

SQUARE FUNCTIONS FOR BI-LIPSCHITZ MAPS AND DIRECTIONAL OPERATORS

FRANCESCO DI PLINIO, SHAOMING GUO, CHRISTOPH THIELE, AND PAVEL ZORIN-KRANICH

ABSTRACT. First we prove a Littlewood-Paley diagonalization result for bi-Lipschitz perturbations of the identity map on the real line. This result entails a number of corollaries for the Hilbert transform along lines and monomial curves in the plane. Second, we prove a square function bound for a single scale directional operator. As a corollary we give a new proof of part of a theorem of Katz on direction fields with finitely many directions.

1. INTRODUCTION

This paper grew out of a study of variable directional operators in the plane. We present two main results together with some corollaries.

It is a folklore conjecture and discussed by several authors, for example [Ste93], [LL10], [Guo17], that Lipschitz is the critical regularity assumption on a direction field to yield L^p boundedness of some associated directional operators. Possibly at the heart of positive results in this direction appears to be a one dimensional Littlewood-Paley diagonalization estimate for bi-Lipschitz maps, which is our first main theorem.

Theorem 1.1. *Let $A : \mathbb{R} \rightarrow \mathbb{R}$ be a Lipschitz function with $\|A\|_{\text{Lip}} \leq 1/100$ and consider the change of variable $T_A f(x) := f(x + A(x))$.*

*Let ψ be a Schwartz function on \mathbb{R} such that $\widehat{\psi}$ identically equals 1 on $\pm[\frac{99}{100}, \frac{103}{100}]$ and vanishes outside of $\pm[\frac{98}{100}, \frac{104}{100}]$. Let Ψ be another Schwartz function on \mathbb{R} such that $\widehat{\Psi}$ is supported on $\pm[1, 101/100]$. Let $P_t f := \psi_t * f$ be the Littlewood-Paley operators associated to ψ , where $\psi_t(x) = t^{-1}\psi(t^{-1}x)$. Then*

$$\left\| \sum_{t \in 2^{\mathbb{Z}}} |(1 - P_t)T_A(\Psi_t * f)| \right\|_p \lesssim_{p, \psi, \Psi} \|A\|_{\text{Lip}} \|f\|_p, \quad 1 < p < \infty.$$

Note that when the Lipschitz norm of A becomes too large, then in general T_A fails to be a bijection and the estimate of the theorem breaks down. By rescaling with $c > 0$ and a convexity argument the estimate of the theorem remains true for the following expressions in place of the left hand side:

$$\left\| \sum_{t \in 2^{\mathbb{Z}}} |(1 - P_{ct})T_A(\Psi_{ct} * f)| \right\|_p, \quad \left\| \int_0^\infty |(1 - P_t)T_A(\Psi_t * f)| \frac{dt}{t} \right\|_p.$$

2010 *Mathematics Subject Classification.* 42B25.

Key words and phrases. Littlewood-Paley diagonalization, bi-Lipschitz maps, directional square functions, directional Hilbert transform.

FDP was partially supported by NSF grants DMS-1500449 and DMS-1650810, by the Severo Ochoa Program SEV-2013-0323 and by Basque Government BERG Program 2014-2017. SG and CT acknowledge support by the NSF under grant DMS-1440140 through participation in the harmonic analysis program at MSRI in Spring 2017. CT and PZK acknowledge support by DFG-SFB 1060 and the Hausdorff Center for Mathematics in Bonn.

We call this result a Littlewood-Paley diagonalization result, since it compares for suitable normalization of ψ and Ψ

$$T_A f = \int_0^\infty T_A(\Psi_t * f) \frac{dt}{t} = \int_0^\infty \int_0^\infty P_s T_A(\Psi_t * f) \frac{dt}{t} \frac{ds}{s}$$

with the diagonal term

$$\int_0^\infty P_t T_A(\Psi_t * f) \frac{dt}{t}.$$

The diagonal term by Littlewood-Paley theory and the Fefferman-Stein maximal theorem can and typically will be controlled in L^p norm by that of any of the following square functions

$$(1.2) \quad \left(\int_0^\infty |P_t T_A(\Psi_t * f)|^2 \frac{dt}{t} \right)^{1/2}, \left(\int_0^\infty |M T_A(\Psi_t * f)|^2 \frac{dt}{t} \right)^{1/2}, \left(\int_0^\infty |T_A(\Psi_t * f)|^2 \frac{dt}{t} \right)^{1/2},$$

where M denotes the Hardy-Littlewood maximal operator.

An application of Theorem 1.1 is to the directional Hilbert transform in the plane defined for measurable $u : \mathbb{R}^2 \rightarrow [-1, 1]$ as

$$H_u f(x, y) := \text{p. v.} \int_{-1}^{+1} f(x+r, y+u(x, y)r) \frac{dr}{r}.$$

Corollary 1.3. *Assume that $u(x, \cdot)$ has Lipschitz constant $\leq 1/100$ for almost every $x \in \mathbb{R}$. With notation as in Theorem 1.1, we have*

$$\left\| \sum_{t \in 2^{\mathbb{Z}}} |(1 - P_t) H_u(\Psi_t * f)| \right\|_p \lesssim_{p, \phi, \Psi} \|f\|_p, \quad 1 < p < \infty.$$

Here P_t and convolution with Ψ_t act in the second variable.

As outlined above, this theorem reduces bounds for H_u to bounds for a square function. The L^2 part of the following corollary is then immediate.

Corollary 1.4. *Let $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ be such that $u(x, \cdot)$ has Lipschitz constant $\leq 1/100$ for almost every $x \in \mathbb{R}$. Assume further with notation as in Theorem 1.1 that*

$$(1.5) \quad \sup_{0 < t < t_0} \|H_u(\Psi_t * f)\|_{p_0} \lesssim \|f\|_{p_0}$$

for some $1 < p_0 \leq 2$ and $t_0 > 0$. If $p_0 = 2$, then

$$(1.6) \quad \|H_u f\|_2 \lesssim \|f\|_2.$$

If $1 < p_0 < 2$, then

$$(1.7) \quad \|H_u f\|_p \lesssim \|f\|_p, \quad 1 + \frac{1}{3 - p_0} < p < \infty.$$

Lacey and Li [LL10] proved (1.5) for all $2 < p_0 < \infty$ (including a weak type (2, 2) endpoint), and they stated a condition [LL10, Conjecture 1.14] on u under which they extended (1.5) to all p_0 in a neighborhood of 2. That condition is known to hold for analytic vector fields, and more generally for a class of vector fields previously considered by Bourgain [Bou89]. Lacey and Li have also deduced (1.6) from (1.5) with $p_0 = 2$ with Lipschitz assumption on the vector field replaced by $C^{1+\eta}$.

The estimate (1.5) for all $1 < p_0 < \infty$ is known for 1-parameter vector fields [Bat13] and vector fields constant along Lipschitz curves [Guo17]. In these cases the conclusion (1.7) has been obtained in [BT13] and [Guo17], respectively. Our argument for the corollary follows closely [BT13], the main additional observation being that (1.5) can be used as a black box, whereas in [BT13] elements of the proof of this estimate for one-parameter vector fields have been used.

We also recall that the Lipschitz regularity hypothesis in Corollary 1.4 cannot be substantially relaxed. Once the segments of integration emanating from the points of a fixed vertical line start to overlap, they may do so in a bad way and one can disprove L^p boundedness by testing on characteristic functions of Perron trees, see e.g. [Ste93, Section X.1].

Adding curvature to the picture by defining

$$(1.8) \quad H_u^{(\alpha)} f(x) := \int_{-1}^1 f(x+r, y+u(x,y)r^\alpha) \frac{dr}{r},$$

where r^α may be interpreted either as $|r|^\alpha$ or $\text{sgn}(r)|r|^\alpha$, we may argue similarly as above but remove the conditionality thanks to the results in [Guo+16]. We obtain

Corollary 1.9. *For every $0 < \alpha < \infty$, $\alpha \neq 1$, and every $1 < p < \infty$, there exists $\epsilon_0 > 0$ such that for every Lipschitz function u with $\|u\|_{\text{Lip}} \leq \epsilon_0$, we have*

$$(1.10) \quad \|H_u^{(\alpha)} f\|_p \lesssim \|f\|_p.$$

Our second main result concerns bounds for the square function of the single scale directional operator

$$(1.11) \quad A_{u,\phi} f(x,y) := \int_{-\infty}^{+\infty} \phi(r) f(x+r, y+u(x,y)r) dr$$

associated to a Schwartz function ϕ .

Theorem 1.12. *Let $u : \mathbb{R}^2 \rightarrow [-1, 1]$ be a measurable function. Then*

$$(1.13) \quad \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |A_{u,\phi} P_t f|^2 \right)^{1/2} \right\|_p \lesssim_{p,\phi} \|f\|_p, \quad 2 < p < \infty.$$

The operator $A_{u,\phi}$ is in general not bounded on L^p unless $p = \infty$. Even if we assume u to be Lipschitz in the vertical direction, we may not apply our first main theorem if ϕ does not have suitable compact support, and the operator $A_{u,\phi}$ remains unbounded in general.

As an application of this result, we elaborate on a remark made by Demeter in [Dem10].

Corollary 1.14. *Assume the measurable function $u : \mathbb{R}^2 \rightarrow [-1, 1]$ takes at most N different values. Then*

$$(1.15) \quad \|A_{u,\phi} f\|_p \lesssim_{p,\phi} \log(N+2)^{1/2} \|f\|_p, \quad 2 < p < \infty.$$

Indeed, Demeter proves the sharper endpoint version of this estimate for $p = 2$, reproducing an earlier result by Katz [Kat99]. Demeter proposes an alternative proof of this result using an inequality by Chang, Wilson, and Wolff [CWW85], in the same vein as in his proof of [Dem10, Theorem 2]. Theorem 1.12 allows to follow through with this proposal, albeit only for $p > 2$. For the operator obtained by replacing ϕ in (1.11) with a one-dimensional singular integral kernel, the same quantitative estimate as (1.15), up to ϵ -losses in the power of $\log N$ when $p > 2$ is sufficiently close to 2, holds when the finite range of u is assumed to have additional structure [DD14]. For instance, one may take $u(\mathbb{R}^2) = \{2k/N : k = -N/2, \dots, N/2\}$. Thus, it is of interest whether the methods behind Corollary 1.14 may be applied to the singular integral case, with the aim of lifting the structure restrictions appearing in [DD14].

2. LIPSCHITZ VECTOR FIELDS

2.1. Carleson embeddings with compactly supported test functions. We refer to [DT15, Section 2 and 3] for the general theory of outer measure spaces. In this section

we use the outer measure space $X = \mathbb{R}^d \times (0, \infty)$ with the collection of distinguished sets \mathbf{E} consisting of the *tents*

$$T(x, s) = \{(y, t) : \|x - y\| + t \leq s\}$$

and an outer measure μ generated by $\sigma(T(x, s)) = s^d$.

Let ω be a *Dini modulus of continuity*, that is, $\omega : [0, \infty) \rightarrow [0, \infty)$ is a function that is subadditive in the sense

$$u \leq s + t \implies \omega(u) \leq \omega(s) + \omega(t)$$

and has finite Dini norm $\|\omega\|_{\text{Dini}} = \int_0^1 \omega(t) \frac{dt}{t}$. Let \mathcal{C} be the class of testing functions $\phi : \mathbb{R}^d \rightarrow \mathbb{C}$ that satisfy

$$(2.1) \quad \int \phi(z) dz = 0,$$

$$(2.2) \quad \text{supp } \phi \subset B(0, 1)$$

$$(2.3) \quad |\phi(z) - \phi(z')| \leq \omega(\|z - z'\|) \quad \text{for all } z, z' \in \mathbb{R}^d.$$

For locally integrable functions f we define the embeddings

$$\begin{aligned} \mathcal{A}_c f(x, t) &:= t^{-d} \int_{B(x, t)} |f|, \\ \mathcal{D}_c f(x, t) &:= \sup_{\phi \in \mathcal{C}} \left| t^{-d} \int f(z) \phi(t^{-1}(y - z)) dz \right|. \end{aligned}$$

Theorem 2.4 (cf. [DT15, Theorem 4.1]). *For every $1 < p \leq \infty$ we have*

$$\begin{aligned} \|\mathcal{A}_c f\|_{L^p(S^\infty)} &\lesssim \|f\|_{L^p(\mathbb{R}^d)}, \\ \|\mathcal{D}_c f\|_{L^p(S^2)} &\lesssim \|f\|_{L^p(\mathbb{R}^d)}. \end{aligned}$$

Moreover, we have the endpoint estimates

$$\begin{aligned} \|\mathcal{A}_c f\|_{L^{1, \infty}(S^\infty)} &\lesssim \|f\|_{L^1(\mathbb{R}^d)}, \\ \|\mathcal{D}_c f\|_{L^{1, \infty}(S^2)} &\lesssim \|f\|_{L^1(\mathbb{R}^d)}. \end{aligned}$$

The main difference from [DT15, Theorem 4.1] is the supremum over $\phi \in \mathcal{C}$ in the definition of \mathcal{D}_c , whereas [DT15, Theorem 4.1] uses a fixed ϕ . This supremum does not affect the proof strongly, but is important for our application. The precise choice of the class of test functions \mathcal{C} is not important for this application, but the Dini regularity condition appears naturally in the proof.

We linearize the supremum in the definition of $\mathcal{D}_c f$ by choosing for each pair (y, t) a function $\phi \in \mathcal{C}$ for which the supremum is almost attained. Denote then $\phi_{y, t}(z) = t^{-d} \phi(t^{-1}(y - z))$. This is an L^1 normalized wave packet at scale t . The almost orthogonality of these wave packets is captured by the following estimate.

Lemma 2.5. *If $t \leq t'$ then*

$$|\langle \phi_{y, t}, \phi_{y', t'} \rangle| \lesssim (t')^{-d} \omega(t/t')$$

Proof. Using the cancellation condition (2.1) and the support condition we write

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \phi_{y, t}(z) \phi_{y', t'}(z) dz \right| &= \left| \int_{B(y, t)} \phi_{y, t}(z) (\phi_{y', t'}(z) - \phi_{y', t'}(y)) dz \right| \\ &\leq \int_{B(y, t)} |\phi_{y, t}(z)| (t')^{-d} \omega(t/t') dz \lesssim (t')^{-d} \omega(t/t'). \quad \square \end{aligned}$$

We use the almost orthogonality statement in Lemma 2.5 to deduce a square function estimate for $p = 2$.

