Rev. Mat. Iberoam. **29** (2013), no. 3, 1021–1069
DOI 10.4171/RMI/748

Single annulus L^p estimates for Hilbert transforms along vector fields

Michael Bateman

Abstract. We prove L^p estimates, $p \in (1, \infty)$, on the Hilbert transform along a one variable vector field acting on functions with frequency support in an annulus. Estimates when $p > 2$ were proved by Lacey and Li. This paper also contains key technical ingredients for a companion paper with Christoph Thiele in which L^p estimates are established for the full Hilbert transform. The operators considered here are singular integral variants of maximal operators arising in the study of planar differentiation problems.

1. Introduction

The subject of this paper is an operator related to two difficult and well-known problems in harmonic analysis. The first is the conjecture of Zygmund on the almost everywhere differentiation of Lipschitz vector fields. More generally, this problem is connected to the family of Kakeya-type geometric problems in multidimensional Euclidean space. The second problem is that of almost everywhere convergence of Fourier integrals. The Carleson–Hunt theorem establishes this convergence for L^p functions of a single real variable, $p \in (1, \infty)$. The two-dimensional question is still unsettled: even in L^2, where convergence in norm follows from Plancherel's theorem, almost everywhere convergence is unknown. Further, Fefferman's result on the unboundedness of the disc multiplier for $p \neq 2$ leaves the almost everywhere L^2-convergence problem every bit in question.

1.1. More on the connection to Kakeya-type problems

The existence of the Besicovitch construction obstructs the L^p boundedness of directional maximal operators of the form

$$M_K f(x) = \sup \frac{1}{|L|} \int_L f,$$

Mathematics Subject Classification (2010): Primary 42B20; Secondary 42B25.
Keywords: Carleson's theorem, time-frequency analysis, Stein's conjecture, Zygmund's conjecture, differentiation of vector fields, Hilbert transform in direction of a vector field.

where the sup is over line segments containing x. An alternative point of view is to define

$$M_K f(x) = \sup_{|v|=1} \sup_{\epsilon} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} f(x+vt)\,dt,$$

where here the sup is over unit vectors v. By defining a vector field $v(x)$ such that v almost attains the supremum in the last equation, we can instead focus on linearized operators

$$M_v f(x) = \sup_{\epsilon} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} f(x+v(x)t)\,dt$$

for a fixed vector field v, and ask for L^p estimates on M_v. Of course L^p estimates are sufficient to establish almost everywhere convergence of appropriate averages along the vector field, analogous to the implication of Lebesgue's differentiation theorem from L^p estimates on the Hardy–Littlewood maximal operator.

The Besicovitch construction shows no such estimate can hold independent of v. Specifically, for every $\epsilon > 0$ there exists a collection $\mathcal{R}_\epsilon$ such that $|\cup_{R\in\mathcal{R}_\epsilon} R| \leq \epsilon$ but $|\cup_{R\in\mathcal{R}_\epsilon} 3R| \geq 1$. By taking $f_\epsilon = \mathbf{1}_{\cup_{R\in\mathcal{R}_\epsilon} R}$, we get $M_v f_\epsilon(x) \gtrsim 1$ for $x \in \cup_{R\in\mathcal{R}_\epsilon} 3R$, for an appropriate choice of v. This gives $\|M_v f\|_p^p \gtrsim 1 >>> \epsilon \geq \|f\|_p^p$.

This led a number of authors to consider several subclasses of v for which one might hope to obtain meaningful results. We consider several natural examples below.

1) v with finite range. Here the goal was to obtain L^p estimates with optimal dependence on the size of the range. Stromberg [12] first established that the L^2 norm of a maximal operator over N directions is logarithmic in N; the sharp exponent on the logarithm was obtained by Katz in [5], following work of a number of others.

2) v with infinite range. Here the result is essentially "If the range of v is contained in a lacunary set (of finite order), then M_v is bounded on L^p, $p > 1$. Otherwise, M_v is not bounded on L^p for any $p < \infty$." This statement follows directly from work of Nagel, Stein, and Wainger [9], who obtained boundedness in the lacunary case; Sjolin and Sjogren [10], who extended this to the higher order lacunary case; and the author [1], who established the converse. Further [1] proves that if a set is not contained in a lacunary set of finite order, then it has a Cantor-type structure, and that in such a case there exist Besicovitch type sets of rectangles whose slopes are in the Cantor-type set.

3) v with some kind of smoothness. It is rather simple to use the Besicovitch construction to show that no uniform bound on M_v can exist if we only assume v is Hölder continuous with index < 1. This is because the Hölder continuity condition is not scale invariant. More concretely, given a Besicovitch set of size $< \epsilon$ as above, one can find a continuous vector field v such that $M_v f$ is large on $\cup 3R$. We may not know anything about the modulus of continuity of v, but we know it exists. By shrinking our collection of rectangles, and dilating v as well, we can get v to be Hölder continuous. It is essentially a conjecture of Zygmund that M_v is bounded as soon as v is Lipschitz (i.e., Hölder continuous with index $= 1$), with estimates depending only on $\|v\|_{\mathrm{Lip}}$. (One small note: we consider in this case only line segments shorter than $1/(10\|v\|_{\mathrm{Lip}})$, to rule out uninteresting radial counterexamples with long line segments, which play no role in the differentiation question.) Progress on this conjecture has been difficult to obtain. Bourgain showed that M_v is bounded when v is real analytic, but no stronger result is known.

1.2. Relation of this paper to smoothness assumptions on v

It is in this third direction that this paper makes some contribution. In addition to the maximal operator M_v, there is a natural singular integral variant H_v, given by

$$H_v f(x) = p.v. \int \frac{f(x - tv(x))}{t}\, dt.$$

Results on the boundedness of H_v are equally scarce, with unboundedness for Hölder v following the same lines as for the case of M_v. Stein and Street [10] have recently established L^p-boundedness for H_v in the real analytic case.

In this paper we consider v that are constant in one variable, but arbitrary in the other, i.e., $v(x_1, x_2) = v(x_1, 0)$, but $v(x_1, 0)$ is merely measurable. This paper has a companion [4] in which there is proved

$$\|H_v f\|_p \lesssim \|f\|_p, \quad p \in \big(\tfrac{3}{2}, \infty\big) \tag{1.1}$$

for such vector fields. This paper proves the same estimate for $p \in (1, \infty)$, under the additional assumption that f have frequency support in an annulus. The present paper is somewhat subservient to its companion because the primary interest in establishing the single-annulus estimate is in establishing (1.1). (We actually transport important subresults to [4] rather than the main theorem itself.) Nevertheless, the tools developed in the present paper are likely to find application in further two-dimensional problems requiring time-frequency analysis, not the least interesting of which is the study of Fourier series as discussed above. It is worth mentioning that this paper also uses heavily ideas of Lacey and Li, which are two-dimensional adaptations of time-frequency methods of Lacey and Thiele.

1.3. More on Fourier series

It is also worth mentioning that the single-annulus version of (1.1) is already quite substantial. The proof of its weak-L^2 boundedness, given by Lacey and Li in [6], is isomorphic to the Lacey–Thiele proof of Carleson's theorem. In fact, the L^2 estimate for Carleson's operator follows from the L^2 estimate for H_v (see [7]) and vice versa (see [4]). The extra work required here for L^p-boundedness arises from the need to appeal to maximal-type theorems when $p < 2$. We comment more on this below.

Further, a serious study of two-dimensional Fourier series is likely to require an understanding of one-dimensional Fourier series as well as an understanding of the interplay between Fourier-analytic and geometric issues. This paper contains some basic ingredients in this direction.

2. The main theorem

Let v be a nonvanishing vector field that depends on one variable, i.e., $v\colon \mathbb{R}^2 \to \mathbb{R}^2 \setminus \{0\}$ and $v(x_1, x_2) = v(x_1)$. In this paper we prove L^p estimates on the Hilbert transform along v precomposed with frequency restriction to an almost annular region. More specifically, define

$$H_v f(x) = p.v. \int \frac{f(x - tv(x))}{t}\, dt.$$

Because of the structure of the Hilbert kernel, the magnitude of v is irrelevant, provided it is nonzero. For this reason we may assume that $v(x_1, x_2) = (1, u(x_1))$. We will further assume that the slope of v is bounded by 1. This will be helpful in this paper for some technical reasons, but our main interest is in the action of H_v on arbitrary functions (i.e., those not necessarily having frequency support in an annulus); in this more general case, the operator is invariant under dilations in the vertical variable. See [4] for more on the symmetries of this problem. This invariance allows us to assume, in that case, that the slope of v is bounded by 1. (This is mostly a technical convenience, that allows us to think of rectangles and parallelograms as being the same kinds of objects.) Since this general problem is the primary motivation for this paper, we adopt the restriction on the slope here as well. The general problem is addressed in a companion paper with Christoph Thiele [4]. The present paper is logically prior to the other, and is therefore self-contained. Fix $w \geq 0$, and define τ to be the trapezoid with corners $(-1/w, 1/w)$, $(1/w, 1/w)$, $(-2/w, 2/w)$, and $(2/w, 2/w)$. Also define

$$\widehat{\Pi_\tau f}(\xi) = \mathbf{1}_\tau(\xi)\hat{f}(\xi).$$

Here we prove the following:

Theorem 2.1. *Let v be a vector field depending on one variable with slope bounded by 1. Let $p \in (1, \infty)$. Then*

$$\|(H_v \circ \Pi_\tau) f\|_p \lesssim \|\Pi_\tau f\|_p.$$

This is just a linearized version of the maximal inequality

$$\Big\| \sup_{|v_2|=1} \|H_{(1,v_2)} \circ \Pi_\tau\|_{L^p_{x_2}} \Big\|_{L^p_{x_1}} \lesssim \|\Pi_\tau f\|_p.$$

We remark that the estimate in this theorem is independent of the parameter w in the definition of τ, which comes as no surprise given the dilation invariance of the problem. Further, the restriction to a trapezoid specifically is nothing to take seriously. Using the assumption on the slope of the vector field we can already assume $\operatorname{supp} \hat{f}$ lies in a two-ended cone near the vertical axis, because H_v acts trivially on functions with support outside this cone. More precisely, if $\hat{f}$ is supported in a cone close to the horizontal axis, then we have, for the constant vector field $(1, 0)$,

$$H_v f(x, y) = H_{(1,0)} f(x, y)\,, \tag{2.1}$$

because $H_{(1,0)}$ is a multiplier corresponding to right and left half-planes. But $H_{(1,0)}$ is trivially bounded, justifying our claim. Finally, a trapezoid is the restriction of the cone to a horizontal frequency band. We could have equally well stated the theorem for functions with support in the full band, and reduced it to the trapezoidal case. Alternatively, we could have worked with an annular region, or an annular region intersected with a cone. Our methods work equally well in these cases. We chose the horizontal band (rather than an annulus) because of the special structure of one-variable vector fields, but for other vector fields an annular region may be more appropriate.

Perhaps the biggest contribution of this paper (aside from its applicability to [4]) is a more streamlined and mechanized collection of two-dimensional time-frequency tools. Building heavily on important earlier work of Lacey–Li (see [6] and [7]), we clarify the relationship between the density-related maximal operators (see Lemma 6.2) and the more classical time-frequency tools. Specifically, a key sublemma in [2], combined with this more efficient understanding, allows us to obtain the full range of exponents $p \in (1, \infty)$ here. Further, although the results are stated only for one-variable vector fields, it is clear how to combine a maximal theorem for a different vector field with the methods of this paper. We should remark that time-frequency analysis in two-dimensions is rather less well-developed than in one-dimension, with work of Lacey–Li being the only natural precursor to this paper. We therefore strove to make the paper self-contained and to include proofs of a number of lemmas that are standard in one-dimension, but whose proofs in the two-dimensional situation do not seem to appear in the literature.

2.1. More details on related work

Study of such problems is motivated by the obvious connection to the problem of estimating the Hilbert transform on functions that have not been Fourier localized. Stein, for example, conjectured that if v is Lipschitz, then H_v (or rather, a truncated version of it) is a bounded operator on L^2. We note that when v depends on only one variable, the L^2 boundedness of H_v is a rather immediate consequence of Carleson's theorem, as shown in [7]. Stein's conjecture is the singular integral variant of Zygmund's well-known conjecture on the differentiation of Lipschitz vector fields. For a more complete history, see [7]. More recently, Thiele and the author proved a range of L^p estimates on the full Hilbert transform along a one variable vector field, using some key lemmas from the present paper. It is known that the operator H_v is related to the return times theorem from ergodic theory; see [4] for more on this connection.

We remark that the operator H_v is quite similar to Carleson's operator (i.e., the maximal Fourier partial sum operator). The argument in [6] is also quite similar to the Lacey–Thiele proof of Carleson's theorem (see [8]). The argument here draws on ingredients from [6], but obtaining L^p estimates for $p < 2$ in this situation requires more effort, partly because the relevant maximal operators are more complicated, but also because making use of the maximal theory is more complicated. In the one-dimensional situation, exceptional sets are unions of intervals; nothing so simple is the case here.

For $p > 2$, Theorem 2.1 was proved for arbitrary vector fields by Lacey and Li in [6]. (In fact, they proved a weak L^2 result.) The same authors, in [7], introduced a method for obtaining L^p, $p < 2$, estimates on $H_v \circ \Pi_\tau$ when a certain maximal theorem is available for the vector field v in question. (The story is a bit technical: they proved a theorem contingent on the existence of a certain maximal theorem in the case of truncated Hilbert kernels. However the method had little to do with the truncation of the kernel, allowing us to extend it here.) The author proved such an L^p maximal theorem when v depends on one variable in [2]. Given this result, it is not surprising that the method from [7] yields a result for some $p < 2$, but the value of p obtained from the method in [7] seemed far from sharp. (At the very least, the method seemed nonsharp. Of course, this was not important for the authors there.) It was clear, for example that new ideas would be required to even reach p close to $3/2$. The author recently improved the estimates in this maximal theorem to the (essentially) best possible in [3]. Because of this, the author decided to investigate the precise range of p for which Theorem 2.1 holds.

