TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 369, Number 4, April 2017, Pages 2493–2519
http://dx.doi.org/10.1090/tran/6750
Article electronically published on May 25, 2016

HILBERT TRANSFORM ALONG MEASURABLE VECTOR FIELDS CONSTANT ON LIPSCHITZ CURVES: L^p BOUNDEDNESS

SHAOMING GUO

ABSTRACT. We prove the L^p ($p > 3/2$) boundedness of the directional Hilbert transform in the plane relative to measurable vector fields which are constant on suitable Lipschitz curves.

1. STATEMENT OF THE MAIN RESULT

In [15] we proved that the Hilbert transforms along measurable vector fields which are constant on a suitable family of Lipschitz curves are bounded in L^2. The main goal of this paper is to generalize the above L^2 bounds to L^p for p other than 2 in the same setting.

Theorem 1.1 (Main Theorem). *There exists an $\epsilon_0 < 1$ such that for any vector field $v : \mathbf{R}^2 \to \mathbf{R}^2$ of the form $(1, u(h))$ where $h : \mathbf{R}^2 \to \mathbf{R}$ is a Lipschitz function such that*

$$\|\nabla h - (1,0)\|_\infty \leq \epsilon_0, \tag{1.1}$$

and $u : \mathbf{R} \to \mathbf{R}$ is a measurable function such that

$$\|u\|_\infty \leq 1, \tag{1.2}$$

the associated Hilbert transform, which is defined as

$$H_v f(x) := \int_{\mathbf{R}} f(x - tv(x))dt/t, \tag{1.3}$$

is bounded in L^p for all $p > 3/2$.

The above result is a Lipschitz perturbation of the following result by Bateman and Thiele in [5], which is further based on Bateman [3], [4], Lacey and Li [24], [25]:

Theorem 1.2 ([5]). *Let $v : \mathbf{R}^2 \to \mathbf{R}^2$ be a one-variable vector field, i.e. a vector field of the form*

$$v(x_1, x_2) = (1, u(x_1)), \tag{1.4}$$

for some measurable function u. Then the associated Hilbert transform is bounded in L^p for all $p > 3/2$.

Received by the editors January 10, 2015 and, in revised form, April 13, 2015.
2010 *Mathematics Subject Classification.* Primary 42B20, 42B25.
Key words and phrases. Singular integral, differentiation theorem, Jones' beta numbers, one-parameter family of paraproducts, Littlewood-Paley theory on Lipschitz curves.

In our Main Theorem, if we take $h(x_1, x_2) = x_1$, then the vector field becomes $(1, u(h(x_1, x_2))) = (1, u(x_1))$, which is a one-variable vector field. However, we have one more assumption that $\|u\|_\infty \leq 1$. To recover the result in Theorem 1.2, we just need to apply the anisotropic scaling

$$x_1 \to x_1, x_2 \to \lambda x_2 \tag{1.5}$$

and a simple limiting argument.

As we state our main result as a Lipschitz perturbation of the one-variable vector fields, in the following we will explain separately why the one-variable vector fields are interesting and why we do the perturbation at the level of the Lipschitz regularity but not others.

First of all, there is an interesting connection between the Hilbert transform along the one-variable vector fields and Carleson's maximal operator, which was observed by Coifman and El Kohen. We review the discussion as presented in [5]. Take a one-variable vector field $v(x_1, x_2) = (1, u(x_1))$, consider the associated Hilbert transform, which is given by

$$H_v f(x_1, x_2) = \int_{\mathbf{R}} f(x_1 - t, x_2 - tu(x_1)) \frac{dt}{t}. \tag{1.6}$$

Denoting by $\widehat{f}$ the partial Fourier transform in the second variable we obtain formally

$$\begin{aligned} &\int f(x_1 - t, x_2 - u(x_1)t) \frac{dt}{t} \\ &= \int e^{ix_2\xi_2} \int \widehat{f}(x_1 - t, \xi_2) e^{iu(x_1)t\xi_2} \frac{dt}{t} d\xi_2. \end{aligned} \tag{1.7}$$

By the Plancherel theorem,

$$\|H_v f\|_2 = \| \int \widehat{f}(x_1 - t, \xi_2) e^{iu(x_1)t\xi_2} \frac{dt}{t} \|_2. \tag{1.8}$$

For each fixed ξ_2, we recognize this to essentially be the linearization of Carleson's maximal operator

$$(Cf)(x) := \sup_{N \in \mathbf{R}} | \int_{\mathbf{R}} f(x - t) e^{iNt} \frac{dt}{t} |. \tag{1.9}$$

Hence the right hand side of (1.8) can be bounded by

$$\|C\hat{f}(x_1, \xi_2)\|_2 \leq C_0 \|\hat{f}(x_1, \xi_2)\|_2 \leq C_0 \|f\|_2, \tag{1.10}$$

for some constant $C_0 > 0$. Moreover, by choosing the function u properly in (1.6), the L^2 boundedness of H_v also implies the L^2 boundedness of Carleson's maximal operator.

Secondly, the class of the one-variable vector fields is also very natural from the viewpoint of the scaling symmetries. We leave the detailed discussion to the next section, where it will also become clear that the equivalence of the L^2 bounds of H_v and Carleson's maximal operator is due to the fact that they enjoy the same symmetries, especially the modulation symmetry.

Next we will explain the appearance of the Lipschitz regularity. For a vector field v and a constant $\epsilon_0 > 0$, if one truncates (1.3) as

$$H_{v,\epsilon_0} f(x) := \int_{-\epsilon_0}^{\epsilon_0} f(x - tv(x)) dt/t, \tag{1.11}$$

then it is reasonable to ask for a pure regularity assumption on v in order to bound H_{v,ϵ_0}. Indeed, a counterexample in [25] based on the Perron tree construction of the Besicovitch-Kakeya set (see [26] and [12]) shows that no bounds are possible for v being Hölder continuous of an exponent less than one, and it is a long-standing open problem in harmonic analysis whether Lipschitz regularity suffices.

At the regularity scale, the only known result is for an analytic vector field, due to Stein and Street [27]. The maximal variant of (1.11) in the same setting was proved much earlier by Bourgain [7]. In the same direction, a prior result for smooth vector fields under certain geometric assumptions appeared in [10]. For some other partial results, see [8], [17], [18].

Lacey and Li [24], exploiting the connection between the Hilbert transform along vector fields and Lacey and Thiele's proof for the boundedness of the bilinear Hilbert transform [21], [22] and Carleson's maximal operator [23], proved that for any measurable vector field v, the operator $H_v P_k$, which is the composition of the Hilbert transform along v with a Littlewood-Paley projection operator P_k for some fixed k, maps L^2 to weak L^2 and L^p to L^p for $p > 2$, uniformly in k. Moreover, conditioning on the boundedness of what they called the Lipschitz-Kakeya maximal operator, Lacey and Li [25] also proved that for any $C^{1+\alpha}$ vector field v with $\alpha > 0$, the operator H_{v,ϵ_0} is bounded in L^2 for some properly chosen ϵ_0.

To our knowledge, the result in the present paper is the first that handles a certain class of Lipschitz vector fields. Indeed, as has also been mentioned in [15], our result has the following corollary, which includes a large class of Lipschitz vector fields:

Corollary 1.3. *For a measurable unit vector field $v_0 : \mathbf{R}^2 \to S^1$, suppose that*
i) *there exists a bi-Lipschitz map $g_0 : \mathbf{R}^2 \to \mathbf{R}^2$ s.t.*

$$v_0(g_0(x_1, x_2)) \text{ is constant in } x_2; \tag{1.12}$$

ii) *there exists $d_0 > 0$ s.t. $\forall x_1 \in \mathbf{R}$,*

$$\angle(\partial_2 g_0(x_1, x_2), \pm v_0(g_0(x_1, x_2))) \geq d_0 \text{ for } x_2\text{-a.e. in } \mathbf{R}. \tag{1.13}$$

Then the associated Hilbert transform H_{v_0} is bounded in L^p for all $p > 3/2$, with the operator norm depending only on p, d_0 and the bi-Lipschitz norm of g_0.

Remark 1.4. Without the assumption that $d_0 > 0$, the operator H_{v_0} might be unbounded in L^p for any $p > 1$. This unboundedness is caused by doing no truncation of the Hilbert transform (in the sense of (1.11)). The counter-example is based on the Besicovitch-Kakeya set construction, which can be found on page 1022 in [4] and on page 7 in [25].

Remark 1.5. The structure theorem for Lipschitz functions by Azzam and Schul in [2] states exactly that any Lipschitz function $u : \mathbf{R}^2 \to \mathbf{R}$ (any Lipschitz unit vector field v_0 in our case) can be precomposed with a bi-Lipschitz function $g_0 : \mathbf{R}^2 \to \mathbf{R}^2$ such that $u \circ g_0$ is Lipschitz in the first coordinate and constant in the second coordinate when restricted to a "large" portion of the domain.

In the end, let us mention the new ingredients that will be used to extend the L^2 bounds in [15]. Recall that in the L^2 case, the crucial ingredients are the use of Jones' beta numbers and the adapted L^2-Littlewood-Paley theory, which is in the spirit of the work on the Cauchy integral on Lipschitz curves (for example see [11]). The techniques used in [15] are the Hilbert space techniques, as we need to use some facts like taking the L^2 norm that work trivially with certain square functions. For this reason, only L^2 bounds are obtained.

In the L^p case for p other than 2, one novelty is that we discovered a new paraproduct, which is indeed a one-parameter family of paraproducts, with each paraproduct living on one Lipschitz level curve of the vector field v. To prove the L^p bounds for the one-parameter family of paraproducts, the difficulty is how to embed each paraproduct into two dimensions without losing orthogonality. To overcome this difficulty, we need to develop an adapted L^p-Littlewood-Paley theory, which again requires a new square function as an intermediate step. This new two-dimensional square function shares some common features with the bi-parameter square function. See the following crucial Lemma 6.1 and Claim 6.7.

Another difference from the L^2 case in [15] is that we will write the proof by using the δ-calculus, which has been used intensively in the Fourier restriction estimates; see [19], [14] and [9] for example. One significant advantage of the δ-calculus, which we will see shortly in the proof, is that it allows us to express everything in terms of the function h from the Main Theorem instead of going back and forth between h and its inverse as in [15]. For example, this can be seen by comparing the crucial definition of the adapted Littlewood-Paley operator associated to the vector fields, namely by comparing Definition 3.3 in [15] with Definition 3.4 in the current paper.

Organization of the paper. In Section 2, we will review the symmetries that were discussed in [5] for the Hilbert transforms along the one-variable vector fields. Moreover, we will introduce one more symmetry which appears only after we allow Lipschitz perturbation of the one-variable vector fields.

In Section 3 we will state the strategy of the proof for the Main Theorem. If we denote by P_k a Littlewood-Paley operator in the second variable, the main observation in Bateman and Thiele's proof is that H_v commutes with P_k. In our case, this is no longer true. To recover the orthogonality, an adapted Littlewood-Paley operator was introduced by the author in [15] (see the following Definition 3.4), which allows us to split the operator H_v into a main term and a commutator term

$$\sum_{k\in\mathbf{Z}} H_v P_k(f) = \sum_{k\in\mathbf{Z}} (H_v P_k(f) - \tilde{P}_k H_v P_k(f) + \tilde{P}_k H_v P_k(f)). \tag{1.14}$$

The new symmetry is used in the definition of $\tilde{P}_k$.