Lemma 2.6.

$$(2.7) \quad \int_{\mathbb{R}^d \times \mathbb{R}_{>0}} |\mathcal{D}_c f(y, t)|^2 dy \frac{dt}{t} \lesssim \|f\|_2^2.$$

Proof. We begin with a measurable selection of functions $\phi_{y,t}$ that almost extremize $\mathcal{D}_c f(y, t)$. Expand the square of the left hand side of (2.7)

$$\begin{aligned} \left(\int |\langle f, \phi_{y,t} \rangle|^2 dy \frac{dt}{t} \right)^2 &= \left(\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d \times \mathbb{R}_{>0}} \langle f, \phi_{y,t} \rangle \phi_{y,t}(z) dy \frac{dt}{t} \right) \overline{f(z)} dz \right)^2 \\ &\leq \left\| \int_{\mathbb{R}^d \times \mathbb{R}_{>0}} \langle f, \phi_{y,t} \rangle \phi_{y,t}(z) dy \frac{dt}{t} \right\|_2^2 \|f\|_2^2 \end{aligned}$$

We further expand the square from the former term

$$\begin{aligned} &= \iint \langle f, \phi_{y,t} \rangle \langle \phi_{y,t}, \phi_{y',t'} \rangle \langle \phi_{y',t'}, f \rangle dy \frac{dt}{t} dy' \frac{dt'}{t'} \|f\|_2^2 \\ &\leq \int |\langle f, \phi_{y,t} \rangle|^2 \int |\langle \phi_{y,t}, \phi_{y',t'} \rangle| dy' \frac{dt'}{t'} dy \frac{dt}{t} \|f\|_2^2, \end{aligned}$$

using the estimate

$$2|\langle f, \phi_{y,t} \rangle \langle \phi_{y',t'}, f \rangle| \leq |\langle f, \phi_{y,t} \rangle|^2 + |\langle f, \phi_{y',t'} \rangle|^2$$

in the last inequality. It suffices to verify

$$\sup_{y,t} \int |\langle \phi_{y,t}, \phi_{y',t'} \rangle| dy' \frac{dt'}{t'} < \infty.$$

By Lemma 2.5 and using bounded support of the $\phi_{y,t}$'s we have

$$\begin{aligned} \int |\langle \phi_{y,t}, \phi_{y',t'} \rangle| dy' \frac{dt'}{t'} &\lesssim \int_{t \leq t'} \int_{\|y-y'\| \leq t+t'} (t')^{-d} \omega(t/t') dy' \frac{dt'}{t'} \\ + \int_{t > t'} \int_{\|y-y'\| \leq t+t'} t^{-d} \omega(t'/t) dy' \frac{dt'}{t'} \\ &\lesssim \int_{t \leq t'} \omega(t/t') \frac{dt'}{t'} + \int_{t > t'} \omega(t'/t) \frac{dt'}{t'} \lesssim \|\omega\|_{\text{Dini}}. \end{aligned}$$

This finishes the proof of Lemma 2.6. \square

Proof of Theorem 2.4. We may assume that the superlevel sets $\{Mf > \lambda\}$, where M is the uncentered Hardy–Littlewood maximal function, have finite measure for all $\lambda > 0$, since otherwise the right-hand side of the conclusion is infinite.

Let $\{Q_i\}_i$ be a Whitney decomposition of the superlevel set $\{Mf > \lambda\}$. Let x_i denote the center and r_i the diameter of Q_i . Let

$$(2.8) \quad E := \bigcup_i T(x_i, 3\sqrt{d}r_i)$$

and note that

$$\mu(E) \lesssim |\{Mf > \lambda\}|.$$

The claim of the theorem will therefore follow from the more precise results

$$(2.9) \quad \|\mathcal{A}_c f \mathbf{1}_{E^c}\|_{L^\infty(S^\infty)} \lesssim \lambda,$$

$$(2.10) \quad \|\mathcal{D}_c f \mathbf{1}_{E^c}\|_{L^\infty(S^2)} \lesssim \lambda.$$

Let $(x, t) \in E^c$. Then no ball $B(y, t/\sqrt{d})$ with $\|x - y\| \leq t$ is contained in a Whitney cube. It follows that, for some constant C that depends only on the dimension, the ball $B(x, Ct)$ is not contained in $\{Mf > \lambda\}$. Hence

$$\mathcal{A}_c f(x, t) \leq t^{-d} \int_{B(x, Ct)} |f| \lesssim \lambda.$$

This completes the proof of (2.9). Now we show (2.10). The Calderón–Zygmund decomposition $f = g + b$, $b = \sum_i b_i$ associated to the Whitney decomposition $\{Q_i\}_i$ has the properties

- (1) $\|g\|_\infty \lesssim \lambda$,
- (2) $\text{supp } b_i \subset Q_i$,
- (3) $\int b_i = 0$,
- (4) $|Q_i|^{-1} \int |b_i| \lesssim \lambda$.

Using the bounded support condition on the wave packets and Lemma 2.6 we obtain

$$\begin{aligned} S^2(\mathcal{D}_c g)(T(x, s)) &= \left(\frac{1}{s^d} \int_{T(x, s)} |\mathcal{D}_c g(y, t)|^2 dy \frac{dt}{t} \right)^{1/2} \\ &= \left(\frac{1}{s^d} \int_{T(x, s)} |\mathcal{D}_c (g \mathbf{1}_{B(x, 2s)})(y, t)|^2 dy \frac{dt}{t} \right)^{1/2} \\ &\lesssim s^{-d/2} \|g \mathbf{1}_{B(x, 2s)}\|_2 \lesssim \|g\|_\infty. \end{aligned}$$

Hence (2.10) holds with f replaced by g . By sublinearity of the embedding map \mathcal{D} and subadditivity of the outer $L^\infty(S^2)$ norm it remains to show (2.10) holds with f replaced by b . More explicitly, for every tent $T = T(x, r)$ we want to show

$$S^2(\mathcal{D}_c b \mathbf{1}_{E^c})(T) \lesssim \lambda.$$

We know

$$S^\infty(\mathcal{D}_c b \mathbf{1}_{E^c})(T) \lesssim S^\infty(\mathcal{D}_c f \mathbf{1}_{E^c})(T) + S^\infty(\mathcal{D}_c g \mathbf{1}_{E^c})(T) \lesssim S^\infty(\mathcal{A}_c f \mathbf{1}_{E^c})(T) + \lambda \lesssim \lambda.$$

By logarithmic convexity of S^p sizes it therefore suffices to show

$$(2.11) \quad S^1(\mathcal{D}_c b \mathbf{1}_{E^c})(T) \lesssim \lambda.$$

Claim 2.12. $\int_{t>r_i} \mathcal{D}_c b_i(x, t) dx \frac{dt}{t} \lesssim \lambda r_i^d$.

Proof of Claim 2.12. Notice that, due to support constraints, $\mathcal{D}_c b_i(x, t)$ can only be non-zero if $\|x - x_i\| \leq r_i/2 + t$. Moreover, under this condition and choosing $\phi_{x, t}$ that almost extremizes $\mathcal{D}_c b_i(x, t)$ we obtain

$$\begin{aligned} \left| \int b_i(z) \phi_{x, t}(z) dz \right| &= \left| \int_{\|z - x_i\| \leq r_i/2} b_i(z) (\phi_{x, t}(z) - \phi_{x, t}(x_i)) dz \right| \\ &\leq t^{-d} \omega(r_i/(2t)) \int_{\|z - x_i\| \leq r_i/2} |b_i(z)| dz \\ &\lesssim t^{-d} \omega(r_i/(2t)) \lambda r_i^d. \end{aligned}$$

Hence

$$\begin{aligned} \int_{t>r_i} \mathcal{D}_c b_i(x, t) dx \frac{dt}{t} &\leq \int_{t>r_i, \|x - x_i\| \leq t + r_i/2} \mathcal{D}_c b_i(x, t) dx \frac{dt}{t} \\ &\lesssim \int_{t>r_i} \omega(r_i/(2t)) \lambda r_i^d \frac{dt}{t} \lesssim \lambda r_i^d \|\omega\|_{\text{Dini}}. \end{aligned}$$

This finishes the proof of Claim 2.12. \square

In order to show (2.11) notice that only the Whitney cubes $Q_i \subset B(x, 10r)$ contribute to $\mathcal{D}_c b \mathbf{1}_{T \setminus E}$.

$$\begin{aligned} S^1(\mathcal{D}_c b \mathbf{1}_{E^c})(T) &= r^{-d} \int_{T \setminus E} \mathcal{D}_c b(z, t) dz \frac{dt}{t} \\ &\leq r^{-d} \sum_{i: Q_i \subset B(x, 10r)} \int_{T \setminus E} \mathcal{D}_c b_i(z, t) dz \frac{dt}{t} \\ &\leq r^{-d} \sum_{i: Q_i \subset B(x, 10r)} \int_{t > r_i} \mathcal{D}_c b_i(z, t) dz \frac{dt}{t} \end{aligned}$$

using Claim 2.12

$$\lesssim r^{-d} \lambda \sum_{i: Q_i \subset B(x, 10r)} |Q_i|$$

by disjointness of Whitney cubes

$$\lesssim r^{-d} \lambda |B(x, 10r)| \lesssim \lambda.$$

This finishes the proof of Theorem 2.4. \square

2.2. Carleson embeddings with tails. It is possible to adapt the proofs in Section 2.1 to embeddings defined using test functions with tails. Since we do not need testing functions with sharp decay rates for tails, we will instead estimate such embeddings by averaging the results in Section 2.1.

In this section we work in dimension $d = 1$ and consider the following embedding maps:

$$(2.13) \quad \mathcal{A}f(x, t) := \int t^{-1} (1 + |x - y|/t)^{-5} |f(y)| dy,$$

$$(2.14) \quad \mathcal{D}f(x, t) := \sup_{\phi \in \Phi} \left| \int t^{-1} \phi((x - y)/t) f(y) dy \right|,$$

where

$$\Phi = \{ \phi : \mathbb{R} \rightarrow \mathbb{C}, \int \phi = 0, |\phi(x)| \leq (1 + |x|)^{-10}, |\phi'(x)| \leq (1 + |x|)^{-10} \}.$$

The smoothness and decay conditions in these embeddings are not optimal, but they suffice for our purposes. Decomposing the testing functions $(1 + |x|)^{-5}$ and $\phi \in \Phi$ into series of compactly supported bump functions as in [Mus+06, Lemma 3.1], see also Lemma 3.4 in this article, we can deduce the embeddings

$$(2.15) \quad \|\mathcal{A}f\|_{L^p(S^\infty)} \lesssim \|f\|_p,$$

$$(2.16) \quad \|\mathcal{D}f\|_{L^p(S^2)} \lesssim \|f\|_p$$

for $1 < p \leq \infty$ from Theorem 2.4.

2.3. Jones beta numbers. Let $A : \mathbb{R} \rightarrow \mathbb{C}$ be a Lipschitz function and let a be its distributional derivative, so that $\|a\|_\infty = \|A\|_{\text{Lip}}$. Let ψ be a compactly supported bump function with

$$(2.17) \quad \int \psi(x) dx = \int x \psi(x) dx = 0$$

and

$$\int_0^\infty \hat{\psi}(\xi \xi_0) \frac{d\xi}{\xi} = 1 \quad \text{for } \xi_0 \neq 0.$$

Let $\psi_t = t^{-1}\psi(t^{-1}\cdot)$ be an L^1 normalized mean zero bump function at scale t . Let

$$(2.18) \quad \alpha(x, t) := \int_t^\infty a * \psi_s(x) \frac{ds}{s}$$

be the average slope of A near x at scale t and let

$$(2.19) \quad \beta_n(x, t) := \sup_{x_0, x_1, x_2 \in B(x, 2^n \cdot 3t), 2^{-n}t \leq \tilde{t} \leq 2^n t} t^{-1} |A(x_2) - A(x_1) - \alpha(x_0, \tilde{t})(x_2 - x_1)|.$$

This definition includes the supremum over the range of uncertainty around (x, t) , which seems convenient.

Lemma 2.20. *With the notation (2.19) we have*

$$\beta_n(x, t) \lesssim t^{-1} \int_0^{2^n t} \left(\int_{|y-x| \lesssim 2^n t} \mathcal{D}a(y, s)^2 s^{-1} dy \right)^{1/2} ds + 2^{2n} \int_{2^n t}^\infty \mathcal{D}a(x, s) \frac{t ds}{s^2}$$

Proof. Let $x_0, x_1, x_2 \in B(x, 2^n \cdot 3t)$, $2^{-n}t \leq \tilde{t} \leq 2^n t$. By the fundamental theorem of calculus and Calderón's reproducing formula for a we can write

$$\begin{aligned} t^{-1}(A(x_2) - A(x_1) - \alpha(x_0, \tilde{t})(x_2 - x_1)) &= t^{-1} \int_{x_1}^{x_2} a(y) dy - t^{-1} \alpha(x_0, \tilde{t})(x_2 - x_1) \\ &= t^{-1} \int_{x_1}^{x_2} \int_0^\infty a * \psi_s(y) \frac{ds}{s} dy - t^{-1} \int_{x_1}^{x_2} \int_{\tilde{t}}^\infty a * \psi_s(x_0) \frac{ds}{s} dy. \end{aligned}$$

Splitting the integral in s in the former term at \tilde{t} we further obtain

$$\begin{aligned} &= t^{-1} \int_{x_1}^{x_2} \int_0^{\tilde{t}} a * \psi_s(y) \frac{ds}{s} dy + t^{-1} \int_{x_1}^{x_2} \int_{\tilde{t}}^\infty (a * \psi_s(y) - a * \psi_s(x_0)) \frac{ds}{s} dy \\ &=: I + II. \end{aligned}$$

We estimate the two terms on the right-hand side separately. In the first term we note $\psi_s = s(\tilde{\psi}_s)'$, where $\tilde{\psi}_s$ is also an L^1 normalized mean zero bump function at scale s , by assumption (2.17). Therefore

$$\begin{aligned} I &\leq t^{-1} \int_0^{2^n t} \left| \int_{x_1}^{x_2} a * \psi_s(y) dy \right| \frac{ds}{s} \\ &= t^{-1} \int_0^{2^n t} |a * \tilde{\psi}_s(x_2) - a * \tilde{\psi}_s(x_1)| ds \\ &\lesssim t^{-1} \int_0^{2^n t} \sup_{|y-x| \lesssim 2^n t} \mathcal{D}a(y, s) ds. \end{aligned}$$

Since $\mathcal{D}a(\cdot, s)$ is almost constant at scale s , this can be further estimated by

$$I \lesssim t^{-1} \int_0^{2^n t} \left(\int_{|y-x| \lesssim 2^n t} \mathcal{D}a(y, s)^2 s^{-1} dy \right)^{1/2} ds.$$

We split the second term $II \leq II_a + II_b$ via

$$(2.21) \quad \int_{\tilde{t}}^\infty \leq \int_{2^{-n}t}^{2^n t} + \int_{2^n t}^\infty.$$

Then

$$II_a \leq t^{-1} \int_{2^{-n}t}^{2^n t} \sup_{|y-x| \lesssim 2^n t} |a * \psi_s(y)| \frac{ds}{s},$$

and this can be absorbed into the estimate for I . The latter term from (2.21) is bounded by

$$II_b \lesssim t^{-1} \int_{x_1}^{x_2} \int_{2^{nt}}^{\infty} |a * [\psi_s(\cdot - x + y) - \psi_s(\cdot - x + x_0)](x)| \frac{ds}{s} dy.$$

Since $|x - y|, |x - x_0| \lesssim 2^n t \lesssim s$, the function in the square brackets is a mean zero L^1 normalized bump function at scale s with constant $\lesssim 2^n t/s$ by the fundamental theorem of calculus, so

$$II_b \lesssim t^{-1} \int_{x_1}^{x_2} \int_{2^{nt}}^{\infty} \mathcal{D}a(x, s) \frac{2^n t ds}{s^2} dy \lesssim 2^{2n} \int_{2^{nt}}^{\infty} \mathcal{D}a(x, s) \frac{t ds}{s^2}.$$

This finishes the proof of Lemma 2.20. \square

Lemma 2.22 (cf. [Jon89, Lemma 3]). $\|\beta_n\|_{L^\infty(S^2)} \lesssim 2^{3n/2} \|a\|_\infty$.