2.2. New ideas

The novelties in this paper that allow us to obtain the full range of p claimed in Theorem 2.1 are a simplification of the approach in [7], and a more efficient appeal to the maximal theorems.

We elaborate a bit more on these points for readers already familiar with the argument in [7].

Regarding the first point: in [7], tiles are sorted into trees via standard density and orthogonality (size) lemmas. An important additional observation made in [7] is that if $\mathcal{T}$ is a collection of trees such that for each $T \in \mathcal{T}$ the "size" of T is about σ and the "density" of the top of T is about δ, then we can control $\sum_{T\in\mathcal{T}} |\mathbf{top}(T)|$ by using an appropriate maximal theorem. Their argument, however, requires an additional twist to handle trees with large size whose tiles have density $\sim \delta$, but whose tops have density much less than δ. Here we use an organization of the tiles that admits a more straightforward argument. This organization is carried out in Section 9, which contains more discussion as well.

Regarding the second point: A rather simple observation allows us to appeal to a key ingredient in the proof of the maximal theorem, rather than the theorem itself. This strengthens estimates on $\sum_{T\in\mathcal{T}} |\mathbf{top}(T)|$ for trees as mentioned in the last paragraph. This observation allows us to obtain the full range of p. This observation uses the proof of [2], and hence does not even take advantage of the sharp L^p estimates on the maximal operator obtained in [3]. See Lemma 6.2.

2.3. Organization of paper

Readers familiar with time-frequency analysis, having a bit of faith, and wanting an executive summary should follow this outline: Skip to the definition of the model operator in Section 3.4. Then (possibly after skimming Section 4 to review essentially standard definitions,) read Sections 5, 6, and 9. Those wanting to check the numerology should also read Section 7. A comprehensive outline is below.

In Section 3, we reduce the theorem to an analogous one for a model operator.

In Section 4, we present some key definitions needed for the organization of our set of tiles. (Recall that the operators in question are model sums over tiles.)

In Section 5, we make the main decomposition of the collection of tiles and state several key estimates that follow from the decomposition.

In Section 6, we state the main lemmas needed to prove the estimates stated in Section 5.

In Section 7, we balance these various estimates to prove the main theorem. There is no serious content here.

In Section 8, we prove the density lemma, which estimates $\sum_{T\in\mathcal{T}}|\mathbf{top}(T)|$ for certain collections $\mathcal{T}$ by using elementary covering ideas.

In Section 9, we prove the maximal estimate, which controls $\sum_{T\in\mathcal{T}}|\mathbf{top}(T)|$ for certain collections $\mathcal{T}$ by using more sophisticated techniques in combination with L^p and BMO-type estimates on a square function related to the "projection" operator associated to trees.

In Section 10, we compare the **size** of a tree to its intersection with the function in the definition of **size**.

In Section 11, we prove the tree lemma, which controls the contribution to the model sum from one tree. The proof mirrors that of the (more) classical one-dimensional tree lemma, with a small bit of extra work required to handle two-dimensional tail terms.

In Section 12, we prove the size lemma, which estimates $\sum_{T\in\mathcal{T}}|\mathbf{top}(T)|$ for certain collections $\mathcal{T}$ by using orthogonality.

In Section 13, we prove a refined Bessel inequality that allows us to control tail terms in the size and tree lemmas, as well as in the proof of localized L^p estimates for the square function mentioned above.

In Section 14, we prove localized (to the top of a tree) L^p estimates for a square function associated to a tree. Once again, we follow a relatively standard argument and appeal to the refined Bessel inequality to handle some two-dimensional technicalities.

In Section 15, we prove that higher L^p norms of the square function are controlled by lower L^p norms by using standard BMO techniques.

In the appendix, we recall the proof in [6] of the L^p, $p > 2$, case of our main theorem.

Acknowledgments. The author thanks Ciprian Demeter and Christoph Thiele for helpful discussions, and especially Christoph Thiele for making many comments on an early version of this paper. The author also thanks Francesco Di Plinio for pointing out an important typo. Thanks also to Shaoming Guo for catching a typo. And lastly, thanks to the referee for helpful suggestions.

3. Reductions

In this section we reduce the L^p estimates in Theorem 2.1 to restricted weak-type estimates on a model operator. The model operator should look familiar to readers familiar with developments in time-frequency analysis from the last ten to fifteen years: it is a sum over "tiles" of wave packets. The model operator arises from decomposing

1. the Hilbert kernel $1/t$ into (smoothly cutoff) dyadic intervals on the frequency side; for technical reasons we make these annuli rather thin, resulting in two summation indices for the Hilbert kernel. In fact, we actually decompose the projection operator into positive frequencies, and write the Hilbert transform as a linear combination of this operator and the identity operator.
2. Given any integer $l \geq 0$, $\hat{f}$ on τ into $\sim 2^l$ pieces; again, the "$\sim$" here comes from another summation introduced to provide strict orthogonality between the various pieces.

3.1. Discretizing the kernel

In this section we decompose the operator $H \circ \Pi_\tau$ into a sum of model operators.

We begin by selecting a Schwartz function $\psi_0^{(0)}$ such that $\psi_0^{(0)}$ is supported on $[98/100, 102/100]$ and equal to 1 on $[99/100, 101/100]$. Let $\psi_l^{(0)}(t) = \psi^{(0)}(2^l t)$. Now define $\psi^{(0)} = \sum_{l \in \mathbb{Z}} \psi_l^{(0)}$. By appropriately defining $\psi_0^{(i)}$ with similarly sized support, and defining $\psi_l^{(i)}(t) = \psi_0^{(i)}(2^l t)$, we can construct a partition of unity for $\mathbb{R}^+$; i.e.,

$$\mathbf{1}_{(0,\infty)} = \sum_{i=0}^{99} \psi^{(i)}.$$

This gives us the Hilbert kernel as a linear combination of 100 model kernels and the identity. More precisely, let

$$H_l^{(i)} g(x,y) = \int \check{\psi}_l^{(i)}(t)\, g(x-t, y-tu(x))\, dt.$$

Then writing I for the identity operator,

$$c_1 H \circ \Pi_\tau f(x,y) + I \circ \Pi_\tau f(x,y) = c_2 \sum_{l \in \mathbb{Z}} \sum_{i=0}^{99} H_l^{(i)} \circ \Pi_\tau f(x,y).$$

By the triangle inequality, we have

$$\|H \circ \Pi_\tau f\|_p \lesssim \|I \circ \Pi_\tau f\|_p + \sum_{i=0}^{99} \|H^{(i)} \circ \Pi_\tau f\|_p,$$

where $H^{(i)} = \sum_l H_l^{(i)}$. We note that $H_l \circ \Pi_\tau f = 0$ for $l \leq \log(1/w) + c$, because of the Fourier support of the kernel of the operator H_l.

3.2. Discretizing the function

We next focus on discretizing the function f. For $l \geq 0$, we write $\mathcal{D}_l$ to denote the collection of dyadic intervals of length 2^{-l} contained in $[-2,2]$. Fix a smooth positive function $\beta\colon \mathbb{R} \to \mathbb{R}$ such that $\beta(x) = 1$ for $x \in [-1,1]$ and such that $\beta(x) = 0$ when $|x| \geq 2$. Also assume that $\sqrt{\beta}$ is a smooth function. This point will become relevant for the definition of φ immediately before Lemma 3.1. Now fix an integer c (whose exact value is unimportant) and for each $\omega \in \mathcal{D}_l$, define

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

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

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

Note that

$$\beta_l(x + 2^{-l}) = \beta_l(x)$$

for $x \in [-2, 2 - 2^{-l}]$. Now define

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

Because of the local periodicity mentioned above, we have that $\gamma_l(x)$ is constant for $x \in [-1,1]$; say $\gamma_l(x) = \delta$, where δ is a constant independent of l. Hence

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

Define yet another multiplier $\tilde{\beta}\colon \mathbb{R} \to \mathbb{R}$ with support in $[1/2, 5/2]$, and $\tilde{\beta}(x) = 1$ for $x \in [1,2]$. Just as γ_l is an average over translates of β_l, so each $H^{(i)}$ is an average of model operators. We define the corresponding multipliers on $\mathbb{R}^2$:

$$\widehat{m_\omega}(\xi,\eta) = \tilde{\beta}(\eta)\,\beta_\omega\Big(\frac{\xi}{\eta}\Big)$$
$$\widehat{m_{l,t}}(\xi,\eta) = \tilde{\beta}(\eta)\,\beta_l\Big(t + \frac{\xi}{\eta}\Big)$$
$$\widehat{\mathfrak{m}_l}(\xi,\eta) = \tilde{\beta}(\eta)\,\gamma_l\Big(\frac{\xi}{\eta}\Big).$$

We know that for each l,

$$\mathfrak{m}_l(\xi,\eta)\,\mathbf{1}_\tau(\xi,\eta) = \mathbf{1}_\tau(\xi,\eta)$$

for $(\xi,\eta) \in \tau$. Note that for each i,

$$\begin{aligned}\|H^{(i)}(\Pi_\tau \circ f)\|_p &= \Big\| \sum_l (H_l^{(i)} \circ \Pi_\tau)\,(\tfrac{1}{\delta}\,\mathfrak{m}_l * f)\Big\|_p \\ &= \Big\| \frac{1}{2}\int_{-1}^{1} \sum_l (H_l^{(i)} \circ \Pi_\tau)\,(\tfrac{1}{\delta}\,m_{l,t} * f)dt\Big\|_p \\ &\le \frac{1}{2}\int_{-1}^{1} \Big\| \sum_l (H_l^{(i)} \circ \Pi_\tau)\,(\tfrac{1}{\delta}\,m_{l,t} * f)\Big\|_p dt,\end{aligned}$$

so it is enough to consider the discretized projections $m_{l,t}$. In what follows, we will assume, without loss of generality, that $t = 0 = i$ and omit the dependence on t and i.

3.3. Constructing the tiles

For each $\omega \in \mathcal{D}$ with $l \geq 0$, let $\mathcal{U}_\omega$ be a partition of $\mathbb{R}^2$ by parallelograms of width w and length $w/|\omega|$ whose long side has slope θ, where $\tan\theta = c(\omega)$ and where $c(\omega)$ is the center of the interval ω, and whose projection onto the x-axis is a dyadic interval. We remark that $l < 0$ need not be considered. (See the remark immediately prior to Section 3.2. Note that the index l plays a slightly different role there.) Briefly, the parts of the Hilbert kernel whose frequency support is outside the interval $[-1/w, 1/w] \subseteq \mathbb{R}$ ((i.e., ψ_l for $l < \log(1/w)$) have no interaction with our function f whose frequency support is contained in the annulus of radius $1/w$. Finally, let

$$\mathcal{U} = \bigcup_{\omega \in \mathcal{D}} \mathcal{U}_\omega.$$

If $s \in \mathcal{U}_\omega$, we will write $\omega_s := \omega$.

An element of $\mathcal{U}$ is called a "tile". The following lemma, stated in essentially this form in [6], allows us to further discretize our operator into a sum over tiles. Let R_ω denote an element of $\mathcal{U}_\omega$ containing the origin. Suppose φ_ω is such that $|\widehat{\varphi_\omega}|^2 = \widehat{m_\omega}$. Note that φ_ω is smooth, by our assumption on the function β mentioned above. Further, each region

$$\Big\{(\xi, \eta) \colon \frac{\xi}{\eta} \in \omega, \quad \eta \in [1, 2]\Big\}$$

can be obtained by a linear transformation of the trapezoid with corners $(-1, 1)$, $(1, 1)$, $(-2, 2)$, and $(2, 2)$, which ensures that the functions φ_ω, with $\omega \in \mathcal{D} := \cup_{l \geq 0} \mathcal{D}_l$, satisfy uniform decay conditions. To see this, consider the transformations

$$A = \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix}, \quad B = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad C = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}.$$

A composition of these three takes the trapezoid bounded by $(-1, 1)$, $(1, 1)$, $(-2, 2)$, $(2, 2)$ to the trapezoid bounded by $(M(\epsilon+\lambda), M)$, $(M(-\epsilon+\lambda), M)$, $(2M(\epsilon+\lambda), 2M)$, $(2M(-\epsilon+\lambda), 2M)$, which is precisely the area of support for φ_ω when M, ϵ, and λ are chosen appropriately. Define

$$\varphi_s(p) = \sqrt{|s|}\, \varphi_\omega(p - c(s)).$$

Note that the functions m_ω are L^1 normalized, so the functions φ_s are L^2 normalized.

Lemma 3.1. *Using notation above, we have*

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

Proof. We compute directly:

$$\begin{aligned} f * m_\omega(x) &= \int_{z\in\mathbb{R}^2} f(z) \int_{p\in\mathbb{R}^2} \overline{\varphi_\omega(p)}\, \varphi_\omega(p+x-z)\, dp\, dz \\ &= \int_{z\in\mathbb{R}^2} f(z) \sum_{s\in\mathcal{U}_\omega} \int_{p\in s} \overline{\varphi_\omega(p+z)}\, \varphi_\omega(p+x)\, dp\, dz \\ &= \sum_{s\in\mathcal{U}_\omega} \int_{p\in s} \int_{z\in\mathbb{R}^2} f(z)\, \overline{\varphi_\omega(p+z)}\, dz\, \varphi_\omega(p+x)\, dp \\ &= \sum_{s\in\mathcal{U}_\omega} \int_{p\in s} \langle f, \varphi_\omega(p+\cdot)\rangle \varphi_\omega(p+x)\, dp \\ &= \sum_{s\in\mathcal{U}_\omega} \frac{1}{|R_\omega|} \int_{p\in R_\omega} \langle f, \varphi_s(p+\cdot)\rangle\, \varphi_s(p+x)\, dp \\ &= \lim_{N\to\infty} \frac{1}{4N^2} \int_{[-N,N]^2} \sum_{s\in\mathcal{U}_\omega} \langle f, \varphi_s(p+\cdot)\rangle\, \varphi_s(p+x)\, dp. \end{aligned}$$

To see the last equality, note that the integrand is periodic in p, and the error (which arises from the fact that $[-N,N]^2$ will not exactly agree with the boundaries of the tiles s) goes to zero as $N \to \infty$. □

This lemma allows us to conclude (using the dominated convergence theorem) that

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

This allows us to restrict attention to the model operator that we define shortly. Define

$$\psi_s = \psi_{\log(\mathrm{length}(s))}$$

and

$$\phi_s(x_1, x_2) = \int \check{\psi}_s(t)\, \varphi_s(x_1 - t, x_2 - tv(x))\, dt.$$

We record the following fact for use in the proof of the tree lemma in Section 11.2.2.