The L^p $(p > 3/2)$ bounds of the main term $\sum_{k\in\mathbf{Z}} \tilde{P}_k H_v P_k(f)$ can be proved essentially by the same argument as in Bateman and Thiele [5], with just minor modifications that we will state in Section 4.

The main novelty is the L^p boundedness of the commutator term

$$\sum_{k\in\mathbf{Z}} (H_v P_k(f) - \tilde{P}_k H_v P_k(f)). \tag{1.15}$$

To achieve this, we will first review the time-frequency decomposition of the operator and the functions in Section 5, and then prove in Section 6 that (1.15) is bounded in L^p for all $p > 1$.

Notation. Throughout this paper, we will write $x \ll y$ to mean that $x \le y/10$, $x \lesssim y$ to mean that there exists a universal constant C s.t. $x \le Cy$, and $x \sim y$ to mean that $x \lesssim y$ and $y \lesssim x$. $\mathbb{1}_E$ will always denote the characteristic function of the set E.

2. Discussion on the symmetries

In this section we will discuss various symmetries that the Hilbert transforms along vector fields have. We will start from the most general case, i.e. the case of measurable vector fields, and then introduce more and more assumptions suggested by the symmetries.

Given an arbitrary measurable vector field $v(x_1, x_2) : \mathbf{R}^2 \to \mathbf{R}^2$, by a renormalization, we assume that it is of the form $v(x_1, x_2) = (1, u(x_1, x_2))$ for some measurable function $u : \mathbf{R}^2 \to \mathbf{R}$. Consider the associated Hilbert transform along this vector field, which is defined as

$$H_v f(x_1, x_2) := \int_{\mathbf{R}} f(x_1 - t, x_2 - tu(x_1, x_2))dt/t. \tag{2.1}$$

Suppose for the moment that we would like to prove the ideal estimate

$$\|H_v f\|_p \leq C\|f\|_p, \tag{2.2}$$

for some $p > 1$ and some universal constant C. We start by studying the symmetries of the above operators, which are, for example, translation, dilation and rotation.

First, it is simple to see that this operator is invariant under translation:

$$x_1 \to x_1 + x_{1,0}, x_2 \to x_2 + x_{2,0}, \tag{2.3}$$

with the vector field being changed to $(1, u(x_1 - x_{1,0}, x_2 - x_{2,0}))$, which is still a measurable function.

Next, we consider the dilation and rotation given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & b \\ d & e \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

which are also supposed to be non-degenerate. By the decomposition of 2×2 matrices, it is not difficult to see that there are in total four generators:

$$A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ \lambda_3 & 1 \end{pmatrix}, D = \begin{pmatrix} 1 & \lambda_4 \\ 0 & 1 \end{pmatrix}.$$

Symmetry $\mathcal{A}$: This is the dilation in the x variable

$$x_1 \to \lambda_1 x_1, x_2 \to x_2. \tag{2.4}$$

Under this change of variables, the vector field is changed to $(1, \frac{1}{\lambda_1} u(\frac{x_1}{\lambda_1}, x_2))$.

Symmetry $\mathcal{B}$: This is the dilation in the y variable

$$x_1 \to x_1, x_2 \to \lambda_2 x_2. \tag{2.5}$$

Under this change of variables, the vector field is changed to $(1, \frac{1}{\lambda_2} u(x_1, \frac{x_2}{\lambda_2}))$.

Symmetry $\mathcal{C}$: This is what Bateman and Thiele called "shearing transformation" in [5]:

$$x_1 \to x_1, x_2 \to x_2 + \lambda_3 x_1, \tag{2.6}$$

with the vector field being changed to $(1, u(x_1, x_2 - \lambda_3 x_1) + \lambda_3)$.

Remark 2.1. In frequency, the change of variables (2.6) corresponds to

$$\xi_1 \to \xi_1 - \lambda_3 \xi_2, \xi_2 \to \xi_2. \tag{2.7}$$

Notice that if we restrict ξ_2 to one single frequency band, say $\xi_2 \sim 1$, then roughly we have

$$\xi_1 \to \xi_1 - \lambda_3, \tag{2.8}$$

which is the translation in the frequency variable ξ_1. Indeed, it will become clear later in the time-frequency decomposition in Section 5 that this is the same as the modulation invariance in Carleson's maximal operator.

Symmetry $\mathcal{D}$: This is the shearing transformation with x_1 and x_2 being exchanged:

$$x_1 \to x_1 + \lambda_4 x_2, x_2 \to x_2, \tag{2.9}$$

with the vector field being changed to $(1, u(x_1 - \lambda_4 x_2, x_2) + \lambda_4)$.

So far we have shown that if we only assume the vector field to be measurable, then the operator (2.1) satisfies the translation symmetry and the Symmetries $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$. Unfortunately, even for Hölder continuous vector field (with exponent less than one), the operator (2.1) might not be bounded in L^p for any $p \geq 1$.

However, if we eliminate the Symmetry $\mathcal{D}$ (or equivalently the Symmetry $\mathcal{C}$) from the class of the measurable vector fields, then it is not difficult to see that a very natural choice is the class of the one-variable measurable vector fields, which enjoys all the other symmetries. Moreover, as has been pointed out before, by exploring the translation symmetry and the symmetries $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, Bateman [4], Bateman and Thiele [5] have proved that the operator (2.1) is bounded in L^p ($\forall p > 3/2$) for arbitrary measurable one-variable vector fields.

Let us explain a bit more from the viewpoint of symmetries why we expect the operator in (2.1), when composed with the Littlewood-Paley projection operator P_k in the vertical variable, to be bounded for the one-variable vector fields. For a one-variable vector $v(x_1, x_2) = (1, u(x_1))$, what Bateman has proved in [4] is

$$\|H_v P_k f\|_p \lesssim \|P_k f\|_p, \forall p \in (1, \infty), \tag{2.10}$$

with the bound being independent of $k \in \mathbf{Z}$. Notice that the operator $H_v P_k$ has the translation symmetry, symmetry $\mathcal{A}$ and $\mathcal{C}$, which corresponds to the translation, dilation and modulation symmetries for Carleson's maximal operator. In this sense, we say that Bateman's result is equivalent to the boundedness of Carleson's maximal operator (the precise calculation is done in (1.7)-(1.10)).

However, we still want to go beyond the one-variable vector fields. Notice that for the Hilbert transform along a general Lipschitz vector field, both Symmetry $\mathcal{C}$ and Symmetry $\mathcal{D}$ might still appear at the same time: for $v(x) = (1, u(x_1, x_2))$ with u being Lipschitz, by applying Symmetry $\mathcal{C}$ with λ_3 being small and Symmetry $\mathcal{D}$ with λ_4 being small, what we get is

$$(1, u(x_1 - \lambda_3 x_2, x_2 - \lambda_4 x_1 - \lambda_3\lambda_4 x_2) + \lambda_3 + \lambda_4), \tag{2.11}$$

which is still a Lipschitz vector field with a comparable Lipschitz constant.

Indeed, by including a Lipschitz perturbation of the one-variable vector fields (see the assumption of the Main Theorem), we bring Symmetry $\mathcal{D}$ into the problem, with the cost that all the symmetries $\mathcal{A}$-$\mathcal{D}$ become "quasi-symmetries". Let us explain what we mean by this: the vector fields that we can handle are of the form $(1, u(h(x)))$, where $\|u\|_\infty \leq 1$ and h is a Lipschitz function satisfying

$$\|\nabla h - (1, 0)\|_\infty \leq \epsilon_0 \ll 1. \tag{2.12}$$

If we apply Symmetry $\mathcal{D}$ with

$$\lambda_4 \ll 1, \tag{2.13}$$

the new vector field $(1, u_{\lambda_4}(h_{\lambda_4}(x)))$ will satisfy

$$\|u_{\lambda_4}\| \leq 2, \|\nabla h_{\lambda_4} - (1,0)\|_\infty \leq 2\epsilon_0; \tag{2.14}$$

i.e. under the action of the symmetry, the assumption on the vector field is preserved up to a factor of two. This explains the notion of "quasi-symmetry".

The new quasi-symmetry $\mathcal{D}$ will be used implicitly in Definition 3.2; hence it will also be used in the crucial Definition 3.4 of the adapted Littlewood-Paley projection operators.

3. Strategy of the proof of the Main Theorem

We first observe that if we denote by Γ the two-ended cone which forms an angle less than $\pi/4$ with the vertical axis, then by the assumption that $|u| \leq 1$, we can w.l.o.g. assume that

$$\text{supp } \hat{f} \subset \Gamma. \tag{3.1}$$

As for functions f with frequency supported on $\mathbf{R}^2 \setminus \Gamma$, we have that

$$H_v f(x) = H_{(1,0)} f(x), \tag{3.2}$$

which is the Hilbert transform along the constant vector field $(1, 0)$. But $H_{(1,0)}$ is bounded by Fubini's theorem and the L^2 boundedness of the Hilbert transform.

The rest of the proof consists of two relatively independent steps. The first step will just be an adaption of Bateman and Thiele's argument in [5] to our case. Our key observation is that both covering lemmas used there (Lemma 7 and Lemma 8) indeed hold true in our setting, from which we can derive the following proposition as a corollary by repeating the rest of the argument in [5].

Proposition 3.1. *Under the same assumptions as in the Main Theorem, we have the square function estimate*

$$\left\| \left(\sum_{k\in\mathbf{Z}} (H_v P_k(f))^2 \right)^{1/2} \right\|_p \lesssim \|f\|_p, \forall p > 3/2, \tag{3.3}$$

where P_k is the k-th Littlewood-Paley projection operator in the vertical direction.

The proof of Proposition 3.1 is postponed to Section 4. The operator P_k used in (3.3) is defined in the following way: if we denote by ψ_0 a smooth function with support on $[-5/2, -1/2] \cup [1/2, 5/2]$ such that

$$\sum_{k\in\mathbf{Z}} \psi_k(t) = 1, \forall t \neq 0, \tag{3.4}$$

and

$$\psi_k(t) := \psi_0(2^{-k}t), \tag{3.5}$$

then

$$P_k f(x_1, x_2) := \int_{\mathbf{R}} f(x_1, x_2 - y_2)\check{\psi}_k(y_2)dy_2. \tag{3.6}$$

For the one-variable vector fields, i.e. vector fields of the form $v(x, y) = (1, u(x))$ for some measurable function u, Bateman and Thiele in [5] used (3.3) and the crucial observation that

$$H_v P_k = P_k H_v \tag{3.7}$$

to conclude the boundedness of H_v. In our case, the identity (3.7) is no longer true; i.e. the orthogonality between $H_v P_k f$ for different $k \in \mathbf{Z}$ is missing.

To recover the orthogonality, an adapted Littlewood-Paley operator along the level curves of the vector field was introduced by the author in [15]. This operator is in the spirit of prior work on the Cauchy integral on Lipschitz curves, but more of a bi-parameter type as we have one-parameter family of level curves.

Here we give an equivalent definition of the operator $\tilde{P}_k$ by using the language of the δ-calculus. The advantage of this new definition is, compared with the one in [15], that it does not necessitate either the change of coordinates or the parametrization of the Lipschitz curves, both of which can be replaced by introducing the following auxiliary function. To do this, we need some notation: for $t \in \mathbf{R}$ we define

$$\Gamma_t := \{x \in \mathbf{R}^2 : h(x) = t\}. \tag{3.8}$$

Moreover, we denote by v_t the value of the vector field v, which is a constant along Γ_t.