Proof. We have to show

$$\frac{1}{t_0} \int_{t < t_0} \int_{|x-x_0| < t_0} \beta_n(x, t)^2 dx \frac{dt}{t} \lesssim 2^{3n} \|a\|_\infty^2$$

with the implicit constant independent of $(x_0, t_0) \in \mathbb{R} \times \mathbb{R}_+$.

We estimate the S^2 size on the tent centered at x_0 with height t_0 separately for the two terms in the conclusion of Lemma 2.20. For the first term we consider the square of the S^2 size:

$$\frac{1}{t_0} \int_{t < t_0} \int_{|x-x_0| < t_0} \left(t^{-1} \int_0^{2^{nt}} \left(\int_{|y-x| \lesssim 2^{nt}} \mathcal{D}a(y, s)^2 s^{-1} dy \right)^{1/2} ds \right)^2 dx \frac{dt}{t}$$

Apply Hölder's inequality in the s -variable

$$\leq \frac{1}{t_0} \int_{t < t_0} \int_{|x-x_0| < t_0} \int_0^{2^{nt}} \left(\int_{|y-x| \lesssim 2^{nt}} \mathcal{D}a(y, s)^2 dy \right) \frac{ds}{s^{1/2}} \cdot \int_0^{2^{nt}} \frac{ds}{s^{1/2}} dx \frac{dt}{t^3}$$

Change the order of integration

$$\begin{aligned} &\lesssim \frac{2^{n/2}}{t_0} \int_{s \leq 2^n t_0} \int_{2^{-n}s < t < t_0} \int_{|y-x_0| \lesssim 2^n t_0} \int_{|x-y| \lesssim 2^n t} dx \mathcal{D}a(y, s)^2 dy \frac{dt}{t^{5/2}} \frac{ds}{s^{1/2}} \\ &\lesssim \frac{2^{2n}}{t_0} \int_{s \leq 2^n t_0} \int_{|y-x_0| \lesssim 2^n t_0} \mathcal{D}a(y, s)^2 dy \frac{ds}{s} \lesssim 2^{3n} \|\mathcal{D}a\|_{L^\infty(S^2)}^2. \end{aligned}$$

For the second term we consider the S^2 size

$$2^{2n} \left(\frac{1}{t_0} \int_{t < t_0} \int_{|x-x_0| < t_0} \left(\int_{2^{nt}}^{\infty} \mathcal{D}a(x, s) \frac{t ds}{s^2} \right)^2 dx \frac{dt}{t} \right)^{1/2}$$

By applying a change of variable $s \rightarrow t\tau$ and Minkowski's integral inequality:

$$\begin{aligned} &\leq 2^{2n} \int_{2^n}^{\infty} \left(\frac{1}{t_0} \int_{t < t_0} \int_{|x-x_0| < t_0} \mathcal{D}a(x, t\tau)^2 dx \frac{dt}{t} \right)^{1/2} \frac{d\tau}{\tau^2} \\ &= 2^{2n} \int_{2^n}^{\infty} \left(\frac{1}{\tau t_0} \int_{s < \tau t_0} \int_{|x-x_0| < t_0} \mathcal{D}a(x, s)^2 dx \frac{ds}{s} \right)^{1/2} \frac{d\tau}{\tau^{3/2}} \\ &\lesssim 2^{2n} \int_{2^n}^{\infty} \|\mathcal{D}a\|_{L^\infty(S^2)} \frac{d\tau}{\tau^{3/2}} \lesssim 2^{3n/2} \|\mathcal{D}a\|_{L^\infty(S^2)}. \end{aligned}$$

The conclusion follows from (2.16). \square

Corollary 2.23 (cf. [Jon89, Lemma 4]). *Let $\epsilon > 0$ and*

$$\beta(x, t) = \sup_{x_0, x_1, x_2 \in \mathbb{R}, \tilde{t} > 0} \left(1 + \frac{\max_i(|x_i - x|)}{t} + \frac{\tilde{t}}{t} + \frac{t}{\tilde{t}}\right)^{-3/2-\epsilon} \frac{|A(x_2) - A(x_1) - \alpha(x_0, \tilde{t})(x_2 - x_1)|}{t}.$$

Then

$$\|\beta\|_{L^\infty(S^2)} \lesssim \|a\|_\infty.$$

The difference from the original formulation of Jones's beta number estimate is that we take a supremum over an uncertainty region in all available parameters.

2.4. Littlewood–Paley diagonalization of Lipschitz change of variables.

Proof of Theorem 1.1. Since the Lipschitz norm of A is strictly smaller than 1, the change of variable $x \mapsto x + A(x)$ is invertible and bi-Lipschitz. Denote its inverse function by b , so that $z = b(z) + A(b(z))$.

Write

$$T_A(\Psi_t * f)(x) = T_A(\Psi_t * P_t f)(x) = \int_{-\infty}^{\infty} P_t f(z) \Psi_t(x + A(x) - z) dz.$$

This integral is a linear combination of the functions $x \mapsto \Psi_t(x + A(x) - z)$ that we view as non-linear deformations of wave packets centered at $b(z)$. The main idea is to replace the non-linear change of variable $x \mapsto x + A(x) - z$ in the argument of Ψ_t by the linear change of variable $x \mapsto (1 + \alpha(b(z), t))(x - b(z))$, where α is the average slope of the function A in the sense of (2.18). Since $|\alpha| \leq \|A\|_{\text{Lip}}$, the function

$$x \mapsto \int_{-\infty}^{\infty} P_t f(z) \Psi_t((1 + \alpha(b(z), t))(x - b(z))) dz$$

has Fourier support inside $t^{-1}[99/100, 103/100]$, so it is annihilated by $I - P_t$.

It remains to estimate the error that has been made in approximating the non-linear change of coordinates in the argument of Ψ_t by a linear one. To this end we compute the difference of the arguments:

$$(2.24) \quad |x + A(x) - z - (1 + \alpha(b(z), t))(x - b(z))| = |(A(x) - A(b(z)) - \alpha(b(z), t)(x - b(z)))|$$

By the Lipschitz property of A and since $|\alpha| \leq \|A\|_{\text{Lip}}$ we have

$$(2.24) \leq \frac{1}{2}|x - b(z)|,$$

and it follows that both $x + A(x) - z$ and $(1 + \alpha(b(z), t))(x - b(z))$ have (signed) distance of the order $\approx x - b(z)$ from zero. Therefore

$$\begin{aligned} & |\Psi_t(x + A(x) - z) - \Psi_t((1 + \alpha(b(z), t))(x - b(z)))| \\ & \lesssim t^{-2}(1 + |x - b(z)|/t)^{-20} \cdot (2.24) && \text{by decay of } \Psi_t' \\ & \lesssim t^{-1}\beta(b(z), t)(1 + |x - b(z)|/t)^{-10} && \text{by definition of } \beta \text{ numbers.} \end{aligned}$$

It follows that

$$\begin{aligned} & \sum_{t \in 2^{\mathbb{Z}}} |(1 - P_t)T_A(\Psi_t * f)| \\ & = \sum_{t \in 2^{\mathbb{Z}}} \left| (1 - P_t) \int P_t f(z) (\Psi_t(x + A(x) - z) - \Psi_t((1 + \alpha(b(z), t))(x - b(z)))) dz \right| \\ & \lesssim \sum_{t \in 2^{\mathbb{Z}}} (\delta_0 + t^{-1}(1 + |\cdot|/t)^{-10}) * \int |P_t f(z)| t^{-1}\beta(b(z), t)(1 + |x - b(z)|/t)^{-10} dz \\ & \lesssim \sum_{t \in 2^{\mathbb{Z}}} \int |P_t f(z)| t^{-1}\beta(b(z), t)(1 + |x - b(z)|/t)^{-5}. \end{aligned}$$

Multiplying this with a function $g \in L^{p'}(\mathbb{R})$ and integrating in x we obtain the estimate

$$\sum_{t \in 2^{\mathbb{Z}}} \int \mathcal{D}f(z, t) \beta(b(z), t) \mathcal{A}g(b(z), t) dz.$$

The sum over t can be dominated by $\int_0^\infty \frac{dt}{t}$ since all functions $\mathcal{D}, \beta, \mathcal{A}$ are almost (up to a multiplicative factor) constant on Carleson boxes $B(x, t) \times [t, 2t]$. By [DT15, Proposition 3.6] and outer Hölder inequality [DT15, Proposition 3.4] this is bounded by

$$\|\mathcal{D}f\|_{L^p(S^2)} \|\beta(b(\cdot), \cdot)\|_{L^\infty(S^2)} \|\mathcal{A}g(b(\cdot), \cdot)\|_{L^{p'}(S^\infty)}.$$

Since the function b is bi-Lipschitz, it does not affect outer norms up to a multiplicative constant. To see this note that

$$\|F \mathbf{1}_{(\cup_i T(x_i, s_i))^c}\|_{L^\infty(S^q)} \leq \lambda \implies \|F(b(\cdot), \cdot) \mathbf{1}_{(\cup_i T(b^{-1}(x_i), 2s_i))^c}\|_{L^\infty(S^q)} \leq C\lambda$$

for a sufficiently large constant C .

Thus we obtain the estimate

$$\|\mathcal{D}f\|_{L^p(S^2)} \|\beta\|_{L^\infty(S^2)} \|\mathcal{A}g\|_{L^{p'}(S^\infty)}.$$

Estimating the first term using (2.16), the middle term using Corollary 2.23, and the last term using (2.15) we obtain the claim. \square

2.5. Application to truncated directional Hilbert transforms.

Proof of Corollary 1.3. By Minkowski's integral inequality we obtain

$$\begin{aligned} & \left\| \sum_{t \in 2^{\mathbb{Z}}} |(1 - P_t) T_u(\Psi_t * f)| \right\|_{L^p(\mathbb{R}^2)} \\ & \leq \int_{r=-1}^1 \left(\int_{\mathbb{R}} \left\| \sum_{t \in 2^{\mathbb{Z}}} |(1 - P_t)(\Psi_t * f(x + r, \cdot + ru(\cdot)))| \right\|_{L^p(\mathbb{R})}^p dx \right)^{1/p} \frac{dr}{|r|} \end{aligned}$$

By Theorem 1.1, we further obtain

$$\begin{aligned} & \lesssim \int_{r=-1}^1 \left(\int_{\mathbb{R}} \|ru(x, \cdot)\|_{\text{Lip}}^p \|f(x + r, \cdot)\|_{L^p(\mathbb{R})}^p dx \right)^{1/p} \frac{dr}{|r|} \\ & \lesssim \int_{r=-1}^1 \left(\int_{\mathbb{R}} \|f(x + r, \cdot)\|_{L^p(\mathbb{R})}^p dx \right)^{1/p} dr \lesssim \int_{r=-1}^1 \|f\|_{L^p(\mathbb{R}^2)} dr \lesssim \|f\|_{L^p(\mathbb{R}^2)}. \end{aligned}$$

This finishes the proof of Corollary 1.3. \square

In the remaining part of this section we prove Corollary 1.4. As an initial reduction observe that it suffices to estimate the restriction of H_u to a vertical strip; more precisely we need an estimate of the form

$$\|H_u f\|_{L^p([N-1, N+2] \times \mathbb{R})} \lesssim \|H_u f\|_{L^p([N, N+1] \times \mathbb{R})}$$

for functions f supported in the vertical strip $[N, N+1] \times \mathbb{R}$. This reduction will be important in the case $p_0 < 2$. Also, it is easy to see that we may replace H_u by the smoothly truncated operator

$$(2.25) \quad \tilde{H}_u f(x, y) := \text{p.v.} \int_{\mathbb{R}} f(x + r, y + u(x, y)r) \phi(r) \frac{dr}{r},$$

where ϕ is a smooth even function with $\phi(0) = 1$, $\int \phi(x) dx = \int x \phi(x) dx = \dots = \int x^N \phi(x) dx = 0$ for some large N and $\text{supp } \phi \subset [-1, 1]$. This is possible because the maps $(x, y) \mapsto (x + r, y + u(x, y)r)$ are uniformly bi-Lipschitz for $r \in [-1, 1]$, so $f \mapsto f(x + r, y + u(x, y)r)$ is a bounded operator on L^p .

We note that the operators $f \mapsto H_u(\Psi_t * f)$ (as well as the analogous ones obtained with \tilde{H}_u from (2.25) in place of H_u) are also trivially bounded in L^p uniformly in $t \geq t_0$. To see this split

$$H_u f(x, y) = \int_{-1}^1 f(x+r, y) \frac{dr}{r} + \int_{-1}^1 (f(x+r, y+u(x, y)r) - f(x+r, y)) \frac{dr}{r}.$$

The first term is a one-dimensional truncated Hilbert transform on each horizontal line, and therefore bounded on any L^p , $1 < p < \infty$. The second term can be written as

$$\int_{-1}^1 \int_{s=0}^{u(x, y)r} \partial_2 f(x+r, y+s) ds \frac{dr}{r}$$

This is in turn bounded by

$$\int_{-1}^1 M_2 \partial_2 f(x+r, y) dr \leq M_1 M_2 \partial_2 f(x, y),$$

where M_i denotes the Hardy–Littlewood maximal function in the i -th variable. The differential operator ∂_2 is L^p bounded on the subspace of functions with $\hat{f}(\xi, \eta) = 0$ for $|\eta| > 2/t_0$ and therefore we obtain L^p estimates for this term.

Remark 2.26. The same argument can be used to estimate H_u on functions with small horizontal frequencies, thus simplifying an argument in [GT16, Section 3].

Below, we work with \tilde{H}_u from (2.25) in place of H_u , and omit the tilde for simplicity of notation. By linearity and the Calderón reproducing formula it suffices to estimate the operator

$$f \mapsto \int_0^\infty H_u(\Psi_t * f) \frac{dt}{t}$$

in L^p . By superposition of Corollary 1.3 we obtain L^p estimates for the off-diagonal term

$$f \mapsto \int_0^\infty (1 - P_t) H_u(\Psi_t * f) \frac{dt}{t},$$

so it suffices to estimate the diagonal term

$$f \mapsto \int_0^\infty P_t H_u(\Psi_t * f) \frac{dt}{t}.$$

By discretization and Littlewood–Paley theory it suffices to show

$$\left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |H_u(\Psi_t * f)|^2 \right)^{1/2} \right\|_p \lesssim \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |P_t * f|^2 \right)^{1/2} \right\|_p,$$

or, more generally,

$$\left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |H_u(\Psi_t * f_t)|^2 \right)^{1/2} \right\|_{L^p([N-1, N+2] \times \mathbb{R})} \lesssim \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |f_t|^2 \right)^{1/2} \right\|_p$$

for arbitrary functions f_t supported in the strip $[N, N+1] \times \mathbb{R}$. In the case $p = p_0 = 2$ this follows immediately from the single band hypothesis (1.5) and Fubini’s theorem.

In order to obtain the larger range of p ’s in the case $1 < p_0 < 2$ we use the technique for proving vector-valued estimates introduced in [BT13] (see also [DS15] for more applications of this technique).