Lemma 3.2. *We have $\phi_s(x) = 0$ unless $v(x) \in \omega_{s,2}$.*

Proof. Use Plancherel's theorem and the Fourier supports of ψ_s and φ_s. □

3.4. The model operator

We can finally define our model operator:

$$\mathbf{C}f = \sum_{s\in\mathcal{U}} \langle f, \varphi_s\rangle\, \phi_s.$$

For readers following the executive summary: φ_s is a standard wave packet adapted to the tile s, and ϕ_s is the appropriate scale of the Hilbert transform acting on φ_s.

A good mental shortcut is to imagine $\phi_s(x) = \varphi_s(x)\mathbf{1}_{\omega_{s,2}}(u(x))$, an expression quite similar to one appearing in the Lacey–Thiele proof of Carleson's theorem. By Lemma 3.1, each operator $H^{(i)}$ is an average of models of the form $\mathbf{C}$. Hence it is enough to prove the following theorem.

Theorem 3.3. *With* $\mathbf{C}$ *defined as above, and* $p \in (1, \infty)$*, we have*

$$\|\mathbf{C}f\|_p \lesssim \|f\|_p. \tag{3.1}$$

By appealing to restricted weak-type interpolation, it suffices to prove

$$|\langle \mathbf{C}\mathbf{1}_F, \mathbf{1}_E\rangle| \lesssim |E|^{1-1/p}\,|F|^{1/p}$$

for arbitrary $E, F \subseteq \mathbb{R}^2$ and $p \in (1, \infty)$. Of course by the triangle inequality it suffices to prove the following inequality:

$$\sum_{s\in\mathcal{S}} |\langle \mathbf{1}_F, \varphi_s\rangle\langle \mathbf{1}_E, \phi_s\rangle| \lesssim |E|^{1-1/p}\,|F|^{1/p}$$

for any $p \in (1, \infty)$, any $E, F \subseteq \mathbb{R}^2$, and any finite $\mathcal{S} \subseteq \mathcal{U}$. This is our task for the rest of the paper. Lacey and Li have already proved this estimate for arbitrary vector fields when $p \geq 2$. We discuss this proof in the appendix. Note that for $p \leq 2$, we have

$$|E|^{1-1/p}\,|F|^{1/p} = |E|^{1/2}\,|F|^{1/2}\Big(\frac{|F|}{|E|}\Big)^{1/p-1/2} \gtrsim |E|^{1/2}\,|F|^{1/2}$$

whenever $|F| \gtrsim |E|$ because $1/p - 1/2 > 0$. Hence our estimate is already proved when $|F| \gtrsim |E|$, so we restrict attention to the case $|F| \leq c|E|$ for some small constant c.

4. Key definitions

Definition 4.1. Given a parallelogram R, we write CR to denote the parallelogram with the same center as R but dilated by a factor of C.

Definition 4.2. Given two parallelograms R_1 and R_2 in $\mathcal{U}$, we will write $R_1 \leq R_2$ whenever $R_1 \subseteq CR_2$ and $\omega_{R_2} \subseteq \omega_{R_1}$.

Recall that ω_R is defined in Section 3.3. The exact value of C in the last Definitions is not important: 10 is enough. We need that if $R_1 \cap R_2 \neq \emptyset$ and $\omega_{R_1} \subseteq \omega_{R_2}$, then $R_2 \leq R_1$.

Definition 4.3. A *tree* is a collection T of parallelograms with a top parallelogram, denoted $\mathbf{top}(T)$, with $\mathbf{top}(T) \in \mathcal{U}$, such that for all $s \in T$, we have $s \leq \mathbf{top}(T)$. A tree T is a j-tree if $\omega_{\mathbf{top}(T)} \cap \omega_{s,j} = \emptyset$. Given a tree T, we will write T_j to denote the maximal j-tree contained in T.

Recall that $\omega_{s,1}$ is the right half of ω_s and $\omega_{s,2}$ is the left half. The following definitions will help us organize our collections of tiles. Recall that our vector field v is defined on a set E; this set plays a role in the definitions of **dense** and $\overline{\textbf{dense}}$ below. Similarly, the definition of **size** depends on our other set F.

For $x \in \mathbb{R}^2$, let

$$\chi(x) = \frac{1}{1 + |x|^{100}}.$$

For any parallelogram s, let $\chi_s^{(p)}$ be an L^p normalized version of χ adapted to the parallelogram s.

Definition 4.4. For a parallelogram s and a collection of parallelograms $\mathcal{S}$, define the following:

$$E_s = \{(x, y) \in E \colon u(x) \in \omega_s\}$$

$$\textbf{dense}(s) = \int_{E_s} \chi_s^{(1)}$$

$$\overline{\textbf{dense}}(s) = \sup_{s' \geq s, s' \in \mathcal{U}} \textbf{dense}(s')$$

$$\textbf{size}(\mathcal{S}) = \sup_{\text{1-trees } T \subseteq \mathcal{S}} \left(\frac{1}{|\textbf{top}(T)|} \sum_{s \in T} |\langle \mathbf{1}_F, \varphi_s \rangle|^2 \right)^{1/2}.$$

We remark that the function χ is needed for density since the wave packets φ_s have Schwartz tails. See the proofs of the tree and density lemmas. The extra technicality involved in defining $\overline{\textbf{dense}}$ (as opposed to just **dense**) is needed for our proof of the tree lemma (just as it is in the one-dimensional theory of [8]). The cost is rather high: a density estimate (see Estimate 5.4 below) is still easily obtainable, but the maximal estimate becomes much more difficult to prove. If $\overline{\textbf{dense}}(s)$ were equal to $\textbf{dense}(s)$ for every tile s, then the tops of the trees constructed in Section 5 are already prepared for an application of maximal technology. Unfortunately this is not the case, and this difficulty prompts our consideration of the collections $\mathcal{R}_j$ in Section 9. See also the delicate sorting algorithm in Lacey–Li [7], where the authors wrestle with the same issue.

5. Organization

In this section we carry out the main decomposition of the collection of tiles. We sort a given collection of tiles into subsets of tiles of approximately constant density, and further into trees of approximately constant size. The relevance of trees is shown in the following:

Lemma 5.1 (Tree lemma). *Let T be a tree. Suppose* $\overline{\textbf{dense}}(T) \leq \delta$. *Suppose* $\textbf{size}(T) \leq \sigma$. *Then*

$$\sum_{s \in T} |\langle \mathbf{1}_F, \varphi_s \rangle \langle \mathbf{1}_E, \phi_s \rangle| \lesssim \delta\sigma\, |\textbf{top}(T)|.$$

This is the "Tree Lemma" from [6], which is the two-dimensional version of the similar lemma in [8]. We prove it in Section 11. It reduces (3.4) to proving, for each $0 < \epsilon < 1$,

$$\sum_{\delta} \sum_{\sigma} \sum_{T \in \mathcal{T}_{\delta,\sigma}} \delta\sigma |\mathbf{top}(T)| \lesssim |F|^{1-\epsilon} |E|^{\epsilon}.$$

We can already prove this with the Estimates 5.3, 5.4, and 5.5 (appearing in the next lemma) and some bookkeeping – this is carried out in Section 7.

Lemma 5.2 (Organizational Lemma). *Let $\mathcal{S}$ be a finite collection of tiles. Then there exist a partition of $\mathcal{S}$ into trees $\mathcal{T}_{\delta,\sigma}$ where δ and σ are dyadic with $\delta \lesssim 1$, (i.e., $\mathcal{S} = \bigcup_{\delta,\sigma} \bigcup_{T \in \mathcal{T}_{\delta,\sigma}} T$) such that the following estimates hold:*

Estimate 5.3. (Orthogonality)

$$\sum_{T \in \mathcal{T}_{\delta,\sigma}} |\mathbf{top}(T)| \lesssim \frac{|F|}{\sigma^2}.$$

Estimate 5.4. (Density)

$$\sum_{T \in \mathcal{T}_{\delta,\sigma}} |\mathbf{top}(T)| \lesssim \frac{|E|}{\delta}.$$

Estimate 5.5. (Maximal) *For any $\epsilon > 0$,*

$$\sum_{T \in \mathcal{T}_{\delta,\sigma}} |\mathbf{top}(T)| \lesssim \frac{|F|^{1-\epsilon} |E|^{\epsilon}}{\delta \sigma^{1+\epsilon}}.$$

Remark 5.6. In fact we can take $\sigma \lesssim 1$, which we need (and prove) in the appendix.

In the remainder of this section we construct the collections of trees $\mathcal{T}_{\delta,\sigma}$. In the following sections we prove the estimates above. Estimate 5.3 follows from the construction of the trees $\mathcal{T}_{\delta,\sigma}$, and the proof of the standard **size** lemma; we give a proof in Section 12. We prove Estimates 5.4 and 5.5 in Section 9. We remark that we make these claims about the same family of trees. This is in contrast to [8], [6], and [7], in which the argument has the form "There exists a family $\mathcal{T}_{\text{size}}$ such that $\mathcal{S}_\delta = \cup_{T \in \mathcal{T}_{\text{size}}} T$ and such that the size estimate holds for the collection $\mathcal{T}_{\text{size}}$; further there is a (potentially different!) family $\mathcal{T}_{\text{density}}$ such that $\mathcal{S}_\delta = \cup_{T \in \mathcal{T}_{\text{density}}} T$ and such that the density estimate holds for the collection $\mathcal{T}_{\text{density}}$."

First, we sort the tiles by density: Let

$$\mathcal{S}_\delta = \{s \in \mathcal{S} \colon \overline{\mathbf{dense}}(s) \in (\tfrac{1}{2}\delta, \delta]\}$$

for dyadic δ. By the definition of **dense**, we need only consider $\delta \leq \|\chi\|_1 \lesssim 1$.

We next sort each collection $\mathcal{S}_\delta$ into families of trees with comparable size. The following algorithm is a slight variant of the sorting algorithm used in [8] and in [6]. We want to ensure that $\mathbf{top}(T) \in T$ for each tree T in our construction. There are some small technicalities that arise in the two-dimensional situation due to the nontransitivity of the relation "$\leq$". Without loss of generality, we may assume our collection of tiles $\mathcal{S}$ is finite, so we know there exists $\sigma_{\max}$ such that $\mathbf{size}(\mathcal{S}) \leq \sigma_{\max}$ for every $T \subseteq \mathcal{S}_\delta$. This gives us a starting point for the following lemma.

Lemma 5.7. *Let $\mathcal{S}$ be a collection of tiles satisfying $\mathbf{size}(\mathcal{S}) < \sigma$. Then there exists a disjoint collection of trees $\mathcal{T}_\sigma$ such that for all $T \in \mathcal{T}_\sigma$, we have $\mathbf{top}(T) \in T$, and*

$$\mathbf{size}\Big(\mathcal{S} \setminus \bigcup_{T \in \mathcal{T}_\sigma} T\Big) < \frac{\sigma}{2}.$$

Finally, we have the estimate

$$\sum_{T \in \mathcal{T}_\sigma} |\mathbf{top}(T)| \lesssim \frac{|F|}{\sigma^2}, \tag{5.1}$$

where here F is the set used in the definition of **size***.*

Remark 5.8. Having $\mathbf{top}(T) \in T$ will be helpful in Section 9. See in particular the construction of the rectangles R_T and the collections $\mathcal{T}_R$.

Proof. Initialize

$$\begin{aligned} \text{STOCK} &= \mathcal{S} \\ \mathcal{T}_\sigma &= \emptyset. \end{aligned}$$

In the following scheme we write C to denote the constant used in the definition of a tree (see Definition 4.3), which we assume is somewhat large. While there is a 1-tree $T \subseteq \text{STOCK}$ with

$$\sqrt{\frac{1}{\mathbf{top}(T)} \sum_{s \in T} |\langle \mathbf{1}_f, \varphi_s \rangle|^2} \geq \frac{\sigma}{C}$$

and with $\mathbf{top}(T) \in T$, choose T with $c(\omega_{\mathbf{top}(T)})$ most clockwise, let $\tilde{T}$ be the maximal tree with top equal to $\mathbf{top}(T)$, and update

$$\begin{aligned} \text{STOCK} &:= \text{STOCK} \setminus \tilde{T} \\ \mathcal{T}_\sigma &:= \mathcal{T}_\sigma \cup \{\tilde{T}\}. \end{aligned}$$

(Again, we write $c(\omega_{\mathbf{top}(T)})$ to denote the center of $\omega_{\mathbf{top}(T)}$.)

Remark 5.9. We remark that our choice of $c(\omega_{\mathbf{top}(T)})$ as the most clockwise will be used in the proof of Estimate 5.1 in Section 12. See specifically Claim 12.2.

