v}BD_{l-\frac{w}{\alpha},l\alpha-w+v}P_k^{(1)}P_m^{(2)}f(z). \tag{4.4.32}$$

Since $D_{a,b}P_k^{(1)}P_m^{(2)} = P_{k+a}^{(1)}P_{m+b}^{(2)}D_{a,b}$, the right hand side in the previous display can be written as

$$D_{-l+\frac{w}{\alpha},-l\alpha+w-v}BP_{k+l-\frac{w}{\alpha}}^{(1)}P_{m+l\alpha-w+v}^{(2)}D_{l-\frac{w}{\alpha},l\alpha-w+v}f(z). \tag{4.4.33}$$

Since conjugation by $D_{a,b}$ is an L^p isometry, the L^p norm of (4.4.33) equals

$$\|BP^{(1)}_{k+l-\frac{w}{\alpha}}P^{(2)}_{m+l\alpha-w+v}f\|_p. \tag{4.4.34}$$

The claim now follows from Theorem 4.6.1 once we notice that the frequency support of $P^{(1)}_{k+l-\frac{w}{\alpha}}P^{(2)}_{m+l\alpha-w+v}f$ is contained in the annulus $\|(\xi,\eta)\|\sim 2^M$. □

4.4.2 The low frequency case

In this subsection, we bound the first term in (4.4.16). We write it as

$$\begin{aligned} I_{k_0} &= \sum_{k\in\mathbb{Z}}\sum_{l\le\max\{-k,-k_0/\alpha\}} T_{k_0,l}P^{(1)}_k P^{(2)}_0 f \\ &= \sum_{k\in\mathbb{Z}}\sum_{l\le-\frac{k_0}{\alpha}} T_{k_0,l}P^{(1)}_k P^{(2)}_0 f + \sum_{k\le\frac{k_0}{\alpha}}\sum_{-\frac{k_0}{\alpha}\le l\le -k} T_{k_0,l}P^{(1)}_k P^{(2)}_0 f \\ &=: I^{(1)}_{k_0} + I^{(2)}_{k_0}. \end{aligned} \tag{4.4.35}$$

The first term can be bounded by the strong maximal function and the maximally truncated Hilbert transform in the x-variable. Indeed, comparing $I^{(1)}_{k_0}$ with

$$\sum_{l\le-\frac{k_0}{\alpha}}\int_{\mathbb{R}} P^{(2)}_0 f(x-t,y)\psi_l(t)\frac{dt}{t},$$

we find that their difference is bounded by the strong maximal function. This follows by the same argument as in Section 4.2.1.

We pass now to the treatment of the second term, $I^{(2)}_{k_0}$. For this purpose we first define the "one-dimensional" operator

$$U_{k_0,l}f(x,y) := \int_{\mathbb{R}} f(x,y-u_z^{(k_0)}[t]^\alpha)\psi_l([u_z^{(0)}]^\beta t)\frac{dt}{t}. \tag{4.4.36}$$

Notice that when $[t]^\alpha$ is an even function this operator is identically zero. We next rewrite the second term as

$$I^{(2)}_{k_0} = \sum_{l=0}^{\infty}\sum_{k\le\frac{k_0}{\alpha}-l}(T_{k_0,-\frac{k_0}{\alpha}+l}P^{(1)}_k P^{(2)}_0 f - U_{k_0,-\frac{k_0}{\alpha}+l}P^{(1)}_k P^{(2)}_0 f) + \sum_{l=0}^{\infty}\sum_{k\le\frac{k_0}{\alpha}-l} U_{k_0,-\frac{k_0}{\alpha}+l}P^{(1)}_k P^{(2)}_0 f. \tag{4.4.37}$$

For the contribution coming from the latter term, by the triangle inequality, it suffices to prove that

$$\Big\|\sup_{k_0\in\mathbb{Z}}\Big|\sum_{k\le\frac{k_0}{\alpha}-l_0}\int_{\mathbb{R}}(P^{(1)}_k P^{(2)}_0 f)(x,y-u_z^{(k_0)}[t]^\alpha)\psi_{-\frac{k_0}{\alpha}+l_0}([u_z^{(0)}]^\beta t)\frac{dt}{t}\Big|\Big\|_p \lesssim 2^{-\gamma_p\cdot l_0}\|f\|_p, \tag{4.4.38}$$

for each $l_0\in\mathbb{N}$, and for a constant γ_p depending only on p. This further follows from the pointwise bound

$$\Big|\sum_{k\le\frac{k_0}{\alpha}-l_0}\int_{\mathbb{R}}(P_k^{(1)}P_0^{(2)}f)(x,y-u_z^{(k_0)}[t]^\alpha)\psi_{-\frac{k_0}{\alpha}+l_0}([u_z^{(0)}]^\beta t)\frac{dt}{t}\Big|\lesssim 2^{-\gamma\cdot l_0}M_Sf(x,y),\tag{4.4.39}$$

which is again a consequence of the mean zero property of $\psi_{-\frac{k_0}{\alpha}+l_0}(t)\cdot\frac{1}{t}$. Indeed, (4.4.39) follows from classical Calderón-Zygmund theory. We leave the details for the interested reader.

Turning our attention towards the contribution from the former term on the right hand side of (4.4.37), we notice that it is enough to show that

$$\Big\|\sup_{k_0\in\mathbb{Z}}\Big|\sum_{k\le\frac{k_0}{\alpha}-l_0}\big(T_{k_0,-\frac{k_0}{\alpha}+l_0}P_k^{(1)}P_0^{(2)}f-U_{k_0,-\frac{k_0}{\alpha}+l}P_k^{(1)}P_0^{(2)}f\big)\Big|\Big\|_p\lesssim 2^{-\gamma_p\cdot l_0}\|f\|_p,\tag{4.4.40}$$

for each $l_0\in\mathbb{N}$. This in turn follows from

Claim 4.4.2. *In the above setting, for each $l_0, j_0\in\mathbb{N}$ and $k_0\in\mathbb{Z}$ the following estimate holds uniformly:*

$$\Big\|\sup_{k_0\in\mathbb{Z}}\Big|T_{k_0,-\frac{k_0}{\alpha}+l_0}P^{(1)}_{\frac{k_0}{\alpha}-l_0-j_0}P_0^{(2)}f-U_{k_0,-\frac{k_0}{\alpha}+l}P^{(1)}_{\frac{k_0}{\alpha}-l_0-j_0}P_0^{(2)}f\Big|\Big\|_p\lesssim 2^{-\gamma_p\cdot\max\{l_0,\frac{j_0}{9}\}}\|f\|_p\,.\tag{4.4.41}$$

Proof. We first look at the case $j_0\le 9\cdot l_0$. Based on the treatment of $I_{k_0}^{(2)}$, it is enough to show that

$$\|T_{k_0,-\frac{k_0}{\alpha}+l_0}P^{(1)}_{\frac{k_0}{\alpha}-l_0-j_0}P_0^{(2)}f\|_p\lesssim 2^{-\gamma_p\cdot l_0}\|f\|_p,\tag{4.4.42}$$

for each fixed k_0, l_0 and j_0. This follows from Lemma 4.4.1.

Next we consider the case $j_0\ge 9\cdot l_0$. It suffices to prove the pointwise bound

$$\big|T_{k_0,-\frac{k_0}{\alpha}+l_0}P^{(1)}_{\frac{k_0}{\alpha}-l_0-j_0}P_0^{(2)}f-U_{k_0,-\frac{k_0}{\alpha}+l}P^{(1)}_{\frac{k_0}{\alpha}-l_0-j_0}P_0^{(2)}f\big|\lesssim 2^{-\gamma\cdot j_0}\cdot M_S(f),\tag{4.4.43}$$

for some positive $\gamma>0$. By scaling it suffices to look at the case $k_0=0$. We now observe that the left hand side of (4.4.43) can be written as

$$\Big|\int_{\mathbb{R}}\Big[(P^{(1)}_{-l_0-j_0}P_0^{(2)}f)(x-t,y-u_z^{(0)}[t]^\alpha)-(P^{(1)}_{-l_0-j_0}P_0^{(2)}f)(x,y-u_z^{(0)}[t]^\alpha)\Big]\psi_{l_0}([u_z^{(0)}]^\beta t)\frac{dt}{t}\Big|.\tag{4.4.44}$$

Now our claim follows by applying the fundamental theorem of calculus in the first variable of $P^{(1)}_{-l_0-j_0}P_0^{(2)}f$. □

4.5 Proof of Theorem 4.1.1

The proof is organized as follows:

- In Section 4.5.1 we reduce our proof to the exponential decay estimate (4.5.6).
- In Section 4.5.2 we prove this decay estimate for $p>2$, by only assuming u to be a measurable function. This recovers the result of Marletta and Ricci [MR98].

- In Section 4.5.3 we prove the estimate (4.5.6) for $p \leq 2$. This relies on the Lipschitz assumption of the function u, a condition that is used when applying a suitable change of variables (see (4.5.31)). To enable that change of variables, we will introduce several auxiliary functions and make use of Lemma 4.5.1 on the boundedness of maximal operators along curves in lacunary directions. The proof of this lemma is provided in Section 4.5.4.

4.5.1 Preliminaries

Since we are dealing with a positive operator we may assume without loss of generality that $f \geq 0$. Furthermore, we may assume that $u(x, y) > 0$ for all $(x, y) \in \mathbb{R}^2$. We will adopt the notation (4.4.4) – (4.4.6) from the previous section.

For a positive real number u and $l \in \mathbb{Z}$ we define

$$A_{u,l}f(z) = \int_{\mathbb{R}} f(x - t, y - u[t]^\alpha)\psi_l(2^{\frac{v}{\alpha}}(2^{-v}u)^\beta t)\frac{dt}{|t|},$$

where $v \in \mathbb{Z}$ is such that $u \in [2^v, 2^{v+1})$ and $\beta := \frac{1}{\alpha-1}$. In particular, $2^{-v}u \in [1, 2)$. Observe that

$$\mathcal{M}^{(\alpha)}_{u,\varepsilon_0}f(z) \lesssim \sup_{l\,:\,v_z \in E_l} A_{u_z,l}f(z),$$

where

$$E_l := \{v \in \mathbb{Z} \,:\, 2^l \leq c_\alpha \varepsilon_0 2^{\frac{v}{\alpha}}\}, \tag{4.5.1}$$

for $v \in \mathbb{Z}$ and c_α is a fixed constant depending only on α. Linearizing the supremum we introduce the operator

$$\mathcal{M}f(z) := A_{u_z,l_z}f(z),$$

where $z \mapsto l_z$ is an arbitrary measurable map $\mathbb{R}^2 \to \mathbb{Z}$ such that $v_z \in E_{l_z}$ for all z. To prove our theorem it suffices to show that

$$\|\mathcal{M}f\|_p \lesssim_{\alpha,p} \|f\|_p$$

for all $1 < p < \infty$. By the Fourier inversion formula we have

$$A_{u_z,l}f(z) = \int_{\mathbb{R}^2} \widehat{f}(\xi, \eta)e^{ix\xi+iy\eta}m_l(z, \xi, \eta)d(\xi, \eta),$$

with the symbol $m_l(z, \xi, \eta)$ given by

$$m_l(z, \xi, \eta) := \int_{\mathbb{R}} e^{-i(\xi t+\eta u_z[t]^\alpha)}\psi_l(2^{\frac{v_z}{\alpha}}(u_z^{(0)})^\beta t)\frac{dt}{|t|}, \tag{4.5.2}$$

where $u_z^{(0)} := 2^{-v_z}u_z$, as defined in (4.4.5). As in our previous analysis, our approach relies on decomposing our operator relative to the behavior of the phase function of the multiplier. Our initial focus will be on the behavior of the η component of the phase which corresponds in the spatial variable to the component containing the vector field u. Thus, we will discuss the following two cases:

- the $\eta-$low frequency regime: $|\eta|2^{\alpha l} \leq 1$;
- the $\eta-$high frequency regime: $|\eta|2^{\alpha l} > 1$.