Theorem 2.27. *Let $1 < p, q < \infty$ and let $T_k : L^{p,1}(\Omega) \rightarrow L^{p,\infty}(\Omega')$ be a sequence of subadditive operators. Let $0 \leq c < 1$ and suppose that for every pair of (non-null, finite measure) measurable sets $H \subset \Omega$, $G \subset \Omega'$ with $0 < |H|, |G| < \infty$ there exist subsets $H' \subset H$, $G' \subset G$ with*

$$\left(\frac{|G \setminus G'|}{|G|} \right)^{1-1/p} + \left(\frac{|H \setminus H'|}{|H|} \right)^{1/p} \leq c$$

for every k and every function f supported on H' we have

$$(2.28) \quad \|T_k f\|_{L^q(G')} \lesssim (|G|/|H|)^{1/q-1/p} \|f\|_{L^q(H')}.$$

Then for any functions $f_k \in L^{p,1}(\Omega)$ we have

$$\|(\sum_k |T_k f_k|^q)^{1/q}\|_{L^{p,\infty}(\Omega')} \lesssim \|(\sum_k |f_k|^q)^{1/q}\|_{L^{p,1}(\Omega)}.$$

Proof. By the monotone convergence theorem it suffices to consider a finite sequence of operators as long as we obtain estimates that do not depend on its length. The hypothesis (2.28) continues to hold for the operator $T(\vec{f}) := (\sum_k |T_k f_k|^q)^{1/q}$ defined on ℓ^q -valued functions, and we know

$$\|Tf\|_{L^{p_0,\infty}(\Omega')} \lesssim \|f\|_{L^{p_0,1}(\Omega,\ell^q)}$$

with some constant given by the qualitative boundedness assumption on T_k 's and depending on the length of the sequence of operators. By duality of Lorentz spaces this is equivalent to

$$\int_G |Tf| \leq B|H|^{1/p} |G|^{1-1/p}$$

for all finite measure sets H, G and all functions $f : \Omega \rightarrow \ell^q$ with $|f| \leq \mathbf{1}_H$. We have to find a universal upper bound for B .

Let G, H be measurable sets with finite measure and G', H' be the major subsets given by the hypothesis. Then for any function $f : \Omega \rightarrow \ell^q$ with $|f| \leq \mathbf{1}_{H'}$ we have

$$\begin{aligned} \int_{G'} |Tf| &\leq \|Tf\|_{L^q(G')} \|\mathbf{1}_G\|_{L^{q'}} \\ &\lesssim (|G|/|H|)^{1/q-1/p} \|f\|_{L^q(H',\ell^q)} |G|^{1/q'} \\ &\lesssim |H|^{1/p} |G|^{1-1/p} \end{aligned}$$

by Hölder's inequality and the hypothesis. It follows that for any function $f : \Omega \rightarrow \ell^q$ with $|f| \leq \mathbf{1}_H$ we have

$$\begin{aligned} \int_G |Tf| &\leq C|H|^{1/p} |G|^{1-1/p} + \int_{G \setminus G'} |Tf| + \int_{G'} |T(f \mathbf{1}_{H \setminus H'})| \\ &\leq C|H|^{1/p} |G|^{1-1/p} + B|H|^{1/p} |G \setminus G'|^{1-1/p} + B|H \setminus H'|^{1/p} |G|^{1-1/p} \\ &\leq (C + cB)|H|^{1/p} |G|^{1-1/p}. \end{aligned}$$

Taking a supremum over H, G we obtain $B \leq C/(1-c)$. \square

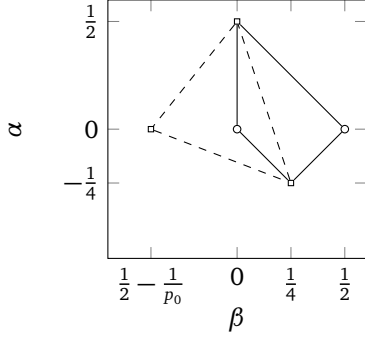
Corollary 1.4 will be obtained via an application of Theorem 2.27 to the operators $T_k f = H_u(\Psi_{2^k} *_2 f)$, with the choice $q = 2$. The corresponding assumption (2.28) in Theorem 2.27 will follow by interpolation of the estimates

$$(2.29) \quad \int (T_k(\mathbf{1}_{H'} \mathbf{1}_F)) \mathbf{1}_{G'} \mathbf{1}_E \lesssim |E|^{1/2} |F|^{1/2} (|G|/|H|)^\alpha (|E|/|F|)^\beta,$$

where $H \subset [N, N+1] \times \mathbb{R}$, $G \subset [N-1, N+1] \times \mathbb{R}$, $H' \subset H$ and $G' \subset G$ are as in Theorem 2.27, $F, E \subset \mathbb{R}^2$ are arbitrary measurable subsets, $\alpha = 1/2 - 1/p$, and β is in a neighborhood of 0.

The set of pairs (α, β) for which the estimate (2.29) holds is clearly convex. Hence it suffices to establish (2.29) near the vertices of the dashed triangle in Figure 1. The intersection of the line $\beta = 0$ with this triangle corresponds to the range of p 's claimed in (1.7).

We will use Estimates 16, 17, 21, and 22 from [BT13], which do not rely on the single parameter assumption on the vector field made in [Bat13; BT13]. One twist is in the proof of Estimate 21, where we have to use a version of [BT13, Theorem 8] for



The estimate (2.29) is known unconditionally in the interior of the solid polygon: the line $\alpha = 0$ corresponds to the non-localized estimates in [Bat13] and the other two endpoints are the localized estimates in [BT13].

In the proof of Corollary 1.4 we use estimates in the interior of the dashed triangle, whose leftmost vertex is the hypothesis (1.5).

FIGURE 1. Localized estimates for the single band directional Hilbert transform

Lipschitz vector fields. This result goes back to [LL06]; a slightly simplified version of the proof of the required covering lemma in [BT13] is presented in Appendix A. The covering lemma for Lipschitz vector fields only holds for parallelograms of bounded length. This is the reason for restricting the operator H_u to a vertical strip: we can apply the covering lemma to the intersection of parallelograms with this vertical strip. The other difficulty is that we are dealing with a (smooth) truncation of the Hilbert kernel, so the results of [Bat13] do not directly apply. The easiest way to work around this seems to be running the argument in [Bat13] with more general wave packets which can be used to assemble also the truncated Hilbert kernel $\phi(r)/r$.

2.5.1. *Using the single band estimate below L^2 .* The hypothesis (1.5) shows in particular that (2.29) holds with $(\alpha, \beta) = (0, 1/2 - 1/p_0)$.

2.5.2. *Using the Córdoba–Fefferman covering argument.* By Estimates 16, 17, and 22 in [BT13] we can estimate the left-hand side of (2.29) by

$$\sum_{\delta} \sum_{\sigma \lesssim \delta^{-n} (|G|/|H|)^{n-1}} \min(|F| \delta \sigma^{-1}, |E| \sigma)$$

for any integer $n \geq 2$, where both sums are over positive dyadic numbers.

The (geometric) sum over σ has two critical points: $\sigma \sim \delta^{-n} (|G|/|H|)^{n-1}$ and $\sigma \sim (\delta |F|/|E|)^{1/2}$. This gives the estimate

$$\sum_{\delta} \min((\delta |F|/|E|)^{1/2}, |E| \delta^{-n} (|G|/|H|)^{n-1}).$$

The sum over δ has a critical point with $\delta_0^{2n+1} \sim (|G|/|H|)^{2n-2} (|E|/|F|)$, and we obtain the estimate

$$(\delta_0 |F|/|E|)^{1/2} \sim (|F|/|E|)^{1/2} (|G|/|H|)^{(n-1)/(2n+1)} (|E|/|F|)^{1/(4n+2)}.$$

This proves the claim with $\alpha = (n-1)/(2n+1)$, $\beta = 1/(4n+2)$. We can make (α, β) approach $(1/2, 0)$ by choosing n suitably large.

2.5.3. *Using the Lacey–Li covering argument.* By Estimates 16, 17, and 21 from [BT13] we can estimate the left-hand side of (2.29) by

$$\sum_{\delta} \sum_{\sigma} \min(|F| \delta \sigma^{-1}, |E| \sigma, |E| (|H|/|G|)^{1/2} \sigma^{-\epsilon} \delta^{-1/2-\epsilon})$$

The sum over σ now has two critical points with $\sigma \sim (\delta |F|/|E|)^{1/2}$ and with $\sigma^{1+\epsilon} \sim (|H|/|G|)^{1/2} \delta^{-1/2-\epsilon}$ and is dominated by the minimum of the two corresponding terms,

so we have the estimate

$$\sum_{\delta \leq 1} \min(|E|((|H|/|G|)^{1/2} \delta^{-1/2-\epsilon})^{1/(1+\epsilon)}, (\delta|F||E|)^{1/2})$$

The sum over δ has a critical point at $\delta_0^{2+3\epsilon} \sim (|E|/|F|)^{1+\epsilon}(|H|/|G|)$. This gives the estimate

$$(\delta_0|F||E|)^{1/2} \sim (|F||E|)^{1/2}((|E|/|F|)^{1+\epsilon}(|H|/|G|))^{1/(4+6\epsilon)}.$$

Making ϵ small we can make (α, β) approach $(-1/4, 1/4)$. This completes the proof of Corollary 1.4.

Remark 2.30. The upper part of the solid polygon in Figure 1 yields the hypothesis of Theorem 2.27 for any $2 < q < p < \infty$. This implies that the operator H_u maps $L^p(\mathbb{R}^2)$ into a directional Triebel–Lizorkin space of type $F_{p,q}^0$ (provided that u is Lipschitz in the vertical direction). More precisely,

$$\left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |P_t H_u f|^q \right)^{1/q} \right\|_{L^p(\mathbb{R}^2)} \lesssim \|f\|_{L^p(\mathbb{R}^2)}, \quad 2 < p, q < \infty.$$

Indeed, the left-hand side is monotonically decreasing in q , so it suffices to consider $2 < q < p < \infty$. With a suitable choice of Ψ we may write $f = \sum_{t \in 2^{\mathbb{Z}/100}} \Psi_t *_2 f$. For notational simplicity we consider only the contribution of $t \in 2^{\mathbb{Z}}$. By the Fefferman–Stein maximal inequality we may replace P_t by larger Littlewood–Paley projections such that $\sum_{t \in 2^{\mathbb{Z}}} P_t = \text{id}$.

In the diagonal term we use the Fefferman–Stein maximal inequality, the vector-valued estimate provided by Theorem 2.27 with $p > 2$, monotonicity of ℓ^q norms, and Littlewood–Paley theory to estimate

$$\begin{aligned} \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |P_t H_u(\Psi_t *_2 f)|^q \right)^{1/q} \right\|_{L^p(\mathbb{R}^2)} &\lesssim \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |H_u(\Psi_t *_2 f)|^q \right)^{1/q} \right\|_{L^p(\mathbb{R}^2)} \\ &\lesssim \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |\Psi_t *_2 f|^q \right)^{1/q} \right\|_{L^p(\mathbb{R}^2)} \\ &\leq \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |\Psi_t *_2 f|^2 \right)^{1/2} \right\|_{L^p(\mathbb{R}^2)} \\ &\lesssim \|f\|_{L^p(\mathbb{R}^2)}. \end{aligned}$$

In the off-diagonal term we use monotonicity of ℓ^q norms, Littlewood–Paley theory, and Corollary 1.3 to estimate

$$\begin{aligned} \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |P_t H_u \left(\sum_{t' \neq t} \Psi_{t'} *_2 f \right)|^q \right)^{1/q} \right\|_{L^p(\mathbb{R}^2)} &\leq \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |P_t H_u \left(\sum_{t' \neq t} \Psi_{t'} *_2 f \right)|^2 \right)^{1/2} \right\|_{L^p(\mathbb{R}^2)} \\ &\lesssim \left\| \sum_{t' \in 2^{\mathbb{Z}}} \left(\sum_{t \neq t'} P_t \right) H_u(\Psi_{t'} *_2 f) \right\|_{L^p(\mathbb{R}^2)} \\ &= \left\| \sum_{t' \in 2^{\mathbb{Z}}} (1 - P_{t'}) H_u(\Psi_{t'} *_2 f) \right\|_{L^p(\mathbb{R}^2)} \\ &\lesssim \|f\|_{L^p(\mathbb{R}^2)}. \end{aligned}$$

2.6. Application to Hilbert transforms along Lipschitz variable parabolas.

Proof of Corollary 1.9. In the following, we will assume for notational convenience that $0 < u \leq 1$ almost everywhere. The region that $-1 \leq u < 0$ can be handled similarly, while the region $u = 0$ is trivial by Fubini as the operator acts only in the first variable. By the trivial analogue of Corollary 1.3, it suffices to show

$$(2.31) \quad \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |H_u^{(\alpha)}(\Psi_t *_2 f)|^2 \right)^{1/2} \right\|_p \lesssim \|f\|_p.$$

We use $P_t(\Psi_t *_2 f) = \Psi_t *_2 f$ where P_t is as defined before acting in the second variable. We note that for

$$|r|^\alpha u(x, y)/t \leq 1$$

we have by an application of the fundamental theorem of calculus

$$|P_t(\Psi_t *_2 f)(x+r, y+u(x, y)r^\alpha) - P_t(\Psi_t *_2 f)(x+r, y)| \leq u(x, y)|r|^\alpha t^{-1} M_2(\Psi_t *_2 f)(x+r, y).$$

Hence we have for the integral over small values of r

$$(2.32) \quad \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} \int_{|r|^\alpha u(x, y)/t \leq 1} P_t(\Psi_t *_2 f)(x+r, y+u(x, y)r^\alpha) \frac{dr}{r} \right)^2 \right\|_{L^p(x, y)} \\ \lesssim \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} \int_{|r|^\alpha u(x, y)/t \leq 1} P_t(\Psi_t *_2 f)(x+r, y) \frac{dr}{r} \right)^2 \right\|_{L^p(x, y)}$$

$$(2.33) \quad + \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} \int_{|r|^\alpha u(x, y)/t \leq 1} u(x, y)|r|^\alpha t^{-1} M_2(\Psi_t *_2 f)(x+r, y) \frac{dr}{|r|} \right)^2 \right\|_{L^p(x, y)}.$$

The former term (2.32) can be estimated using the vector-valued estimate for the maximally truncated Hilbert transform. Using integrability of $|r|^{\alpha-1}$ near zero we estimate the latter term (2.33) by

$$\left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |M_1 M_2(\Psi_t *_2 f)(x, y)|^2 \right)^{1/2} \right\|_{L^p(x, y)} \lesssim \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} |\Psi_t *_2 f(x, y)|^2 \right)^{1/2} \right\|_{L^p(x, y)} \lesssim \|f\|_p.$$

Here we have used the Fefferman–Stein maximal inequality and Littlewood-Paley theory.

We turn to the remaining part of the kernel with $|r|^\alpha u(x, y)/t \geq 1$ and $|r| \leq 1$. Note we may restrict the summation over t to $t \leq 1$, as for $t > 1$ the domain of integration is empty. We will break up the integral into lacunary pieces parametrized by $s \in 2^{\alpha\mathbb{N}}$ and estimate the pieces separately, with suitable power decay in s allowing to geometrically sum the estimates.