When no such trees remain, we have the collection of trees $\mathcal{T}_\sigma$ described in the statement of the lemma. By construction we see that $\mathbf{top}(\tilde{T}) \in \tilde{T}$ and that $\mathbf{size}(\tilde{T}) \geq \sigma/C$ for each $\tilde{T} \in \mathcal{T}_\sigma$. The estimate (5.1) follows rather standard arguments; we present the proof in Section 12. It remains to prove the following:

Claim 5.10.

$$\mathbf{size}\,(\text{STOCK}) < \frac{\sigma}{2}.$$

Consider a tree $T \subseteq \textsc{stock}$. Without loss of generality, T is a 1-tree (since the definition of **size** only takes into consideration 1-tree subtrees of T anyway). We will partition T into a collection $\mathcal{T}_T$ of subtrees of T, each of which contains its top, as follows: Initialize

$$\textsc{pantry} := T$$
$$T_{\max} := \emptyset.$$

While $\textsc{pantry}$ is nonempty, choose a tile t of maximal length in $\textsc{pantry}$, let T_t be the maximal subset of $\textsc{pantry}$ such that $s \le t$ for $s \in T_t$, and update

$$\textsc{pantry} := \textsc{pantry} \setminus T_t$$
$$T_{\max} := T_{\max} \cup \{t\}.$$

It is clear that this construction exhausts all of T; i.e., eventually $\textsc{pantry}$ becomes empty. Since the tiles $t \in T_{\max}$ all satisfy $\omega_{\mathbf{top}(T)} \subseteq \omega_t$, and since each is maximal with respect to "$\le$", we know these tiles are pairwise disjoint. On the other hand, they are all contained in $C\mathbf{top}(T)$, and $t = \mathbf{top}(T_t)$, so

$$\sum_{t \in T_{\max}} |\mathbf{top}(T_t)| \le C|\mathbf{top}(T)|.$$

Further, since each tree T_t for $t \in T_{\max}$ contains its top, we know

$$\sqrt{\frac{1}{\mathbf{top}(T)} \sum_{s \in T} |\langle \mathbf{1}_f, \varphi_s \rangle|^2} \le \frac{\sigma}{C},$$

for otherwise T_t would have been selected and put into $\mathcal{T}_\sigma$. Hence

$$\sum_{s \in T} |\langle f, \varphi_s \rangle|^2 = \sum_{t \in T_{\max}} \sum_{s \in T_t} |\langle f, \varphi_s \rangle|^2 \le \sum_{t \in T_{\max}} |\mathbf{top}(T_t)| \frac{\sigma^2}{C^2} \le \frac{\sigma^2 |\mathbf{top}(T)|}{C}.$$

This implies

$$\mathbf{size}(T) \le \frac{\sigma}{\sqrt{C}},$$

which proves the claim provided $C \ge 4$. □

By applying the lemma iteratively to each collection $\mathcal{S}_\delta$, we obtain collections $\mathcal{S}_{\delta,\sigma}$ and $\mathcal{T}_{\delta,\sigma}$ such that

$$\mathcal{S}_{\delta,\sigma} = \bigcup_{T \in \mathcal{T}_{\delta,\sigma}} T$$

where the union is disjoint, such that $\overline{\mathbf{dense}}(s) \sim \delta$ for $s \in \mathcal{S}_{\delta,\sigma}$, and such that

$$\mathbf{size}(T) \sim \sigma \sim \sqrt{\frac{1}{\mathbf{top}(T)} \sum_{s \in T} |\langle \mathbf{1}_f, \varphi_s \rangle|^2}$$

for $T \in \mathcal{T}_{\delta,\sigma}$. This proves Lemma 5.2, except for Estimates 5.4 and 5.5. Note that Estimate 5.3 follows from (5.1).

6. Main Lemmas

Here we present the main lemmas needed to prove Estimates 5.4 and 5.5.

Lemma 6.1. *Suppose $\mathcal{R}$ is a collection of pairwise incomparable (under "$\leq$") parallelograms of uniform width such that* $\mathbf{dense}(R) \geq \delta$ *for $R \in \mathcal{R}$. Then*

$$\sum_{R\in\mathcal{R}} |R| \lesssim \frac{|E|}{\delta}.$$

Lemma 6.1 is nothing more than the Density Lemma from [8] with straightforward modifications for the two-dimensional setting.

Lemma 6.2. *Suppose $\mathcal{R}$ is a collection of pairwise incomparable (under "$\leq$") parallelograms of uniform width such that for each $R \in \mathcal{R}$, we have*

$$\frac{|E \cap u^{-1}(\omega_R) \cap R|}{|R|} \geq \delta \tag{6.1}$$

and

$$\frac{1}{|R|}\int_R \mathbf{1}_F \geq \lambda. \tag{6.2}$$

Then for each $\epsilon > 0$,

$$\sum_{R\in\mathcal{R}} |R| \lesssim \frac{|F|}{\delta\lambda^{1+\epsilon}}.$$

The proof of Lemma 6.2 is contained in Section 3 of [2]. More specifically, see estimate (3.10) on page 959, as well as the construction of the collection of parallelograms called $\mathcal{R}_1$ there. Note that this last lemma requires an assumption of the form

$$\frac{1}{|R|}\int_R \mathbf{1}_F > \lambda;$$

on the other hand, our assumption on $T \in \mathcal{T}_{\delta,\sigma}$ is that $\mathbf{size}(T) \lesssim \sigma$ and

$$\Big(\frac{1}{|\mathbf{top}(T)|}\sum_{s\in T_1} |\langle \mathbf{1}_F, \varphi_s\rangle|^2\Big)^{1/2} \gtrsim \sigma,$$

where T_1 is the maximal 1-tree in T. The following lemma shows that the second kind of fact implies the first without much loss:

Lemma 6.3. *Let $F \subseteq \mathbb{R}^2$. Suppose T is a tree with* $\mathbf{size}(T) \lesssim \sigma$ *and*

$$\Big(\frac{1}{|\mathbf{top}(T)|}\sum_{s\in T_1} |\langle \mathbf{1}_F, \varphi_s\rangle|^2\Big)^{1/2} \gtrsim \sigma,$$

where T_1 is the maximal 1-tree in T. Then for any $\epsilon > 0$,

$$\frac{|\sigma^{-\epsilon}\,\mathbf{top}(T) \cap F|}{|\sigma^{-\epsilon}\,\mathbf{top}(T)|} \gtrsim \sigma^{1+\epsilon}.$$

Lemma 6.3 is proved in Section 10; it follows from L^p and BMO-type estimates on a square function related to the notion of **size**.

Estimate 5.5 deserves more prominent mention. An estimate in this spirit was proved in [7]. However here we have much better dependence on the parameter δ due to a rather simple observation. The argument in [7] follows essentially the argument of the density lemma, with an appeal to a maximal theorem to control $|\{M_\delta \mathbf{1}_F > \lambda\}|$. In our case of a vector field depending on only one variable, the relevant maximal operator was studied by the author in [2] and [3]. However this approach is inefficient. Instead of combining the density argument with a maximal function estimate (each of which costs in terms of $1/\delta$), we appeal to an argument made in [2], which directly estimates

$$\sum_{R \in \mathcal{R}} |R| \lesssim \frac{|F|}{\delta \lambda^{1+\epsilon}}$$

for any $\epsilon > 0$. In fact, this estimate was established en route to a covering lemma which implies the maximal theorem. Interestingly, the improved L^2 estimates established in [3], which interpolate to give improved L^p estimates, are unhelpful in this setting, precisely because they are estimates on the operator norm, rather than on a sum like the one appearing immediately above.

7. Balancing the estimates

In this section we carry out some computations which allow us to prove (3.4), and hence the main theorem. We now estimate

$$\sum_{\delta} \sum_{\sigma} \sum_{T \in \mathcal{T}_{\delta,\sigma}} \delta\sigma \, |\mathbf{top}(T)|.$$

We have two cases. Recall that E and F are sets with $|F| \leq |E|$.

7.1. Case 1: $\delta \geq |F|/|E|$

A quick computation shows that (up to additive $O(\epsilon)$ terms in the exponents)

- the maximal estimate is more efficient when $\sigma \geq |F|/|E|$;
- the density lemma is more efficient when $\sigma \leq |F|/|E|$.

Remark 7.1. The maximal estimate is more effective than the size estimate for $\delta \geq |F|/|E|$ and σ close to $|F|/|E|$. Without this, we would not be able to obtain L^p estimates for any $p < 2$.

For the first range, with δ fixed, we have for any $\epsilon > 0$

$$\sum_{\sigma \geq \frac{|F|}{|E|}} \sum_{T \in \mathcal{T}_{\delta,\sigma}} \delta\sigma |\mathbf{top}(T)| \lesssim \sum_{\sigma \geq \frac{|F|}{|E|}} \delta\sigma \frac{|F|^{1-\epsilon}|E|^{\epsilon}}{\delta\sigma^{1+\epsilon}} = |F|^{1-\epsilon}|E|^{\epsilon} \sum_{\sigma \geq \frac{|F|}{|E|}} \frac{1}{\sigma^{\epsilon}} \sim |F|^{1-2\epsilon}|E|^{2\epsilon}.$$

Summing this over dyadic $1 \gtrsim \delta \geq |F|/|E|$ gives us a total of $\lesssim |F|^{1-3\epsilon}|E|^{3\epsilon}$.

For the second range, with δ fixed, we have

$$\sum_{\frac{|F|}{|E|}\geq\sigma}\sum_{T\in\mathcal{T}_{\delta,\sigma}}\delta\sigma|\mathbf{top}(T)|\lesssim\sum_{\frac{|F|}{|E|}\geq\sigma}\delta\sigma\frac{|E|}{\delta}=\sum_{\frac{|F|}{|E|}\geq\sigma}\sigma|E|\sim|F|.$$

Again, summing this over dyadic $1\gtrsim\delta\geq|F|/|E|$ gives us a total of $\lesssim|F|^{1-\epsilon}|E|^{\epsilon}$.

7.2. Case 2: $\delta\leq|F|/|E|$

In this case, the size and density estimates alone will be enough for us. A quick computation shows that

- the size estimate is most efficient when $\sigma\geq\sqrt{\delta|F|/|E|}$;
- the density estimate is most efficient when $\sigma\leq\sqrt{\delta|F|/|E|}$.

We decompose our sum over σ into these two ranges. For the first range, we have

$$\sum_{\sigma\geq\sqrt{\delta|F|/|E|}}\delta\sigma\frac{|F|}{\sigma^2}=|F|\delta\sum_{\sigma\geq\sqrt{\delta|F|/|E|}}\frac{1}{\sigma}\lesssim\sqrt{|F||E|\delta}.$$

Summing over $\delta\leq|F|/|E|$ gives us a total of $\lesssim|F|\lesssim|F|^{1-\epsilon}|E|^{\epsilon}$, since $|F|\leq|E|$.

For the second range, we have

$$\sum_{\sigma\leq\sqrt{\delta|F|/|E|}}\delta\sigma\frac{|E|}{\delta}\sim|E|\sum_{\sigma\leq\sqrt{\delta|F|/|E|}}\sigma\sim\sqrt{|F||E|\delta}.$$

Once again, summing over $\delta\leq|F|/|E|$ gives us a total of $\lesssim|F|\lesssim|F|^{1-\epsilon}|E|^{\epsilon}$, since $|F|\leq|E|$.

This completes the proof of the main estimate (3.4) modulo the proofs of the lemmas, which are given in the following sections.

8. Density lemma

In this section we prove Lemma 6.1. Let $\mathcal{R}$ be as in the hypotheses of the lemma. For $k=0,1,2,\dots$, let $\mathcal{R}_k$ be the collection of $R\in\mathcal{R}$ such that

$$|u^{-1}(\omega_R)\cap 2^kR\cap E|\geq\frac{1}{100}\,\delta\,2^{20k}\,|2^kR|,$$

and such that k is the least integer with this property. Note $\mathcal{R}=\cup_k\mathcal{R}_k$, since if $R\in\mathcal{R}$ but $R\notin\cup_k\mathcal{R}_k$, then

$$\begin{aligned}\mathbf{dense}(R)&\leq\int_{E_R}\chi_R^{(1)}\leq\sum_{k=0}^{\infty}\left|u^{-1}(\omega_R)\cap 2^kR\cap E\right|2^{-100k}\,\frac{1}{|R|}\\&\leq\frac{1}{100}\,\frac{\delta}{|R|}\sum_{k=0}^{\infty}2^{25k}\,|R|\,2^{-100k}\leq\frac{\delta}{50}.\end{aligned}$$

We now run an iterative selection procedure to find a subset of $\mathcal{R}_k$ such that the parallelograms $2^k R$ are disjoint: Initialize

$$\begin{aligned} \text{STOCK} &= \mathcal{R}_k \\ \widetilde{\mathcal{R}_k} &= \emptyset. \end{aligned}$$

While STOCK $\neq \emptyset$, choose R with maximal length, let

$$\mathcal{A}_R = \{R' \in \text{STOCK} \colon 2^k R' \cap 2^k R \neq \emptyset \ \text{ and } \ \omega_{R'} \cap \omega_R \neq \emptyset\},$$

and update

$$\begin{aligned} \text{STOCK} :&= \mathcal{R}_k \setminus \mathcal{A}_R \\ \widetilde{\mathcal{R}_k} &= \widetilde{\mathcal{R}_k} \cup \{R\}. \end{aligned}$$

Note that the parallelograms in $\mathcal{A}_R$ are pairwise disjoint by the pairwise incomparability of parallelograms in $\mathcal{R}$, and because $\omega_{R'} \cap \omega_R \neq \emptyset$ for $R' \in \mathcal{A}_R$. Hence, using the definition of $\mathcal{R}_k$, we have

$$\begin{aligned} \sum_{R \in \mathcal{R}_k} |R| = \sum_{R \in \widetilde{\mathcal{R}_k}} \sum_{R' \in \mathcal{A}_R} |R'| &\lesssim 2^{2k} \sum_{R \in \widetilde{\mathcal{R}_k}} |R| \\ &\lesssim 2^{2k}\, 2^{-20k}\, \frac{1}{\delta} \sum_{R \in \widetilde{\mathcal{R}_k}} |u^{-1}(\omega_R) \cap 2^k R \cap E| \lesssim 2^{-18k}\, \frac{1}{\delta}\, |E|, \end{aligned}$$

where in the last inequality we have used the fact that the parallelograms $2^k R$ are pairwise incomparable, and that $\omega_R = \omega_{2^k R}$, so that the sets $\{u^{-1}(\omega_R) \cap 2^k R\}$ are disjoint. Finally, we sum over k to obtain the result.