Definition 3.2 (Auxiliary function)**.** For every $t \in R$, we define a new function $h_t : \mathbf{R}^2 \to \mathbf{R}$ in such a way that, if for some $y \in \Gamma_t$ we have

$$z - y = d \cdot v_t \tag{3.9}$$

for some $d \in \mathbf{R}$, then we set $h_t(z) = d$.

Lemma 3.3. *Under the above notation, it holds that*

$$|\nabla h_t| \sim 1, \ \textit{a.e. in } \mathbf{R}^2, \tag{3.10}$$

where the constant is independent of $t \in \mathbf{R}$.

Proof of Lemma 3.3. For a given $t \in \mathbf{R}$, we first rotate the coordinate such that $v_t = (1, 0)$. In this new coordinate, we parametrize the curve Γ_t by $x = g_t(y)$ with $y \in \mathbf{R}$. Hence by Definition 3.2, for a point $z = (z_1, z_2) \in \mathbf{R}^2$, we have

$$h_t(z) = z_1 - g_t(z_2). \tag{3.11}$$

It has been proved in Lemma 3.2 [15] that $\|g_t(\cdot)\|_{Lip} \lesssim 1$; hence the estimate (3.10) follows from a direct calculation of the gradient of h_t. □

Definition 3.4 (Adapted Littlewood-Paley operator)**.** For $x \in \mathbf{R}^2$, we denote $t = h(x)$. We then define the adapted Littlewood-Paley projection operator $\tilde{P}_k$ restricted on the curve Γ_t by

$$\tilde{P}_k f(x) := \int_{\mathbf{R}^2} \delta(h_t(y)) f(y) \check{\psi}_k((x - y) \cdot v_t^{\perp}) dy, \tag{3.12}$$

where $\psi_k(\cdot)$ is given by (3.5).

In the above definition, the symbol δ stands for the Dirac measure, for which we have the rule that for any continuous function φ:

$$\int_{\mathbf{R}} \delta(x) \varphi(x) dx = \varphi(0). \tag{3.13}$$

Now we explain the expression (3.12) by applying this rule. For a fixed $t \in \mathbf{R}$, the two vectors v_t and $v_t^{\perp}$ form an orthogonal coordinate system of the plane. Write $y \in \mathbf{R}^2$ in this new system as

$$y = y_1 v_t + y_2 v_t^{\perp}, \tag{3.14}$$

and for the sake of simplicity we will still use the notation $y = (y_1, y_2)$. This changes the expression in (3.12) to

$$\begin{aligned} &\int_{\mathbf{R}^2} f(y_1, y_2)\delta(h_t(y_1, y_2))\check{\psi}_k(x_2 - y_2)dy \\ &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f(y_1, y_2)\delta(h_t(y_1, y_2))dy_1 \right) \check{\psi}_k(x_2 - y_2)dy_2. \end{aligned} \tag{3.15}$$

Hence if we use the same parametrization as the one in Definition 3.3 in [15], which is

$$\Gamma_t = \{y_2 v_t^{\perp} + g_t(y_2)v_t | y_2 \in \mathbf{R}\}, \tag{3.16}$$

then by the definition of the function h_t in Definition 3.2, which implies

$$\int_{\mathbf{R}} \delta(h_t(x))dx_1 = 1, \tag{3.17}$$

the right hand side of (3.15) will equal

$$\int_{\mathbf{R}} f(g_t(y_2), y_2)\check{\psi}_k(x_2 - y_2)dy_2. \tag{3.18}$$

This becomes a one-dimensional integration. Moreover, we see that (3.18) is exactly the one given by Definition 3.3 in [15]; namely these two definitions are equivalent.

Lemma 3.5 (Adapted Littlewood-Paley theory)**.** *For $p \in (1, \infty)$, we have the following variants of the Littlewood-Paley theorem:*

$$\|(\sum_{k\in\mathbf{Z}} |\tilde{P}_k f|^2)^{1/2}\|_p \lesssim \|f\|_p, \tag{3.19}$$

$$\|(\sum_{k\in\mathbf{Z}} |\tilde{P}_k^* f|^2)^{1/2}\|_p \lesssim \|f\|_p. \tag{3.20}$$

Proof of Lemma 3.5. By the Fubini theorem, we obtain

$$\int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} |\tilde{P}_k f|^2 \right)^{p/2} = \int_{\mathbf{R}} \int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} |\tilde{P}_k f|^2 \right)^{p/2} \delta(h(x) - t)dxdt. \tag{3.21}$$

When integrating against dx, by doing the change of variables $h(x) - t \to h_t(x)$, we can write the right hand side of the above expression as

$$\int_{\mathbf{R}} \int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} |\tilde{P}_k f|^2 \right)^{p/2} \delta(h_t(x)) \frac{|\nabla h_t(x)|}{|\nabla h(x)|} dxdt. \tag{3.22}$$

By the bound on ∇h_t in (3.10) and our assumption on ∇h in (1.1) that

$$|\nabla h| \sim 1, \text{ a.e. in } \mathbf{R}^2, \tag{3.23}$$

it suffices to show that

$$\int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} |\tilde{P}_k f|^2 \right)^{p/2} \delta(h_t(x))dx \lesssim \int_{\mathbf{R}^2} |f(x)|^p \delta(h_t(x))dx, \tag{3.24}$$

with a bound being independent of $t \in \mathbf{R}$.

We substitute the definition of $\tilde{P}_k$ into the left hand side of the last expression to obtain

$$\int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} \left| \int_{\mathbf{R}^2} \delta(h_t(y)) f(y) \check{\psi}_k((x-y)\cdot v_t^{\perp}) dy \right|^2 \right)^{p/2} \delta(h_t(x))dx. \tag{3.25}$$

The above expression can be viewed as a two-dimensional Littlewood-Paley operator with the singular measure $\delta(h_t(\cdot))$; hence heuristically it is bounded by

$$\int_{\mathbf{R}^2} |f(x)|^p \delta(h_t(x))dx. \tag{3.26}$$

To make the above argument rigorous, we introduce the change of variables

$$x \to x_1 v_t + x_2 v_t^{\perp}, y \to y_1 v_t + y_2 v_t^{\perp}. \tag{3.27}$$

For the sake of simplicity, after the change of variables, we will still write $x = (x_1, x_2)$ and $y = (y_1, y_2)$. The expression in (3.25) hence becomes

$$\begin{aligned} &\int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} \left| \int_{\mathbf{R}^2} \delta(h_t(y)) f(y) \check{\psi}_k(x_2 - y_2) dy \right|^2 \right)^{p/2} \delta(h_t(x))dx \\ &= \int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} \left| \int_{\mathbf{R}} \left(\int_{\mathbf{R}} \delta(h_t(y)) f(y) dy_1 \right) \check{\psi}_k(x_2 - y_2) dy_2 \right|^2 \right)^{p/2} \delta(h_t(x))dx. \end{aligned} \tag{3.28}$$

Notice that for any $x_2 \in \mathbf{R}$, we have

$$\int_{\mathbf{R}} \delta(h_t(x))dx_1 = 1. \tag{3.29}$$

Hence the right hand side of the last display becomes

$$\int_{\mathbf{R}} \left(\sum_{k\in\mathbf{Z}} \left| \int_{\mathbf{R}} \left(\int_{\mathbf{R}} \delta(h_t(y)) f(y) dy_1 \right) \check{\psi}_k(x_2 - y_2) dy_2 \right|^2 \right)^{p/2} dx_2. \tag{3.30}$$

It is not difficult to see that the above is just a one-dimensional Littlewood-Paley square function for the function

$$\int_{\mathbf{R}} \delta(h_t(y)) f(y) dy_1; \tag{3.31}$$

hence it can be bounded by

$$\int_{\mathbf{R}} \left| \int_{\mathbf{R}} \delta(h_t(y)) f(y) dy_1 \right|^p dy_2 = \int_{\mathbf{R}^2} \delta(h_t(x)) |f(x)|^p dx. \tag{3.32}$$

So far we have finished the proof of (3.24), thus (3.19). For the second equivalence relation (3.20), the proof is similar, hence we leave it out. □

To proceed, we will split the operator into two terms,

$$\sum_{k\in\mathbf{Z}} H_v P_k(f) = \sum_{k\in\mathbf{Z}} (H_v P_k(f) - \tilde{P}_k H_v P_k(f) + \tilde{P}_k H_v P_k(f)). \tag{3.33}$$

Then by the triangle inequality, we have

$$\|\sum_{k\in\mathbf{Z}} H_v P_k(f)\|_p \lesssim \|\sum_{k\in\mathbf{Z}} (H_v P_k(f) - \tilde{P}_k H_v P_k(f))\|_p + \|\sum_{k\in\mathbf{Z}} \tilde{P}_k H_v P_k(f)\|_p. \tag{3.34}$$

We call the second term the main term, and the first term the commutator term.

To bound the main term, we first use duality to write the L^p norm into

$$\begin{aligned}\|\sum_{k\in\mathbf{Z}} \tilde{P}_k H_v P_k(f)\|_p &= \sup_{\|g\|_{p'}=1} |\langle \sum_{k\in\mathbf{Z}} \tilde{P}_k H_v P_k(f), g\rangle| \\ &= \sup_{\|g\|_{p'}=1} |\sum_{k\in\mathbf{Z}} \langle H_v P_k(f), \tilde{P}_k^*(g)\rangle|.\end{aligned}$$

Then by Cauchy-Schwartz and Hölder's inequality, we bound the right hand side by

$$\begin{aligned}&\sup_{\|g\|_{p'}=1} \int (\sum_{k\in\mathbf{Z}} |H_v P_k(f)|^2)^{1/2} (\sum_{k\in\mathbf{Z}} |\tilde{P}_k^*(g)|^2)^{1/2} \\ &\lesssim \sup_{\|g\|_{p'}=1} \|(\sum_{k\in\mathbf{Z}} |H_v P_k(f)|^2)^{1/2}\|_p \|(\sum_{k\in\mathbf{Z}} |\tilde{P}_k^*(g)|^2)^{1/2}\|_{p'}.\end{aligned} \tag{3.35}$$

In the end, by applying Proposition 3.1 to the former term in the last expression and Lemma 3.5 to the latter term, we get the desired bound

$$(3.35) \lesssim \sup_{\|g\|_{p'}=1} \|f\|_p \|g\|_{p'} = \|f\|_p. \tag{3.36}$$

Now we turn to the commutator term. Before explaining the idea of estimating the commutator term, we recall some notation from [15]. Select a Schwartz function ψ_0 such that ψ_0 is supported on $[\frac{1}{2}, \frac{5}{2}]$, and let

$$\psi_l(t) := \psi_0(2^{-l}t). \tag{3.37}$$

By choosing ψ_0 properly, we can construct a partition of unity for $\mathbf{R}^+$, i.e.

$$\mathbb{1}_{(0,\infty)} = \sum_{l\in\mathbf{Z}} \psi_l. \tag{3.38}$$

Let

$$H_l f(x) := \int \check{\psi}_l(t) f(x - tv(x)) dt. \tag{3.39}$$

Then the operator H_v can be decomposed into the sum

$$H_v = -\mathbb{1} + 2\sum_{l\in\mathbf{Z}} H_l. \tag{3.40}$$

We continue to explain the strategy of proving the L^p boundedness of the commutator term, which is

$$\|\sum_{k\in\mathbf{Z}} (H_v P_k(f) - \tilde{P}_k H_v P_k(f))\|_p \lesssim \|f\|_p. \tag{3.41}$$

By the dyadic decomposition in (3.40), this is equivalent to bounding the following:

$$\sum_{k\in\mathbf{Z}}\sum_{l\in\mathbf{Z}}(H_l P_k f - \tilde{P}_k H_l P_k f). \tag{3.42}$$

Notice that by definition, $H_l P_k f$ vanishes for $l > k$, which simplifies the last expression to

$$\sum_{l\geq 0}\sum_{k\in\mathbf{Z}}(H_{k-l} P_k f - \tilde{P}_k H_{k-l} P_k f). \tag{3.43}$$

So by the triangle inequality it suffices to prove

Proposition 3.6. *Under the same assumptions as in the Main Theorem, for any $p \in (1,\infty)$, there exists a constant $\gamma_p > 0$ such that*

$$\|\sum_{k\in\mathbf{Z}}(H_{k-l} P_k(f) - \tilde{P}_k H_{k-l} P_k(f))\|_p \lesssim 2^{-\gamma_p l}\|f\|_p, \tag{3.44}$$

with the constant being independent of $l \in \mathbf{N}$.