Following the above description, we perform a Littlewood-Paley decomposition in the y-variable and write

$$A_{u_z,l_z} = \sum_{k\in\mathbb{Z}} A_{u_z,l_z} P_k^{(2)} = \sum_{k\le -\alpha l_z} A_{u_z,l_z} P_k^{(2)} + \sum_{k>-\alpha l_z} A_{u_z,l_z} P_k^{(2)}. \tag{4.5.3}$$

Case 1. $\eta-$low frequency.

The sum over $k \le -\alpha l_z$ can be estimated by the strong maximal operator and is therefore bounded on all L^p, $1 < p \le \infty$. This follows by the same arguments as in Section 4.2.1.

Case 2. $\eta-$high frequency.

The remaining term equals[6]

$$\sum_{k\in\mathbb{N}} A_{u_z,l_z} P^{(2)}_{-\alpha l_z+k} = \sum_{k\in\mathbb{N}} \sum_{m\in\mathbb{Z}} A_{u_z,l_z} P^{(1)}_m P^{(2)}_{-\alpha l_z+k}. \tag{4.5.4}$$

Fix a $k \in \mathbb{N}$, and for the following heuristic, let us imagine for a moment that u_z is a constant. Then from (4.5.2), the symbol of $A_{u_z,l_z} P^{(2)}_{-\alpha l_z+k}$ becomes

$$\int_{\mathbb{R}} e^{i(2^{l_z} u_z^{-\frac{1}{\alpha}} \xi t + \eta 2^{\alpha\cdot l_z}[t]^\alpha)} \psi_0(t) \frac{dt}{|t|}.$$

From the stationary phase principle it is plausible to expect an exponential decay for the L^p norm of the term (4.5.4) in terms of $k \in \mathbb{N}$. Indeed, this is the case for Hilbert transforms along one-variable curves (see the estimate (4.3.10)). However, in the present situation we do not know how to exhibit such an exponential decay. To remedy this we first remove from the term (4.5.4) the $\xi-$low frequency component; the remaining part then admits an exponential decay estimate in $k \in \mathbb{N}$.

We split the term in (4.5.4) into two components corresponding to

- the $\xi-$low frequency regime: $|2^{l_z} u_z^{-\frac{1}{\alpha}} \xi| < 1$;
- the $\xi-$high frequency regime: $|2^{l_z} u_z^{-\frac{1}{\alpha}} \xi| \ge 1$.

That is, we write (4.5.4) as

$$\sum_{k\in\mathbb{N}} A_{u_z,l_z} \sum_{m \le -l_z + \frac{v_z}{\alpha}} P^{(1)}_m P^{(2)}_{-\alpha l_z+k} + \sum_{k\in\mathbb{N}} \sum_{m\in\mathbb{N}} A_{u_z,l_z} P^{(1)}_{-l_z+\frac{v_z}{\alpha}+m} P^{(2)}_{-\alpha l_z+k}. \tag{4.5.5}$$

Case 2.1. $\xi-$low frequency.

This corresponds to the first term in (4.5.5). The contribution from this term can be controlled by the strong maximal function using the same argument as in the η-low frequency case. We omit the details.

[6]Here and in the following our notation will ignore the issue that $-\alpha l_z$ might not be an integer as this can be easily addressed by setting $P^{(i)}_\alpha = P^{(i)}_{\lfloor\alpha\rfloor}$.

Case 2.2. $\xi-$high frequency.

To handle the latter term on the right hand side of (4.5.5), it suffices to show that for every $p > 1$ there exists $\gamma_p > 0$ such that

$$\Big\| \sum_{m\in\mathbb{N}} A_{u_z,l_z} P^{(1)}_{-l_z+\frac{v_z}{\alpha}+m} P^{(2)}_{-\alpha l_z+k} f \Big\|_p \lesssim 2^{-\gamma_p k} \|f\|_p. \tag{4.5.6}$$

In view of the Littlewood-Paley square function estimate, the argument splits naturally into two cases: $p > 2$ and $p \leq 2$.

4.5.2 The case $p > 2$

In this section we reprove the result of Marletta and Ricci [MR98] and, in particular, we only require the function u to be measurable. We also do not need to consider a truncated maximal function; under the measurability assumption, the truncated case is equivalent to the untruncated case, by a scaling argument. Hence we take $\varepsilon_0 = \infty$.

We first remark that the left hand side of (4.5.6) is bounded from above by

$$\Big\| \Big(\sum_{l\in\mathbb{Z}} \Big| \sum_{m\in\mathbb{N}} A_{u_z,l} P^{(1)}_{-l+\frac{v_z}{\alpha}+m} P^{(2)}_{-\alpha l+k} f \Big|^p \Big)^{1/p} \Big\|_p. \tag{4.5.7}$$

Hence, it suffices to prove that for every $l \in \mathbb{Z}$ one has

$$\Big\| \sum_{m\in\mathbb{N}} A_{u_z,l} P^{(1)}_{-l+\frac{v_z}{\alpha}+m} P^{(2)}_{-\alpha l+k} f \Big\|_p \lesssim 2^{-\gamma_p k} \|f\|_p\,. \tag{4.5.8}$$

For a fixed $v \in \mathbb{Z}$, recall the definition of $u_z^{(v)}$ from (4.4.6). Then the left hand side of (4.5.8) can be bounded by

$$\Big\| \Big(\sum_{v\in\mathbb{Z}} \Big| \sum_{m\in\mathbb{N}} A_{u_z^{(v)},l} P^{(1)}_{-l+\frac{v}{\alpha}+m} P^{(2)}_{-\alpha l+k} f \Big|^p \Big)^{1/p} \Big\|_p. \tag{4.5.9}$$

Taking into account (4.5.7) –(4.5.9), we see that it suffices to show that for every $p > 2$ there exists $\gamma_p > 0$ such that for every $k, m \in \mathbb{N}$ and $l, v \in \mathbb{Z}$ we have

$$\Big\| A_{u_z^{(v)},l} P^{(1)}_{-l+\frac{v}{\alpha}+m} P^{(2)}_{-\alpha l+k} f(z) \Big\|_p \lesssim 2^{-\gamma_p\, \max\{k,m\}} \|f\|_p\,. \tag{4.5.10}$$

Noting that $A_{u_z^{(v)},l} = A_{2^v u_z^{(0)},v,l}$ we recognize this as precisely the conclusion of Lemma 4.4.1.

4.5.3 L^p estimates for $p \leq 2$

In this subsection we will prove (4.5.6) for all $1 < p \leq 2$. Note that we are working here with two extra assumptions, as compared with the previous subsection:

- First, we assume that u is Lipschitz. This is because if u is only assumed to be measurable, then it is possible for $\mathcal{M}^{(\alpha)}_{u,\epsilon_0}$ to be unbounded for each $p \leq 2$. One may verify this by simply plugging the characteristic function of the unit ball into the inequality.

- Second, the truncation $\epsilon \leq \epsilon_0$ will play a crucial role. That is, in effect we will have to take into account the range restriction of l_z expressed by (4.5.1).

Both of these assumptions will only be used to ensure the validity of the Lipschitz change of variables in (4.5.31).

Let us begin with the proof. By interpolation with the case $p > 2$, it suffices to prove an estimate without decay; that is, it suffices to show that

$$\Big\| \sum_{m > -l_z + \frac{v_z}{\alpha}} A_{u_z, l_z} P_m^{(1)} P_{-\alpha l_z + k}^{(2)} f(z) \Big\|_{L_z^p} \lesssim \|f\|_p. \tag{4.5.11}$$

Again we get rid of the z-dependence of l_z by inserting a sum over all possible values of l_z, similar to (4.5.7). In particular, we bound the expression inside the L^p norm on the left hand side of (4.5.11) by[7]

$$\Big(\sum_{l \,:\, v_z \in E_l} \Big| \sum_{m > -l + \frac{v_z}{\alpha}} A_{u_z, l} P_m^{(1)} P_{-\alpha l + k}^{(2)} f(z) \Big|^2 \Big)^{1/2}. \tag{4.5.12}$$

We write (4.5.12) as

$$\Big(\sum_{l \in \mathbb{Z}} \Big| \mathbb{1}_{E_l}(v_z) \sum_{m > -l + \frac{v_z}{\alpha}} A_{u_z, l} P_m^{(1)} P_{-\alpha l + k}^{(2)} f(z) \Big|^2 \Big)^{1/2}. \tag{4.5.13}$$

To estimate (4.5.13) we will run an interpolation argument which requires one to consider estimates of the form

$$\Big\| \Big(\sum_{l \in \mathbb{Z}} \Big| \mathbb{1}_{E_l}(v_z) \sum_{m > -l + \frac{v_z}{\alpha}} A_{u_z, l} P_m^{(1)} P_{-\alpha l + k}^{(2)} f(z) \Big|^q \Big)^{1/q} \Big\|_p \lesssim \Big\| \Big(\sum_{l \in \mathbb{Z}} \Big| P_{-\alpha l + k}^{(2)} f \Big|^q \Big)^{1/q} \Big\|_p. \tag{4.5.14}$$

When $q = \infty$, it is not difficult to see that the expression inside the L^p norm on the left hand side of the last display is bounded by

$$\mathcal{M}_{u, \epsilon_0}^{(\alpha)} (M_S f)(z). \tag{4.5.15}$$

Here M_S denotes the strong maximal operator. Recall that we already proved the L^p boundedness of $\mathcal{M}_{u, \epsilon_0}^{(\alpha)}$ for all $p > 2$ in Section 4.5.2. This implies that

$$\text{(4.5.14) holds for all } p > 2 \text{ and } q = \infty. \tag{4.5.16}$$

We will interpolate estimate (4.5.16) with

$$\Big\| \mathbb{1}_{E_l}(v_z) \sum_{m > -l + \frac{v_z}{\alpha}} A_{u_z, l} P_m^{(1)} P_{-\alpha l + k}^{(2)} f(z) \Big\|_p \lesssim \|f\|_p, \tag{4.5.17}$$

[7] At this point we fix ϵ_0 possibly depending on α so that this estimate becomes valid.

for all $p > 1$. Suppose for the moment that (4.5.17) holds. Therefore

$$\text{(4.5.14) holds for all } p = q > 1. \tag{4.5.18}$$

Then, interpolating (4.5.18) with (4.5.16) we obtain that

$$\|(4.5.13)\|_p \lesssim \|f\|_p, \tag{4.5.19}$$

for all $p > \frac{4}{3}$. Combining this with (4.5.15) implies that

$$\text{(4.5.14) holds for } p > \frac{4}{3} \text{ and } q = \infty. \tag{4.5.20}$$

We interpolate (4.5.20) with (4.5.17) to see that (4.5.19) holds for all $p > \frac{8}{7}$. Repeating this interpolation procedure sufficiently many times, we obtain that (4.5.13) is bounded on L^p for all $p > 1$. We learned this interpolation trick from Nagel, Stein and Wainger [NSW78].