9. Proofs of maximal and density estimates

We now look more closely at the collections $\mathcal{T}_{\delta,\sigma}$. For the remainder of this section we regard δ and σ as fixed. Notation in this section is understood to depend on both δ and σ. (So, for example, $\mathcal{T} = \mathcal{T}_{\delta,\sigma}$.) We begin by isolating a collection of tiles with density δ. First, let

$$\tilde{\mathcal{R}} = \{R \in \mathcal{U} \colon \mathbf{dense}(R) \sim \delta\}.$$

We now find a maximal subset of $\tilde{\mathcal{R}}$ whose elements are pairwise incomparable. Initialize:

$$\begin{aligned} \text{STOCK} &= \tilde{\mathcal{R}} \\ \mathcal{R} &= \emptyset. \end{aligned}$$

While STOCK $\neq \emptyset$, choose R of maximal length in STOCK. Define

$$\mathcal{A}_R = \{R' \in \text{STOCK} \colon R' \leq R\},$$

and update

$$\text{STOCK} = \text{STOCK} \setminus \mathcal{A}_R$$
$$\mathcal{R} = \mathcal{R} \cup \{R\}.$$

When the loop terminates, elements of $\mathcal{R}$ are pairwise incomparable (under $\leq$), and $\mathcal{R}$ is maximal with respect to this property.

Remark 9.1. Recall that for $T \in \mathcal{T}$ we have $\overline{\mathbf{dense}}(\mathbf{top}(T)) \sim \delta$, but maybe $\mathbf{dense}(\mathbf{top}(T))$ is much less than δ. This makes the maximal Lemma 6.2 unavailable to us. Note that several ingredients are required, and $\mathbf{top}(T)$ may lack the **dense** required. The work in this section is dedicated to organizing the trees in such a way that we can legitimately appeal to Lemma 6.2.

Next we associate to each tree $T \in \mathcal{T}$ a parallelogram $R_T \in \mathcal{R}$. This requires a few steps. Note that for each $s \in \cup_{T \in \mathcal{T}} T$, we have $\overline{\mathbf{dense}}(s) \sim \delta$. By Lemma 5.7, we know that $\mathbf{top}(T) \in T$ for each $T \in \mathcal{T}$. Hence $\overline{\mathbf{dense}}(\mathbf{top}(T)) \sim \delta$. This means there exists a parallelogram $\tilde{R} \in \tilde{\mathcal{R}}$ such that $\mathbf{dense}(R) \sim \delta$ and such that $\mathbf{top}(T) \leq \tilde{R}$. (This is the reason why it is convenient to have $\mathbf{top}(T) \in t$). Further, for each $\tilde{R} \in \tilde{\mathcal{R}}$, there is $R \in \mathcal{R}$ (again, possibly not unique) such that $\tilde{R} \leq R$. Hence we may assign to each $T \in \mathcal{T}$ some $R \in \mathcal{R}$, and there is $\tilde{R}$ such that $\mathbf{top}(T) \leq \tilde{R} \leq R$. (Of course there may be more than one R to choose from for each T; choose one!) Call this parallelogram R_T. Now for each $R \in \mathcal{R}$, define

$$\mathcal{T}_R = \{T \in \mathcal{T} \colon R_T = R\}.$$

By construction,

$$\mathcal{T} = \cup_{R \in \mathcal{R}} \mathcal{T}_R.$$

Our goal now is to control

$$\sum_{R \in \mathcal{R}} \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)|.$$

First, we shall show that for all $R \in \mathcal{R}$,

$$\sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \lesssim |R|.$$

The collection $\{\mathbf{top}(T) \colon T \in \mathcal{T}_R\}$ need not be pairwise disjoint, but we do have the following satisfactory substitute.

Claim 9.2. *There exists $\overline{\mathcal{T}_R} \subseteq \mathcal{T}_R$ such that $\{\mathbf{top}(T) \colon T \in \overline{\mathcal{T}_R}\}$ is pairwise disjoint and such that*

$$\sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \lesssim \sum_{T \in \overline{\mathcal{T}_R}} |\mathbf{top}(T)|.$$

Proof. Initialize

$$\text{STOCK} = \mathcal{T}_R$$
$$\overline{\mathcal{T}_R} = \emptyset.$$

While $\textsc{stock} \neq \emptyset$, choose $T \in \textsc{stock}$ such that $\mathbf{top}(T)$ is of maximal length. Then define

$$\mathcal{A}_T = \{T' \in \textsc{stock}\colon \mathbf{top}(T') \cap \mathbf{top}(T) \neq \emptyset\},$$

and update

$$\textsc{stock} := \textsc{stock} \setminus \mathcal{A}_T$$
$$\overline{\mathcal{T}_R} := \overline{\mathcal{T}_R} \cup \{T\}.$$

We stop when $\textsc{stock}$ is empty. By construction, the tops of the trees in $\overline{\mathcal{T}_R}$ are pairwise disjoint. Now we show that

$$\sum_{T' \in \mathcal{A}_T} |\mathbf{top}(T')| \leq C' |\mathbf{top}(T)|.$$

With this we will know that

$$\sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| = \sum_{T \in \overline{\mathcal{T}_R}} \sum_{T' \in \mathcal{A}_T} |\mathbf{top}(T')| \leq C' \sum_{T \in \overline{\mathcal{T}_R}} |\mathbf{top}(T)|.$$

Suppose not. Define $S = \cup_{T' \in \mathcal{A}_T} T_1'$, where for a tree T, T_1 is defined to be the maximal 1-tree contained in T. We claim S can be partitioned into a small number of trees S_j, $j = 1, \ldots, 10C^2$, with each a 1-tree. To see that they are 1-trees, suppose $s \in T' \in \mathcal{A}_T$. Then $\omega_{s,2} \supseteq \omega_{\mathbf{top}(T')} \supseteq \omega_{\mathbf{top}(T)}$, so $\omega_{s,1} \cap \omega_{\mathbf{top}(T)} = \emptyset$. To see that we only need a few trees, just note that for each $T' \in \mathcal{A}_T$, $\mathbf{top}(T') \subseteq C(\mathbf{top}(T))$. Then since each $s \in T'$ satisfies $s \subseteq C(\mathbf{top}(T'))$, we know that S can be partitioned into $\sim C^2$ subtrees S_j by considering (possibly overlapping) tiles in $C^2\,\mathbf{top}(T)$ of height w and length the same as length of $\mathbf{top}(T)$.

Hence,

$$\sum_{j=1}^{10C^2} \sum_{s \in S_j} |\langle f, \varphi_s \rangle|^2 \geq \sum_{T' \in \mathcal{A}_T} \sum_{s \in T_1'} |\langle f, \varphi_s \rangle|^2 \geq \frac{1}{4} \sum_{T' \in \mathcal{A}_T} \sigma^2 |\mathbf{top}(T')| \geq \sigma^2 \frac{C'}{4} |\mathbf{top}(T)|.$$

Provided C' is taken large enough (with respect to a universal constant C mentioned in Section 4), one of the trees S_j satisfies $\mathbf{size}(S_j) \geq 10\sigma$, which is impossible since the trees $T \in \mathcal{T}_R$ were chosen from a collection with **size** less than σ. This proves the second claim about $\overline{\mathcal{T}_R}$. □

9.1. Proof of the density estimate

We are already in position to prove Estimate 5.4. Note that the collection $\mathcal{R}$ constructed above is of pairwise incomparable parallelograms of uniform width and **dense** $\sim \delta$. Hence the previous claim, together with Lemma 6.1, implies

$$\sum_{R \in \mathcal{R}} \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \lesssim \sum_{R \in \mathcal{R}} |R| \lesssim \frac{|E|}{\delta}.$$

9.2. Proof of the maximal estimate

The proof of Estimate 5.5 is a bit more involved. For the rest of this section, fix $\epsilon > 0$. The first key step is to sort the parallelograms in $\mathcal{R}$ by how heavily they are covered by the trees in $\mathcal{T}_R$. Specifically, for integers $j \geq 0$, define

$$\mathcal{R}_j = \Big\{ R \in \mathcal{R}\colon \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \sim 2^{-j} |R| \Big\}.$$

Since our goal is to control

$$\sum_{R \in \mathcal{R}} \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \sim \sum_j \sum_{R \in \mathcal{R}_j} \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \sim \sum_j 2^{-j} \sum_{R \in \mathcal{R}_j} |R|,$$

it is enough to estimate $\sum_{R \in \mathcal{R}_j} |R|$ with suitable dependence on j.

In order to apply maximal technology (in the form of Lemma 6.2), we must find parallelograms R that heavily intersect F, and that also contain a large subset on which v points in the direction of R. Because of the Schwartz tails in the definition of **dense**, we do not know that each $R \in \mathcal{R}_j$ satisfies

$$|u^{-1}(\omega_R) \cap E \cap R| \gtrsim \delta\, |R|.$$

Rather, we know that

$$|u^{-1}(\omega_R) \cap E \cap 2^k R| \gtrsim 2^{20k}\, \delta\, |R| \tag{9.1}$$

for some integer $k \geq 0$, as in Section 8. Define $\mathcal{R}_{j,k}$ to be the set of $R \in \mathcal{R}_j$ such that condition (9.1) holds for R but such that it does not hold with any smaller k. Similarly, we cannot conclude that R itself intersects F heavily. Recall that Lemma 6.3 guarantees that F intersects $\sigma^{-\epsilon}\mathbf{top}(T)$ heavily, whenever $T \in \mathcal{T}_{\delta,\sigma}$; we cannot however, conclude that F intersects $\mathbf{top}(T)$ itself. This causes some minor differences in the treatment of the cases $2^k \geq \sigma^{-\epsilon}$ and $2^k \leq \sigma^{-\epsilon}$ that the reader should not take too seriously. It suffices then to control sums like

$$\sum_{R \in \mathcal{R}_{j,k}} |R|$$

with suitable dependence on k and j.

9.2.1. Case 1: $2^k \geq \sigma^{-\epsilon}$. We want to apply Lemma 6.2 to the collection $\mathcal{R}_{j,k}$. The defining condition of $\mathcal{R}_{j,k}$ gives us the kind of information needed by the hypothesis (6.1). The following claim gives us the kind of information needed by the hypothesis (6.2).

Claim 9.3. *For $R \in \mathcal{R}_{j,k}$*

$$\frac{|F \cap 2^k R|}{|2^k R|} \gtrsim 2^{-j}\, \sigma^{1+3\epsilon} \Big(\frac{\sigma^{-\epsilon}}{2^k} \Big)^2.$$

We postpone the proof of the claim until the end of this section. With the claim, the only ingredient still needed to apply Lemma 6.2 is the pairwise incomparability of the parallelograms in question. We arrange this with the usual type of sorting algorithm. Initialize

$$\textsc{stock} = \mathcal{R}_{j,k}$$
$$\widetilde{\mathcal{R}_{j,k}} = \emptyset.$$

While $\textsc{stock} \neq \emptyset$, choose R with maximal length, let

$$\mathcal{A}_R = \{R' \in \textsc{stock} : 2^k R' \cap 2^k R \neq \emptyset \text{ and } \omega_{R'} \cap \omega_R \neq \emptyset\},$$

and update

$$\textsc{stock} := \mathcal{R}_{j,k} \setminus \mathcal{A}_R$$
$$\widetilde{\mathcal{R}_{j,k}} = \widetilde{\mathcal{R}_{j,k}} \cup \{R\}.$$

(Note $\omega_R = \omega_{CR}$ for any C). Since the parallelograms $R' \in \mathcal{A}_R$ are pairwise incomparable, we know they are in fact disjoint (see earlier in Section 9 for a similar argument), so

$$\sum_{R' \in \mathcal{A}_R} |R'| \lesssim |2^k R|.$$

Hence

$$\sum_j \sum_k \sum_{R \in \mathcal{R}_{j,k}} \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \lesssim \sum_j \sum_k \sum_{R \in \mathcal{R}_{j,k}} 2^{-j}|R|$$
$$\lesssim \sum_j \sum_k \sum_{R \in \widetilde{\mathcal{R}_{j,k}}} \sum_{R' \in \mathcal{A}_R} 2^{-j}|R'| \lesssim \sum_j \sum_k \sum_{R \in \widetilde{\mathcal{R}_{j,k}}} 2^{-j}|2^k R|.$$

We now focus our attention on

$$\sum_{R \in \widetilde{\mathcal{R}_{j,k}}} 2^{-j}|2^k R|.$$

Claim 9.3 together with the defining condition for parallelograms in $\mathcal{R}_{j,k}$ allows us to apply Lemma 6.2, with "δ" in (6.1) being $2^{20k}\delta$ and "λ" in (6.2) being $2^{-j}\,2^{-2k}\,\sigma^{1+O(\epsilon)}$, as in Claim 9.3. The huge gain in k from (9.1) allows us to sum the contributions from the various $\mathcal{R}_{j,k}$. More specifically, Lemma 6.2 yields

$$\sum_{R \in \widetilde{\mathcal{R}_{j,k}}} |2^k R| \lesssim \frac{1}{2^{20k}\,\delta}\,\frac{|F|}{(\sigma^{1+\epsilon}\,2^{-2k}\,2^{-j})^{1+\epsilon}}$$

This obviously sums in k to prove

$$\sum_{R \in \mathcal{R}_j} \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \lesssim \sum_{R \in \mathcal{R}_j} 2^{-j}|R| \lesssim \frac{1}{\delta}\,\frac{2^{\epsilon j}|F|}{(\sigma^{1+\epsilon})^{1+\epsilon}};$$

this estimate is effective for small j. Estimate 5.4 tells us that for any j,

$$\sum_{R \in \mathcal{R}_j} \sum_{T \in \mathcal{T}_R} |\mathbf{top}(T)| \lesssim \sum_{R \in \mathcal{R}_j} 2^{-j}\,|R| \lesssim 2^{-j}\,\frac{|E|}{\delta};$$

this estimate is effective for large j.