The proof of Proposition 3.6 is postponed to Section 6. The idea of proving endpoint estimates like the $L^\infty \to BMO$ estimate will probably not work, as the output of the operator H_v is so rough that it is only measurable across the family of Lipschitz level curves. In other words, the orthogonality between different tiles is missing.

To recover the orthogonality at the level of the L^2 estimate, the argument in [15] relies heavily on the fact that taking the L^2 norm works perfectly (also trivially) with the square function (see the calculation (5.16)-(5.22) in [15]). Hence we could expand certain square summation and apply Hölder's inequality to turn the problem into the analysis on every single Lipschitz curve.

However, in the L^p estimate for $p \neq 2$, this strategy does not work. Instead we will invoke a new square function as an intermediate step (see Lemma 6.8). This square function is akin to the square function in the product space $\mathbf{R}\times\mathbf{R}$ (indeed the stronger version of the square function estimate in Lemma 6.8, which is (6.9) by Demeter and Di Plinio [13], enjoys the full bi-parameter flavor).

Remark 3.7. Although the endpoint $L^\infty \to BMO$ estimate for (3.44) in Proposition 3.6 might not work with the classical BMO space, we still hope that there would be some variants, possibly similar to the fiber wise Hardy and BMO spaces in [6] and [20], which will act as the right substitutes for the endpoint theory.

Remark 3.8. For the one-variable vector fields $v(x_1, x_2) = (1, u(x_2))$, it was proved in [8], under some convexity and curvature assumptions on the function $u : \mathbf{R} \to \mathbf{R}$, that the associated Hilbert transform and maximal function map $H^1_{prod}(\mathbf{R}\times\mathbf{R})$ to L^1, where $H^1_{prod}(\mathbf{R}\times\mathbf{R})$ denotes the product Hardy space.

However, it was also pointed out that this might not be the right endpoint theory, and some new underlying Calderon-Zygmund theory is to be expected. See Remark (*iii*) on page 597 in [8].

4. Boundedness of the main term: Proof of Proposition 3.1

The goal of this section is to make an observation that Bateman and Thiele's square function estimate (see (2.1) in [5]) for the one-variable vector fields, which is

$$\|(\sum_{k\in\mathbf{Z}}(H_vP_k(f))^2)^{1/2}\|_p \lesssim \|f\|_p, \forall p > 3/2, \tag{4.1}$$

works equally well for our case, with just minor modifications. Indeed, the proof of the estimate (4.1) is reduced by Bateman and Thiele in [5] to three covering lemmas (Lemma 7 and Lemma 8 in [5], Lemma 6.2 in [4]), and our observation is that all these covering lemmas still hold true for the case where the vector fields are constant only on Lipschitz curves instead of vertical lines.

Before stating the covering lemmas and the modification that we will make in the proof, we first need to introduce several definitions.

Definition 4.1. For a rectangle $R \subset \mathbf{R}^2$, with l_R its length, w_R its width, we define its uncertainty interval $EX(R) \subset \mathbf{R}$ to be the interval of width w_R/l_R and centered at slope(R). Denote by $E(R)$ the collection of the points $x \in R$ s.t. the vector $v(x) = (1, u(h(x)))$ points roughly in the same direction as the long side of R:

$$E(R) = \{x \in R : u(h(x)) \in EX(R)\}. \tag{4.2}$$

Then the popularity of the rectangle R is defined to be

$$pop_R := |\{x \in \mathbf{R}^2 : u(h(x)) \in EX(R)\}|/|R|. \tag{4.3}$$

Here u and h are the two functions in the Main Theorem.

Definition 4.2. Given two rectangles R_1 and R_2 in $\mathbf{R}^2$, we write $R_1 \le R_2$ whenever $R_1 \subset CR_2$ and $EX(R_2) \subset EX(R_1)$, where C is some properly chosen large constant, and CR_2 is the rectangle with the same center as R_2 but dilated by the factor C.

Now we are ready to state the key covering lemmas:

Lemma 4.3 (Lemma 6.2 in [4]; see also Lemma 4.3 in [15]). *Suppose $\mathcal{R}_0$ is a collection of pairwise incomparable (under "$\le$") rectangles of uniform width such that for each $R \in \mathcal{R}_0$, we have*

$$pop_R \ge \delta \text{ and } \frac{1}{|R|}\int_R \mathbb{1}_F \ge \lambda. \tag{4.4}$$

Then under the same assumptions on u and h as in the Main Theorem, we have for each $p > 1$ that

$$\sum_{R\in\mathcal{R}_0} |R| \lesssim \frac{|F|}{\delta\lambda^p}. \tag{4.5}$$

Lemma 4.4 (Lemma 7 in [5]). *Under the same assumptions as in the Main Theorem, let $\delta > 0$ and $q > 1$, let $G \subset \mathbf{R}^2$ be a measurable set and let $\mathcal{R}$ be a finite collection of rectangles such that*

$$|E(R) \cap G| \ge \delta|G| \tag{4.6}$$

for each $R \in \mathcal{R}$. Then

$$|\bigcup_{R\in\mathcal{R}} R| \lesssim \delta^{-q}|G|. \tag{4.7}$$

Lemma 4.5 (Lemma 8 in [5]). *Under the same assumptions as in the Main Theorem, let $0 < \sigma, \delta \leq 1$, let H be a measurable set, and let $\mathcal{R}$ be a finite collection of rectangles such that for each $R \in \mathcal{R}$ we have*

$$pop_R \geq \sigma, |H \cap R| \geq \delta |R|. \tag{4.8}$$

Then

$$|\bigcup_{R \in \mathcal{R}} R| \lesssim \sigma^{-1}\delta^{-2}|H|. \tag{4.9}$$

To prove these covering lemmas, one just needs to replace the classical rectangles by the following "rectangles" adapted to the vector fields, and run the same argument as in Bateman and Thiele in [5].

Definition 4.6 (Rectangles adapted to the vector field). For a rectangle $R \subset \mathbf{R}^2$, with its two long sides lying on the parallel lines $x_2 = kx_1 + b_1$ and $x_2 = kx_1 + b_2$ for some $k \in [-1, 1]$ and $b_1, b_2 \in \mathbf{R}$, define $\tilde{R}$ to be the adapted version of R, which is given by the set

$$\{x \in \mathbf{R}^2 : h(x) \in h(R)\} \cap \{(x_1, kx_1 + b) : x_1 \in \mathbf{R}, b \in [b_1, b_2]\}, \tag{4.10}$$

where $h : \mathbf{R}^2 \to \mathbf{R}$ is the function from the Main Theorem.

From R to $\tilde{R}$, the length and the width of the rectangle are preserved up to a constant, and the same also holds true for the "popularity". Moreover, the proofs in [5] are "stable" under bi-Lipschitz mapping. Hence we will leave out the details and refer to [5] and [15].

These two lemmas were used to give an upper bound on the size of the exceptional sets (the sets H on page 1585 and set G on page 1588 in [5]) around which the rectangles have either large size or large density. After excluding the exceptional sets, the argument in [5], together with [4] (which also works equally well for our case as has been pointed out in [15]), will lead to the square function estimate, i.e. Proposition 3.1.

5. Time-frequency decomposition

The content of this section is the same as Subsection 5.1 in [15]. We still include the notation here for the sake of completeness.

Discretizing the functions. Fixing $l \geq 0$, we write $\mathcal{D}_l$ as the collection of the dyadic intervals of length 2^{-l} contained in $[-2, 2]$. Fix a smooth positive function $\beta : \mathbf{R} \to \mathbf{R}$ s.t.

$$\beta(x) = 1, \forall |x| \leq 1; \beta(x) = 0, \forall |x| \geq 2. \tag{5.1}$$

Also choose β such that $\sqrt{\beta}$ is a smooth function. Then fixing an integer c (whose exact value is unimportant), for each $\omega \in \mathcal{D}_l$, define

$$\beta_\omega(x) = \beta(2^{l+c}(x - c_{\omega_1})), \tag{5.2}$$

where ω_1 is the right half of ω and c_{ω_1} is its center.

Define

$$\beta_l(x) = \sum_{\omega \in \mathcal{D}_l} \beta_\omega(x), \tag{5.3}$$

and note that

$$\beta_l(x + 2^{-l}) = \beta_l(x), \forall x \in [-2, 2 - 2^{-l}]. \tag{5.4}$$

Define

$$\gamma_l = \frac{1}{2}\int_{-1}^{1}\beta_l(x+t)dt. \tag{5.5}$$

Because of the above periodicity, we know that γ_l is constant for $x \in [-1,1]$, independent of l. Say $\gamma_l(x) = \delta > 0$, hence

$$\frac{1}{\delta}\gamma_l(x)\mathbb{1}_{[-1,1]}(x) = \mathbb{1}_{[-1,1]}(x). \tag{5.6}$$

Define another multiplier $\tilde{\beta} : \mathbf{R} \to \mathbf{R}$ with support in $[\frac{1}{2}, \frac{5}{2}]$ and $\tilde{\beta}(x) = 1$ for $x \in [1,2]$. We define the corresponding multiplier on $\mathbf{R}^2$:

$$\begin{aligned}
\hat{m}_{k,\omega}(\xi_1,\xi_2) &= \tilde{\beta}(2^{-k}\xi_2)\beta_\omega(\frac{\xi_1}{\xi_2}),\\
\hat{m}_{k,l,t}(\xi_1,\xi_2) &= \tilde{\beta}(2^{-k}\xi_2)\beta_l(t+\frac{\xi_1}{\xi_2}),\\
\hat{m}_{k,l}(\xi_1,\xi_2) &= \tilde{\beta}(2^{-k}\xi_2)\gamma_l(\frac{\xi_1}{\xi_2}).
\end{aligned}$$

Then what we need to bound can be written as

$$\begin{aligned}
\|\sum_{k\in\mathbf{Z}}\sum_{l\in\mathbf{Z}} H_l P_k(f)\|_p &= \|\int_{-1}^{1}\sum_{k\in\mathbf{Z}}\sum_{l\geq 0} H_{k-l}(\frac{1}{\delta}m_{k,l} * f)dt\|_p\\
&\leq \int_{-1}^{1}\|\sum_{k\in\mathbf{Z}}\sum_{l\geq 0} H_{k-l}(\frac{1}{\delta}m_{k,l,t} * f)\|_p dt,
\end{aligned}$$

where the terms $H_l P_k$ for $l > k$ in the sum vanish as explained before.