We introduce Littlewood-Paley projections in the first variable and write $P_t^{(1)}$ and $P_t^{(2)}$ to distinguish projections in first and second variable. Consider the averaging operator

$$E_s^{(1)} = \int_s^\infty P_t^{(1)} \frac{dt}{t}.$$

We note similarly to above for the averaged part of the integral pieces:

$$(2.34) \quad \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} \int_{s \leq |r|^\alpha u(x, y)/t \leq 2^{\alpha s}} E_s^{(1)} P_t^{(2)}(\Psi_t *_2 f)(x+r, y+u(x, y)r^\alpha) \frac{dr}{r} \right)^2 \right\|_{L^p(x, y)} \\ \lesssim \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} \int_{s \leq |r|^\alpha u(x, y)/t \leq 2^{\alpha s}} E_s^{(1)} P_t^{(2)}(\Psi_t *_2 f)(x, y+u(x, y)r^\alpha) \frac{dr}{r} \right)^2 \right\|_{L^p(x, y)} \\ (2.35) \quad + \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} \left(\int_{s \leq |r|^\alpha u(x, y)/t \leq 2^{\alpha s}} s^{-1} M_1 P_t^{(2)}(\Psi_t *_2 f)(x, y+u(x, y)r^\alpha) \frac{dr}{|r|} \right)^2 \right)^{1/2} \right\|_{L^p(x, y)}$$

The factor $(st/u)^{1/\alpha}$ in the index of the averaging operator is chosen because it is roughly $|r|$ in the domain of integration. In the former term (2.34) we change variables, replacing $u(x, y)r^\alpha$ by r on the positive and similarly on the negative axis and do a partial integration in r , noting that by the mean zero property the primitive of the kernel of $P_t^{(2)}$ is a bump function again, to estimate this term by

$$\lesssim \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} \left(\int_{s \leq |r|/t \leq 2^{\alpha s}} t M_1 M_2(\Psi_t *_2 f)(x, y+r) \frac{dr}{|r|^2} \right)^2 \right)^{1/2} \right\|_{L^p(x, y)}$$

$$\lesssim s^{-1} \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} (M_2 M_1 M_2 (\Psi_t *_2 f)(x, y))^2 \right)^{1/2} \right\|_{L^p(x, y)}$$

plus two similar boundary terms, which are all estimated by the Fefferman-Stein maximal inequality with power decay in s . The latter term (2.35) above is estimated by the same change of variables by

$$\begin{aligned} & s^{-1} \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} \left(\int_{s \leq |r|/t \leq 2^\alpha s} M_1 P_t^{(2)} (\Psi_t *_2 f)(x, y+r) \frac{dr}{|r|} \right)^2 \right)^{1/2} \right\|_{L^p(x, y)} \\ & \lesssim s^{-1} \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} (M_2 M_1 P_t^{(2)} (\Psi_t *_2 f)(x, y))^2 \right)^{1/2} \right\|_{L^p(x, y)} \end{aligned}$$

which is again estimated by the Fefferman-Stein maximal inequality with decay in s .

A similar estimate can be obtained if instead of the sharp cut-off $s \leq |r|^\alpha u(x, y)/t \leq 2^\alpha s$ we use a smooth cut-off. More precisely, we will choose cut-off functions as defined in the following operator:

$$(2.36) \quad A_s f(x, y) = \int_{\mathbb{R}} f(x+r, y+u(x, y)r^\alpha) \chi(s^{-1}r^\alpha v(x, y)t^{-1}(u(x, y)v^{-1}(x, y))^{\alpha/(\alpha-1)}) \frac{dr}{r},$$

where χ is smooth and supported on $\pm[2^{-\alpha}, 2^\alpha]$ and $\sum_{s \in 2^{\alpha\mathbb{N}}} \chi(s^{-1}x) = 1$ for $x \neq 0$, and where $v(x, y)$ is the largest integer power of 2 less than $u(x, y)$. Note the auxiliary factor u/v is bounded above and below respectively by 2 and 1.

Then, with the above arguments, it suffices to estimate the rough part of each piece with some $\gamma > 0$ that may depend on p as follows:

$$(2.37) \quad \left\| \left(\sum_{t \in 2^{-\mathbb{N}}} |A_s(1 - E_{s(\frac{st}{u(x, y)})^{1/\alpha}})^{(1)} P_t^{(2)} (\Psi_t *_2 f)|^2 \right)^{1/2} \right\|_{L^p} \lesssim s^{-\gamma} \|f\|_p.$$

Here we point out that this estimate has essentially been established in [Guo+16]. First of all, we recognize that the left hand side of (2.37) is essentially the term (5.13) in [Guo+16], there one has a large power of s in the index of E but this makes their bound only stronger. By the local smoothing estimates and a certain interpolation argument, the L^p bounds of (2.37) for all $1 < p \leq 2$ have been established in Subsection 5.3 in [Guo+16]. To prove L^p bounds for all $p > 2$, we cite the pointwise estimate (3.19) in [Guo+16], which implies for these p that

$$\left\| \left(\sum_{t \in 2^{-\mathbb{N}}} |A_s(1 - E_{s(\frac{st}{u(x, y)})^{1/\alpha}})^{(1)} P_t^{(2)} (\Psi_t *_2 f)|^2 \right)^{1/2} \right\|_{L^p} \lesssim \log(1+s)^4 \|f\|_p.$$

A further interpolation gives the desired estimate (2.37) for all $1 < p < \infty$ for slightly smaller γ . This finishes the proof of the square function estimate (2.31). \square

3. SINGLE SCALE OPERATOR

In this section we prove Theorem 1.12. The strategy is to use duality and outer Hölder inequality to reduce the estimate to two estimates of Carleson embedding flavor; the “energy embedding” in Section 3.2 and the “mass embedding” in Section 3.3.

3.1. Tiles and the outer measure space. We subdivide the parameter space into *tiles*. Each tile can be represented in three equivalent ways:

(1) by a *shearing matrix*

$$A = \begin{pmatrix} 2^{k_1} & 0 \\ l2^{k_1} & 2^{k_2} \end{pmatrix}, \quad k_1, k_2, l \in \mathbb{Z}$$

and the spatial location $(2^{-k_1}n_1, 2^{-k_2}n_2)$, $n_1, n_2 \in \mathbb{Z}$.

(2) by the corresponding *spatial parallelogram*

$$P = A^{-1}([0, 1] \times [0, 1]) + (2^{-k_1}n_1, 2^{-k_2}n_2),$$

(3) or by the corresponding *frequency parallelogram* $A^*([0, 1] \times [1, 2])$ and the spatial location

$$(2^{-k_1}n_1, 2^{-k_2}n_2).$$

Figure 2 shows the spatial and the frequency parallelograms of a tile (with $n_1 = n_2 = 0$). The frequency picture also includes the symmetric parallelograms $A^*([0, 1] \times [-2, -1])$ (in a lighter shade of gray), because the Fourier transforms of the wave packets associated to tiles will concentrate on both these parallelograms. However, for combinatorial purposes it suffices to consider only the upper parallelogram. The *slope* of a tile is the number $-l2^{-k_2+k_1}$. It is the slope of the lower and the upper side of the corresponding spatial parallelogram. The spatial parallelogram seems to be the most concise description of a tile, so we denote tiles by the letter P (for “parallelogram”).

The fact that we are dealing with a single scale operator in Section 3 is reflected in that we define an outer measure on a finite set X of tiles with $k_1 = 0$, that is, tiles with the fixed horizontal scale 1. (The restriction to finite sets of tiles avoids technicalities associated with infinite sums. All estimates will be independent of the specific finite set, so we can pass to the set of all tiles at the end of the argument.) The outer measure is generated by a function σ whose domain \mathbf{E} is the collection of all non-empty subsets of X . We denote by CP the parallelogram with the same slope and center as P but side lengths multiplied by C . For $\mathcal{R} \in \mathbf{E}$ set

$$(3.1) \quad \sigma(\mathcal{R}) := \sup_{L \geq 1} L^{-C} \left| \bigcup_{R \in \mathcal{R}} LR \right|,$$

where C is a large number to be chosen later. The three sizes that we need are

$$\begin{aligned} S^1(F)(\mathcal{R}) &:= \sigma(\mathcal{R})^{-1} \sum_{R \in \mathcal{R}} |R| |F(R)|, \\ S^2(F)(\mathcal{R}) &:= \left(\sigma(\mathcal{R})^{-1} \sum_{R \in \mathcal{R}} |R| |F(R)|^2 \right)^{1/2} = S^1(F^2)(\mathcal{R})^{1/2}, \\ S^\infty(G)(\mathcal{R}) &:= \sup_{R \in \mathcal{R}} |G(R)|. \end{aligned}$$

3.2. Wave packets and the energy embedding. Let $\Phi = \Phi_C$ be the set of functions on \mathbb{R}^2 that satisfy

$$|\partial^\alpha \phi(x)| \leq (1 + |x|)^{-C}, \quad \|\alpha\|_{\ell^1} \leq C,$$

for some sufficiently large C that will be chosen later and

$$\int_{\mathbb{R}} x_2^n \phi(x_1, x_2) dx_2 = 0, \quad x_1 \in \mathbb{R}, \quad n = 0, \dots, C-2.$$

We think of ϕ as morally supported on $[0, 1]^2$ and of $\hat{\phi}$ as morally supported on $[0, 1] \times [1, 2]$ for $\phi \in \Phi$.

The L^∞ normalized wave packets associated to a tile $P = (A, n_1, n_2)$ are the functions of the form

$$\phi_p^{(\infty)}(x) = \phi(A(x_1 - 2^{-k_1}n_1, x_2 - 2^{-k_2}n_2)), \quad \phi \in \Phi.$$

The L^p normalized wave packets, $1 \leq p < \infty$, are the functions $\phi_p^{(p)} = \det(A)^{1/p} \phi_p^{(\infty)}$. Note that $\widehat{\phi(A \cdot)}(\xi) = (\det A)^{-1} \hat{\phi}(A^{-*}\xi)$. The spatial and the frequency parallelograms of a tile correspond to the moral space/frequency support of the wave packets associated to this tile.

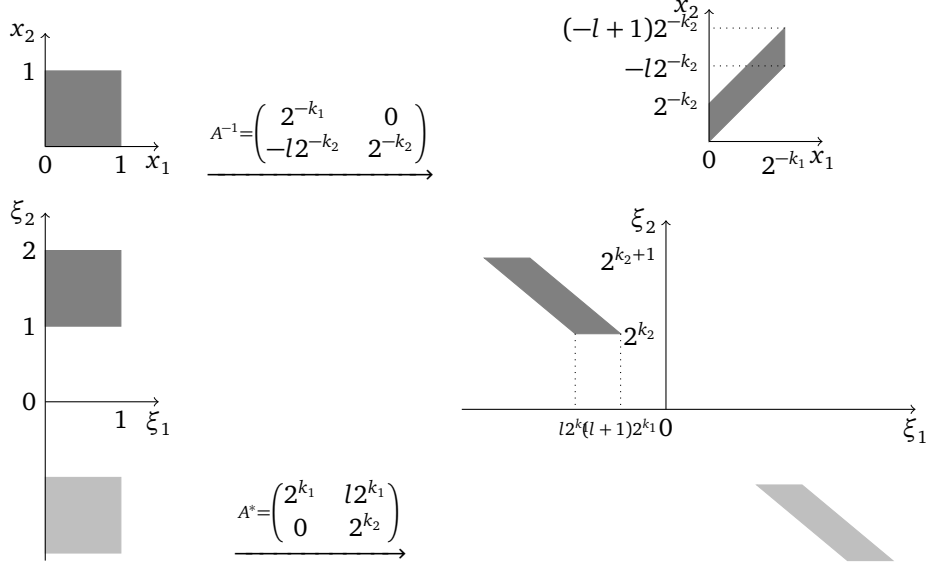


FIGURE 2. Spatial and frequency parallelograms of a tile

3.2.1. *Almost orthogonality.* The fundamental property of the wave packets is their almost orthogonality for tiles with different scales or slopes.

Lemma 3.2.

$$|\langle \phi_p^{(2)}, \phi_{p'}^{(2)} \rangle| \lesssim \min(1, (2^{\max(k_2, k'_2)} |2^{-k_2} l - 2^{-k'_2} l'|)^{-C}, 2^{-C|k_2 - k'_2|}),$$

where C can be made arbitrarily large provided that the order of decay in the definition of Φ is sufficiently large.

Proof. Without loss of generality suppose $k_2 \geq k'_2$. We will estimate

$$\int_{\mathbb{R}^2} |\widehat{\phi(A \cdot)}| |\widehat{\phi'(A' \cdot)}|$$

for $\phi, \phi' \in \Phi$. This is sufficient because the spatial location of the tiles only affects the phase of the Fourier transforms of the associated wave packets, but not their magnitude.

Correlation decay due to shearing. Let $0 < \epsilon \ll 1$ and $S_N = \{-N, N\} \times \mathbb{R}$ be a vertical strip of width $N \geq 1$. The critical intersection $A^* S_N \cap (A')^* S_N$ is a parallelogram centered at zero of width $\sim N$ and height $\sim N/|2^{-k_2} l - 2^{-k'_2} l'|$. By the vanishing moments assumption we have

$$|\widehat{\phi(A \cdot)}| \lesssim 2^{-k_2} (2^{-k_2} N / |2^{-k_2} l - 2^{-k'_2} l'|)^C$$

on the critical intersection. Using the fact that the Fourier transforms $\widehat{\phi(A \cdot)}$ and $\widehat{\phi'(A' \cdot)}$ are L^1 normalized functions and the decay of these Fourier transforms at infinity we obtain

$$\begin{aligned} \int_{\mathbb{R}^2} |\widehat{\phi(A \cdot)}| |\widehat{\phi'(A' \cdot)}| &\leq \int_{\mathbb{R}^2 \setminus A^* S_N} + \int_{\mathbb{R}^2 \setminus (A')^* S_N} + \int_{A^* S_N \cap (A')^* S_N} \\ &\leq \sup_{\mathbb{R}^2 \setminus A^* S_N} |\widehat{\phi(A \cdot)}| + \sup_{\mathbb{R}^2 \setminus (A')^* S_N} |\widehat{\phi'(A' \cdot)}| + \sup_{A^* S_N \cap (A')^* S_N} |\widehat{\phi(A \cdot)}| \\ &\lesssim 2^{-k_2} N^{-C(1/\epsilon-1)} + 2^{-k'_2} N^{-C(1/\epsilon-1)} + 2^{-k_2} (2^{-k_2} N / |2^{-k_2} l - 2^{-k'_2} l'|)^C. \end{aligned}$$

Choosing $N = 2^{\epsilon(k_2 - k'_2)/C} (2^{k_2} |2^{-k_2} l - 2^{-k'_2} l'|)^\epsilon$ as we may provided that $|l - 2^{k_2 - k'_2} l'| \geq 1$, we obtain

$$\int_{\mathbb{R}^2} |\widehat{\phi(A \cdot)}| |\widehat{\phi'(A' \cdot)}| \lesssim 2^{-k_2 + \epsilon(k_2 - k'_2)} (2^{k_2} |2^{-k_2} l - 2^{-k'_2} l'|)^{-(1-\epsilon)C},$$

and this gives the second estimate in the conclusion of the lemma.