It remains to balance these two estimates:

$$\sum_{j\geq 0}\sum_{R\in\mathcal{R}_j}\sum_{T\in\mathcal{T}_R}|\mathbf{top}(T)| = \sum_{j\leq \log\frac{|E|\sigma}{|F|}}\sum_{R\in\mathcal{R}_j} 2^{-j}|R| + \sum_{j\geq \log\frac{|E|\sigma}{|F|}}\sum_{R\in\mathcal{R}_j} 2^{-j}|R|$$

$$\lesssim \sum_{j\leq \log\frac{|E|\sigma}{|F|}} 2^{\epsilon j}\frac{|F|}{\delta\,\sigma^{(1+\epsilon)^2}} + \sum_{j\geq \log\frac{|E|\sigma}{|F|}} 2^{-j}\frac{|E|}{\delta} \lesssim \frac{|F|^{1-\epsilon}\,|E|^{\epsilon}}{\delta\,\sigma^{(1+\epsilon)^2}} \lesssim \frac{|F|^{1-5\epsilon}\,|E|^{5\epsilon}}{\delta\,\sigma^{1+5\epsilon}}.$$

Remark 9.4. Of course the first sum above is empty when $\sigma \leq |F|/|E|$; in this case we recover the density estimate. Recalling Section 7, we see that for this range of σ we have no need for the maximal estimate anyway.

This completes the proof of the maximal estimate, except for the proof of Claim 9.3, to which we turn now.

Proof of Claim 9.3. For each $T \in \overline{\mathcal{T}_R}$, Lemma 6.3 tells us that

$$\frac{|\sigma^{-\epsilon}\,\mathbf{top}(T)\cap F|}{|\sigma^{-\epsilon}\,\mathbf{top}(T)|} \geq \sigma^{1+\epsilon}.$$

One minor technical problem is that the parallelograms $\sigma^{-\epsilon}\,\mathbf{top}(T)$ might not be disjoint. However, since all parallelograms $\{\mathbf{top}(T) : T \in \overline{\mathcal{T}_R}\}$ have (essentially) the same orientation, we may use a standard covering argument to select a subset $\widetilde{\mathcal{T}_R}$ of $\overline{\mathcal{T}_R}$ such that

$$\{\sigma^{-\epsilon}\,\mathbf{top}(T)\}_{T\in\widetilde{\mathcal{T}_R}}$$

is pairwise disjoint, and such that

$$\Big|\bigcup_{T\in\widetilde{\mathcal{T}_R}}\sigma^{-\epsilon}\,\mathbf{top}(T)\Big| \gtrsim \Big|\bigcup_{T\in\overline{\mathcal{T}_R}}\sigma^{-\epsilon}\,\mathbf{top}(T)\Big|.$$

Hence

$$|F\cap C\sigma^{-\epsilon}R| \gtrsim \Big|\bigcup_{T\in\widetilde{\mathcal{T}_R}}\sigma^{-\epsilon}\,\mathbf{top}(T)\cap F\Big| \overset{\text{by disjointness}}{=} \sum_{T\in\widetilde{\mathcal{T}_R}}|\sigma^{-\epsilon}\,\mathbf{top}(T)\cap F|$$

$$\overset{\text{by Lemma 6.3}}{\gtrsim} \sigma^{1+\epsilon}\sum_{T\in\widetilde{\mathcal{T}_R}}|\sigma^{-\epsilon}\,\mathbf{top}(T)| \gtrsim \sigma^{1+\epsilon}\Big|\bigcup_{T\in\widetilde{\mathcal{T}_R}}\sigma^{-\epsilon}\,\mathbf{top}(T)\Big|$$

$$\gtrsim \sigma^{1+\epsilon}\Big|\bigcup_{T\in\overline{\mathcal{T}_R}}\sigma^{-\epsilon}\,\mathbf{top}(T)\Big| \gtrsim \sigma^{1+\epsilon}\Big|\bigcup_{T\in\overline{\mathcal{T}_R}}\mathbf{top}(T)\Big|$$

$$\overset{\text{by disjointness}}{\gtrsim} \sigma^{1+\epsilon}\sum_{T\in\overline{\mathcal{T}_R}}|\mathbf{top}(T)| \overset{\text{by Claim 9.2}}{\gtrsim} \sigma^{1+\epsilon}\sum_{T\in\mathcal{T}_R}|\mathbf{top}(T)|$$

$$\overset{\text{by definition of } \mathcal{R}_j}{\gtrsim} \sigma^{1+\epsilon}\,2^{-j}\,|R|.$$

This finishes the proof of Claim 9.3. □

9.2.2. Case 2: $2^k \le \sigma^{-\epsilon}$. This section is very similar to the previous section. As in the last section, we verify the hypotheses of Lemma 6.2 for a suitable collection.

We consider all of these collections $\mathcal{R}_{j,k}$ together. Let

$$\mathcal{R}_{j,\mathrm{small}} = \bigcup_{0\le k\le \log \sigma^{-\epsilon}} \mathcal{R}_{j,k}.$$

Now we sort the tiles as before. Initialize

$$\begin{aligned} \textsc{stock} &= \mathcal{R}_{j,\mathrm{small}} \\ \widetilde{\mathcal{R}_{j,\mathrm{small}}} &= \emptyset. \end{aligned}$$

While $\textsc{stock} \ne \emptyset$, choose R with maximal length, let

$$\mathcal{A}_R = \{R' \in \textsc{stock} : \sigma^{-\epsilon}R' \cap \sigma^{-\epsilon}R \ne \emptyset \ \text{ and } \ \omega_{R'} \cap \omega_R \ne \emptyset\},$$

and update

$$\begin{aligned} \textsc{stock} :&= \mathcal{R}_{\mathrm{small}} \setminus \mathcal{A}_R \\ \widetilde{\mathcal{R}_{j,\mathrm{small}}} &= \widetilde{\mathcal{R}_{j,\mathrm{small}}} \cup \{R\}. \end{aligned}$$

As before, we have

$$\sum_{R\in\mathcal{R}_{j,\mathrm{small}}} |R| \le \sum_{R\in\widetilde{\mathcal{R}_{j,\mathrm{small}}}} \sum_{R'\in\mathcal{A}_R} |R'| \le \sum_{R\in\widetilde{\mathcal{R}_{j,\mathrm{small}}}} |\sigma^{-\epsilon}R|.$$

We again note several properties of the parallelograms in $\widetilde{\mathcal{R}_{j,\mathrm{small}}}$. First, they are pairwise incomparable. Second, they satisfy the estimate

$$\frac{|\sigma^{-\epsilon}R \cap E \cap u^{-1}(\omega_{\sigma^{-\epsilon}R})|}{|\sigma^{-\epsilon}R|} \gtrsim \sigma^{2\epsilon}\,\delta.$$

This gives us the density estimate

$$\sum_{R\in\widetilde{\mathcal{R}_{j,\mathrm{small}}}} |\sigma^{-\epsilon}R| \lesssim \frac{|E|}{\sigma^{2\epsilon}\delta}, \tag{9.2}$$

from a direct application of Lemma 6.1. Third, just as in Claim 9.3, they satisfy the estimate

$$\frac{|\sigma^{-\epsilon}R \cap F|}{|\sigma^{-\epsilon}R|} \gtrsim 2^{-j}\,\sigma^{1+\epsilon}.$$

So, by Lemma 6.2, we have

$$\sum_{R\in\widetilde{\mathcal{R}_{j,\mathrm{small}}}} |\sigma^{-\epsilon}R| \lesssim \frac{|F|}{\delta\,(2^{-j}\sigma^{1+\epsilon})^{1+\epsilon}}. \tag{9.3}$$

As before, we split the sum into large and small j and use (9.2) and (9.3), respectively, to obtain

$$\begin{aligned}
&\sum_{j\geq 0}\sum_{R\in\mathcal{R}_{j,\mathrm{small}}}\sum_{T\in\mathcal{T}_R}|\mathbf{top}(T)| \\
&\quad = \sum_{j\leq \log\frac{|E|\sigma}{|F|}}\sum_{R\in\mathcal{R}_{j,\mathrm{small}}} 2^{-j}|R| + \sum_{j\geq \log\frac{|E|\sigma}{|F|}}\sum_{R\in\mathcal{R}_{j,\mathrm{small}}} 2^{-j}\,|R| \\
&\quad \lesssim \sum_{j\leq \log\frac{|E|\sigma}{|F|}} 2^{\epsilon j}\frac{|F|}{\delta\,\sigma^{(1+\epsilon)^2}} + \sum_{j\geq \log\frac{|E|\sigma}{|F|}} 2^{-j}\,\frac{|E|}{\sigma^{2\epsilon}\,\delta} \lesssim \frac{|F|^{1-5\epsilon}\,|E|^{5\epsilon}}{\delta\,\sigma^{1+5\epsilon}},
\end{aligned}$$

which is what we needed, since ϵ is arbitrary.

10. Large size implies large intersection with *F*

Remark 10.1. The title of the section is technically a bit misleading, since $\mathbf{size}(T)$ is actually the supremum over all subtrees of T of an l^2-type norm; nevertheless, the trees obtained through the selection procedure in Section 5 all satisfy the property that the full tree (essentially) achieves this supremum.

To prove Lemma 6.3, we need the following notation. For a fixed 1-tree T, define the operator

$$\Delta(f) = \Big(\sum_{s\in T}|\langle f,\varphi_s\rangle|^2\,\frac{\mathbf{1}_s}{|s|}\Big)^{1/2}.$$

We need the following facts about Δ.

Lemma 10.2. *For any $N>0$, we have*

$$\|\Delta f\|_p \lesssim \|f\beta_{N,T}\|_p$$

for $p\in(1,\infty)$, where

$$\beta_N(x_1,x_2) = \frac{1}{1+|x_1|^N+|x_2|^N},$$

and $\beta_{N,T}$ is an L^∞-normalized version of β_N adapted to $\mathbf{top}(T)$*. The implicit constant depends on N but not on T.*

We prove Lemma 10.2 in Section 14. Of course proving $\|\Delta f\|_2 \lesssim \|f\|_2$ is straightforward; indeed, it is an easy special case of Lemma 13.1. The work is in inserting the smooth cutoff β_N, which is the point of Lemma 13.1, and moving below L^2. Second, we have:

Lemma 10.3.

$$\|\Delta f\|_2 \lesssim \frac{1}{|\mathbf{top}(T)|^{1/2}}\int_{C\mathbf{top}(T)}\Delta f,$$

provided that T satisfies the following uniform size estimate:

$$\sup_{\text{1-trees } T' \subseteq T} \left(\frac{1}{|\mathbf{top}(T')|} \sum_{s \in T'} |\langle f, \varphi_s \rangle|^2 \right)^{1/2} \lesssim \left(\frac{1}{|\mathbf{top}(T)|} \sum_{s \in T} |\langle f, \varphi_s \rangle|^2 \right)^{1/2}.$$

The condition in Lemma 10.3 is the one mentioned in Remark 10.1. We prove Lemma 10.3 in Section 15. The point of these lemmas is that $\|\Delta f\|_2$ is closely related to $\mathbf{size}(T)$. Indeed,

$$\|\Delta f\|_2^2 = \sum_{s \in T} |\langle f, \varphi_s \rangle|^2.$$

On the other hand, we want information about $|F \cap \mathbf{top}(T)|$ (or possibly $|F \cap M\mathbf{top}(T)|$ for a dilate $M\mathbf{top}(T)$ of $\mathbf{top}(T)$, which is actually what we will obtain below), which is much more closely related to $\|\Delta f\|_p$ for p close to 1, as we see below. Combining these two lemmas and Hölder's inequality gives us

$$\begin{aligned}\left(\frac{1}{|\mathbf{top}(T)|} \sum_{s \in T} |\langle f, \varphi_s \rangle|^2 \right)^{1/2} &= \frac{1}{|\mathbf{top}(T)|^{1/2}} \|\Delta f\|_2 \lesssim \frac{1}{|\mathbf{top}(T)|} \int_{C\mathbf{top}(T)} \Delta f \\ &\lesssim \left(\frac{1}{|\mathbf{top}(T)|} \int (\Delta f)^{1+\epsilon} \right)^{\frac{1}{1+\epsilon}} \lesssim \left(\frac{1}{|\mathbf{top}(T)|} \int (f \beta_{N,T})^{1+\epsilon} \right)^{\frac{1}{1+\epsilon}}.\end{aligned}$$

Applying this with $f = \mathbf{1}_F$ and a tree T such that $\left(\frac{1}{|\mathbf{top}(T)|} \sum_{s \in T} |\langle f, \varphi_s \rangle|^2 \right)^{1/2} \sim \sigma$ gives us for any N,

$$\begin{aligned}\sigma^{1+\epsilon} |\mathbf{top}(T)| &\lesssim \int \mathbf{1}_F (\beta_{N,T})^{1+\epsilon} \\ &\lesssim |\sigma^{-\epsilon} \mathbf{top}(T) \cap F| + \sigma^{(N-2)\epsilon} |\sigma^{-\epsilon} \mathbf{top}(T)|\end{aligned}$$

This proves Lemma 6.3 since N can be chosen arbitrarily large with respect to ϵ.