Before we turn to the proof of (4.5.17) we need to introduce some new notation. For $n \in \mathbb{Z}$ and $z \in \mathbb{R}^2$ we define

$$u_n(z) := \begin{cases} u_z & \text{if } v_z = n, \\ 2^n & \text{if } v_z \neq n. \end{cases}$$

Note carefully that $u_n(z)$ is different from $u_z^{(n)}$. Both take values in the interval $[2^n, 2^{n+1})$. However, the function u_n retains more of the regularity of u than $z \mapsto u_z^{(n)}$. This will be important during the proof of (4.5.17) (see (4.5.29), (4.5.38)).

Eliminating the z-dependence of v_z by introducing another sum we estimate the left hand side of (4.5.17) by

$$\Big\|\Big(\sum_{v\in E_l}\Big|\sum_{m>-l+\frac{v}{\alpha}} A_{u_v(z),l}P_m^{(1)}P_{-\alpha l+k}^{(2)}f(z)\Big|^2\Big)^{1/2}\Big\|_p. \tag{4.5.21}$$

By the triangle inequality this is no greater than

$$\sum_{m>0}\Big\|\Big(\sum_{v\in E_l}\Big|A_{u_v(z),l}P_{m-l+\frac{v}{\alpha}}^{(1)}P_{-\alpha l+k}^{(2)}f(z)\Big|^2\Big)^{1/2}\Big\|_p. \tag{4.5.22}$$

Thus (4.5.17) will follow if we can show that there exist constants $\gamma_p > 0$ such that for every $k, m \in \mathbb{N}, l \in \mathbb{Z}$ we have

$$\Big\|\Big(\sum_{v\in E_l}\Big|A_{u_v(z),l}P_{m-l+\frac{v}{\alpha}}^{(1)}P_{-\alpha l+k}^{(2)}f(z)\Big|^2\Big)^{1/2}\Big\|_p \lesssim 2^{-\gamma_p \max\{k,m\}}\|f\|_p. \tag{4.5.23}$$

We first prove (4.5.23) for $p = 2$. By Lemma 4.4.1 we obtain that for all $p > 2$ there exists $\gamma_p > 0$ such that

$$\Big\|A_{u_v(z),l}P_{m-l+\frac{v}{\alpha}}^{(1)}P_{-\alpha l+k}^{(2)}f(z)\|_p \lesssim 2^{-\gamma_p \max\{k,m\}}\|f\|_p. \tag{4.5.24}$$

Thus, in order to show (4.5.23) for $p = 2$ it will be enough to show the estimate

$$\left\| A_{u_v(z),l} P^{(1)}_{m-l+\frac{v}{\alpha}} P^{(2)}_{-\alpha l+k} f(z) \right\|_p \lesssim \|f\|_p. \tag{4.5.25}$$

for an arbitrary $p < 2$. In the following we will prove something stronger, namely that

$$\left\| A_{u_v(z),l} f(z) \right\|_p \lesssim \|f\|_p \tag{4.5.26}$$

holds for all $p > 1$ provided that $v \in E_l$. Before commencing the proof of (4.5.26), we need to introduce another auxiliary function. If at a point $z = (x, y)$ we have $v_z = n$, then let $\tilde{u}_n(z) := u_z$. We now extend $\tilde{u}_n$ to the whole space, requiring only that

$$\tilde{u}_n(z) \in [2^n, 2^{n+1}) \text{ and } \|\tilde{u}_n\|_{\text{Lip}} \leq 2\|u\|_{\text{Lip}}. \tag{4.5.27}$$

The advantage is that, unlike u_n, the function $\tilde{u}_n$ is Lipschitz continuous with controlled Lipschitz norm. This will later be of key importance in ensuring the validity of the Lipschitz change of variables (4.5.31). Most importantly, the way that $\tilde{u}_n$ and u_n are designed is such that they stay very "close" to each other, in the sense that if we define the following maximal operator:

$$\mathcal{M}_{\text{dyad}} f(z) := \sup_{j \in \mathbb{Z}} \sup_{\epsilon > 0} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} |f(x - t, y - 2^j [t]^\alpha)| dt. \tag{4.5.28}$$

then we have the pointwise estimate

$$A_{u_v(z),l} f(z) \lesssim \mathcal{M}_{\text{dyad}} f(z) + A_{\tilde{u}_v(z),l} f(z) \tag{4.5.29}$$

for all $v, l \in \mathbb{Z}$, $z \in \mathbb{R}^2$. The maximal operator $\mathcal{M}_{\text{dyad}}$ is bounded in L^p for all $p > 1$.

Lemma 4.5.1. *For each $p > 1$, we have*

$$\|\mathcal{M}_{\text{dyad}} f\|_p \lesssim \|f\|_p. \tag{4.5.30}$$

The implicit constant depends only on p.

We postpone the proof of this fact until the end of this subsection and turn our attention toward the second term in (4.5.29). By Minkowski's integral inequality, the L^p norm of $A_{\tilde{u}_v(z),l} f(z)$ is at most

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^2} (f(x - t, y - \tilde{u}_v(z)[t]^\alpha) \psi_l(2^{\frac{v}{\alpha}} r_z t))^p dz \right)^{1/p} \frac{dt}{|t|}.$$

where $r_z := (2^{-v} \tilde{u}_n(z))^\beta \approx 1$. To bound this term, we apply the change of variables

$$\begin{cases} x_1 & := x - t \\ y_1 & := y - \tilde{u}_v(x, y)\, [t]^\alpha \end{cases} . \tag{4.5.31}$$

This is the only point at which we use the Lipschitz regularity of u and it is also where we need to exploit the range restriction on l, i.e. that $2^l \leq c_\alpha \epsilon_0 2^{\frac{v}{\alpha}}$. Indeed, computing the determinant of the corresponding Jacobian J, we have

$$\det J = \left|\frac{\partial(x_1,y_1)}{\partial(x,y)}\right| = \begin{vmatrix} 1 & 0 \\ -\frac{\partial \tilde{u}_v}{\partial x}\,[t]^\alpha & 1-\frac{\partial \tilde{u}_v}{\partial y}\,[t]^\alpha \end{vmatrix} = 1-\frac{\partial \tilde{u}_v}{\partial y}\,[t]^\alpha\,. \tag{4.5.32}$$

Observe that t in (4.5.31) obeys

$$|[t]^\alpha| \approx 2^{l-\frac{v}{\alpha}} \le c_\alpha \epsilon_0$$

and hence

$$|\det J| = \left|1-\frac{\partial \tilde{u}_n}{\partial y}\,[t]^\alpha\right| \ge 1-\tilde{c}_\alpha\|\tilde{u}_n\|_{\mathrm{Lip}}\epsilon_0 \ge \frac{1}{2}, \tag{4.5.33}$$

where we have chosen ϵ_0 to be smaller than $(2\tilde{c}_\alpha\|u\|_{\mathrm{Lip}})^{-1}$. Here $\tilde{c}_\alpha$ is a constant depending only on α. This shows that the change of variables (4.5.31) is valid and therefore

$$\|A_{\tilde{u}_v(z),l}f\|_p \lesssim \|f\|_p \tag{4.5.34}$$

holds for all $p>1$. Hence we can infer that also (4.5.26) holds. This concludes the proof of (4.5.23) in the case $p=2$.

It remains to prove (4.5.23) for the remaining values of p. By interpolation with (4.5.23) at $p=2$, it suffices to prove that we have the estimate without decay,

$$\Big\|\Big(\sum_{v\in E_l}\Big|A_{u_v(z),l}P^{(1)}_{m-l+\frac{v}{\alpha}}P^{(2)}_{-\alpha l+k}f(z)\Big|^2\Big)^{1/2}\Big\|_p \lesssim \|P^{(2)}_{-\alpha l+k}f\|_p \sim \Big\|\Big(\sum_{v\in\mathbb{Z}}\Big|P^{(1)}_{-l+m+\frac{v}{\alpha}}P^{(2)}_{-\alpha l+k}f\Big|^2\Big)^{1/2}\Big\|_p \tag{4.5.35}$$

for all $p>1$. We will again use the interpolation trick from Nagel, Stein and Wainger [NSW78]. Thus we consider the more general estimate

$$\Big\|\Big(\sum_{v\in E_l}\Big|A_{u_v(z),l}P^{(1)}_{m-l+\frac{v}{\alpha}}P^{(2)}_{-\alpha l+k}f(z)\Big|^q\Big)^{1/q}\Big\|_p \lesssim \Big\|\Big(\sum_{v\in\mathbb{Z}}\Big|P^{(1)}_{-l+m+\frac{v}{\alpha}}P^{(2)}_{-\alpha l+k}f\Big|^q\Big)^{1/q}\Big\|_p \tag{4.5.36}$$

for $1<p<\infty$ and $1<q\le\infty$.

Recall that in Section 4.5.2 we proved that

$$\|\mathcal{M}^{(\alpha)}_{u,\epsilon_0}f\|_p \lesssim \|f\|_p \text{ for all } p>2. \tag{4.5.37}$$

Moreover, we have the pointwise bound

$$|A_{u_v(z),l}f(z)| \lesssim \mathcal{M}_{\mathrm{dyad}}f(z)+\mathcal{M}^{(\alpha)}_{u,\epsilon_0}f(z) \text{ for all } v,\, l\in\mathbb{Z}. \tag{4.5.38}$$

Note that this pointwise estimate does not hold if u_v is replaced by $\tilde{u}_v$, because we do not know how $\tilde{u}_v$ behaves outside of the region where it coincides with the original Lipschitz function u.