So it suffices to prove a uniform bound on $t \in [-1,1]$. Without loss of generality we will just consider the case $t = 0$, which is

$$\sum_{k\in\mathbf{Z}}\sum_{l\geq 0} H_{k-l}(m_{k,l,0} * f) = \sum_{k\in\mathbf{Z}}\sum_{l\geq 0} H_{k-l}([\tilde{\beta}(2^{-k}\xi_2)\beta_l(\frac{\xi_1}{\xi_2})] * f). \tag{5.7}$$

Constructing the tiles. For each $k \in \mathbf{Z}$ and $\omega \in \mathcal{D}_l$ with $l \geq 0$, let $\mathcal{U}_{k,\omega}$ be a partition of $\mathbf{R}^2$ by rectangles of width 2^{-k} and length 2^{-k+l}, whose long side has slope $-c(\omega)$, where $c(\omega)$ is the center of the interval ω. If $s \in \mathcal{U}_{k,\omega}$, we will write $\omega_s := \omega$, with $\omega_{s,1}$ the right half of ω and $\omega_{s,2}$ the left half.

An element of $\mathcal{U}_{k,\omega}$ for some $\omega \in \mathcal{D}_l$ is called a "tile". Define $\varphi_{k,\omega}$ such that

$$|\hat{\varphi}_{k,\omega}|^2 = \hat{m}_{k,\omega}. \tag{5.8}$$

Then $\varphi_{k,\omega}$ is smooth by our assumption on β mentioned above.

For a tile $s \in \mathcal{U}_{k,\omega}$, define

$$\varphi_s(p) := \sqrt{|s|}\varphi_{k,\omega}(p - c(s)), \tag{5.9}$$

where $c(s)$ is the center of s. Notice that

$$\|\varphi_s\|_2^2 = \int_{\mathbf{R}^2}|s|\varphi_{k,\omega}^2 = |s|\int_{\mathbf{R}^2}\hat{m}_{k,\omega} = 1; \tag{5.10}$$

i.e. φ_s is L^2 normalized.

The construction of the tiles above by the uncertainty principle is to localize the function further in space. For this purpose we need

Lemma 5.1 ([4])**.**

$$f * m_{k,\omega}(x) = \lim_{N\to\infty} \frac{1}{4N^2} \int_{[-N,N]^2} \sum_{s\in\mathcal{U}_{k,\omega}} \langle f, \varphi_s(p+\cdot)\rangle \varphi_s(p+x)dp. \tag{5.11}$$

The above lemma allows us to pass to the model sum

$$\sum_{k\in\mathbf{Z}}\sum_{l\geq 0} H_{k-l}(f * m_{k,l,0}) = \sum_{k\in\mathbf{Z}}\sum_{l\geq 0}\sum_{\omega\in\mathcal{D}_l}\sum_{s\in\mathcal{U}_{k,\omega}} \langle f, \varphi_s\rangle H_{k-l}(\varphi_s).$$

Define

$$\psi_s = \psi_{-\log(length(s))} \tag{5.12}$$

and

$$\phi_s(x) := \int \check{\psi}_s(t)\varphi_s(x - tv(x))dt. \tag{5.13}$$

Then the model sum turns to

$$\sum_{k\in\mathbf{Z}}\sum_{l\geq 0}\sum_{\omega\in\mathcal{D}_l}\sum_{s\in\mathcal{U}_{k,\omega}} \langle f, \varphi_s\rangle \phi_s. \tag{5.14}$$

Lemma 5.2. *We have that $\phi_s(x) = 0$ unless $-u(h(x)) \in \omega_{s,2}$.*

The proof of the above lemma is by the Plancherel theorem; we just need to observe that the frequency support of ψ_s and $\hat{\varphi}_s$ will be disjoint at the point x unless $-u(h(x)) \in \omega_{s,2}$.

6. Boundedness of the commutator term: Proof of Proposition 3.6

In this section we intend to prove that for any $p > 1$, there exists $\gamma_p > 0$ such that

$$\left\|\sum_{k\in\mathbf{Z}} \Big(H_{v,k-l}P_k(f) - \tilde{P}_k H_{v,k-l}P_k(f)\Big)\right\|_p \lesssim 2^{-\gamma_p l}\|f\|_p. \tag{6.1}$$

If we expand the left hand side of the last expression to a model sum by the notation in Section 5, (6.1) becomes

$$\left\|\sum_{k\in\mathbf{Z}}\sum_{\omega\in\mathcal{D}_l}\sum_{s\in\mathcal{U}_{k,\omega}} \langle f, \varphi_s\rangle(\phi_s - \tilde{P}_k\phi_s)\right\|_p \lesssim 2^{-\gamma_p l}\|f\|_p. \tag{6.2}$$

Observe that for a fixed point $x \in \mathbf{R}^2$, by Lemma 5.2, the expression

$$\sum_{k\in\mathbf{Z}}\sum_{s\in\mathcal{U}_{k,\omega}} \langle f, \varphi_s\rangle(\phi_s - \tilde{P}_k\phi_s)(x) \tag{6.3}$$

can be non-zero for at most one $\omega \in \mathcal{D}_l$, which implies that

$$
\begin{aligned}
&\Big\| \sum_{k\in\mathbf{Z}} \sum_{\omega\in\mathcal{D}_l} \sum_{s\in\mathcal{U}_{k,\omega}} \langle f,\varphi_s\rangle(\phi_s - \tilde{P}_k\phi_s) \Big\|_p \\
&\lesssim \Big(\sum_{\omega\in\mathcal{D}_l} \int_{\mathbf{R}^2} \Big| \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} \langle f,\varphi_s\rangle(\phi_s - \tilde{P}_k\phi_s) \Big|^p \Big)^{1/p}.
\end{aligned} \tag{6.4}
$$

From the right hand side of the above inequality, we see that (6.2) is reduced to separate $\omega \in \mathcal{D}_l$. Hence we just need to do the estimate for each ω separately. To be precise, we will prove

Lemma 6.1. *Under the above notation, we have*

$$
\Big\| \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} \langle f,\varphi_s\rangle(\phi_s - \tilde{P}_k\phi_s) \Big\|_p \lesssim 2^{-l}\|P_\omega f\|_p, \tag{6.5}
$$

where P_ω is the frequency projection operator given by

$$
\mathcal{F}P_\omega f(\xi_1,\xi_2) = \beta_\omega(\frac{\xi_1}{\xi_2})\mathcal{F}f(\xi_1,\xi_2), \tag{6.6}
$$

and the constant in (6.5) *is independent of $\omega \in \mathcal{D}_l$.*

Lemma 6.2. *We have the following bounds for the multiplier β_ω:*

$$
\|P_\omega f\|_p \lesssim \|f\|_p, \tag{6.7}
$$

for all $p \in (1,\infty)$, with the constant being independent of ω.

Finishing the proof of Proposition 3.6. We substitute the estimates in Lemma 6.1 and Lemma 6.2 into (6.4) to obtain

$$
\begin{aligned}
&\Big\| \sum_{k\in\mathbf{Z}} \sum_{\omega\in\mathcal{D}_l} \sum_{s\in\mathcal{U}_{k,\omega}} \langle f,\varphi_s\rangle(\phi_s - \tilde{P}_k\phi_s) \Big\|_p \\
&\lesssim \Big(\sum_{\omega\in\mathcal{D}_l} 2^{-pl}\|P_\omega f\|_p^p \Big)^{1/p} \lesssim 2^{-\frac{p-1}{p}\cdot l}\|f\|_p,
\end{aligned} \tag{6.8}
$$

which finishes the proof of Proposition 3.6. ☐

Remark 6.3. It has been proved by Demeter and Di Plinio in [13] that

$$
\Big(\sum_{\omega\in\mathcal{D}_l} \|P_\omega f\|_p^p \Big)^{1/p} \lesssim \|f\|_p, \tag{6.9}
$$

for $p \geq 2$, with the constant being independent of $l \in \mathbf{N}$. This will provide a better exponential decay in l in the last inequality in (6.8). However, here we do not need such an orthogonality estimate but simply a triangle inequality.

6.1. **Proof of Lemma 6.2.** We first reduce the estimate (6.7) to one single $\omega \in \mathcal{D}_l$ by applying the shearing transform. Suppose for the moment that we have proved (6.7) for $\omega = [0, 2^{-l}]$. By doing the following change of variables:

$$x_1 \to x_1, x_2 \to x_2 + \lambda x_1, \tag{6.10}$$

for the function f, the frequency variables are transformed into

$$\xi_1 \to \xi_1 - \lambda \xi_2, \xi_2 \to \xi_2. \tag{6.11}$$

This linear change of variables turns

$$P_{\omega'} f(\xi_1, \xi_2) = \mathcal{F}^{-1}\left(\beta_{\omega'}(\frac{\xi_1}{\xi_2})\hat{f}(\xi_1, \xi_2)\right), \tag{6.12}$$

which is the term on the left hand side of (6.7), into

$$\mathcal{F}^{-1}\left(\beta_{\omega'}(\frac{\xi_1}{\xi_2})\hat{f}(\xi_1 - \lambda\xi_2, \xi_2)\right). \tag{6.13}$$

If we denote

$$\tilde{\xi_1} := \xi_1 - \lambda \xi_2, \tilde{\xi_2} := \xi_2, \tag{6.14}$$

the multiplier in (6.13) turns into

$$\beta_{\omega'}(\frac{\tilde{\xi_1} + \lambda\tilde{\xi_2}}{\tilde{\xi_2}}) = \beta(2^{l+c}\frac{\tilde{\xi_1}}{\tilde{\xi_2}} + \lambda 2^{l+c} - 2^{l+c}c_{\omega_1'}). \tag{6.15}$$

So far it becomes clear that by taking λ in (6.15) properly, we can apply the change of variables (6.10) to turn the projection operator $P_{\omega'} f$ for an arbitrary $\omega \in \mathcal{D}_l$ into $P_\omega f$, where $\omega = [0, 2^{-l}]$.

Next, we will reduce the estimate for all $l \in \mathbf{N}$ to the one simply for $l = 0$. This can be done by applying the anisotropic scaling symmetry

$$x_1 \to \lambda x_1, x_2 \to x_2, \tag{6.16}$$

for the function f. Under the above change of variables, the Fourier transform of f is transformed from $\hat{f}(\xi_1, \xi_2)$ to

$$\frac{1}{\lambda}\hat{f}(\frac{\xi_1}{\lambda}, \xi_2). \tag{6.17}$$

Correspondingly, the function $P_\omega f$ is changed to

$$\begin{aligned} &\int \beta_\omega(\frac{\xi_1}{\xi_2})\frac{1}{\lambda}\hat{f}(\frac{\xi_1}{\lambda}, \xi_2)e^{ix_1\xi_1 + ix_2\xi_2}d\xi_1 d\xi_2 \\ &= \int \beta_\omega(\frac{\lambda\xi_1}{\xi_2})\hat{f}(\xi_1, \xi_2)e^{i\lambda x_1\xi_1 + ix_2\xi_2}d\xi_1 d\xi_2. \end{aligned} \tag{6.18}$$

Hence the multiplier $\beta_\omega(\xi_1/\xi_2)$ has the same L^p norm with $\beta_\omega(\lambda\xi_1/\xi_2)$. However, by the definition of β_ω, we have

$$\beta_\omega(\frac{\lambda\xi_1}{\xi_2}) = \beta(\frac{2^{l+c}\lambda\xi_1}{\xi_2} - 2^{l+c}c_{\omega_1}), \tag{6.19}$$

which means that if we take $\lambda = 2^{-l}$, the right hand side of the last expression becomes $\beta_{\omega_0}(\xi_1/\xi_2)$ where $\omega_0 = [0, 1]$.