Correlation decay for separated scales. Let $2^{k'_2} \ll N \ll 2^{k_2}$. Using again the fact that the Fourier transforms $\widehat{\phi(A \cdot)}$ and $\widehat{\phi'(A' \cdot)}$ are L^1 normalized functions and the decay of Fourier transforms near $\xi_2 = 0$ and at infinity we obtain

$$\begin{aligned} \int_{\mathbb{R}^2} |\widehat{\phi(A \cdot)}| |\widehat{\phi'(A' \cdot)}| &\leq \int_{|\xi_2| \leq N} + \int_{|\xi_2| \geq N} \\ &\leq \sup_{|\xi_2| \leq N} |\widehat{\phi(A \cdot)}| + \sup_{|\xi_2| \geq N} |\widehat{\phi'(A' \cdot)}| \\ &\lesssim 2^{-k_2} (N/2^{k_2})^C + 2^{-k'_2} (N/2^{k'_2})^{-(C+1)/\epsilon}. \end{aligned}$$

Choosing $N \sim 2^{k'_2 + \epsilon(k_2 - k'_2)}$ we obtain

$$\int_{\mathbb{R}^2} |\widehat{\phi(A \cdot)}| |\widehat{\phi'(A' \cdot)}| \lesssim 2^{-k_2 - C(1-\epsilon)(k_2 - k'_2)} = 2^{-k_2/2 - k'_2/2 - (C+1/2-\epsilon')(k_2 - k'_2)},$$

and this gives the third estimate in the conclusion of the lemma. \square

3.2.2. Bessel inequality.

Lemma 3.3. *For each tile P fix an L^2 normalized wave packet ϕ_P adapted to P . Then*

$$\sum_P |\langle f, \phi_P \rangle|^2 \lesssim \|f\|_2^2.$$

Proof. Schur's test

$$\begin{aligned} \sum_P |\langle f, \phi_P \rangle|^2 &= \langle f, \sum_P \phi_P \langle \phi_P, f \rangle \rangle \\ &\leq \|f\|_2 \left\| \sum_P \phi_P \langle \phi_P, f \rangle \right\|_2 \\ &= \|f\|_2 \left(\sum_{P, P'} |\langle f, \phi_P \rangle \langle \phi_P, \phi_{P'} \rangle \langle \phi_{P'}, f \rangle| \right)^{1/2} \\ &\leq \|f\|_2 \left(\sum_P |\langle f, \phi_P \rangle|^2 \sum_{P'} |\langle \phi_P, \phi_{P'} \rangle| \right)^{1/2} \end{aligned}$$

shows that it suffices to prove

$$\sup_P \sum_{P'} |\langle \phi_P, \phi_{P'} \rangle| < \infty.$$

For a fixed tile P we split the above sum according to the shearing matrix A' of the tile P' . For a given shearing matrix A' we distinguish the cases $k_2 \leq k'_2$ and $k_2 > k'_2$.

In the case $k_2 \leq k'_2$ the tile P has larger scale than P' , so the tail of the associated wave packet is more important. For $L \in 2^{\mathbb{N}}$ let

$$\tilde{\mathcal{R}}_L := \{P' \text{ with shearing matrix } A' \text{ such that } LP \cap P' \neq \emptyset\}$$

and let $\mathcal{R}_1 := \tilde{\mathcal{R}}_1$, $\mathcal{R}_L := \tilde{\mathcal{R}}_L \setminus \tilde{\mathcal{R}}_{L/2}$ for $L \geq 2$. Then

$$|\tilde{\mathcal{R}}_L| \lesssim L(L2^{-k_2} + |2^{-k_2} l - 2^{-k'_2} l'|)/2^{-k'_2},$$

and

$$\sum_{L \in 2^{\mathbb{N}}} \sum_{P' \in \mathcal{R}_L} |\langle \phi_P, \phi_{P'} \rangle| \lesssim \sum_{L \in 2^{\mathbb{N}}} |\tilde{\mathcal{R}}_L| \min(L^{-C}, 2^{-C(k'_2 - k_2)}, (2^{k'_2} |2^{-k_2} l - 2^{-k'_2} l'|)^{-C}),$$

where the first estimate inside the minimum is due to spatial separation and the other two estimates come from Lemma 3.2. Summing this over $k'_2 \geq k_2$ and l' we obtain

$$\begin{aligned} & \sum_{L \in 2^{\mathbb{N}}, k'_2 \geq k_2, l' \in \mathbb{Z}} L(L2^{-k_2} + |2^{-k_2}l - 2^{-k'_2}l'|) / 2^{-k'_2} \min(L^{-C}, 2^{-C(k'_2 - k_2)}, (2^{k'_2}|2^{-k_2}l - 2^{-k'_2}l'|)^{-C}) \\ & \lesssim \sum_{L \in 2^{\mathbb{N}}, k \geq 0, l' \in \mathbb{Z}} L(L2^k + |2^k l - l'|) \min(L^{-C}, 2^{-Ck}, |2^k l - l'|^{-C}) \\ & \lesssim \sum_{L \in 2^{\mathbb{N}}, k \geq 0, l' \in \mathbb{Z}} L(L2^k + |2^k l - l'|)(L + 2^k + |2^k l - l'|)^{-C} \leq C. \end{aligned}$$

In the region $k_2 \geq k'_2$ we make a similar decomposition with

$$\tilde{\mathcal{R}}_L := \{P' \text{ with shearing matrix } A' \text{ such that } P \cap LP' \neq \emptyset\}.$$

The resulting estimate is similar to the above with the roles of k_2 and k'_2 reversed. \square

3.2.3. *Splitting into compactly supported wave packets.* In order to obtain a localized Bessel inequality we decompose wave packets into compactly supported parts as in [Mus+06, Lemma 3.1].

Lemma 3.4. *For every C there exists C' such that if $C' \phi \in \Phi_{C'}$, then there exists a decomposition*

$$\phi = \sum_{k \geq 0} 2^{-Ck} \phi_k, \quad \phi_k \in \Phi_C, \text{ supp } \phi_k \subset B(0, 2^k).$$

Sketch of proof. Let ψ be a smooth function supported on $B(0, 1/2)$ and identically equal to 1 on $B(0, 1/4)$. Write $\psi_k(x) = \psi(2^{-k}x)$ for its L^∞ dilates. Let also $\eta^{(0)}, \dots, \eta^{(C-2)}$ be smooth functions supported on $[-1/2, 1/2]$ with

$$\int x^n \eta^{(m)}(x) dx = \mathbf{1}_{n=m}.$$

For $k \in \mathbb{N}$ and $x_1 \in \mathbb{R}$ let

$$m_k^{(n)}(x_1) := \int_{\mathbb{R}} x_2^n \phi(x_1, x_2) \psi_k(x_1, x_2) dx_2,$$

then for $|\alpha| \leq C$ and $n < C$ we have

$$|\partial^\alpha m_k^{(n)}(x_1)| = \left| \int_{\mathbb{R}} x_2^n \partial_1^\alpha \phi(x_1, x_2) (\psi_k(x_1, x_2) - 1) dx_2 \right| \lesssim 2^{-Ck} (1 + |x_1|)^{-C}$$

provided that C' is sufficiently large. The claimed splitting is given by

$$\phi_k := \begin{cases} \phi(\psi_k - \psi_{k-1}) - \sum_{n=0}^{C-2} (m_k^{(n)} - m_{k-1}^{(n)}) \otimes \eta^{(n)}, & k > 0, \\ \phi\psi_0 - \sum_{n=0}^{C-2} m_0^{(n)} \otimes \eta^{(n)}, & k = 0. \end{cases} \quad \square$$

3.2.4. *Energy embedding.* The energy embedding is defined by

$$F(R) := \sup_{\phi_R^{(1)}} |\langle f, \phi_R^{(1)} \rangle|, \quad R \in X,$$

where the supremum is taken over all L^1 normalized wave packets adapted to R with a sufficiently large order of decay C' .

Lemma 3.5. $\|F\|_{L^{2,\infty}(S^2)} \lesssim \|f\|_2$.

Proof. Let \mathcal{R} be a maximal collection of tiles with $S^2(F)(\mathcal{R}) \geq \lambda$. If $\mathcal{R}' \subset X \setminus \mathcal{R}$ also has size $\geq \lambda$, then using subadditivity of σ it is easy to see that $\mathcal{R} \cup \mathcal{R}'$ also has size

$\geq \lambda$, contradicting maximality. Hence by maximality we have $\text{outsup}_{X \setminus \mathcal{R}} S^2(F) \leq \lambda$. On the other hand,

$$\sigma(\mathcal{R}) \leq \lambda^{-2} \sum_{R \in \mathcal{R}} |R| |F(R)|^2 \lesssim \lambda^{-2} \|f\|_2^2$$

by Lemma 3.3. □

Lemma 3.6. $\|F\|_{L^\infty(S^2)} \lesssim \|f\|_\infty$.

Proof. Let $\mathcal{R} \in \mathbf{E}$ and let $\phi_R, R \in \mathcal{R}$, be wave packets that almost extremize $F(R)$. Splitting the corresponding members of $\Phi_{C'}$ using Lemma 3.4 we obtain decompositions $\phi_R = \sum_{k \geq 0} 2^{-Ck} \phi_{R,k}$, where each $\phi_{R,k}$ is an L^1 normalized wave packet adapted to R (with a lower order of decay C) and supported on $2^k R$.

By Lemma 3.3 and the support condition we have

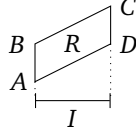
$$\begin{aligned} \sum_{R \in \mathcal{R}} |R| |\langle f, \phi_{R,k} \rangle|^2 &\lesssim 2^{-2Ck} \|f \mathbf{1}_{\cup\{2^k R : R \in \mathcal{R}\}}\|_2^2 \\ &\leq 2^{-2Ck} \|f\|_\infty^2 |\cup_{R \in \mathcal{R}} 2^k R| \\ &\leq 2^{(C_{3.1} - 2C)k} \|f\|_\infty^2 \sigma(R), \end{aligned}$$

and summing in k we obtain

$$\sum_{R \in \mathcal{R}} |R| |\langle f, \phi_R \rangle|^2 \lesssim \|f\|_\infty^2 \sigma(R),$$

so that $S^2(F)(\mathcal{R}) \lesssim \|f\|_\infty$ as required. □

3.3. Covering lemma for parallelograms and the mass embedding. For completeness we include a slightly streamlined proof of a covering lemma from [BT13]. Covering lemmas of this type go back to [CF75]. We consider parallelograms with two vertical edges as shown below:



The *height* $H(R)$ is the common length of AB and CD . The *shadow* $I(R)$ is the projection of R onto the horizontal axis. The *slope* $s(R)$ is the common slope of the edges BC and AD . The *uncertainty interval* $U(R) \subset \mathbb{R}$ is the interval between the slopes of BD and AC . It is the interval of length $2H(R)/|I(R)|$ centered at $s(R)$.

Lemma 3.7 (cf. [BT13, Lemma 7]). *Let \mathcal{R} a finite collection of parallelograms with vertical edges and dyadic shadow. Then there exists $\mathcal{G} \subset \mathcal{R}$ such that*

$$(3.8) \quad \left| \bigcup_{R \in \mathcal{R}} R \right| \lesssim \sum_{R \in \mathcal{G}} |R|$$

and for every $n \in \mathbb{N}$ we have

$$(3.9) \quad \sum_{R_1, \dots, R_n \in \mathcal{G} : U(R_1) \cap \dots \cap U(R_n) \neq \emptyset} |R_1 \cap \dots \cap R_n| \lesssim_n \sum_{R \in \mathcal{G}} |R|.$$

In particular, for every measurable function $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ the sets

$$E(R) := \{(x, y) \in R : u(x, y) \in U(R)\}.$$

satisfy

$$(3.10) \quad \int \left(\sum_{R \in \mathcal{G}} \mathbf{1}_{E(R)} \right)^q \lesssim_q \sum_{R \in \mathcal{G}} |R|, \quad 0 < q < \infty.$$

In [BT13] the conclusion (3.10) is stated for one-variable vector fields, but this structural assumption is not used in the proof.

In the proof of Lemma 3.7 we denote by CR the parallelogram with the same center, slope, and shadow as R but height $CH(R)$ (this definition of CR is used only here). We need the following geometric observation:

Lemma 3.11. *Let R, R' be two parallelograms with $I(R) = I(R')$, $U(R) \cap U(R') \neq \emptyset$, and $R \cap R' \neq \emptyset$. If $7H(R) \leq H(R')$, then $7R \subseteq 7R'$.*

Let M_V denote the Hardy–Littlewood maximal operator in the vertical direction:

$$(3.12) \quad M_V f(x, y) = \sup_{y \in J} |J|^{-1} \int_J |f(x, z)| dz,$$

where the supremum is taken over all intervals J containing y .

Proof of Lemma 3.7. We select \mathcal{G} using the following iterative procedure. Initialize

$$\begin{aligned} STOCK &:= \mathcal{R} \\ \mathcal{G} &:= \emptyset. \end{aligned}$$

While $STOCK \neq \emptyset$, choose an $R \in STOCK$ with maximal $|I(R)|$. Update

$$\begin{aligned} \mathcal{G} &:= \mathcal{G} \cup \{R\}, \\ STOCK &:= STOCK \setminus \{R \in STOCK : R \subset \{M_V(\sum_{R' \in \mathcal{G}} \mathbf{1}_{7R'}) \geq 10^{-4}\}\}. \end{aligned}$$

This procedure terminates after finitely many steps since at each step at least the selected parallelogram R is removed from $STOCK$.

By construction

$$(3.13) \quad \bigcup_{R \in \mathcal{R}} R \subset \{x : M_V(\sum_{R \in \mathcal{G}} \mathbf{1}_{7R})(x) \geq 10^{-4}\},$$

and (3.8) follows by the weak (1, 1) inequality for M_V .

We prove (3.9) by induction on n . For $n = 1$ the statement clearly holds. Suppose that (3.9) holds for a given n , we will show that it also holds with n replaced by $n + 1$. For each $R' \in \mathcal{G}$ let

$$\mathcal{G}(R') := \{R \in \mathcal{G} \text{ chosen prior to } R' \text{ with } R \cap R' \neq \emptyset, U(R) \cap U(100R') \neq \emptyset\}.$$

All terms in (3.9) in which some R_i occurs at least twice are estimated by the inductive hypothesis. In the remaining terms we may arrange the R_i 's in the order reverse to the selection order (losing a factor $(n + 1)!$), and omitting some vanishing terms we obtain the estimate

$$(3.14) \quad \sum_{R_0 \in \mathcal{G}, R_1 \in \mathcal{G}(R_0), \dots, R_n \in \mathcal{G}(R_{n-1})} |R_0 \cap \dots \cap R_n| \leq \sum_{R_0 \in \mathcal{G}, R_1 \in \mathcal{G}(R_0), \dots, R_n \in \mathcal{G}(R_{n-1})} |I(R_0)| \cdot |H(R_n)|.$$

We claim that for every $R' \in \mathcal{G}$ we have

$$(3.15) \quad \sum_{R \in \mathcal{G}(R')} H(R) \leq H(R').$$

To see this let $R \in \mathcal{G}(R')$, so that in particular $I(R') \subset I(R)$ and $U(R) \cap U(10R') \neq \emptyset$. If $H(R') \leq H(R)$, then $7H(10R') \leq H(70R)$, and Lemma 3.11 shows that $70R' \subset 490R$, so that $M_V(\mathbf{1}_R) \geq 490^{-1}$ on R' , contradicting $R' \in \mathcal{G}$. Therefore $H(R') > H(R)$, so $7H(R) \leq H(10R')$, and Lemma 3.11 shows that

$$7R \cap (I(R') \times \mathbb{R}) \subset 70R'.$$

The inequality (3.15) follows, since otherwise $M_V(\sum_{R \in \mathcal{G}(R')} \mathbf{1}_R) \geq 70^{-1}$ on R' , contradicting $R' \in \mathcal{G}$. Hence

$$\begin{aligned} (3.14) &\leq \sum_{R_0 \in \mathcal{G}, R_1 \in \mathcal{G}(R_0), \dots, R_{n-1} \in \mathcal{G}(R_{n-2})} |I(R_0)| \cdot |H(R_{n-1})| \\ &\leq \dots \leq \sum_{R_0 \in \mathcal{G}} |I(R_0)| \cdot |H(R_0)| = \sum_{R_0 \in \mathcal{G}} |R_0|. \end{aligned}$$

This completes the proof of (3.9). In order to see (3.10) observe that its left-hand side is monotonically increasing in q , so it suffices to consider integer values $q = n$, and in this case the left-hand side of (3.10) is dominated by the left-hand side of (3.9). \square

3.3.1. *Mass embedding.* The mass embedding is given by

$$G(R) := |R|^{-1} \int_{E_R} |g|, \quad R \in X.$$

Lemma 3.16. *Let $1 < q < \infty$. If the constant C in the definition of σ is sufficiently large depending on q , then $\|G\|_{L^{q,\infty}(S^\infty)} \lesssim \|g\|_q$.*

Recall that CP now again denotes the parallelogram P expanded by the factor C both in the horizontal and in the vertical direction.