From (4.5.37), (4.5.38) and Lemma 4.5.1 we see that (4.5.36) holds for $q=\infty$ and $p>2$. Moreover, by (4.5.26) we know that (4.5.36) holds for all $q=p>1$. By interpolation we obtain (4.5.36) for $q=2$ and $p>4/3$. This in turn implies that

$$\|\mathcal{M}^{(\alpha)}_{u,\epsilon_0}f\|_p \lesssim \|f\|_p, \tag{4.5.39}$$

for all $p > 4/3$. Iterating this process sufficiently many times we prove (4.5.35) for every $p > 1$. This finishes the proof of (4.5.23) and thereby also concludes the proof of Theorem 4.1.1.

We close this subsection with several remarks regarding the last part of the proof:

1. In the proof of (4.5.26), we did not need the full strength of Lemma 4.5.1. For that purpose it would have been sufficient to take a fixed j in the definition of $\mathcal{M}_{\text{dyad}}$ instead of taking the supremum over $j \in \mathbb{Z}$. The full strength of Lemma 4.5.1 is only needed during the interpolation procedure that is used for proving (4.5.23) for $p < 2$.

2. One might wonder why we did not use the auxiliary function $\tilde{u}_v$ right away, rather than first introducing the auxiliary function u_v. Again, the point is that this would cause the interpolation argument for $p < 2$ to fail, because the pointwise estimate (4.5.38) would no longer be available.

4.5.4 Proof of Lemma 4.5.1

Before we start the proof of this lemma, we emphasize that this lemma was first established by Hong, Kim and Yang [HKY09]. Indeed, their results have a much wider scope, in the sense that they considered general polynomials in all dimensions. Due to such generality, their proof is significantly more intricate. For the sake of completeness, we provide here the proof of Lemma 4.5.1.

We will present the argument for $\alpha = 2$ and $[t]^2 = t^2$; the general case follows *mutatis mutandis.*

For k, $j \in \mathbb{Z}$ we set

$$M_k^j f(x,y) := \Big|\frac{1}{2^k}\int_{2^k}^{2^{k+1}} f(x-t, y-2^{-j}t^2)dt\Big|. \tag{4.5.40}$$

Fix $j \in \mathbb{Z}$ consider the maximal operator along the parabola $(t, 2^{-j}t^2)$, which is given by

$$M^j f(x,y) := \sup_{k\in\mathbb{Z}} M_k^j f(x,y). \tag{4.5.41}$$

Hence

$$\mathcal{M}_{\text{dyad}} f(x,y) = \sup_j M^j f(x,y) = \sup_{k,j} M_k^j f(x,y). \tag{4.5.42}$$

A key idea to prove the L^2 bounds for $\mathcal{M}_{\text{dyad}}$ is to compare its components M_k^j with some suitable smoother versions $\widetilde{M}_k^j$ and prove bounds for the corresponding square function

$$\Big(\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} |M_k^j - \widetilde{M}_k^j|^2\Big)^{1/2}. \tag{4.5.43}$$

As a first attempt for finding a good candidate for $\widetilde{M}_k^j$ one may consider the *linearized* model

$$\bar{M}_k^j f(x,y) := \frac{1}{2^k\cdot 3\cdot 2^{2k}}\int_{2^k}^{2^{k+1}}\int_{2^{2k}}^{2^{2k+2}} f(x-t, y-2^{-j}\tau)dtd\tau. \tag{4.5.44}$$

It turns out that while the operator $\left(\sum_{k\in\mathbb{Z}}|M_k^j - \bar{M}_k^j|^2\right)^{1/2}$ is indeed bounded on L^2, one does not have a good control over the double indexed sum in (4.5.43). For this reason, one needs to choose a variant that is closer in spirit to (4.5.40), which is simultaneously smoother and preserves the *quadratic* nature of the initial operator:

$$\widetilde{M}_k^j f(x,y) := \frac{1}{2^k \cdot 2^k}\int_{2^k}^{2^{k+1}}\int_{2^k}^{2^{k+1}} f(x-t, y-2^{-j}\tau^2)\,dt\,d\tau. \tag{4.5.45}$$

It is not difficult to see that we have the pointwise bound

$$\widetilde{M}_k^j f(x,y) \lesssim M_S f(x,y)\,, \tag{4.5.46}$$

and thus, in order to prove the L^p bound of $\mathcal{M}_{\text{dyad}}$, it suffices to prove the L^p bound for

$$\Big(\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}|M_k^j - \widetilde{M}_k^j|^2\Big)^{1/2}. \tag{4.5.47}$$

As opposed to the general L^p case, the proof for the L^2 boundedness of (4.5.47) is significantly simpler and thus we choose to present it first.

4.5.4.1 L^2 boundedness

In order to prove the L^2 bounds we rely on Plancherel's theorem. Naturally, we start by analyzing the multipliers for the corresponding operators M_k^j and $\widetilde{M}_k^j$.

The multiplier of M_k^j is given by

$$m_k^j(\xi,\eta) := \int_1^2 e^{i2^k t\xi + i2^{2k-j}t^2\eta}\,dt\,, \tag{4.5.48}$$

while the multiplier of the operator (4.5.45) is given by

$$\widetilde{m}_k^j(\xi,\eta) := \int_1^2\int_1^2 e^{i2^k t\xi + i2^{2k-j}\tau^2\eta}\,dt\,d\tau\,. \tag{4.5.49}$$

By Plancherel's theorem, it suffices to prove that

$$\sum_{j\in\mathbb{Z},k\in\mathbb{Z}}\left|m_k^j(\xi,\eta) - \widetilde{m}_k^j(\xi,\eta)\right|^2 \le C, \tag{4.5.50}$$

for all $(\xi,\eta)\in\mathbb{R}^2$ and some universal constant $C>0$. This in turn follows from

$$\sum_{j\in\mathbb{Z},k\in\mathbb{Z}}\left|m_k^j(\xi,\eta) - \widetilde{m}_k^j(\xi,\eta)\right| \lesssim 1,\ \text{for all } \xi\sim 1 \text{ and } \eta\sim 1. \tag{4.5.51}$$

- **Case I:** $k+10\ge j\ge 0$**.** Applying van der Corput's lemma, we have

$$\left|m_k^j(\xi,\eta) - \widetilde{m}_k^j(\xi,\eta)\right| \lesssim 2^{-(k-\frac{j}{2})}. \tag{4.5.52}$$

The last display is easily seen to be summable within the range $k\ge j\ge 0$.

- **Case II:** $4k \geq j > k + 10 \geq 0$ **or** $k \geq 0, j \leq 0$ **or** $j \leq 4k \leq 0$**.** For simplicity, we only detail the case $4k \geq j > k \geq 0$. Under this assumption, the phase functions in (4.5.49) and (4.5.48) do not admit any critical point. Thus,

$$\left| m_k^j(\xi, \eta) - \widetilde{m}_k^j(\xi, \eta) \right| \lesssim 2^{-k}. \tag{4.5.53}$$

  By summing over $10 + k < j \leq 4k$, we obtain the upper bound $3k \cdot 2^{-k}$, which is summable in $k \in \mathbb{N}$.

- **Case III:** $j \geq 4k \geq 0$ **or** $k \leq 0, j \geq 0$ **or** $k \leq j \leq 0$**.** Again we only detail one case, that of $j \geq 4k \geq 0$. By definition, we have

$$\begin{aligned} &m_k^j(\xi, \eta) - \widetilde{m}_k^j(\xi, \eta) = \int_1^2 e^{i2^k t\xi + i2^{2k-j}t^2\eta} dt - \int_1^2 \int_1^2 e^{i2^k t\xi + i2^{2k-j}\tau^2\eta} dt d\tau \\ &= \int_1^2 e^{i2^k t\xi} \left(e^{i2^{2k-j}t^2\eta} - 1 \right) dt - \int_1^2 \int_1^2 e^{i2^k t\xi} \left(e^{i2^{2k-j}\tau^2\eta} - 1 \right) dt d\tau. \end{aligned} \tag{4.5.54}$$

  By the fundamental theorem of calculus, the last display can be bounded by 2^{2k-j}, which is summable in $j \geq 4k \geq 0$.

  Here we mention that the term $\bar{M}_k^j f$ from (4.5.44) would also work in this case.

- **Case IV:** $4k \leq j \leq k \leq 0$**.** In this case we will see the main difference between $\bar{M}_k^j$ and $\widetilde{M}_k^j$. By definition,

$$\begin{aligned} &m_k^j(\xi, \eta) - \widetilde{m}_k^j(\xi, \eta) = \int_1^2 e^{i2^k t\xi + i2^{2k-j}t^2\eta} dt - \int_1^2 \int_1^2 e^{i2^k t\xi + i2^{2k-j}\tau^2\eta} dt d\tau \\ &= \int_1^2 \left(e^{i2^k \tau} - 1 \right) e^{i2^{2k-j}\tau^2} d\tau - \int_1^2 \int_1^2 \left(e^{i2^k t} - 1 \right) e^{i2^{2k-j}\tau^2} dt d\tau. \end{aligned} \tag{4.5.55}$$

  By the fundamental theorem of calculus, we bound the last display by 2^k. Summing over $4k \leq j \leq k$, we obtain $|k| \cdot 2^k$, which is summable for $k \leq 0$.

This finishes the proof of the L^2 boundedness of our operator defined in (4.5.42).

4.5.4.2 L^p boundedness

In what follows we will make use of some ideas from Nagel, Stein and Wainger [NSW78] and Carlsson, Christ et al. [CCC+86]. We denote

$$M_k^j f := \mu_k^j * f, \text{ with } \widehat{\mu_k^j}(\xi, \eta) := m_k^j(\xi, \eta). \tag{4.5.56}$$

A key insight in [CCC+86], is to compare μ_k^j with σ_k^j, where

$$\sigma_k^j := \mu_k^j * [(\phi_k^j - \delta) \otimes (\tilde{\phi}_k^j - \delta)]. \tag{4.5.57}$$

Here ϕ and $\tilde{\phi}$ are two non-negative smooth functions supported on $[-1,1]$ having mean one, while

$$\phi_k^j(t) := 2^{-k}\phi(2^{-k}t)\,, \qquad \tilde{\phi}_k^j(t) := 2^{-2k-2+j}\tilde{\phi}(2^{-2k-2+j}t)\,, \tag{4.5.58}$$

and δ is the Dirac point mass at the origin. The meaning of the tensor product in (4.5.57) is that its first component acts on the first variable while its second component acts on the second variable.