After the above reductions, we just need to prove (6.7) with $\omega_0 = [0, 1]$. For $p = 2$, the estimate is trivial due to Plancherel's theorem. For $p \neq 2$, if we denote by P_k a Littlewood-Paley projection operator in the second variable, then by the Littlewood-Paley theory, we obtain

$$\|P_{\omega_0}f\|_p \lesssim \left\| \left(\sum_k |P_k P_{\omega_0} f|^2 \right)^{1/2} \right\|_p . \tag{6.20}$$

By the classical Calderon-Zygmund theory, it is not difficult to prove that

$$\left\| \left(\sum_k |P_k P_{\omega_0} f|^2 \right)^{1/2} \right\|_{BMO} \lesssim \|f\|_\infty \tag{6.21}$$

and

$$\left\| \left(\sum_k |P_k P_{\omega_0} f|^2 \right)^{1/2} \right\|_1 \lesssim \|f\|_{H^1} . \tag{6.22}$$

Hence by interpolation, we obtain the desired estimate for all $p \in (1, \infty)$. So far we have finished the proof of Lemma 6.2. $\square$

6.2. **Proof of Lemma 6.1.** By the same shearing transform as in (6.10), we can reduce the estimate (6.5) for different ω to the one for a fixed ω, say $\omega = [0, 2^{-l}]$. To prove (6.5), by invoking duality, it is equivalent to proving that

$$\int_{\mathbf{R}^2} \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} |\langle f, \varphi_s\rangle| \left| \left(\phi_s - \tilde{P}_k\phi_s\right) \cdot g \right| \lesssim 2^{-l}\|f\|_p , \tag{6.23}$$

where the function g satisfies $\|g\|_{p'} \le 1$. By the Fubini theorem, the left hand side of (6.23) is equal to

(6.24)
$$\begin{aligned} &\int_{\mathbf{R}} \int_{\mathbf{R}^2} \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} |\langle f, \varphi_s\rangle| \left| \left(\phi_s(x) - \tilde{P}_k\phi_s(x)\right) \cdot g(x) \right| \delta(h(x) - t) dx dt \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}^2} \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} |\langle f, \varphi_s\rangle| \left| \left(\phi_s(x) - \tilde{P}_k\phi_s(x)\right) \cdot g(x) \right| \delta(h_t(x)) \frac{|\nabla h_t(x)|}{|\nabla h(x)|} dx dt. \end{aligned}$$

By the bound on ∇h_t in (3.10) and our assumption on ∇h in the Main Theorem, the right hand side of (6.24) can be bounded by

$$\int_{\mathbf{R}} \int_{\mathbf{R}^2} \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} |\langle f, \varphi_s\rangle| \left| \left(\phi_s(x) - \tilde{P}_k\phi_s(x)\right) \cdot g(x) \right| \delta(h_t(x)) dx dt. \tag{6.25}$$

If we denote by $s_{m,n}$ the translation of the tile s by (m, n) units, which is

$$s_{m,n} := s - (m \cdot l_s, n \cdot w_s), \tag{6.26}$$

then the above (6.25) is equal to

(6.27)
$$\sum_{m,n} \int_{\mathbf{R}} \int_{\mathbf{R}^2} \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} |\langle f, \varphi_s\rangle| \mathbb{1}_{s_{m,n}}(x) \left| \left(\phi_s(x) - \tilde{P}_k\phi_s(x)\right) \cdot g(x) \right| \delta(h_t(x)) dx dt.$$

By the notion of the adapted rectangles in Definition 4.6, we can replace $s_{m,n}$ by the slightly enlarged "rectangle" $\tilde{s}_{m,n}$, as from the definition it is clear that

$\tilde{s}_{m,n} \supset s_{m,n}$. Moreover, in the following, we will only focus on the term $m = n = 0$, as the other terms appear as the tail terms by the non-stationary phase method.

To proceed, we need the notion of Jones' beta numbers:

Theorem 6.4 ([16]). *Fix a Lipschitz function $A : \mathbf{R} \to \mathbf{R}$. For each dyadic interval I, there exists a number $\alpha_I(A) \in \mathbf{R}$, such that if we denote*

$$\beta_{j_0}(I) = \sup_{x \in 3(j_0+1)I} \frac{|A(x) - A(c_I) - \alpha_I(A)(x - c_I)|}{|I|}, \tag{6.28}$$

where $j_0 \in \mathbf{N}$ and c_I denotes the center of I, then we will have the following Carleson type condition:

$$\sup_J \frac{1}{|J|} \sum_{I \subset J} \beta_{j_0}^2(I)|I| \lesssim (j_0+1)^3 \|A\|_{Lip}^2. \tag{6.29}$$

$\beta_{j_0}(I)$ and $\alpha_I(A)$ will be called the j_0-th beta number and the "average slope" for the Lipschitz function A near the interval I separately.

The pointwise estimate in the following Lemma 6.5 will play a crucial role in the forthcoming calculation. To state this estimate, we need to make some preparations: for a fixed $t \in \mathbf{R}$, we use the new coordinates system given by $(v_t, v_t^\perp)$. For a tile s, we use $J(t,s)$ to denote the projection of $\Gamma_t \cap \tilde{s}$ on the new vertical axis $v_t^\perp$. Moreover for the interval $J(t,s)$, we let $J^D(t,s)$ denote one of the dyadic intervals (at most two) on the vertical axis such that

$$|J^D(t,s)| \in (8 \cdot |J(t,s)|, 16 \cdot |J(t,s)|] \tag{6.30}$$

and

$$|J^D(t,s) \cap J(t,s)| \geq |J(t,s)|/2. \tag{6.31}$$

For the dyadic interval $J^D(t,s)$, we let $\Phi_{J^D(t,s)}$ denote the associated L^2 normalized Haar function.

Lemma 6.5 ([15]). *Fix $t \in \mathbf{R}$ and $s \in \mathcal{U}_{k,\omega}$ for some $\omega \in \mathcal{D}_l$. For $x \in \Gamma_t \cap \tilde{s}$, we have the pointwise estimate*

$$|\phi_s(x) - \tilde{P}_k \phi_s(x)| \lesssim \sum_{j_0 \in \mathbf{N}} \frac{2^{-3l/2} 2^k \beta_{j_0}(J^D(t,s))}{(j_0+1)^N}, \tag{6.32}$$

where $\beta_{j_0}(J^D(t,s))$ denotes the j_0-th beta number of Γ_t near the dyadic interval $J^D(t,s)$.

Remark 6.6. The proof of the above Lemma 6.5 in [15] relies on those unnecessary parameters and auxiliary functions that we want to avoid by doing δ-calculus. However, as we promised in the introduction that we would carry out the whole argument in the language of δ-calculus completely, we should also be able to prove Lemma 6.5 by doing so. This is postponed to the next subsection.

Substituting the above estimate into the right hand side of (6.27) with $m = n = 0$, we obtain

$$\sum_{j_0} \frac{2^{-l}}{(j_0+1)^N} \int_{\mathbf{R}} \int_{\mathbf{R}^2} \sum_{k \in \mathbf{Z}} \sum_{s \in \mathcal{U}_{k,\omega}} |\langle f, \varphi_s \rangle| 2^k 2^{-l/2} \mathbb{1}_{\tilde{s}}(x) \beta_{j_0}(J^D(t,s)) \cdot |g(x)| \delta(h_t(x)) dx dt. \tag{6.33}$$

To proceed, we need the following.

Claim 6.7. *Fixing $t \in \mathbf{R}$, we have the estimate*

$$\begin{aligned}&\int_{\mathbf{R}^2}\left(\sum_{k\in\mathbf{Z}}\sum_{s\in\mathcal{U}_{k,\omega}} 2^k 2^{-l/2}|\langle f,\varphi_s\rangle|\mathbb{1}_{\tilde{s}}(x)\beta_{j_0}(J^D(t,s))\right)g(x)\delta(h_t(x))dx\\ &\lesssim (j_0+1)^{3/2}\left(\int_{\mathbf{R}^2}\left(\sum_{k\in\mathbf{Z}}\sum_{s\in\mathcal{U}_{k,\omega}}|\langle f,\varphi_s\rangle|^2\chi_s^2(x)\right)^{p/2}\delta(h_t(x))dx\right)^{1/p}\\ &\qquad\cdot\left(\int_{\mathbf{R}_2}|g(x)|^{p'}\delta(h_t(x))dx\right)^{1/p'},\end{aligned}$$

where for $x=(x_1,x_2)$,

$$\chi_s(x_1,x_2) := \frac{|s|^{-1/2}}{(1+(\frac{x_1-c_{s,1}}{l_s})^2+(\frac{x_2-c_{s,2}}{w_s})^2)^5}, \tag{6.34}$$

with $c_s=(c_{s,1},c_{s,2})$ denoting the center of s, $l_s=2^{-k+l}$ the length and $w_s=2^{-k}$ the width.

We postpone the proof of Claim 6.7 till the end of this subsection and continue with the estimate of the term (6.33). By Claim 6.7 and by applying Hölder's inequality to $\int_{\mathbf{R}} dt$, the expression in (6.33) can be bounded by

$$\begin{aligned}&\sum_{j_0}\frac{j_0^{3/2}\cdot 2^{-l}}{(j_0+1)^N}\left(\int_{\mathbf{R}}\int_{\mathbf{R}^2}\left(\sum_{k\in\mathbf{Z}}\sum_{s\in\mathcal{U}_{k,\omega}}|\langle f,\varphi_s\rangle|^2\chi_s^2(x)\right)^{p/2}\delta(h_t(x))dxdt\right)^{1/p}\\ &\lesssim 2^{-l}\cdot\left\|\left(\sum_{k\in\mathbf{Z}}\sum_{s\in\mathcal{U}_{k,\omega}}|\langle f,\varphi_s\rangle|^2\chi_s^2(x)\right)^{1/2}\right\|_p.\end{aligned}\tag{6.35}$$

To bound the last expression, we need the following.

Lemma 6.8. *We have the following variant of the square function estimate:*

$$\left\|\left(\sum_{k\in\mathbf{Z}}\sum_{s\in\mathcal{U}_{k,\omega}}|\langle f,\varphi_s\rangle|^2\chi_s^2(x)\right)^{1/2}\right\|_p \lesssim \|f\|_p. \tag{6.36}$$

Finishing the proof of Lemma 6.1. It is straightforward that, combined with (6.35), Lemma 6.8 finishes the estimate of the expression (6.33), thus the proof of Lemma 6.1. □

Proof of Lemma 6.8. Recall that in the estimate (6.36), we have $\omega=[0,2^{-l}]$. Now we want to reduce the estimate to the case $\omega_0=[0,1]$ by applying the anisotropic scaling

$$x_1\to 2^l x_1, x_2\to x_2. \tag{6.37}$$

Under the above change of variables, as has been explained in the proof of Lemma 6.2, φ_s for some $s\in\mathcal{U}_{k,\omega}$ is changed to $\varphi_{s'}$ for the corresponding $s'\in\mathcal{U}_{k,\omega_0}$ with

$\omega_0 = [0, 1]$. Moreover, the function χ_s will also behave in the same way:

$$\chi_s(2^l x_1, x_2) = \frac{|s|^{-1/2}}{(1 + (\frac{2^l x_1 - c_{s,1}}{l_s})^2 + (\frac{x_2 - c_{s,2}}{w_s})^2)^5} = \frac{|s|^{-1/2}}{(1 + (\frac{x_1 - 2^{-l} c_{s,1}}{2^{-l} l_s})^2 + (\frac{x_2 - c_{s,2}}{w_s})^2)^5}. \tag{6.38}$$

Recall that $l_s = 2^l w_s$; hence the right hand side of (6.38) becomes a bump function with main support on a cube of side length w_s, which means that $\chi_s(2^l x_1, x_2)$ is equal to $\chi_{s'}$ for some $s' \in \mathcal{U}_{k,\omega_0}$ up to a normalization factor.