Proof. Let $\delta > 0$, $g \in L^q(\mathbb{R}^2)$, and let \mathcal{R} be a collection of tiles such that $G(R) \geq \delta$ for $R \in \mathcal{R}$. We have to show

$$(3.17) \quad \sup_{L \geq 1} L^{-C} \left| \bigcup_{R \in \mathcal{R}} LR \right| \lesssim_q \delta^{-q} \|g\|_q^q.$$

Note that the definition of $G(R)$ makes sense for arbitrary parallelograms (not only the dyadic ones that we call tiles). For the enlarged parallelograms LR we still have $G(LR) \geq \delta/L^2$, so it suffices to show (3.17) with $L = 1$ and a collection of arbitrary parallelograms \mathcal{R} , provided that the constant $C_{3.1}$ in the definition of σ is at least $2q$.

Enlarging the parallelograms in such a way that their shadows become intervals in adjacent dyadic grids and the uncertainty intervals stay the same we preserve the hypothesis $G(R) \gtrsim \delta$ up to a multiplicative constant. Hence we may assume that the parallelograms have dyadic shadows.

In view of (3.8) it suffices to consider the parallelograms in the subset $\mathcal{G} \subset \mathcal{R}$ provided by Lemma 3.7. By the density assumption and Hölder's inequality we have

$$\begin{aligned} \sum_{R \in \mathcal{G}} |R| &\leq \sum_{R \in \mathcal{G}} \frac{1}{\delta} \int_{E(R)} |g| \\ &= \frac{1}{\delta} \left\| \sum_{R \in \mathcal{G}} \mathbf{1}_{E(R)} |g| \right\|_1 \\ &\leq \frac{1}{\delta} \left\| \sum_{R \in \mathcal{G}} \mathbf{1}_{E(R)} \right\|_{q'} \|g\|_q \\ &\lesssim \frac{1}{\delta} \left(\sum_{R \in \mathcal{G}} |R| \right)^{1/q'} \|g\|_q, \end{aligned}$$

where in the last passage we have used the estimate (3.10). After division by the middle factor of the right hand side we obtain the claim. \square

3.4. **Estimate for the square function.** We finally prove Theorem 1.12. Note that $A_{u,\phi} P_{2,t} f(x)$ is the integral of f against an L^1 normalized wave packet associated to a

tile that contains x and whose uncertainty interval contains $u(x)$. Hence the left-hand side of (1.13) is bounded by

$$\|(\sum_{R \in X} F(R)^2 \mathbf{1}_{E_R})^{1/2}\|_p = \|\sum_{R \in X} F(R)^2 \mathbf{1}_{E_R}\|_{p/2}^{1/2}.$$

Dualizing with a function $g \in L^{(p/2)'} we obtain$

$$\int \sum_{R \in X} F(R)^2 \mathbf{1}_{E_R} g = \sum_{R \in X} |R| F(R)^2 G(R).$$

For every $\mathcal{R} \in \mathbf{E}$ we have $\sum_{R \in \mathcal{R}} |R| F(R) = \sigma(\mathcal{R}) S^1(F)(\mathcal{R})$. Therefore by [DT15, Proposition 3.6] and outer Hölder inequality [DT15, Proposition 3.4] the above is bounded by

$$\|F^2 G\|_{L^1(S^1)} \lesssim \|F^2\|_{L^{p/2}(S^1)} \|G\|_{L^{(p/2)'}(S^\infty)} = \|F\|_{L^p(S^2)}^2 \|G\|_{L^{(p/2)'}(S^\infty)}.$$

The first term is bounded by $\|f\|_p^2$ by Lemmas 3.5 and 3.6 and interpolation [DT15, Proposition 3.5]. The second term is bounded by $\|g\|_{(p/2)'}$ by Lemma 3.16 and interpolation [DT15, Proposition 3.5].

3.5. Application to a maximal operator with a restricted set of directions. In this section we prove Corollary 1.14.

Although the operator (1.11) is unbounded for general direction fields u , it is clearly bounded (on any L^p , $1 \leq p \leq \infty$) with norm $O(N)$ as long as u is allowed to take at most N values. This trivial estimate has been improved to $O(\sqrt{\log N})$ on L^2 by Katz [Kat99]. Note that we also have the trivial estimate $O(1)$ on L^∞ , and by interpolation one obtains logarithmic dependence on N of the operator norm of (1.11) on L^p also for all $2 < p < \infty$. Demeter [Dem10] gives an alternative proof of Katz's result, and furthermore hints at yet another different proof via reduction to the square function bound Theorem 1.12 by means of the good- λ inequality with sharp constant due to Chang, Wilson, and Wolff [CWW85]. The first appearance of a similar reduction to square function in the context of maximal multipliers goes back to Grafakos, Honzík, and Seeger [GHS06], and analogous approaches have been since used in Demeter [Dem10] and Demeter with the first author [DD14]. We have not been able to reproduce the endpoint $p = 2$ using this technique. However, notice that our square function approach, after interpolation, recovers the result for $p > 2$ up to an arbitrarily small loss in the exponent of the logarithm.

Proof of Corollary 1.14. For $j \in \mathbb{Z}$, define the dyadic martingale averaging operator

$$(3.18) \quad E_j f := \sum 2^{2j} \langle f, \mathbf{1}_Q \rangle \mathbf{1}_Q,$$

where the summation runs over all standard dyadic squares Q in \mathbb{R}^2 with side length 2^{-j} . Further define

$$\begin{aligned} \Delta_j &= E_{j+1} - E_j, \\ \Delta f &:= (\sum_{j \in \mathbb{Z}} |\Delta_j f|^2)^{1/2}. \end{aligned}$$

Let M denote the non-dyadic Hardy–Littlewood maximal operator. Chang, Wilson, and Wolff [CWW85, Corollary 3.1] prove that there are universal constants c_1 and c_2 such that for all $\lambda > 0$ and $0 < \epsilon < 1$

$$(3.19) \quad |\{z : |f(z) - E_0 f(z)| > 2\lambda, \Delta f(z) \leq \epsilon \lambda\}| \leq c_2 e^{-\frac{c_1}{\epsilon^2}} |\{z : M f(z) \geq \lambda\}|.$$

Denote the finitely many values of u by u_i , $1 \leq i \leq N$, and write A_{u_i} for the operator with the constant direction field u_i . Corollary 1.14 follows by Marcinkiewicz interpolation from the weak type inequality

$$|\{z : \sup_i |A_{u_i} f(z)| > 4\lambda\}| \leq C \log(N+2)^{p/2} \lambda^{-p} \|f\|_p^p$$

for $2 < p < \infty$. Gearing up for Chang, Wilson, and Wolff we estimate

$$\begin{aligned}
& |\{z : \sup_i |A_{u_i} f(z)| > 4\lambda\}| \\
&= |\bigcup_i \{z : |A_{u_i} f(z)| > 4\lambda\}| \\
(3.20) \quad &\leq |\bigcup_i \{z : |A_{u_i} f(z) - E_0 A_{u_i} f(z)| > 2\lambda, \Delta A_{u_i} f(z) \leq \epsilon\lambda\}|
\end{aligned}$$

$$(3.21) \quad + |\bigcup_i \{z : |E_0 A_{u_i} f(z)| > 2\lambda\}|$$

$$(3.22) \quad + |\bigcup_i \{z : \Delta A_{u_i} f(z) > \epsilon\lambda\}|$$

Using (3.19) we estimate

$$\begin{aligned}
(3.20) &\leq \sum_i |\{z : |A_{u_i} f(z) - E_0 A_{u_i} f(z)| > 2\lambda, \Delta A_{u_i} f(z) \leq \epsilon\lambda\}| \\
&\leq C \sum_i e^{-\frac{c_1}{\epsilon^2}} |\{z : M(A_{u_i} f)(z) > 2\lambda\}| \\
&\leq C \sum_i e^{-\frac{c_1}{\epsilon^2}} \lambda^{-p} \|A_{u_i} f\|_p^p \\
&\leq CN e^{-\frac{c_1}{\epsilon^2}} \lambda^{-p} \|f\|_p^p \\
&\leq C \lambda^{-p} \|f\|_p^p
\end{aligned}$$

provided $\epsilon \leq c_1^{1/2} \log(N+2)^{1/2}$.

The function $E_0 A_{u_i} f$ in (3.21) is pointwise dominated by the standard Hardy–Littlewood maximal operator, because E_0 and A_{u_i} compose to some averaging operator at scale 0. Therefore

$$(3.21) \leq |\{z : Mf(z) > C\lambda\}| \lesssim \lambda^{-p} \|f\|_p^p.$$

To control (3.22) we introduce a suitable Littlewood–Paley decomposition in the second variable, note that P_{2^k} commutes with A_{u_i} , and estimate pointwise

$$\begin{aligned}
\sup_i \Delta A_{u_i} f &= \sup_i \Delta \left(\sum_{k \in \mathbb{Z}} P_{2^k} A_{u_i} P_{2^k} f \right) \\
&= \sup_i \left(\sum_j |\sum_k \Delta_j P_{2^k} A_{u_i} P_{2^k} f|^2 \right)^{1/2} \\
&\lesssim \sup_i \left(\sum_j \left(\sum_k 2^{-|j-k|/q'} MM_{q,V} A_{u_i} P_{2^k} f \right)^2 \right)^{1/2} \\
&\lesssim \sup_i \left(\sum_j \sum_k 2^{-|j-k|/q'} (MM_{q,V} A_{u_i} P_{2^k} f)^2 \right)^{1/2} \\
&\lesssim \sup_i \left(\sum_t (MM_{q,V} A_{u_i} P_{2^k} f)^2 \right)^{1/2} \\
&\leq \left(\sum_k (MM_{q,V} \sup_i |A_{u_i} P_{2^k} f|)^2 \right)^{1/2},
\end{aligned}$$

where $M_{q,V}$ is the q -maximal operator in the vertical direction $M_{q,V} f = (M_V(f^q))^{1/q}$ for any fixed $1 < q < 2$ with M_V as in (3.12), M is the usual two-dimensional Hardy–Littlewood maximal operator, and the pointwise estimate $|\Delta_j P_{2^k} f| \lesssim 2^{|j-k|/q'} MM_{q,V} f$ follows from [GHS06, Sublemma 4.2] applied in the vertical direction. The Fefferman–Stein maximal inequalities and Theorem 1.12 give

$$\left\| \left(\sum_{t \in 2^{\mathbb{Z}}} (MM_{q,V} \sup_i |A_{u_i} P_t f|)^2 \right)^{1/2} \right\|_p \leq C \left\| \left(\sum_{t \in 2^{\mathbb{Z}}} (\sup_i |A_{u_i} P_t f|)^2 \right)^{1/2} \right\|_p \leq C \|f\|_p.$$

With Tchebysheff we obtain

$$(3.22) = \{|\sup_i \Delta A_{u_i} f(z) > \epsilon \lambda|\} \leq C(\epsilon \lambda)^{-p} \|f\|_p^p \leq C \log(N+2)^{p/2} \lambda^{-p} \|f\|_p^p,$$

and this concludes the proof of Corollary 1.14. \square

APPENDIX A. LACEY-LI COVERING ARGUMENT

Lacey and Li [LL10] have introduced a certain family of maximal operators associated to a vector field u , which they called the ‘‘Lipschitz–Kakeya’’ maximal operator:

$$f \mapsto \sup_{R \in \mathcal{R}_\delta} \langle f, \mathbf{1}_R \rangle \frac{\mathbf{1}_R}{|R|}$$

where, using the notation from Section 3.3, \mathcal{R}_δ is the collection of those parallelograms R with $|E(R)| \geq \delta |R|$; that is, the vector field u points within the uncertainty interval of R on (at least a) δ -portion of R . These authors proved that such maximal operators have weak type $(2, 2)$ operator norm $O(\delta^{-1/2})$ if the vector field is Lipschitz. In the same paper, they have further showed that an L^p bound for this operator for any $p < 2$ implies the L^2 estimate for the single band version of the directional Hilbert transform. Bateman and Thiele [BT13] gave a streamlined proof of the weak type $(2, 2)$ estimate for this maximal operator in the case of a one-variable vector field and used it to obtain square function estimates of the type (1.2) for the directional Hilbert transform.

In this section we further simplify the proof of the weak type $(2, 2)$ estimate for this maximal operator, also taking care of Lipschitz vector fields. We use the notation from Section 3.3 and write $L(R) = |I(R)|$. The main part of the proof is the following covering argument.

Theorem A.1. *Let $0 < \delta \leq 1$ and let \mathcal{R} be a finite collection of parallelograms with vertical edges and dyadic shadow such that for each $R \in \mathcal{R}$ we have*

$$|E(R)| \geq \delta |R|$$

and $L(R) \|v\|_{\text{Lip}} \leq 1/30$. Then there is a subset $\mathcal{G} \subset \mathcal{R}$ such that

$$(A.2) \quad \left| \bigcup_{R \in \mathcal{R}} R \right| \lesssim \sum_{R \in \mathcal{G}} |R|,$$

$$(A.3) \quad \int \left(\sum_{R \in \mathcal{G}} \mathbf{1}_R \right)^2 \lesssim \delta^{-1} \sum_{R \in \mathcal{G}} |R|.$$

The set \mathcal{G} is constructed as in Lemma 3.7, so that (A.2) holds by construction. In the remaining part of this section we will show (A.3). Expanding the square on the left-hand side of (A.3) and using symmetry we obtain the estimate

$$\sum_{R \in \mathcal{G}} |R| + 2 \sum_{(R, R') \in \mathcal{D}} |R \cap R'|,$$

where \mathcal{D} is the set of pairs $(R, R') \in \mathcal{G}^2$ such that $R \cap R' \neq \emptyset$ and R has been chosen before R' . The former term is clearly bounded by the right-hand side of (A.3). In the latter term we notice first that by (3.15) we have

$$\sum_{R' \in \mathcal{G}} \sum_{R \in \mathcal{G}(R')} |R \cap R'| \leq \sum_{R' \in \mathcal{G}} \sum_{R \in \mathcal{G}(R')} L(R') H(R) \leq \sum_{R' \in \mathcal{G}} L(R') H(R'),$$

and this is also bounded by the right-hand side of (A.3). Hence it suffices to estimate

$$(A.4) \quad \sum_{R \in \mathcal{G}} \sum_{R' \in \mathcal{D}(R)} |R \cap R'|,$$

where

$$\mathcal{D}(R) := \{R' : (R, R') \in \mathcal{D}, U(R) \cap 10U(R') = \emptyset\}.$$

First we clarify the position of $U(R')$ relative to $U(R)$ when $R' \in \mathcal{D}(R)$.