The difference $\mu_k^j - \sigma_k^j$ can be bounded by the strong maximal operator, by noticing that the corresponding multiplier has fast decay at infinity. In particular,

$$\sup_{k,j} |(\mu_k^j - \sigma_k^j) * f| \lesssim M_S f. \tag{4.5.59}$$

Thus, it only remains to bound $\sup_{k,j} |\sigma_k^j * f|$. For this, we will perform a conical Littlewood-Paley decomposition for the function f:

$$P_k^{\text{Cone}} f(x,y) := \int_{\mathbb{R}} \hat{f}(\xi,\eta)\psi_k\Big(\frac{\xi}{\eta}\Big) e^{ix\xi+iy\eta} d\xi d\eta, \tag{4.5.60}$$

and write

$$\sup_{k,j} |\sigma_k^j * f| = \sup_{k,j} \Big|\sigma_k^j * \Big(\sum_{l\in\mathbb{Z}} P_{k-j+l}^{\text{Cone}} f\Big)\Big| \tag{4.5.61}$$

Remark 4.5.1. The case $l = 0$ corresponds to the case where the phase function of $\widehat{\mu_k^j} = m_k^j$ has a critical point.

To bound the term (4.5.61) on L^p, by the triangle inequality, it suffices to prove that

$$\| \sup_{k,j} |\sigma_k^j * (P_{k-j+l}^{\text{Cone}} f)| \|_p \lesssim 2^{-\gamma_p |l|} \|f\|_p, \tag{4.5.62}$$

for some $\gamma_p > 0$. This decay in l comes from the fact that away from the case $l = 0$ one never sees the critical point of the phase function of $\widehat{\mu_k^j}$.

In the following, we only focus on the case $l = 0$. The case of general $l \in \mathbb{Z}$ follows a similar approach, with the extra twist of involving the non-stationary phase method (integration by parts). We refer to Carlsson, Christ et al. [CCC+86] for details.

Roughly speaking, estimate (4.5.62) follows from interpolating the L^2 bound of $\mathcal{M}_{\text{dyad}}$ with some simple endpoint bound. This interpolation is again in the spirit of Nagel, Stein and Wainger [NSW78]. To enable this, we need to rewrite (4.5.62) in a slightly different way:

$$\sup_{j,k} |\sigma_k^j * P_{k-j}^{\text{Cone}} f| = \sup_{j,k} |\sigma_k^{k-j} * P_j^{\text{Cone}} f| =: \sup_{j,k} |\tilde{\sigma}_k^j * P_j^{\text{Cone}} f|, \tag{4.5.63}$$

where

$$\tilde{\sigma}_k^j := \sigma_k^{k-j}. \tag{4.5.64}$$

We bound (4.5.63) by the square function:

$$\Big(\sum_j \sup_k |\tilde{\sigma}_k^j * (P_j^{\text{Cone}} f)|^2\Big)^{1/2}. \tag{4.5.65}$$

To prove the L^p boundedness of the last expression, we first place it in a more general framework by considering inequalities of the form

$$\Big\|\Big(\sum_j \sup_k |\tilde{\sigma}_k^j * (P_j^{\text{Cone}} f)|^{q_1}\Big)^{1/q_1}\Big\|_{q_2} \lesssim \|(\sum_j |P_j^{\text{Cone}} f|^{q_1})^{1/q_1}\|_{q_2}. \tag{4.5.66}$$

For the case $q_1 = \infty$ and $q_2 = 2$, by going back to $\sigma = \mu - (\mu - \sigma)$, it is not difficult to see that the left hand side of the last expression can be bounded by $\|\mathcal{M}_{\text{dyad}} f\|_2 + \|M_S f\|_2$, which, by the L^2 bound in the previous part, can be further bounded by $\|f\|_2$.

Hence, it remains to prove that for each fixed j and each $q_1 > 1$, one has

$$\|\sup_k |\tilde{\sigma}_k^j * (P_j^{\text{Cone}} f)|\|\|_{q_1} \lesssim \|P_j^{\text{Cone}} f\|_{q_1}. \tag{4.5.67}$$

That (4.5.65) is bounded on L^p for all $p > 1$ follows from an iterative interpolation argument in the spirit of [NSW78].

Now we prove (4.5.67).

We bound its left hand side by a square function, and prove that

$$\Big\|\Big(\sum_k |\tilde{\sigma}_k^j * (P_j^{\text{Cone}} f)|^2\Big)^{1/2}|\Big\|_{q_1} \lesssim \|P_j^{\text{Cone}} f\|_{q_1}. \tag{4.5.68}$$

By a simple anisotropic scaling argument, it suffices to consider the case $j = 0$.

Notice that working with the projection operator $P_0^{\text{Cone}} f$ means that we are within a frequency cone having $\{(\xi, \eta) : \xi \sim \eta\}$. We continue by performing a finer frequency decomposition, as follows: denote by $P_{0,k}^{\text{Cone}}$ the frequency projection into the region $\xi \sim 2^k, \eta \sim 2^k$; then (4.5.68) is equivalent to

$$\Big\|\Big(\sum_k |\tilde{\sigma}_k^0 * (\sum_{l \in \mathbb{Z}} P_{0,k+l}^{\text{Cone}} f)|^2\Big)^{1/2}|\Big\|_{q_1} \lesssim \|P_0^{\text{Cone}} f\|_{q_1}. \tag{4.5.69}$$

By the triangle inequality, it suffices to prove for some $\lambda_{q_1} > 0$ that

$$\Big\|\Big(\sum_k |\tilde{\sigma}_k^0 * (P_{0,k+l}^{\text{Cone}} f)|^2\Big)^{1/2}|\Big\|_{q_1} \lesssim 2^{-\lambda_{q_1}|l|}\|P_0^{\text{Cone}} f\|_{q_1}. \tag{4.5.70}$$

The above estimate for $q_1 = 2$ follows simply from Plancherel's theorem and the mean zero property of $\tilde{\sigma}_k^0$. Thus, by interpolation and standard Littlewood-Paley theory, it suffices to prove that

$$\Big\|\Big(\sum_k |\tilde{\sigma}_k^0 * (P_{0,k+l}^{\text{Cone}} f)|^2\Big)^{1/2}|\Big\|_{q_1} \lesssim \Big\|\Big(\sum_k |P_{0,k+l}^{\text{Cone}} f|^2\Big)^{1/2}\Big\|_{q_1}. \tag{4.5.71}$$

As before, we first consider this estimate in a more general framework of inequalities of the form

$$\Big\|\Big(\sum_k |\tilde{\sigma}_k^0 * (P_{0,k+l}^{\text{Cone}} f)|^{p_1}\Big)^{1/p_1}|\Big\|_{q_1} \lesssim \Big\|\Big(\sum_k |P_{0,k+l}^{\text{Cone}} f|^{p_1}\Big)^{1/p_1}\Big\|_{q_1}. \tag{4.5.72}$$

The case $p_1 = \infty$ and $q_1 = 2$ follows from the L^2 boundedness of $\mathcal{M}_{\text{dyad}}$ while the case $p_1 = q_1 > 1$ is trivial. Applying the usual interpolation trick we obtain that (4.5.71) holds for all $q_1 > 1$. $\square$

4.6 A local smoothing estimate

Let $\psi_0 : \mathbb{R} \to \mathbb{R}$ be a non-negative smooth function supported on $[1/2, 3]$ with $\psi_0(t) = 1$ for each $t \in [1, 2]$. Moreover, let $\varphi_0 : \mathbb{R} \to \mathbb{R}$ be a non-negative smooth function supported on $[-3, 3]$ with $\varphi_0(t) = 1$ for each $t \in [-1, 1]$. For a given positive real number u, let A_u denote the averaging operator

$$A_u f(x, y) := \int_{\mathbb{R}} f(x - ut, y - ut^\alpha)\psi_0(t)dt. \tag{4.6.1}$$

Here α is a positive real number with $\alpha \neq 1$. For $k \in \mathbb{Z}$, let P_k denote a Littlewood-Paley projection operator on the plane, say

$$P_k f(x, y) := \int_{\mathbb{R}^2} e^{ix\xi + iy\eta} \hat{f}(\xi, \eta)\psi_0\Big(\frac{1}{2^k}\Big(\xi^2 + \eta^2\Big)^{\frac{1}{2}}\Big)d\xi d\eta. \tag{4.6.2}$$

Then we have

Theorem 4.6.1. *Let $k \in \mathbb{N}$ be a positive integer. For each positive $\alpha \neq 1$, and each $p > 2$, there exists $\gamma_{p,\alpha} > 0$ such that*

$$\| \sup_{u \in [1,2]} |A_u P_k f| \|_p \lesssim 2^{-\gamma_{p,\alpha} \cdot k} \|f\|_p. \tag{4.6.3}$$

Here the implicit constant depends only on p and α.

In this section, we will prove Theorem 4.6.1 by reducing it to a decoupling inequality for cones (see Proposition 4.6.3) due to Bourgain [Bou13] and Bourgain and Demeter [BD15]. This follows the approach of Wolff [Wol00]. We will then provide a proof of the relevant decoupling inequality in the next section.

4.6.1 Several reductions

First of all, by applying the change of variable $t^\alpha \to s$, we see that it suffices to consider the case $\alpha > 1$. Second, to simplify our presentation, we will only work on the case $\alpha = 2$. The other values of $\alpha > 1$ can be handled in a similar way.

We take the Fourier transform of $A_u f$:

$$\widehat{A_u f}(\xi, \eta) = \hat{f}(\xi, \eta) \int_{\mathbb{R}} e^{iut\xi + iut^\alpha \eta}\psi_0(t)dt. \tag{4.6.4}$$

By a stationary phase computation (for instance see page 360 in Stein [Ste93] or Lemma 1.2 in Iosevich [Ios94]),

$$\int_{\mathbb{R}} e^{it\xi + it^2\eta}\psi_0(t)dt = a(\xi, \eta)e^{i\frac{\xi^2}{\eta}} + a_\infty(\xi, \eta). \tag{4.6.5}$$

Here $a_\infty(\xi,\eta)$ is a smooth symbol, and $a(\xi,\eta)$ is a symbol belonging to the class $S^{-\frac{1}{2}}$ and is supported on

$$\Big\{(\xi,\eta) : -10 \le \frac{\xi}{\eta} \le -\frac{1}{10}\Big\}. \tag{4.6.6}$$

The contribution from the smooth symbol $a_\infty(\xi,\eta)$ can be handled via a standard argument, see for instance Stein [Ste76]. We omit the details.