After the above reduction, we just need to prove (6.36) for $\omega = [0, 1]$. For the case $p = 2$, by the orthogonality of the wavelet functions, we obtain

$$\left(\int_{\mathbf{R}^2} \sum_{k \in \mathbf{Z}} \sum_{s \in \mathcal{U}_{k,\omega}} |\langle f, \varphi_s \rangle|^2 \chi_s^2 \right)^{1/2} \lesssim \|f\|_2. \tag{6.39}$$

Moreover, by the classical Calderon-Zygmund theory, it is not difficult to prove the following endpoint estimates:

$$\left\| \left(\sum_{k \in \mathbf{Z}} \sum_{s \in \mathcal{U}_{k,\omega}} |\langle f, \varphi_s \rangle|^2 \chi_s^2 \right)^{1/2} \right\|_{BMO} \lesssim \|f\|_\infty \tag{6.40}$$

and

$$\left\| \left(\sum_{k \in \mathbf{Z}} \sum_{s \in \mathcal{U}_{k,\omega}} |\langle f, \varphi_s \rangle|^2 \chi_s^2 \right)^{1/2} \right\|_1 \lesssim \|f\|_{H^1}. \tag{6.41}$$

Hence by interpolation, we can obtain all the expected L^p estimates for (6.36) in the above Lemma 6.8. □

Proof of Claim 6.7. For a fixed $t \in \mathbf{R}$, for the summation on the left hand side of the estimate in Claim 6.7, we observe that

$$\sum_{k \in \mathbf{Z}} \sum_{s \in \mathcal{U}_{k,\omega}} = \sum_{s : s \cap \Gamma_t \neq \emptyset}, \tag{6.42}$$

as the term $\mathbb{1}_{\tilde{s}}(x)$ will vanish if $s \cap \Gamma_t = \emptyset$. We use the new coordinate system $(v_t, v_t^\perp)$, and write $x = x_1 v_t + x_2 v_t^\perp$, which will still be denoted as $x = (x_1, x_2)$ for the sake of simplicity. This turns the left hand side of the estimate in Claim 6.7 into

$$\begin{aligned} &\int_{\mathbf{R}} \int_{\mathbf{R}} \left(\sum_{s : s \cap \Gamma_t \neq \emptyset} |\langle f, \varphi_s \rangle| \mathbb{1}_{\tilde{s}}(x_1, x_2) \beta_{j_0}(J^D(t, s)) 2^k 2^{-l/2} \right) \\ &\qquad \cdot g(x_1, x_2) \delta(h_t(x_1, x_2)) dx_1 dx_2 \\ &= \sum_{s : s \cap \Gamma_t \neq \emptyset} 2^k 2^{-l/2} |\langle f, \varphi_s \rangle| \beta_{j_0}(J^D(t, s)) \\ &\qquad \cdot \int_{\mathbf{R}} \int_{\mathbf{R}} g(x_1, x_2) \mathbb{1}_{\tilde{s}}(x_1, x_2) \delta(h_t(x_1, x_2)) dx_1 dx_2. \end{aligned} \tag{6.43}$$

Notice that the integration on the right hand side of (6.43) can be estimated in the following way:

$$\begin{aligned}&\left|\int_{\mathbf{R}}\int_{\mathbf{R}} g(x_1,x_2)\mathbb{1}_{\tilde{s}}(x_1,x_2)\delta(h_t(x_1,x_2))dx_1dx_2\right|\\ &\lesssim 2^{-k}\left[\int_{\mathbf{R}} g(x_1,\cdot)\delta(h_t(x_1,\cdot))dx_1\right]_{2J^D(t,s)},\end{aligned}$$

where for a function $G:\mathbf{R}\to\mathbf{R}$, $[G(\cdot)]_J$ denotes the average of the function G on the interval $J\subset\mathbf{R}$.

Substituting the above bound into the right hand side of (6.43), we obtain the following bound:

$$\sum_{s:s\cap\Gamma_t\neq\emptyset} 2^{-l/2}|\langle f,\varphi_s\rangle|\beta_{j_0}(J^D(t,s))\left[\int_{\mathbf{R}} g(x_1,\cdot)\delta(h_t(x_1,\cdot))dx_1\right]_{2J^D(t,s)}.$$

To proceed, the idea is to view the above expression as a paraproduct. To do this, we need to find the right function such that it has the wavelet coefficient $2^{-l/2}|\langle f,\varphi_s\rangle|w_s^{-1/2}$, where $w_s=2^{-k}$ denotes the width of the tile s. This can be achieved by defining a function $F_t:\mathbf{R}\to\mathbf{R}$ such that

$$F_t(x_2)=\sum_{s:s\cap\Gamma_t\neq\emptyset} 2^{-l/2}w_s^{-1/2}\langle f,\varphi_s\rangle\Phi_{J^D(t,s)}(x_2), \tag{6.44}$$

where $\Phi_{J^D(t,s)}$ denotes the L^2 normalized Haar function associated to the dyadic interval $J^D(t,s)$.

By the L^p boundedness of the paraproduct (see [1] for example) and Jones' beta number condition that

$$\sup_s \frac{1}{|J^D(t,s)|}\sum_{s':J^D(t,s')\subset J^D(t,s)} \beta_{j_0}^2(J^D(t,s'))w_s\lesssim (j_0+1)^3, \tag{6.45}$$

we obtain for any fixed $t\in\mathbf{R}$ that

(6.46)

$$\begin{aligned}&\sum_{s:s\cap\Gamma_t\neq\emptyset} 2^{-l/2}|\langle f,\varphi_s\rangle|\beta_{j_0}(J^D(t,s))\left[\int_{\mathbf{R}} g(x_1,\cdot)\delta(h_t(x_1,\cdot))dx_1\right]_{2J^D(t,s)}\\ &=\sum_{s:s\cap\Gamma_t\neq\emptyset} 2^{-l/2}w_s^{-1/2}|\langle f,\varphi_s\rangle|\beta_{j_0}(J^D(x,s))w_s^{1/2}\left[\int_{\mathbf{R}} g(x_1,\cdot)\delta(h_t(x_1,\cdot))dx_1\right]_{2J^D(t,s)}\\ &\lesssim (j_0+1)^{3/2}\|F_t(\cdot)\|_p\left\|\int_{\mathbf{R}} g(x_1,\cdot)\delta(h_t(x_1,\cdot))dx_1\right\|_{p'}\\ &\lesssim (j_0+1)^{3/2}\|F_t(\cdot)\|_p\left(\int_{\mathbf{R}_2}|g(x)|^{p'}\delta(h_t(x))dx\right)^{1/p'}.\end{aligned}$$

Hence what remains is to prove the following.

Claim 6.9. *Under the above notation, we have*

$$\|F_t(\cdot)\|_p\lesssim\left(\int_{\mathbf{R}^2}\left(\sum_{k\in\mathbf{Z}}\sum_{s\in\mathcal{U}_{k,\omega}}|\langle f,\varphi_s\rangle|^2\chi_s^2(x)\right)^{p/2}\delta(h_t(x))dx\right)^{1/p}. \tag{6.47}$$

Proof of Claim 6.9. By the square function estimate, we obtain

$$\|F_t\|_p \lesssim \left\| \left(\sum_{s: s\cap\Gamma_t \neq \emptyset} 2^{-l} w_s^{-2} \langle f, \varphi_s\rangle^2 \mathbb{1}_{J^D(t,s)}(\cdot) \right)^{1/2} \right\|_p . \tag{6.48}$$

For the right hand side of (6.47), again we use the new coordinate system $(v_t, v_t^{\perp})$ and denote $x = x_1 v_t + x_2 v_t^{\perp}$ as $x = (x_1, x_2)$ for the sake of simplicity. Then the right hand side of (6.47) becomes

$$\begin{aligned} &\left(\int_{\mathbf{R}^2} \left(\sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} |\langle f, \varphi_s\rangle|^2 \chi_s^2(x_1, x_2) \right)^{p/2} \delta(h_t(x_1, x_2)) dx_1 dx_2 \right)^{1/p} \\ &= \left(\int_{\mathbf{R}} \left(\int_{\mathbf{R}} \sum_{k\in\mathbf{Z}} \sum_{s\in\mathcal{U}_{k,\omega}} |\langle f, \varphi_s\rangle|^2 \chi_s^2(x_1, x_2) \delta(h_t(x_1, x_2)) dx_1 \right)^{p/2} dx_2 \right)^{1/p} . \end{aligned} \tag{6.49}$$

If we compare the right hand sides of (6.48) and (6.49), we observe that the following pointwise estimate in x_2 will finish the proof of the claim: for any $x_2 \in \mathbf{R}$ and any tile s such that $s \cap \Gamma_t \neq \emptyset$, we have

$$2^{-l} w_s^{-2} \langle f, \varphi_s\rangle^2 \mathbb{1}_{J^D(t,s)}(x_2) \lesssim \int_{\mathbf{R}} |\langle f, \varphi_s\rangle|^2 \chi_s^2(x_1, x_2) \delta(h_t(x_1, x_2)) dx_1. \tag{6.50}$$

But this follows easily from the definition of the function χ_s. Thus we have finished the proof of Claim 6.9. □

6.3. **Proof of Lemma 6.5.** Fixing t, we establish a pointwise bound for $x \in \Gamma_t$. Therefore, we can and do assume that the vector field is constant and equal to v_t. That is to say, if we define

$$\phi_s^t(x) := \int_{\mathbf{R}} \varphi_s(x - tv_t) \check{\psi}_{k-l}(t) dt, \forall x \in \mathbf{R}^2, \tag{6.51}$$

we will have

$$\phi_s^t(x) = \phi_s(x), \forall x \in \Gamma_t, \tag{6.52}$$

and the advantage is that the vector field becomes the constant vector field v_t. In the following, we will stick to ϕ_s^t instead of ϕ_s.