Lemma A.5. *Suppose $R' \in \mathcal{P}(R)$. Then*

$$\max(|U(R)|, |U(R')|) \leq \frac{1}{4} \text{dist}(U(R'), U(R)).$$

Proof. We distinguish two cases:

- (1) $|U(R)| \leq |U(R')|$. In this case we use the definition of $\mathcal{P}(R)$.
- (2) $|U(R)| > |U(R')|$. In this case we have

$$H(R') = |U(R')|L(R') < |U(R)|L(R) = H(R),$$

and in particular $7H(R') \leq H(10R)$. If the conclusion was false, then $10U(R) \cap U(R') \neq \emptyset$, and by Lemma 3.11 we obtain $7R' \subset 70R$. This contradicts the hypothesis that R' was added to \mathcal{G} after R . □

The next lemma gives a condition for two parallelograms to have comparable slopes. This is the only place where the Lipschitz hypothesis is used. Denote the projection onto the first coordinate by Π .

Lemma A.6. *Assume $L(R)\|v\|_{\text{Lip}} \leq 1/30$. Suppose $R', R'' \in \mathcal{P}(R)$ and $\Pi E(R') \cap \Pi E(R'') \neq \emptyset$. Then*

$$\text{dist}(U(R'), U(R'')) \leq \frac{1}{8} \text{dist}(U(R), U(R')).$$

Proof. Let $x \in \Pi E(R') \cap \Pi E(R'')$. The distance of the points y', y'' such that $(x, y') \in R'$ and $(x, y'') \in R''$ is bounded above by

$$H(R) + H(R') + H(R'') + L(R') \text{dist}(U(R), U(R')) + L(R'') \text{dist}(U(R), U(R'')).$$

Choosing $(x, y') \in E(R')$ and $(x, y'') \in E(R'')$ and using the Lipschitz hypothesis and Lemma A.5 we obtain

$$\begin{aligned} \text{dist}(U(R'), U(R'')) &\leq \|v\|_{\text{Lip}} \left(H(R) + H(R') + H(R'') \right. \\ &\quad \left. + L(R') \text{dist}(U(R), U(R')) + L(R'') \text{dist}(U(R), U(R'')) \right) \\ &\leq \frac{1}{30} \left(|U(R)| + |U(R')| + |U(R'')| + \text{dist}(U(R), U(R')) + \text{dist}(U(R), U(R'')) \right) \\ &\leq \frac{1}{30} \left(\frac{6}{4} \text{dist}(U(R), U(R')) + \frac{5}{4} \text{dist}(U(R), U(R'')) \right) \\ &\leq \frac{1}{30} \left(\frac{11}{4} \text{dist}(U(R), U(R')) + \frac{5}{4} |U(R')| + \frac{5}{4} \text{dist}(U(R'), U(R'')) \right) \\ &\leq \frac{1}{30} \left(\frac{13}{4} \text{dist}(U(R), U(R')) + \frac{5}{4} \text{dist}(U(R'), U(R'')) \right). \end{aligned}$$

The conclusion follows. □

The basic estimate for the size of the intersection of two parallelograms is the size of the intersection of infinite stripes containing them:

Lemma A.7. *Let $R, R' \in \mathcal{R}$. Then*

$$(A.8) \quad |R \cap R'| \leq \text{dist}(U(R), U(R'))^{-1} H(R) H(R').$$

Proof. By a shearing transformation we may assume that the central line segment of R is horizontal. Let u_0 be the central slope of R' . Then $R \cap R'$ is contained in a parallelogram of height $H(R)$ and base $H(R')u_0^{-1}$. On the other hand, $u_0 \geq \text{dist}(U(R), U(R'))$. □

We decompose the set $\mathcal{P}(R)$ dyadically according to the distance between $U(R)$ and $U(R')$. Specifically, for $k \in \mathbb{N}$ let

$$\mathcal{P}_k(R) := \{R' \in \mathcal{P}(R) : 2^{k-3} < \frac{\text{dist}(U(R), U(R'))}{|U(R)|} \leq 2^{k-2}\}.$$

For a fixed k we will estimate the contribution of $\mathcal{P}_k(R)$ to (A.4) using a stopping time argument. For a dyadic interval I denote $R_I := R \cap (I \times \mathbb{R})$.

Lemma A.9. *Let $I \subseteq I_R$ be a dyadic interval such that there exists $R'' \in \mathcal{P}_k(R)$ with $I_{R''} \subseteq I$. Then*

$$\sum_{R' \in \mathcal{P}_k(R): I \subseteq I_{R'}} |R_I \cap R'| \leq 2|R_I|.$$

Proof. Let $R'' \in \mathcal{P}_k(R)$ be the parallelogram with $I_{R''} \subseteq I$ that has been chosen last. Let

$$\mathcal{Q} := \{R' \in \mathcal{P}_k(R) : I \subseteq I_{R'}, R_I \cap R' \neq \emptyset\} \setminus \{R''\}.$$

Since $|R_I \cap R''| \leq |R_I|$, it suffices to show

$$(A.10) \quad \sum_{R' \in \mathcal{Q}} |R_I \cap R'| \leq 10^{-1}|R_I|.$$

Assume for contradiction that (A.10) fails. Let $U := 2^k U(R)$. By Lemma A.5 we have $U(R'') \subset U$ and thus

$$H(R'') \leq |U| |I_{R''}| \leq |U| |I|.$$

In particular

$$R'' \subset 50(1 + |U| |I| / H(R)) R_I =: \tilde{R}.$$

The parallelogram R'' has been selected for \mathcal{G} after the parallelogram R and the parallelograms $R' \in \mathcal{Q}$. To obtain a contradiction with the construction of \mathcal{G} it suffices to show that

$$M_V(\mathbf{1}_R + \sum_{R' \in \mathcal{Q}} \mathbf{1}_{R'})$$

where M_V is the vertical directional maximal function, is larger than 10^{-3} on the parallelogram \tilde{R} .

First assume there exists $R' \in \mathcal{Q}$ with $H(R') \geq 20|U| |I|$. Note that

$$U(R') \subset U \subset U(\tilde{R}).$$

Applying Lemma 3.11 to the rectangles R'_I and \tilde{R} we obtain

$$M_V(\mathbf{1}_{R'} + \mathbf{1}_R) \geq 7^{-1} H(\tilde{R})^{-1} (\min(H(R'), H(\tilde{R})) + H(R)) > 10^{-3}$$

on \tilde{R} , which proves Lemma A.9 in the given case.

Hence we may assume

$$H(R') \leq 20|U| |I|$$

for every $R' \in \mathcal{Q}$. We then have on \tilde{R} that

$$\begin{aligned} & M_V(\mathbf{1}_R + \sum_{R' \in \mathcal{Q}} \mathbf{1}_{R'}) \\ & \geq H(\tilde{R})^{-1} (H(R) + \sum_{R' \in \mathcal{Q}} H(R')) \\ & \geq H(\tilde{R})^{-1} (H(R) + \sum_{R' \in \mathcal{Q}} |R_I \cap R'| |U| H(R)^{-1}) && \text{by (A.8)} \\ & \geq H(\tilde{R})^{-1} (H(R) + |U| H(R)^{-1} 10^{-1} |R_I|) && \text{since (A.10) fails} \\ & \geq 500^{-1}. \end{aligned}$$

This completes the proof of Lemma A.9. \square

Corollary A.11.

$$\sum_{R' \in \mathcal{P}_k(R)} |R \cap R'| \leq 4H(R) \cdot \left| \bigcup_{R' \in \mathcal{P}_k(R)} \Pi(R') \right|.$$

Proof. Let \mathcal{I} be the set of maximal dyadic intervals contained in $\cup_{R' \in \mathcal{P}_k(R)} \Pi(R')$ that do not contain $I_{R'}$ for any $R' \in \mathcal{P}_k(R)$. For each $I \in \mathcal{I}$ let \tilde{I} denote its dyadic parent. Then by maximality of I and Lemma A.9 we have

$$\begin{aligned} \sum_{R' \in \mathcal{P}_k(R)} |R_I \cap R'| &= \sum_{R' \in \mathcal{P}_k(R): I \not\subseteq I_{R'}} |R_I \cap R'| \\ &\leq \sum_{R' \in \mathcal{P}_k(R): \tilde{I} \subseteq I_{R'}} |R_{\tilde{I}} \cap R'| \\ &\leq 2|R_{\tilde{I}}| \leq 4|R_I|. \end{aligned}$$

The set \mathcal{I} is a covering of $\cup_{R' \in \mathcal{P}_k(R)} \Pi(R')$, so the conclusion of the lemma follows after summing over all intervals in \mathcal{I} . \square

We are now in position to complete the proof of Theorem A.1 by estimating (A.4):

$$\begin{aligned} \sum_{R' \in \mathcal{P}(R)} |R \cap R'| &= \sum_{k \in \mathbb{N}} \sum_{R' \in \mathcal{P}_k(R)} |R \cap R'| \\ &\lesssim H(R) \sum_{k \in \mathbb{N}} |\cup_{R' \in \mathcal{P}_k(R)} \Pi(R')| && \text{by Corollary A.11} \\ &= H(R) \sum_{k \in \mathbb{N}} \sum_{R' \in \mathcal{P}'_k(R)} |\Pi(R')| \\ &\lesssim \delta^{-1} H(R) \sum_{k \in \mathbb{N}} \sum_{R' \in \mathcal{P}'_k(R)} |\Pi E(R')| \\ &\lesssim \delta^{-1} |R|, \end{aligned}$$

where $\mathcal{P}'_k(R) \subset \mathcal{P}_k(R)$ is a system of representatives for maximal intervals $I_{R'}$, in the penultimate step we have used the density hypothesis in the form $|\Pi(R')| \leq |\Pi E(R')|/\delta$, and in the last step we have used Lemma A.6 to conclude that the projections there have bounded overlap.

REFERENCES

- [Bat13] M. Bateman. “Single annulus L^p estimates for Hilbert transforms along vector fields”. In: *Rev. Mat. Iberoam.* 29.3 (2013), pp. 1021–1069. arXiv: [1109.6395 \[math.CA\]](#).
- [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)*. Vol. 137. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 1989, pp. 111–132.
- [BT13] M. Bateman and C. Thiele. “ L^p estimates for the Hilbert transforms along a one-variable vector field”. In: *Anal. PDE* 6.7 (2013), pp. 1577–1600. arXiv: [1109.6396 \[math.CA\]](#).
- [CF75] A. Cordoba and R. Fefferman. “A geometric proof of the strong maximal theorem”. In: *Ann. of Math. (2)* 102.1 (1975), pp. 95–100.
- [CWW85] S.-Y. A. Chang, J. M. Wilson, and T. H. Wolff. “Some weighted norm inequalities concerning the Schrödinger operators”. In: *Comment. Math. Helv.* 60.2 (1985), pp. 217–246.
- [DD14] C. Demeter and F. Di Plinio. “Logarithmic L^p bounds for maximal directional singular integrals in the plane”. In: *J. Geom. Anal.* 24.1 (2014), pp. 375–416. arXiv: [1203.6624 \[math.CA\]](#).
- [Dem10] C. Demeter. “Singular integrals along N directions in \mathbb{R}^2 ”. In: *Proc. Amer. Math. Soc.* 138.12 (2010), pp. 4433–4442. arXiv: [1001.1998 \[math.CA\]](#).
- [DS15] C. Demeter and P. Silva. “Some new light on a few classical results”. In: *Colloq. Math.* 140.1 (2015), pp. 129–147. arXiv: [1311.4092 \[math.CA\]](#).

- [DT15] Y. Do and C. Thiele. “ L^p theory for outer measures and two themes of Lennart Carleson united”. In: *Bull. Amer. Math. Soc. (N.S.)* 52.2 (2015), pp. 249–296. arXiv: [1309.0945 \[math.CA\]](#).
- [GHS06] L. Grafakos, P. Honzik, and A. Seeger. “On maximal functions for Mikhlin-Hörmander multipliers”. In: *Adv. Math.* 204.2 (2006), pp. 363–378. arXiv: [math/0501114](#).
- [GT16] S. Guo and C. Thiele. “Hilbert transforms along Lipschitz direction fields: A lacunary model”. In: *Mathematika* (2016). To appear. arXiv: [1603.03317 \[math.CA\]](#).
- [Guo+16] S. Guo, J. Hickman, V. Lie, and J. Roos. “Maximal operators and Hilbert transforms along variable non-flat homogeneous curves”. In: *Proc. London Math. Soc.* (2016). To appear. arXiv: [1610.05203 \[math.CA\]](#).
- [Guo17] S. Guo. “Hilbert transform along measurable vector fields constant on Lipschitz curves: L^p boundedness”. In: *Trans. Amer. Math. Soc.* 369.4 (2017), pp. 2493–2519. arXiv: [1409.3010 \[math.CA\]](#).
- [Jon89] P. W. Jones. “Square functions, Cauchy integrals, analytic capacity, and harmonic measure”. In: *Harmonic analysis and partial differential equations (El Escorial, 1987)*. Vol. 1384. Lecture Notes in Math. Springer, Berlin, 1989, pp. 24–68.
- [Kat99] N. H. Katz. “Remarks on maximal operators over arbitrary sets of directions”. In: *Bull. London Math. Soc.* 31.6 (1999), pp. 700–710.
- [LL06] M. T. Lacey and X. Li. “Maximal theorems for the directional Hilbert transform on the plane”. In: *Trans. Amer. Math. Soc.* 358.9 (2006), 4099–4117 (electronic).
- [LL10] M. Lacey and X. Li. “On a conjecture of E. M. Stein on the Hilbert transform on vector fields”. In: *Mem. Amer. Math. Soc.* 205.965 (2010), pp. viii+72. arXiv: [0704.0808 \[math.CA\]](#).
- [Mus+06] C. Muscalu, J. Pipher, T. Tao, and C. Thiele. “Multi-parameter paraproducts”. In: *Rev. Mat. Iberoam.* 22.3 (2006), pp. 963–976. arXiv: [math/0411607](#).
- [Ste93] E. M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Vol. 43. Princeton Mathematical Series. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III. Princeton, NJ: Princeton University Press, 1993, pp. xiv+695.

(FDP) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF VIRGINIA, BOX 400137, CHARLOTTESVILLE, VA, 22904-4137, USA

E-mail address: francesco.diplinio@virginia.edu

(SG) INDIANA UNIVERSITY BLOOMINGTON, 831 E THIRD ST, BLOOMINGTON, IN 47405, USA

E-mail address: shaoguo@iu.edu

(CT, PZ) MATHEMATICAL INSTITUTE, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY

E-mail address: thiele@math.uni-bonn.de

E-mail address: pzorin@math.uni-bonn.de