Now we turn to the former term on the right hand side of (4.6.5). Define

$$Tf(u,x,y) := \int_{\mathbb{R}^2} \hat{f}(\xi,\eta)a(u\xi,u\eta)e^{ix\xi+iy\eta+iu\frac{\xi^2}{\eta}}d\xi d\eta, \tag{4.6.7}$$

and

$$T_k f(u,x,y) := \psi_0(u)\cdot T\circ P_k f(u,x,y). \tag{4.6.8}$$

Moreover, we define a variant of T_k given by

$$S_k f(u,x,y) := \psi_0(u)\cdot\int_{\mathbb{R}^2} \widehat{P_k f}(\xi,\eta)\varphi_0\Big(-\frac{\xi}{100\eta}\Big)(1+\xi^2+\eta^2)^{-\frac{1}{4}}e^{ix\xi+iy\eta+iu\frac{\xi^2}{\eta}}d\xi d\eta. \tag{4.6.9}$$

We will prove

Theorem 4.6.2. *Let $k\in\mathbb{N}$ be a positive integer. For each $p>2$, there exists $\gamma_p>0$ such that*

$$\|S_k f\|_{L^p(\mathbb{R}^3)} \lesssim 2^{-(\frac{1}{p}+\gamma_p)k}\|P_k f\|_p. \tag{4.6.10}$$

We will apply Theorem 4.6.2 to prove

$$\|\Delta_u^{\frac{s}{2}} T_k f\|_{L^p(\mathbb{R}^3)} \lesssim 2^{-(\frac{1}{p}+\gamma_p)k+sk}\|f\|_p. \tag{4.6.11}$$

Here

$$\Delta_u^{\frac{s}{2}} F(u,x,y) := \int_{\mathbb{R}} e^{iu\tau}(1+|\tau|^2)^{\frac{s}{2}}\tilde{F}(\tau,x,y)d\tau, \tag{4.6.12}$$

and $\tilde{F}$ denotes the partial Fourier transform taken in the u variable only. The desired estimate (4.6.3) follows from (4.6.11) via the fractional L^∞ Sobolev embedding inequality

$$\|h\|_{L^\infty_u(\mathbb{R})} \lesssim \|\Delta_u^{\frac{s}{2}} h\|_{L^p_u(\mathbb{R})},$$

whenever $2\le p<\infty$ and $s>\frac{1}{p}$.

Now we come to the proof of (4.6.11). By taking the partial Fourier transform of $S_k f$ in the u variable, we see that

$$(S_k f)^\sim(\tau,x,y) = \int_{\mathbb{R}^2} \widehat{P_k f}(\xi,\eta)\varphi_0\Big(-\frac{\xi}{100\eta}\Big)(1+\xi^2+\eta^2)^{-\frac{1}{4}}\check{\psi}_0\Big(\tau-\frac{\xi^2}{\eta}\Big)e^{ix\xi+iy\eta}d\xi d\eta$$

is supported in the region $\{\tau\in\mathbb{R} : \tau\sim 2^k\}$. Hence by Young's inequality and Theorem 4.6.2, we obtain

$$\|\Delta_u^{\frac{s}{2}} S_k f\|_{L^p(\mathbb{R}^3)} \lesssim 2^{sk}\|S_k f\|_{L^p(\mathbb{R}^3)} \lesssim 2^{-(\frac{1}{p}+\gamma_p)k+sk}\|f\|_p. \tag{4.6.13}$$

The estimate (4.6.11) follows from (4.6.13) via the standard Hörmander-Mikhlin multiplier theorem, by realizing that

$$\Big|\partial^\alpha_{\xi,\eta}\frac{a(u\xi,u\eta)}{(1+\xi^2+\eta^2)^{-\frac14}}\Big| \lesssim_\alpha (1+|\xi|+|\eta|)^{-|\alpha|} \text{ for all multi-indices } \alpha\in\mathbb{N}_0^2,$$

uniformly in $u\in[1,2]$.

4.6.2 Proof of Theorem 4.6.2 via a decoupling inequality

Define a rescaled version of the operator S_k from (4.6.9) by

$$E_kf(u,x,y) := \psi_k(u)\cdot\int_{\mathbb{R}^2}\widehat{P_0f}(\xi,\eta)\varphi_0\Big(-\frac{\xi}{100\eta}\Big)(1+\xi^2+\eta^2)^{-\frac14}e^{ix\xi+iy\eta+iu\frac{\xi^2}{\eta}}d\xi d\eta, \tag{4.6.14}$$

where $\psi_k(u) := \psi_0(\frac{u}{2^k})$. By applying a change of variables

$$\xi\to 2^k\xi, \eta\to 2^k\eta \tag{4.6.15}$$

to S_kf and the desired estimate (4.6.10), we see that (4.6.10) is equivalent to

$$\|E_kf\|_{L^p(\mathbb{R}^3)} \lesssim 2^{\frac{k}{2}-\gamma_p\cdot k}\|f\|_p, \text{ for some } \gamma_p>0. \tag{4.6.16}$$

We apply a conical frequency decomposition for P_0f. Let

$$\Sigma_k := \Big\{\frac{j}{2^{\frac{k}{2}}} : j\in\mathbb{Z}, -100\cdot 2^{\frac{k}{2}}\le j\le 100\cdot 2^{\frac{k}{2}}\Big\} \tag{4.6.17}$$

and decompose P_0f by writing

$$\widehat{P_0f} = \sum_{\theta\in\Sigma_k}\widehat{f}_\theta := \sum_{\theta\in\Sigma_k}\psi_0\Big(2^{\frac{k}{2}}\Big(\frac{\xi}{\eta}-\theta\Big)\Big)\cdot\widehat{P_0f}. \tag{4.6.18}$$

The estimate (4.6.10) in Theorem 4.6.2 follows immediately from the following two results.

Proposition 4.6.3 (Bourgain [Bou13], Bourgain and Demeter [BD15])**.** *For each $2\le p\le 4$ and each $\epsilon>0$, we have*

$$\|E_kf\|_{L^p(\mathbb{R}^3)} \lesssim 2^{\frac{k}{2}(\frac12-\frac1p)+\epsilon}(\sum_{\theta\in\Sigma_k}\|E_kf_\theta\|^p_{L^p(\mathbb{R}^3)})^{\frac1p}. \tag{4.6.19}$$

Here the implicit constant depends only on p, α and ϵ.

Lemma 4.6.4. *For each $p\ge 2$, we have*

$$\Big(\sum_{\theta\in\Sigma_k}\|E_kf_\theta\|^p_{L^p(\mathbb{R}^3)}\Big)^{\frac1p} \lesssim 2^{\frac{k}{p}}\|f\|_{L^2(\mathbb{R}^2)}. \tag{4.6.20}$$

The proof of Proposition 4.6.3 can be found in Bourgain [Bou13] and the last section of Bourgain and Demeter [BD15]. Note that here we only rely on a weak form of the decoupling inequality; that is, the exponent p is in the restricted range $2\le p\le 4$, but not $4<p\le 6$. The latter is part of the main content in Bourgain and Demeter [BD15].

As the proof of Proposition 4.6.3 is short and can be made (essentially) self-contained, we will provide it in the next section.

The proof of Lemma 4.6.4 is standard. Here we sketch the proof. By interpolation, it suffices to prove (4.6.20) for $p = 2$ and $p = \infty$. At $p = 2$, the proof is via a simple almost orthogonality argument. For each fixed $u \in [0, 2^k]$, we have

$$\Big(\sum_{\theta\in\Sigma_k} \|E_k f_\theta(u,\cdot)\|^2_{L^2_{x,y}(\mathbb{R}^2)}\Big)^{\frac{1}{2}} \lesssim \Big(\sum_{\theta\in\Sigma_k} \|f_\theta\|^2_{L^2(\mathbb{R}^2)}\Big)^{\frac{1}{2}} \lesssim \|f\|_{L^2(\mathbb{R}^2)}. \tag{4.6.21}$$

In the end, we integrate in u and collect the factor $2^{\frac{k}{2}}$.

At $p = \infty$, by Young's inequality, it suffices to show that

$$\Big\|\int_{\mathbb{R}^2} \varphi_0\Big(-\frac{\xi}{100\eta}\Big)\psi_0(\eta)\psi_0\Big(2^{\frac{k}{2}}\Big(\frac{\xi}{\eta}-\theta\Big)\Big)e^{ix\xi+iy\eta+iu\frac{\xi^2}{\eta}}d\xi d\eta\Big\|_{L^1_{x,y}(\mathbb{R}^2)} \lesssim 1, \tag{4.6.22}$$

uniformly in θ and $u \in [0, 2^k]$. Without loss of generality, we take $\theta = 0$. By the change of variables $2^{\frac{k}{2}}\xi \to \xi$, the last estimate is equivalent to

$$\Big\|\int_{\mathbb{R}^2} \psi_0(\eta)\psi_0\Big(\frac{\xi}{\eta}\Big)e^{ix\xi+iy\eta+iu\frac{\xi^2}{\eta}}d\xi d\eta\Big\|_{L^1_{x,y}(\mathbb{R}^2)} \lesssim 1, \tag{4.6.23}$$

uniformly in $u \in [0, 1]$. However, by the non-stationary phase method, when $|x|+|y| \gg 1$, we always have

$$\Big|\int_{\mathbb{R}^2} \psi_0(\eta)\psi_0\Big(\frac{\xi}{\eta}\Big)e^{ix\xi+iy\eta+iu\frac{\xi^2}{\eta}}d\xi d\eta\Big| \lesssim \frac{1}{|x|^{10}+|y|^{10}}.$$

This further implies the desired estimate (4.6.23).

4.7 The proof of a decoupling inequality

For a dyadic interval $\Delta \subset [0, 1]$, define the extension operator associated with Δ and the parabola (ξ, ξ^2) by

$$E_\Delta g(x) := \int_\Delta g(\xi)e^{ix_1\xi+ix_2\xi^2}d\xi. \tag{4.7.1}$$

In this section we will prove

Theorem 4.7.1 (Bourgain [Bou13]). *For each $2 \le p \le 4$ and $\epsilon > 0$, we have*

$$\|E_{[0,1]}g\|_{L^p(\mathbb{R}^2)} \lesssim \delta^{-(\frac{1}{2}-\frac{1}{p}+\epsilon)}\Big(\sum_{\Delta\subset[0,1];l(\Delta)=\delta} \|E_\Delta g\|^p_{L^p(\mathbb{R}^2)}\Big)^{\frac{1}{p}}. \tag{4.7.2}$$

Proposition 4.6.3 follows from Theorem 4.7.1 via Fubini's theorem and an iteration argument. This iteration first appeared in the work of Pramanik and Seeger [PS07]. We refer to Proposition 8.1 in Bourgain and Demeter [BD15] for the details.