11. Proof of the tree lemma

In this section we present a proof of Lemma 5.1. Recall that we have a fixed tree T in mind. For notational convenience we assume that the slope of the long side of $\mathbf{top}(T)$ is zero. We write $\pi_1(E)$ and $\pi_2(E)$ to denote the vertical and horizontal (respectively) projections of a set E. Of course the width of every tile in T is a fixed number w. Let $\mathcal{J}_1$ be a partition of $\mathbb{R}$ (the horizontal axis) into dyadic intervals such that $3J \times \mathbb{R}$ does not contain any tile $s \in T$, and such that J is maximal with respect to this property. Now let $\mathcal{J}_2$ be a partition of $\mathbb{R}$ (the vertical axis) into intervals of width $|\pi_2(\mathbf{top}(T))|/3$. Let

$$\mathcal{P} = \bigcup_{J_1 \in \mathcal{J}_1} \bigcup_{J_2 \in \mathcal{J}_2} J_1 \times J_2.$$

This is a partition of $\mathbb{R}^2$. The parallelograms $P \in \mathcal{P}$ are the smallest relevant parallelograms for this tree. The parallelograms $P \in \mathcal{P}$ with $\pi_1(P)$ far away from $\mathbf{top}(T)$ are defined so as to still be able to take advantage of the density estimate for tiles in T. Now for each $P \in \mathcal{P}$ we split the operator L into two pieces, one corresponding to tiles with x-projection larger than P, the other to tiles with x-projection smaller than P. Let

$$T_P^+ = \{s \in T\colon |\pi_1(s)| > |\pi_1(P)|\}, \qquad T_P^- = \{s \in T\colon |\pi_1(s)| \le |\pi_1(P)|\}$$
$$L_P^+ = \sum_{s\in T_P^+} \langle f, \varphi_s\rangle\, \phi_s\, \mathbf{1}_E, \qquad L_P^- = \sum_{s\in T_P^-} \langle f, \varphi_s\rangle\, \phi_s\, \mathbf{1}_E.$$

Note that for appropriate ϵ_s with $|\epsilon_s| = 1$, we have

$$\begin{aligned}\sum_{s\in T} |\langle f, \varphi_s\rangle\langle\phi_s \mathbf{1}_E\rangle| &= \sum_{s\in T} \epsilon_s\langle f, \varphi_s\rangle\langle\phi_s \mathbf{1}_E\rangle = \int \sum_{s\in T} \epsilon_s \langle f, \varphi_s\rangle \phi_s \mathbf{1}_E \\ &= \sum_{P\in\mathcal{P}} \int_P \sum_{s\in T} \epsilon_s\langle f, \varphi_s\rangle\phi_s\mathbf{1}_E = \sum_{P\in\mathcal{P}} \int_P L_P^- + \sum_{P\in\mathcal{P}} \int_P L_P^+. \end{aligned} \tag{11.1}$$

The main term will come from parallelograms $P \in \mathcal{P}$ close to $\mathbf{top}(T)$; estimates on parallelograms P away from $\mathbf{top}(T)$ will come with a decay factor. To make things more precise, define, for $k \ge 1$,

$$\mathcal{P}_0 = \{P \in \mathcal{P}\colon \frac{\mathrm{dist}(\pi_2(P), \pi_2(\mathbf{top}(T)))}{|\pi_2(\mathbf{top}(T))|} \le 1\}$$
$$\mathcal{P}_k = \{P \in \mathcal{P}\colon \frac{\mathrm{dist}(\pi_2(P), \pi_2(\mathbf{top}(T)))}{|\pi_2(\mathbf{top}(T))|} \in (2^{k-1}, 2^k]|\}.$$

We focus first on the first term in (11.1). To control it we need only spatial decay in both the horizontal and vertical directions.

11.1. Small tiles

For notational convenience, we further consider for $l \ge 1$,

$$\mathcal{P}_{k,0} = \{P \in \mathcal{P}_k\colon \frac{\mathrm{dist}(\pi_1(P), \pi_1(\mathbf{top}(T)))}{|\pi_1(\mathbf{top}(T))|} \le 1\},$$
$$\mathcal{P}_{k,l} = \{P \in \mathcal{P}_k\colon \frac{\mathrm{dist}(\pi_1(P), \pi_1(\mathbf{top}(T)))}{|\pi_1(\mathbf{top}(T))|} \in (2^{l-1}, 2^l]\}.$$

We divide the sum in the definition of L_P^- into pieces according to how large the tiles are. Specifically, let

$$T_j = \{s \in T_P^-\colon |s| = 2^{-j}|\mathbf{top}(T)|\}.$$

The reason for this is that since the tiles $s \in T_P^-$ are shorter than P, their frequency intervals can be much larger than that of P, meaning we lose control of $|P \cap \mathrm{supp}(L_P^-)|$. We use the extra decay from Schwartz tails to compensate for this. The upper bound of $\mathbf{size}(T) \leq \sigma$ implies that for individual tiles $s \in T$ we have $|\langle f, \varphi_s \rangle| \leq \sigma |s|^{1/2}$. Hence

$$\Big|\sum_{s\in T_j} \langle f, \varphi_s\rangle \phi_s \mathbf{1}_E\Big| \lesssim \sum_{s\in T_j} \sigma\, \chi_s^{(\infty)} \lesssim \sigma\, 2^{-Nk} \sum_{m\geq 2^{j+l}} m^{-N} \lesssim \sigma\, 2^{-Nk}\, 2^{-Nj/2}\, 2^{-Nl/2}.$$

But note that since $\mathbf{dense}(s) \lesssim \delta$, we have

$$\delta \gtrsim \int_{E_s} \chi_s^{(1)} \geq 2^{-100(k+j+l)}\ \frac{|P \cap \mathrm{supp}(\sum_{s\in T_j} \langle f, \varphi_s\rangle \phi_s \mathbf{1}_E)|}{|P|}.$$

This last estimate follows from considering the distance between s and P relative to the length of s. Hence for any $P \in \mathcal{P}_{k,l}$, we have

$$\begin{aligned} \int_P |L_P^-| &\leq \sum_{j\geq 0} \int_P \Big| \sum_{s\in T_j} \langle f, \varphi_s\rangle \phi_s \mathbf{1}_E \Big| \\ &\lesssim \sigma \sum_{j\geq 0} 2^{-Nk}\, 2^{-Nj/2}\, 2^{-Nl/2} \Big| P \cap \mathrm{supp}\Big(\sum_{s\in T_j} \langle f, \varphi_s\rangle \phi_s \mathbf{1}_E\Big)\Big| \lesssim \delta\, |P|\, \sigma\, 2^{-10(l+k)}. \end{aligned}$$

Summing over k, l, and P gives us

$$\begin{aligned} \sum_{P\in\mathcal{P}} \int_P |L_P^-| &\lesssim \sum_{l\geq 0}\sum_{k\geq 0}\sum_{P\in\mathcal{P}_{k,l}} \int_P |L_P^-| \\ &\lesssim \sum_{l\geq 0}\sum_{k\geq 0}\sum_{P\in\mathcal{P}_{k,l}} \sigma\,\delta\,|P|\,2^{-10k}\,2^{-10l} \lesssim \sigma\,\delta\,|\mathbf{top}(T)|, \end{aligned}$$

with the primary contribution coming from P near $\mathbf{top}(T)$ as usual.

11.2. Large tiles

We start by remarking that sorting with respect to the horizontal distance from T (i.e., using the index l, as in the previous subsection) is unnecessary in this subsection. For if $P \in \mathcal{P}_{k,l}$ with $l \geq C$, then T_P^+ is empty, because $|\Pi_1(P)| > |\Pi_1(\mathbf{top}(T))|$. This fact will be used several times in what follows. Next, we show that the term under consideration in this section has small support. Precisely:

Claim 11.1. *For $P \in \mathcal{P}_k$, $L_P^+ \mathbf{1}_E$ is supported on a set of size $\lesssim \delta |P|\, 2^{100k}$.*

The factor 2^{100k} arises from the tail in the definition of **dense** and the fact that P is away from $\mathbf{top}(T)$. Fortunately, the decay in the functions φ_s for $s \in T$ is even greater when P is away from $\mathbf{top}(T)$.

Proof. It is convenient to proceed by contradiction. Assume $L_P^+ \mathbf{1}_E$ has much larger support than $\delta|P|2^{100k}$. By the construction of P, we know that there is some $s \in T$ such that $s \subseteq C2^k P$. But this implies there is R of the same dimensions as P, though located spatially over T, with $\omega_R \subseteq \omega_s$ and such that $\mathbf{dense}(R) \geq 100\delta$, say. Since this implies $s \leq R$, we have contradicted the assumption that $\mathbf{dense}(s) \leq \delta$. □

We now turn our attention to the second term in (11.1). Recall the definitions of 1-trees and 2-trees. Clearly for every $s \in T$, either $\omega_{s,1} \cap \omega_{\mathbf{top}(T)} = \emptyset$ or $\omega_{s,2} \cap \omega_{\mathbf{top}(T)} = \emptyset$, so our tree T can be partitioned as $T = T_1 \cup T_2$, where T_j is a j-tree. Let

$$(T_P^+)_j = T_P^+ \cap T_j$$

for $j = 1, 2$. Of course $(T_P^+)_j$ is still a j-tree. We treat the two cases separately.

11.2.1. The 2-tree case. This case is a bit easier to handle because of the location of the support of the function ϕ_s. More to the point: Since T_2 is a 2-tree, if there exists x such that $\phi_s(x)\phi_t(x) \neq 0$ for $s, t \in T_2$, then $|s| = |t|$. This follows from the fact that $\phi_s(x) = 0$ unless $v(x) \in \omega_{s,2}$, together with the fact that $\omega_{s,1} \supseteq \omega_{\mathbf{top}(T)}$, and similarly for t. (This was mentioned near the definition of ϕ_s in Section 3.) Further, we know that for any tile $s \in T$, we have $|\langle f, \varphi_s \rangle| \leq \sigma |s|^{1/2}$ by the **size** estimate for T. Combining these observations with Claim 11.1 and the rapid decay of ϕ_s in the vertical direction gives us that, for $P \in \mathcal{P}_k$,

$$\int_P \sum_{s \in (T_P^+)_2} \langle f, \varphi_s \rangle \, \phi_s \, \mathbf{1}_E \lesssim \sigma \, \delta \, 2^{-10k} \, |P|,$$

since the integrand is uniformly bounded by $\sigma 2^{-200k}$. As mentioned earlier, if $|\pi_1(s)| \geq |\pi_1(P)|$, then $\pi_1(P) \subseteq C\pi_1(\mathbf{top}(T))$. Hence

$$\sum_k \sum_{P \in \mathcal{P}_k} \int_P \sum_{s \in (T_P^+)_2} \langle f, \varphi_s \rangle \, \phi_s \, \mathbf{1}_E \lesssim \delta \, \sigma \, |\mathbf{top}(T)|.$$

This completes the estimate for T_2.

11.2.2. The 1-tree case. In this case we appeal to orthogonality in the form of the Bessel inequality in Lemma 13.1. For parallelograms $P \in \mathcal{P}$ whose vertical component is large, we need the decay factor from Lemma 13.1. We first introduce some extra functions associated to the tiles. Let

$$\alpha_s(x) = \int \psi_s(t) \, \varphi_s(x_1 - t, x_2) \, dt.$$

The difference between α_s and ϕ_s is that the vector field v makes no explicit appearance in the definition of α_s; rather, the integral is taken over a horizontal line for every x. In ϕ_s, however, the integral is taken over an *almost* horizontal line, where the precise definition of *almost* depends on the length of s. (The line is horizontal because we assumed that the slope of the long side of $\mathbf{top}(T)$ is zero. In the general case it is parallel to $\mathbf{top}(T)$.) We have the obvious equality

$$\int_P \sum_{s \in (T_P^+)_1} \epsilon_s \langle f, \varphi_s \rangle \, \phi_s \, \mathbf{1}_E = \int_P \sum_{s \in (T_P^+)_1} \epsilon_s \langle f, \varphi_s \rangle \, \alpha_s \, \mathbf{1}_E + \int_P \sum_{s \in (T_P^+)_1} \epsilon_s \langle f, \varphi_s \rangle \, (\phi_s - \alpha_s) \, \mathbf{1}_E.$$

This decomposition allows us to reduce our problem to proving the following two claims:

Claim 11.2. *For each $P \in \mathcal{P}$,*

$$\int_P \sum_{s\in(T_P^+)_1} \epsilon_s\langle f,\varphi_s\rangle\alpha_s \mathbf{1}_E \lesssim \delta \sum_{j\geq 0} 2^{-Nj}\frac{1}{|2^jP|}\int_{2^jP}\Big|\sum_{s\in T_1}\langle f,\varphi_s\rangle\alpha_s\Big|.$$

Claim 11.3. *For $P\in\mathcal{P}_k$,*

$$\sum_{s\in(T_P^+)_1} \epsilon_s\langle f,\varphi_s\rangle(\phi_s-\alpha_s)\,\mathbf{1}_E \lesssim 2^{-200k}\,\sigma.$$

Notice that $\operatorname{supp}\widehat{\alpha_s}\subseteq \operatorname{supp}\widehat{\varphi_s}$, since

$$\widehat{\alpha_s}(\xi) = \int \psi_s(t)\, e^{-2\pi i t\xi_1}\,\widehat{\varphi_s}(\xi)\,dt.$$

This will allow us to prove orthogonality statements about the α_s later in the proof. For example, from this we can conclude that

$$\Big\|\sum_{s\in T_1}\epsilon_s\langle f,\varphi_s\rangle\alpha_s\Big\|_2^2 \lesssim \sum_{s\in T_1}|\langle f,\varphi_s\rangle|^2, \tag{11.2}$$

because the fact stated above about the Fourier support of the functions α_s allows us to prove this inequality in the same way we prove the Bessel inequality in Section 13: expand the square, and notice that $\langle\alpha_s,\alpha_t\rangle = 0$ unless $|s|=|t|$.