For a tile s of dimension $w_s \times l_s$ with

$$l_s = 2^l \cdot w_s, \tag{6.53}$$

for a point $x \in \Gamma_t \cap s$ with

$$v_t^{\perp} \in \omega_{s,2}, \tag{6.54}$$

we want to show that

$$|\phi_s^t(x) - \tilde{P}_k \phi_s^t(x)| \lesssim \sum_{j_0\in\mathbf{N}} \frac{2^{-3l/2} \cdot w_s^{-1} \beta_{j_0}(J^D(t,s))}{(j_0+1)^N}. \tag{6.55}$$

To proceed, we again turn to the new coordinate system $(v_t, v_t^{\perp})$ and write

$$x \to x_1 v_t + x_2 v_t^{\perp}. \tag{6.56}$$

By the definition of the operator $\tilde{P}_k$, the left hand side of (6.55) is equal to

$$\phi_s^t(x_1,x_2) - \int_{\mathbf{R}} \left[\int_{\mathbf{R}} \phi_s^t(y_1,y_2)\delta(h_t(y_1,y_2))dy_1 \right] \psi_k(x_2-y_2)dy_2. \tag{6.57}$$

We approximate $\Gamma_t \cap s$ by the line of the "average slope" in the definition of Jones' beta number and call it $l_{s,t}$. Moreover, we define another auxiliary function $L_{s,t}$ associated to the line $l_{s,t}$ in a similar way to h_t:

$$\text{If for some } y \in \Gamma_t \text{ we have } z - y = d \cdot v_t, \text{ then we set } L_{s,t}(z) = d. \tag{6.58}$$

The crucial observation is that

$$\begin{aligned} &\int_{\mathbf{R}} \left[\int_{\mathbf{R}} \phi_s^t(y_1,y_2)\delta(L_{s,t}(y_1,y_2))dy_1 \right] \psi_k(x_2-y_2)dy_2 \\ &= \int_{\mathbf{R}} \phi_s^t(y_1,x_2)\delta(L_{s,t}(y_1,x_2))dy_1. \end{aligned} \tag{6.59}$$

Substitute the above identity into (6.57) to obtain

$$\begin{aligned} &\phi_s^t(x_1,x_2) - \int_{\mathbf{R}} \phi_s^t(y_1,x_2)\delta(L_{s,t}(y_1,x_2))dy_1 \\ &- \int_{\mathbf{R}} \left[\int_{\mathbf{R}} \phi_s^t(y_1,y_2) \left(\delta(h_t(y_1,y_2)) - \delta(L_{s,t}(y_1,y_2))\right) dy_1 \right] \psi_k(x_2-y_2)dy_2. \end{aligned} \tag{6.60}$$

Notice that for $x=(x_1,x_2) \in \Gamma_t$, we have

$$\phi_s^t(x_1,x_2) = \int_{\mathbf{R}} \phi_s^t(y_1,x_2)\delta(h_t(y_1,x_2))dy_1. \tag{6.61}$$

By substituting this into (6.60) we obtain

$$\begin{aligned} &\int_{\mathbf{R}} \phi_s^t(y_1,x_2) \left[\delta(h_t(y_1,x_2)) - \delta(L_{s,t}(y_1,x_2))\right] dy_1 \\ &- \int_{\mathbf{R}} \left[\int_{\mathbf{R}} \phi_s^t(y_1,y_2) \left(\delta(h_t(y_1,y_2)) - \delta(L_{s,t}(y_1,y_2))\right) dy_1 \right] \psi_k(x_2-y_2)dy_2. \end{aligned} \tag{6.62}$$

Observe that the latter term in the above expression is just a Littlewood-Paley projection of the former term; hence it should be expected that these two terms can be handled in a similar way. In [15] this is indeed shown to be the case; hence in the following we will focus on the former term of (6.62).

By the definition of ϕ_s^t in (6.51), we obtain

$$\begin{aligned} &\int_{\mathbf{R}} \phi_s^t(y_1,x_2) \left[\delta(h_t(y_1,x_2)) - \delta(L_{s,t}(y_1,x_2))\right] dy_1 \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} \varphi_s(y_1-t,x_2)\check{\psi}_{k-l}(t)dt \left[\delta(h_t(y_1,x_2)) - \delta(L_{s,t}(y_1,x_2))\right] dy_1. \end{aligned} \tag{6.63}$$

If we denote

$$d := h_t(y_1,x_2) - L_{s,t}(y_1,x_2), \tag{6.64}$$

then the right hand side of (6.63) turns to

$$\begin{aligned} &\int_{\mathbf{R}} \int_{\mathbf{R}} \left(\varphi_s(y_1-t,x_2) - \varphi_s(y_1+d-t,x_2)\right) \check{\psi}_{k-l}(t)dt\delta(h_t(y_1,x_2))dy_1 \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} \varphi_s(y_1-t,x_2) \left(\check{\psi}_{k-l}(t) - \check{\psi}_{k-l}(t+d)\right) dt\delta(h_t(y_1,x_2))dy_1. \end{aligned} \tag{6.65}$$

Hence by the definition of Jones' beta numbers that

$$|d| \lesssim w_s \cdot \beta_0(J^D(t,s)) \tag{6.66}$$

and by applying the fundamental theorem of calculus to $\check{\psi}_{k-l}$, we conclude the desired estimate in Lemma 6.5. □

Acknowledgements

The author would like to thank his advisor, Prof. Christoph Thiele, for his tremendous support. The author also thanks Diogo Oliveira e Silva for discussions on the δ-calculus.

References

[1] P. Auscher, S. Hofmann, C. Muscalu, T. Tao, and C. Thiele, *Carleson measures, trees, extrapolation, and $T(b)$ theorems*, Publ. Mat. **46** (2002), no. 2, 257–325, DOI 10.5565/PUBLMAT_46202_01. MR1934198 (2003f:42019)

[2] Jonas Azzam and Raanan Schul, *Hard Sard: quantitative implicit function and extension theorems for Lipschitz maps*, Geom. Funct. Anal. **22** (2012), no. 5, 1062–1123, DOI 10.1007/s00039-012-0189-0. MR2989430

[3] Michael Bateman, *L^p estimates for maximal averages along one-variable vector fields in $\mathbf{R}^2$*, Proc. Amer. Math. Soc. **137** (2009), no. 3, 955–963, DOI 10.1090/S0002-9939-08-09583-X. MR2457435 (2010e:42023)

[4] Michael Bateman, *Single annulus L^p estimates for Hilbert transforms along vector fields*, Rev. Mat. Iberoam. **29** (2013), no. 3, 1021–1069, DOI 10.4171/RMI/748. MR3090145

[5] Michael Bateman and Christoph Thiele, *L^p estimates for the Hilbert transforms along a one-variable vector field*, Anal. PDE **6** (2013), no. 7, 1577–1600, DOI 10.2140/apde.2013.6.1577. MR3148061

[6] Frédéric Bernicot, *Fiber-wise Calderón-Zygmund decomposition and application to a bi-dimensional paraproduct*, Illinois J. Math. **56** (2012), no. 2, 415–422. MR3161332

[7] J. Bourgain, *A remark on the maximal function associated to an analytic vector field*, Analysis at Urbana, Vol. I (Urbana, IL, 1986), London Math. Soc. Lecture Note Ser., vol. 137, Cambridge Univ. Press, Cambridge, 1989, pp. 111–132. MR1009171 (90h:42028)

[8] Anthony Carbery, Andreas Seeger, Stephen Wainger, and James Wright, *Classes of singular integral operators along variable lines*, J. Geom. Anal. **9** (1999), no. 4, 583–605, DOI 10.1007/BF02921974. MR1757580 (2001g:42026)

[9] Emanuel Carneiro and Diogo Oliveira e Silva, *Some sharp restriction inequalities on the sphere*, Int. Math. Res. Not. IMRN **17** (2015), 8233–8267, DOI 10.1093/imrn/rnu194. MR3404013

[10] Michael Christ, Alexander Nagel, Elias M. Stein, and Stephen Wainger, *Singular and maximal Radon transforms: analysis and geometry*, Ann. of Math. (2) **150** (1999), no. 2, 489–577, DOI 10.2307/121088. MR1726701 (2000j:42023)

[11] R. R. Coifman, Peter W. Jones, and Stephen Semmes, *Two elementary proofs of the L^2 boundedness of Cauchy integrals on Lipschitz curves*, J. Amer. Math. Soc. **2** (1989), no. 3, 553–564, DOI 10.2307/1990943. MR986825 (90k:42017)

[12] Antonio Cordoba, *The Kakeya maximal function and the spherical summation multipliers*, Amer. J. Math. **99** (1977), no. 1, 1–22. MR0447949 (56 #6259)

[13] Ciprian Demeter and Francesco Di Plinio, *Logarithmic L^p bounds for maximal directional singular integrals in the plane*, J. Geom. Anal. **24** (2014), no. 1, 375–416, DOI 10.1007/s12220-012-9340-2. MR3145928

[14] Damiano Foschi, *Maximizers for the Strichartz inequality*, J. Eur. Math. Soc. (JEMS) **9** (2007), no. 4, 739–774, DOI 10.4171/JEMS/95. MR2341830 (2008k:35389)

[15] Shaoming Guo, *Hilbert transform along measurable vector fields constant on Lipschitz curves: L^2 boundedness*, Anal. PDE **8** (2015), no. 5, 1263–1288, DOI 10.2140/apde.2015.8.1263. MR3393679

[16] Peter W. Jones, *Square functions, Cauchy integrals, analytic capacity, and harmonic measure*, Harmonic analysis and partial differential equations (El Escorial, 1987), Lecture Notes in Math., vol. 1384, Springer, Berlin, 1989, pp. 24–68, DOI 10.1007/BFb0086793. MR1013815 (91b:42032)

[17] Nets Hawk Katz, *A partial result on Lipschitz differentiation*, Harmonic analysis at Mount Holyoke (South Hadley, MA, 2001), Contemp. Math., vol. 320, Amer. Math. Soc., Providence, RI, 2003, pp. 217–224, DOI 10.1090/conm/320/05609. MR1979942 (2004e:42030)

[18] Joonil Kim, *Maximal average along variable lines*, Israel J. Math. **167** (2008), 1–13, DOI 10.1007/s11856-008-1037-7. MR2448014 (2009h:42031)

[19] Damiano Foschi and Sergiu Klainerman, *Bilinear space-time estimates for homogeneous wave equations* (English, with English and French summaries), Ann. Sci. École Norm. Sup. (4) **33** (2000), no. 2, 211–274, DOI 10.1016/S0012-9593(00)00109-9. MR1755116 (2001g:35145)

[20] Vjekoslav Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam. **28** (2012), no. 4, 1143–1164, DOI 10.4171/RMI/707. MR2990138

[21] Michael Lacey and Christoph Thiele, *L^p estimates for the bilinear Hilbert transform*, Proc. Nat. Acad. Sci. U.S.A. **94** (1997), no. 1, 33–35, DOI 10.1073/pnas.94.1.33. MR1425870 (98e:44001)

[22] Michael T. Lacey and Christoph M. Thiele, *On Calderón's conjecture for the bilinear Hilbert transform*, Proc. Natl. Acad. Sci. USA **95** (1998), no. 9, 4828–4830 (electronic), DOI 10.1073/pnas.95.9.4828. MR1619285 (99e:42013)

[23] Michael Lacey and Christoph Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. **7** (2000), no. 4, 361–370, DOI 10.4310/MRL.2000.v7.n4.a1. MR1783613 (2001m:42009)

[24] Michael T. Lacey and Xiaochun Li, *Maximal theorems for the directional Hilbert transform on the plane*, Trans. Amer. Math. Soc. **358** (2006), no. 9, 4099–4117 (electronic), DOI 10.1090/S0002-9947-06-03869-4. MR2219012 (2006k:42018)

[25] Michael Lacey and Xiaochun Li, *On a conjecture of E. M. Stein on the Hilbert transform on vector fields*, Mem. Amer. Math. Soc. **205** (2010), no. 965, viii+72, DOI 10.1090/S0065-9266-10-00572-7. MR2654385 (2011c:42019)

[26] Oskar Perron, *Über einen Satz von Besicovitsch* (German), Math. Z. **28** (1928), no. 1, 383–386, DOI 10.1007/BF01181172. MR1544966

[27] Elias M. Stein and Brian Street, *Multi-parameter singular Radon transforms III: Real analytic surfaces*, Adv. Math. **229** (2012), no. 4, 2210–2238, DOI 10.1016/j.aim.2011.11.016. MR2880220

Institute of Mathematics, University of Bonn, Endenicher Allee 60, 53115, Bonn, Germany

Current address: Department of Mathematics, Indiana University, Bloomington, Indiana 47405

E-mail address: `shaoguo@iu.edu`