In the remaining part, we will provide a proof of Theorem 4.7.1. First of all, by a simple localization argument, and by Hölder's inequality, the estimate (4.7.2) follows from

$$\|E_{[0,1]}g\|_{L^p(w_B)} \lesssim \delta^{-\epsilon}\Big(\sum_{\Delta\subset[0,1];l(\Delta)=\delta} \|E_\Delta g\|^2_{L^p(w_B)}\Big)^{\frac{1}{2}}, \tag{4.7.3}$$

for each ball B of radius δ^{-2}. Here w_B is a weight associated with B given by

$$w_B(x) := \Big(1 + \frac{\|x - c_B\|}{\delta^{-2}}\Big)^{-N},$$

for a large integer N which will not be specified. To prove (4.7.3) for all $2 \le p \le 4$, by interpolation with the trivial bound at $p = 2$, it suffices to look at the case $p = 4$. We refer to Garrigós and Seeger [GS10] for the details of such an interpolation argument.

The proof of (4.7.3) for $p = 4$ will be accomplished in three steps, which correspond to the following three subsections.

4.7.1 A bilinear restriction estimate for the parabola

The following proposition follows simply via Plancherel's theorem.

Proposition 4.7.2. *Let $R_1, R_2 \subset [0,1]$ be two dyadic intervals with dist$(R_1, R_2) \ge \nu$ for some $\nu > 0$. We have the bilinear restriction estimate*

$$\Big\||E_{R_1}g_1 E_{R_2}g_2|^{\frac{1}{2}}\Big\|_{L^4(\mathbb{R}^2)} \lesssim_\nu \|g_1\|_2^{\frac{1}{2}}\|g_2\|_2^{\frac{1}{2}}. \tag{4.7.4}$$

The details are left to the interested reader.

4.7.2 A bilinear decoupling inequality

Recall that R_1 and R_2 are two dyadic intervals whose distance is not smaller than ν.

Proposition 4.7.3. *We have a bilinear version of the desired decoupling inequality* (4.7.3)*:*

$$\Big\|(E_{R_1}g_1 E_{R_2}g_2)^{\frac{1}{2}}\Big\|_{L^4(w_B)} \lesssim_\nu \Big(\prod_{j=1}^{2}\sum_{\Delta\subset R_j:l(\Delta)=\delta} \|E_\Delta g_j\|^2_{L^4(w_B)}\Big)^{\frac{1}{4}}, \tag{4.7.5}$$

for each ball B of radius δ^{-2}.

We start by introducing some notation. Let τ_j be the δ^2-neighborhood of the parabola that lies on top of R_j; that is,

$$\tau_j := \{(\xi, \xi^2 + \eta) : \xi \in R_j, |\eta| \le \delta^2\}.$$

We let $\mathcal{P}_j$ be a finitely overlapping cover of τ_j with curved regions θ of the form

$$\theta = \{(\xi, \xi^2 + \eta) : \xi \in [c, c+\delta], |\eta| \le \delta^2\} \text{ for some constant } c.$$

For a function f supported on τ_j, we let f_θ denote the restriction of f to θ.

Let f_1 and f_2 be two functions supported on τ_1 and τ_2, respectively. The bilinear estimate (4.7.4) implies

$$\|(\hat{f}_1\hat{f}_2)^{\frac{1}{2}}\|_{L^4(\mathbb{R}^2)} \lesssim \delta \|f_1\|^{\frac{1}{2}}_{L^2(\mathbb{R}^2)}\|f_2\|^{\frac{1}{2}}_{L^2(\mathbb{R}^2)}. \tag{4.7.6}$$

This can be proven by applying (4.7.4) to each fiber $\{(\xi, \xi^2 + \eta_j)\}$ for fixed η_j $(j = 1, 2)$, and then applying Minkowski's inequality to the variables η_j. Now we apply the L^2 orthogonality and a simple localization argument to further obtain

$$\|(\hat{f}_1\hat{f}_2)^{\frac{1}{2}}\|_{L^4(w_B)} \lesssim \delta\Big(\prod_{j=1}^{2}\sum_{\theta\in\mathcal{P}_j}\|\widehat{f_{j,\theta}}\|^2_{L^2(w_B)}\Big)^{\frac{1}{4}}. \tag{4.7.7}$$

In the end, we apply Hölder's inequality to the right hand side of the last expression

$$\|(\hat{f}_1\hat{f}_2)^{\frac{1}{2}}\|_{L^4(w_B)} \lesssim \Big(\prod_{j=1}^{2}\sum_{\theta\in\mathcal{P}_j}\|\widehat{f_{j,\theta}}\|^2_{L^4(w_B)}\Big)^{\frac{1}{4}}. \tag{4.7.8}$$

This implies the desired estimate in Proposition 4.7.3 by taking $f_j = E_{R_j}g_j$.

4.7.3 Bilinear decoupling implies linear decoupling

We come to the final step of proving the desired decoupling inequality (4.7.3). The idea is that the bilinear decoupling inequality in Proposition 4.7.3 will imply (4.7.3). This is done via a simple version of the Bourgain-Guth argument from [BG11].

We proceed with the details. Fix a large constant $K \ll \delta^{-1}$. We split the interval $[0, 1]$ into smaller intervals of length K^{-1}. We use α to denote such an interval. Then on each ball B_K of radius K, the function $|E_\alpha g|$ behaves like a constant. To sketch the argument we will simply write $|E_\alpha g|(B_K)$ to denote the value of that constant.

For each given B_K, we let α^* denote the interval that maximizes $|E_{\alpha^*}g|(B_K)$. We look at the collection of α, with $dist(\alpha^*, \alpha) \geq \frac{1}{K}$, and

$$|E_\alpha g|(B_K) \geq \frac{1}{10K}|E_{\alpha^*}g(B_k)|.$$

There are two cases. The first case is that this collection is empty. Then

$$|E_{[0,1]}g|(x) \lesssim |E_{\alpha^*}g|(B_k), \text{ for each } x \in B_k. \tag{4.7.9}$$

The second case is that this collection contains at least one element. Call it α^{**}. Then

$$|E_{[0,1]}g|(B_K) \lesssim K^3|E_{\alpha^*}g|^{\frac{1}{2}}(B_K)|E_{\alpha^{**}}g|^{\frac{1}{2}}(B_K). \tag{4.7.10}$$

Putting these two estimates together, we obtain

$$\begin{aligned}\|E_{[0,1]}g\|_{L^4(w_{B_K})} \leq C\Big(\sum_{\alpha:l(\alpha)=\frac{1}{K}} \|E_\alpha g\|^2_{L^4(w_{B_K})}\Big)^{\frac{1}{2}} \\ + C\cdot K^3\Big(\sum_{dist(\alpha_1,\alpha_2)\geq\frac{1}{K}} \Big\||E_{\alpha_1}g|^{\frac{1}{2}}|E_{\alpha_2}g|^{\frac{1}{2}}\Big\|^2_{L^4(w_{B_K})}\Big)^{\frac{1}{2}},\end{aligned} \tag{4.7.11}$$

for a universal constant C. We raise both sides of this estimate to the fourth power, and sum over all balls B_K inside B, a ball of radius δ^{-2}, to obtain that (4.7.11) indeed holds true with B_K being replaced by B. Now we apply the bilinear decoupling inequality that has been proven in the previous subsection to obtain

$$\|E_{[0,1]}g\|_{L^4(w_B)} \leq C\Big(\sum_{\alpha:l(\alpha)=\frac{1}{K}} \|E_\alpha g\|^2_{L^4(w_B)}\Big)^{\frac{1}{2}} + C_K\cdot K^{10}\Big(\sum_{\Delta:l(\Delta)=\delta} \|E_\Delta g\|^2_{L^4(w_B)}\Big)^{\frac{1}{2}}, \tag{4.7.12}$$

for a possibly larger C and a constant C_K depending on K. In the end, by invoking a parabolic rescaling, we iterate the last estimate for $\log_K(\frac{1}{\delta})$ many times, and obtain

$$\|E_{[0,1]}g\|_{L^p(w_B)} \leq C^{\log_K(\frac{1}{\delta})}\cdot C_K\cdot K^{10}\Big(\sum_{\Delta:l(\Delta)=\delta} \|E_\Delta g\|^2_{L^p(w_B)}\Big)^{\frac{1}{2}}. \tag{4.7.13}$$

We just need to observe that by choosing K large enough, compared with C, we can always obtain the desired estimate (4.7.3).

Bibliography

[Bat13] M. Bateman. Single annulus L^p estimates for Hilbert transforms along vector fields. *Rev. Mat. Iberoam.*, 29(3):1021–1069, 2013.

[BD15] J. Bourgain and C. Demeter. The proof of the l^2 decoupling conjecture. *Ann. of Math. (2)*, 182(1):351–389, 2015.

[BG11] J. Bourgain and L. Guth. Bounds on oscillatory integral operators based on multilinear estimates. *Geom. Funct. Anal.*, 21(6):1239–1295, 2011.

[Bou86] J. Bourgain. Averages in the plane over convex curves and maximal operators. *J. Anal. Math.*, 47:69–85, 1986.

[Bou89] J. Bourgain. A remark on the maximal function associated to an analytic vector field. In *Analysis at Urbana, Vol. I (Urbana, IL, 1986–1987)*, volume 137 of *London Math. Soc. Lecture Note Ser.*, pages 111–132. Cambridge Univ. Press, Cambridge, 1989.

[Bou13] J. Bourgain. Moment inequalities for trigonometric polynomials with spectrum in curved hypersurfaces. *Israel J. Math.*, 193(1):441–458, 2013.

[BT13] M. Bateman and C. Thiele. L^p estimates for the Hilbert transforms along a one-variable vector field. *Anal. PDE*, 6(7):1577–1600, 2013.

[Car66] L. Carleson. On convergence and growth of partial sums of Fourier series. *Acta Math.*, 116:135–157, 1966.

[CCC$^+$86] H. Carlsson, M. Christ, A. Córdoba, J. Duoandikoetxea, J. L. Rubio de Francia, J. Vance, S. Wainger, and D. Weinberg. L^p estimates for maximal functions and Hilbert transforms along flat convex curves in $\mathbf{R}^2$. *Bull. Amer. Math. Soc. (N.S.)*, 14(2):263–267, 1986.

[Chr85] M. Christ. Hilbert transforms along curves - I. Nilpotent groups. *Ann. of Math.*, 122:575–596, 1985.

[CNSW99] M. Christ, A. Nagel, E. M. Stein, and S. Wainger. Singular and maximal Radon transforms: analysis and geometry. *Ann. of Math. (2)*, 150(2):489–577, 1999.

[CSWW99] A. Carbery, A. Seeger, S. Wainger, and J. Wright. Classes of singular integral operators along variable lines. *J. Geom. Anal.*, 9(4):583–605, 1999.

[DPL14] F. Di Plinio and A. Lerner. On weighted norm inequalities for the Carleson and Walsh-Carleson operator. *J. Lond. Math. Soc. (2)*, 90(3):654–674, 2014.

[DRdF86] J. Duoandikoetxea and J. L. Rubio de Francia. Maximal and singular integral operators via Fourier transform estimates. *Invent. Math.*, 84(3):541–561, 1986.

[DS15] C. Demeter and P. Silva. Some new light on a few classical results. *Colloq. Math.*, 140(1):129–147, 2015.

[DT10] C. Demeter and C. Thiele. On the two dimensional Bilinear Hilbert Transform. *Amer. J. Math.*, 132(1):201–256, 2010.

[DT15] Y. Do and C. Thiele. Lp theory for outer measures and two themes of Lennart Carleson united. *Bull. Amer. Math. Soc. (N.S.)*, 52:249–296, 2015.

[Duo01] J. Duoandikoetxea. *Fourier analysis*, volume 29 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. Translated and revised from the 1995 Spanish original by D. Cruz-Uribe.

[Fef73] C. Fefferman. Pointwise convergence of Fourier series. *Ann. of Math. (2)*, 98:551–571, 1973.

[FS71] C. Fefferman and E. M. Stein. Some maximal inequalities. *Amer. J. Math.*, 93:107–115, 1971.

[GHLR16] S. Guo, J. Hickman, V. Lie, and J. Roos. Maximal operators and Hilbert transforms along variable non-flat homogeneous curves. 2016. arXiv:1610.05203, to appear in Proc. London Math. Soc., published online under doi: 10.1112/plms.12037.

[GPRY16] S. Guo, L. B. Pierce, J. Roos, and P.-L. Yung. Polynomial Carleson operators along monomial curves in the plane. 2016. arXiv:1605.05812, to appear in J. Geom. Anal., published "Online First" under doi:10.1007/s12220-017-9790-7.

[Gra14a] L. Grafakos. *Classical Fourier Analysis*, volume 249 of *Graduate Texts in Mathematics*. Springer New York, 3rd edition, 2014.

[Gra14b] L. Grafakos. *Modern Fourier Analysis*, volume 250 of *Graduate Texts in Mathematics*. Springer New York, 2014.

[GS10] G. Garrigós and A. Seeger. A mixed norm variant of Wolff's inequality for paraboloids. In *Harmonic analysis and partial differential equations*, volume 505 of *Contemp. Math.*, pages 179–197. Amer. Math. Soc., Providence, RI, 2010.

[GTT04] L. Grafakos, T. Tao, and E. Terwilleger. L^p bounds for a maximal dyadic sum operator. *Math. Z.*, 246(1-2):321–337, 2004.

[Guo15] S. Guo. Hilbert transform along measurable vector fields constant on lipschitz curves: L^2 boundedness. *Anal. PDE*, 8(5):1263–1288, 2015.

[Guo16] S. Guo. Oscillatory integrals related to Carleson's theorem: fractional monomials. *Commun. Pure Appl. Anal.*, 15(3):929–946, 2016.

[Guo17a] S. Guo. Hilbert transform along measurable vector fields constant on Lipschitz curves: L^p boundedness. *Trans. Amer. Math. Soc.*, 369(4):2493–2519, 2017.

[Guo17b] S. Guo. Single annulus estimates for the variation-norm Hilbert transforms along Lipschitz vector fields. *Proc. Amer. Math. Soc.*, 145(2):601–615, 2017.

[HKY09] S. Hong, J. Kim, and C. W. Yang. Maximal operators associated with vector polynomials of lacunary coefficients. *Canad. J. Math.*, 61(4):807–827, 2009.

[Hun68] R. A. Hunt. On the convergence of Fourier series. In *Orthogonal Expansions and their Continuous Analogues (Proc. Conf., Edwardsville, Ill., 1967)*, pages 235–255. Southern Illinois Univ. Press, Carbondale, Ill., 1968.

[Ios94] A. Iosevich. Maximal operators associated to families of flat curves in the plane. *Duke Math. J.*, 76(2):633–644, 1994.

[Kar07] G. A. Karagulyan. On unboundedness of maximal operators for directional Hilbert transforms. *Proc. Amer. Math. Soc.*, 135(10):3133–3141, 2007.

[Kol26] A. N. Kolmogorov. Une série de Fourier-Lebesgue divergente partout. *C. R. Acad. Sci. Paris*, 183:1327–1328, 1926.

[KS24] A. N. Kolmogorov and G. A. Seliverstov. Sur la convergence des séries de Fourier. *C.R. Acad. Sci. Paris*, 178:303–306, 1924.

[KT80] C. Kenig and P. A. Tomas. Maximal operators defined by Fourier multipliers. *Studia Math.*, 68:79–83, 1980.

[Lac04] M. T. Lacey. Carleson's Theorem: Proof, Complements, Variations. *Publ. Mat.*, 48(2):251–307, 2004.

[Lie09] V. Lie. The (weak–L^2) Boundedness of the Quadratic Carleson Operator. *Geom. Funct. Anal.*, 19:457–497, 2009.

[Lie11] V. Lie. The Polynomial Carleson Operator. 2011. arXiv:1105.4504.

[Lie15a] V. Lie. On the boundedness of the bilinear Hilbert transform along "non-flat" smooth curves. *Amer. J. Math.*, 137(2):313–363, 2015.

[Lie15b] V. Lie. On the Boundedness of the Bilinear Hilbert Transform along "non-flat" smooth curves. The Banach Triangle case (L^r, $1 \leq r < \infty$). 2015. arXiv:1512.09356, to appear in Rev. Mat. Iberoam.

[LL06] M. T. Lacey and X. Li. Maximal theorems for the directional Hilbert transform on the plane. *Trans. Amer. Math. Soc.*, 358(9):4099–4117, 2006.

[LL10] M. T. Lacey and X. Li. On a conjecture of E. M. Stein on the Hilbert transform on vector fields. *Mem. Amer. Math. Soc.*, 205(965):viii+72, 2010.

[LT97a] M. T. Lacey and C. Thiele. L^p estimates for the bilinear Hilbert transform. *Proc. Nat. Acad. Sci. U.S.A.*, 94(1):33–35, 1997.

[LT97b] M. T. Lacey and C. Thiele. L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$. *Ann. of Math. (2)*, 146(3):693–724, 1997.

[LT98] M. T. Lacey and C. Thiele. On Calderón's conjecture for the bilinear Hilbert transform. *Proc. Natl. Acad. Sci. USA*, 95(9):4828–4830 (electronic), 1998.

[LT99] M. T. Lacey and C. Thiele. On Calderón's conjecture. *Ann. of Math. (2)*, 149(2):475–496, 1999.

[LT00] M. T. Lacey and C. Thiele. A proof of boundedness of the Carleson operator. *Math. Res. Lett.*, 7(4):361–370, 2000.

[MR98] G. Marletta and F. Ricci. Two-parameter maximal functions associated with homogeneous surfaces in $\mathbf{R}^n$. *Studia Math.*, 130(1):53–65, 1998.

[MSS92] G. Mockenhaupt, A. Seeger, and C. D. Sogge. Wave front sets, local smoothing and Bourgain's circular maximal theorem. *Ann. of Math. (2)*, 136(1):207–218, 1992.

[MST15] M. Mirek, E. M. Stein, and B. Trojan. $\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: Maximal functions and vector-valued estimates. *arXiv:1512.07518*, 2015.

[Mus14] C. Muscalu. Calderón commutators and the Cauchy integral on Lipschitz curves revisited: I. First commutator and generalizations. *Rev. Mat. Iberoam.*, 30(2):727–750, 2014.

[NSW78] A. Nagel, E. M. Stein, and S. Wainger. Differentiation in lacunary directions. *Proc. Nat. Acad. Sci. U.S.A.*, 75(3):1060–1062, 1978.

[NSW79] A. Nagel, E. M. Stein, and S. Wainger. Hilbert transforms and maximal functions related to variable curves. In *Harmonic analysis in Euclidean spaces (Proc. Sympos. Pure Math., Williams Coll., Williamstown, Mass., 1978), Part 1*, Proc. Sympos. Pure Math., XXXV, Part, pages 95–98. Amer. Math. Soc., Providence, R.I., 1979.

[OST+12] R. Oberlin, A. Seeger, T. Tao, C. Thiele, and J. Wright. A variation norm Carleson theorem. *J. Eur. Math. Soc. (JEMS)*, 14(2):421–464, 2012.

[PS07] M. Pramanik and A. Seeger. L^p regularity of averages over curves and bounds for associated maximal operators. *Amer. J. Math.*, 129(1):61–103, 2007.

[PT03] M. Pramanik and E. Terwilleger. A weak L^2 estimate for a maximal dyadic sum operator on $\mathbb{R}^n$. *Illinois J. Math.*, 47(3):775–813, 2003.

[PY15] L. B. Pierce and P.-L. Yung. Polynomial Carleson operators along the paraboloid. 2015. arXiv:1505.03882.

[RdFRT86] J. L. Rubio de Francia, F. J. Ruiz, and J. L. Torrea. Calderón-Zygmund theory for operator-valued kernels. *Adv. in Math.*, 62(1):7–48, 1986.

[See94] A. Seeger. L^2-estimates for a class of singular oscillatory integrals. *Math. Res. Lett.*, 1(1):65–73, 1994.

[Sjö71] P. Sjölin. Convergence almost everywhere of certain singular integrals and multiple Fourier series. *Ark. Mat.*, 9:65–90, 1971.

[SS12] E. M. Stein and B. Street. Multi-parameter singular Radon transforms III: Real analytic surfaces. *Adv. Math.*, 229(4):2210–2238, 2012.

[Ste70] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions.* Monographs in Harmonic Analysis. Princeton University Press, 1970.

[Ste76] E. M. Stein. Maximal functions. I. Spherical means. *Proc. Nat. Acad. Sci. U.S.A.*, 73(7):2174–2175, 1976.

[Ste93] E. M. Stein. *Harmonic Analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of *Princeton Mathematical Series*. Princeton University Press, 1993.

[SW70] E. M. Stein and S. Wainger. The estimation of an integral arising in multiplier transformations. *Studia Math.*, 35:101–104, 1970.

[SW78] E. M. Stein and S. Wainger. Problems in harmonic analysis related to curvature. *Bull. Amer. Math. Soc.*, 84(6):1239–1295, 1978.

[SW01] E. M. Stein and S. Wainger. Oscillatory integrals related to Carleson's theorem. *Math. Res. Lett.*, 8(5-6):789–800, 2001.

[SW03] A. Seeger and S. Wainger. Singular Radon transforms and maximal functions under convexity assumptions. *Rev. Mat. Iberoamericana*, 19(3):1019–1044, 2003.

[Thi06] C. Thiele. *Wave Packet Analysis*. Number 105 in CBMS Regional Conference Series in Mathematics. AMS, 2006.

[Ura16] G. Uraltsev. Variational Carleson embeddings into the upper 3-space. 2016. arXiv:1610.07657.

[Wol00] T. Wolff. Local smoothing type estimates on L^p for large p. *Geom. Funct. Anal.*, 10(5):1237–1288, 2000.