Again we remark that if T_P^+ is nonempty, then $\pi_1(P)\subseteq C\pi_1(\mathbf{top}T)$. Hence in the summation below we can ignore dependence on the parameter l used in the last section. Given these claims, together with Claim 11.1, we control the first term in (11.1) by

$$\begin{aligned}\sum_{P\in\mathcal{P}}\int_P L_P^+ &\lesssim \sum_{P\in\mathcal{P}}\int_P \epsilon_s \sum_{s\in(T_P^+)_1}\langle f,\varphi_s\rangle\alpha_s\mathbf{1}_E + \sum_{P\in\mathcal{P}}\int_P\epsilon_s\sum_{s\in(T_P^+)_1}\langle f,\varphi_s\rangle(\phi_s-\alpha_s)\mathbf{1}_E\\ &\lesssim \sum_k\sum_{P\in\mathcal{P}_k}\delta\int_P\sum_{j\geq 0}2^{-Nj}\frac{1}{|2^jP|}\int_{2^jP}\Big|\sum_{s\in T_1}\langle f,\varphi_s\rangle\alpha_s\Big|\\ &\quad+\sum_k\sum_{P\in\mathcal{P}_k}2^{-200k}\,\sigma\,|P\cap\operatorname{supp}(L_P^+)|.\end{aligned}$$

Note that the second term in the last display is controlled by Claim 11.1. For $P\in\mathcal{P}_k$, it is convenient to split the function $\sum_{s\in T_1}\langle f,\varphi_s\rangle\alpha_s$ into two pieces, using the identity $\mathbf{1}_{\mathbb{R}^2} = \mathbf{1}_{D_{k-5}} + \mathbf{1}_{(D_{k-5})^c}$, where

$$D_k = \{(x,y) : |y| \lesssim 2^k\,|\pi_2(\mathbf{top}(T))|\}.$$

In other words, D_k is a horizontal strip of width $\sim 2^k|\pi_2(\mathbf{top}(T))|$. (Obvious modifications can be made in the case $k\leq 5$.) For the first piece – the one closer to $\mathbf{top}(T)$ – we can use the fact that the tile P is far from $\mathbf{top}(T)$ together with the decay in j to obtain good control. For the second piece – the one away from $\mathbf{top}(T)$ –

we can take advantage of the decay in the wave packets associated to tiles in T in the form of the Bessel inequality in Lemma 13.1. We focus first on the term close to $\mathbf{top}(T)$:

$$\begin{aligned}\sum_k \sum_{P\in\mathcal{P}_k} \delta \int_P \sum_{j\geq 0} 2^{-Nj} \frac{1}{|2^j P|} \int_{2^j P} \Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{D_{k-5}}\Big| \\ = \sum_k \sum_{P\in\mathcal{P}_k} \delta \int_P \sum_{j\geq k} 2^{-Nj} \frac{1}{|2^j P|} \int_{2^j P} \Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{D_{k-5}}\Big| \\ \lesssim \sum_k \sum_{P\in\mathcal{P}_k} \delta 2^{-Nk} \int_P M\Big(\Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{D_{k-5}}\Big|\Big) \\ = \delta \int_{\cup_{l=0}^{C} \cup_{P\in\mathcal{P}_{k,l}} P} 2^{-Nk} M\Big(\Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{D_{k-5}}\Big|\Big) \\ \lesssim \delta 2^{-Nk} \Big| \cup_{l=0}^{C} \cup_{P\in\mathcal{P}_{k,l}} P\Big|^{1/2} \Big(\int \Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{D_{k-5}}\Big|^2\Big)^{1/2}.\end{aligned}$$

This nearly finishes the proof for the first term, since we may estimate this L^2 norm by using orthogonality in the x-variable just as in the proof of Lemma 13.1 below. (Readers uncomfortable with this should look to the proof of Lemma 13.1.) Specifically, we have

$$\begin{aligned}\int \Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{D_{k-5}}\Big|^2 = \sum_{s\in T_1} \sum_{s'\in T_1} \langle f,\varphi_s\rangle \langle f,\varphi_{s'}\rangle \int_{D_k} \alpha_s \alpha_{s'} \\ \lesssim \sum_{s\in T_1} |\langle \varphi_s, f\rangle|^2 \sum_{s' :\ |s|=|s'|} \int |\alpha_s \alpha_{s'}| \lesssim \sum_{s\in T_1} |\langle \varphi_s, f\rangle|^2 \lesssim \sigma^2 |\mathbf{top}(T)|.\end{aligned}$$

We have used symmetry and the x-orthogonality in the first inequality above. This finishes the proof for the first term. To control the second term (the one away from $\mathbf{top}(T)$), we can appeal directly to a Bessel-type inequality. Here we use such an inequality for the functions α_s rather than the functions φ_s, just as in the estimate above, but we also obtain significant decay in k just as in Lemma 13.1. The proof is identical to the proof of Lemma 13.1. Hence

$$\begin{aligned}\sum_k \sum_{l=0}^{C} \sum_{P\in\mathcal{P}_{k,l}} \delta \int_P \sum_{j\geq 0} 2^{-Nj} \frac{1}{|2^j P|} \int_{2^j P} \Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{(D_{k-5})^c}\Big| \\ \lesssim \delta \int_{\cup_{l=0}^{C} \cup_{P\in\mathcal{P}_{k,l}} P} M\Big(\Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{(D_{k-5})^c}\Big|\Big) \\ \lesssim \delta \Big| \cup_{l=0}^{C} \cup_{P\in\mathcal{P}_{k,l}} P\Big|^{1/2} \Big(\int \Big| \sum_{s\in T_1} \langle f,\varphi_s\rangle \alpha_s \mathbf{1}_{(D_{k-5})^c}\Big|^2\Big)^{1/2} \\ \lesssim \delta\, 2^k\, |\mathbf{top}(T)|^{1/2}\, (\sigma^2\, 2^{-100k}\, |\mathbf{top}(T)|)^{1/2} \lesssim 2^{-10k}\, \delta\, \sigma\, |\mathbf{top}(T)|,\end{aligned}$$

which is what we want.

Proof of Claim 11.2. Recall that we are considering a point $x \in P$ for some parallelogram P, and we consider the sum

$$\sum_{s \in T_1 \colon |\pi_1(s)| > |\pi_1(P)|} \langle f, \varphi_s \rangle \phi_s(x).$$

The restriction in the summation already implies that for any x, there is $m(x)$ such that all tiles s that make an appearance in the sum above satisfy $|\pi_1(s)| \geq m(x)$. Further, since we know that $u(x) \in \omega_{s,2}$, we also have $M(x)$ such that all tiles s that appear in the sum above satisfy $|\pi_1(s)| \leq M(x)$. Both of these claims are reversible, so

$$\{s \in T_1 \colon |\pi_1(s)| > |\pi_1(P)|\} = \{s \in T \colon m(x) \leq L(s) \leq M(x)\}.$$

Hence it is our goal to estimate

$$\sum_{s \in T \colon m(x) \leq L(s) \leq M(x)} \langle f, \varphi_s \rangle \alpha_s.$$

Denote by k a Schwartz function such that $\operatorname{supp} \hat{k} \subseteq [-1 - 1/100, 1 + 1/100]^2$, and such that $\hat{k}(\xi) = 1$ for $\xi \in [-1, 1]^2$. Further denote by k_r the function obtained by adapting k to the rectangle $[-1/r, 1/r] \times [-1/w, 1/w]$; i.e., let $k_r(x, y) = k(x/r, y/w)$. With this definition, we know for any N (which appears in the last line of the computation below),

$$\begin{aligned}
\sum_{s \in T_1 \colon m(x) \leq L(s) \leq M(x)} \langle f, \varphi_s \rangle \alpha_s &= \sum_{s \in T_1 \colon m(x) \leq L(s)} \langle f, \varphi_s \rangle \alpha_s - \sum_{s \in T_1 \colon L(s) > M(x)} \langle f, \varphi_s \rangle \alpha_s \\
&= \Big(\sum_{s \in T_1} \langle f, \varphi_s \rangle \alpha_s \Big) * k_{m(x)} - \Big(\sum_{s \in T_1} \langle f, \varphi_s \rangle \alpha_s \Big) * k_{M(x)} \\
&\leq \sum_{j \geq 0} 2^{-Nj} \frac{1}{|2^j P|} \int_{2^j P} \Big| \sum_{s \in T_1} \langle f, \varphi_s \rangle \alpha_s \Big|.
\end{aligned}$$

□

Proof of Claim 11.3. By the argument at the beginning of the proof of Claim 11.2, it suffices to estimate

$$\sum_{s \in T \colon m(x) \leq |\pi_1(s)| \leq M(x)} \langle f, \varphi_s \rangle (\phi_s(x) - \alpha_s(x)) \, \mathbf{1}_{\omega_{s,2}}(u(x)).$$

To do this we first estimate $|\phi_s - \alpha_s|$. By definition, we have

$$|\phi_s(x) - \alpha_s(x)| \leq \int |\psi_s(t)| \, |\varphi_s(x_1 - t, x_2 - tu(x)) - \varphi(x_1 - t, x_2)| \, dt.$$

To compute the difference in the integrand, estimating the following quantity will be helpful:

$$\star := \sup_{z \in [0, tu(x)]} \frac{\partial}{\partial x_2} \varphi_s(x_1 - t, x_2 - z).$$

Fix an integer $j \geq 1$ and consider $|t| \sim 2^j|\pi_1(s)|$. If $(x_1, x_2) \notin 2^{j+10}s$, then $\star \lesssim \chi_s^{(2)}(x_1, x_2)$. If $(x_1, x_2) \in 2^{j+10}s$, then $\star \lesssim 1$. We also have that $\psi_s(t) \lesssim 1/(2^{Nj}|s|)$ for any N. Analogous facts hold when $j = 0$ and $|t| \leq |\pi_1(s)|$. Let $I_j = \{t\colon |t| \sim 2^j|\pi_1(s)|\}$ for $j \geq 1$ and $I_0 = \{t\colon |t| \leq |\pi_1(s)|\}$. Combining these observations gives us for $(x_1, x_2) \notin 2^{j+10}s$ that

$$
\begin{aligned}
|\phi_s(x) - \alpha_s(x)| &\lesssim \sum_{j\geq 0} \int_{I_j} \frac{1}{2^{Nj}|s|}\, 2^j\, |\pi_1(s)|\, \frac{|u(x)|}{w}\, \chi_s^{(2)}(x_1, x_2)\, dt \\
&\lesssim |\pi_1(s)|\, \frac{|u(x)|}{w}\, \chi_s^{(2)}(x_1, x_2).
\end{aligned}
$$

If $(x_1, x_2) \in 2^{j+10}s$, then we have $\star \lesssim 2^{100j}\chi_s^{(2)}$, so

$$
|\phi_s(x) - \alpha_s(x)| \lesssim \sum_{j\geq 0} \int_{I_j} \frac{1}{2^{-Nj}|s|}\, 2^j\, |\pi_1(s)|\, \frac{|u(x)|}{w}\, dt \lesssim |\pi_1(s)|\, \frac{|u(x)|}{w}\, \chi_s^{(2)}(x_1, x_2).
$$

Since $u(x) \in \omega_{s,2}$ for all $s \in T_1$, we know $u(x) \leq w/|\pi_1(s)|$. Combining this with the fact that $|\langle f, \varphi_s\rangle| \lesssim \sigma|s|^{1/2}$ and the estimate immediately above, we have

$$
\begin{aligned}
\Big| \sum_{m(x)\leq|\pi_1(s)|\leq M(x)} \langle f, \varphi_s\rangle(\phi_s - \alpha_s)\Big| &\leq \sum_{|\pi_1(s)|\leq w/u(x)} \sigma\, |s|^{1/2}\, |u(x)|\, \frac{|\pi_1(s)|}{w}\, \chi_s^{(2)}(x_1, x_2) \\
&\lesssim \sigma\, \chi^{(\infty)}_{\mathbf{top}(T)}(x_1, x_2),
\end{aligned}
$$

which is what we claimed. □

12. Proof of size estimate

In this section we write $f = \mathbf{1}_F$; note that we do not use the fact that f is a characteristic function. As with the tree lemma, there are small modifications required from the one-dimensional situation to handle Schwartz tails in the vertical direction. We use the Bessel inequality from Lemma 13.1 to do this. First we note that by assumption,

$$
\begin{aligned}
\sigma^2 \sum_{T\in\mathcal{T}} |\mathbf{top}(T)| &\lesssim \sum_{T\in\mathcal{T}} \sum_{s\in T} |\langle f, \varphi_s\rangle|^2 \\
&= \int f \sum_T \sum_s \langle f, \varphi_s\rangle\varphi_s \leq \|f\|_2 \Big\| \sum_T \sum_s \langle f, \varphi_s\rangle\varphi_s \Big\|_2.
\end{aligned}
$$

It is enough to prove

$$
\Big\| \sum_T \sum_s \langle f, \varphi_s\rangle\varphi_s \Big\|_2 \leq \sigma \sqrt{\sum_{T\in\mathcal{T}} |\mathbf{top}(T)|}.
$$

By expanding the square and using symmetry, we have

$$\begin{aligned}\Big\|\sum_{T\in\mathcal{T}}\sum_{s\in T}\langle f,\varphi_s\rangle\varphi_s\Big\|_2^2 &= \sum_{T\in\mathcal{T}}\sum_{T'\in\mathcal{T}}\sum_{s\in T'}\sum_{s'\in T'}\langle f,\varphi_s\rangle\langle f,\varphi_{s'}\rangle\langle\varphi_s,\varphi_{s'}\rangle\\ &\lesssim \sum_{T\in\mathcal{T}}\sum_{s\in T}\sum_{T'\in\mathcal{T}}\sum_{s'\in T'\colon |s'|=|s|}|\langle f,\varphi_s\rangle\langle f,\varphi_{s'}\rangle\langle\varphi_s,\varphi_{s'}\rangle|\\ &\quad+\Big|\sum_{T\in\mathcal{T}}\sum_{s\in T}\sum_{T'\in\mathcal{T}}\sum_{s'\in T'\colon |s'|<|s|}\langle f,\varphi_s\rangle\langle f,\varphi_{s'}\rangle\langle\varphi_s,\varphi_{s'}\rangle\Big|\\ &= B+C.\end{aligned}$$

Note that

$$\{s'\colon |s'|=|s| \text{ and } \omega_s\cap\omega_{s'}\neq\emptyset\}$$

partitions $\mathbb{R}^2$, so

$$\sum_{|s'|=|s|}|\langle\varphi_s,\varphi_{s'}\rangle|\sim 1.$$

Hence we can estimate the first term, using symmetry again, by

$$\begin{aligned}B &\lesssim \sum_{T\in\mathcal{T}}\sum_{s\in T}\sum_{T'\in\mathcal{T}}\sum_{s'\in T\colon |s'|=|s|}|\langle f,\varphi_s\rangle|^2|\langle\varphi_s,\varphi_{s'}\rangle|\\ &\lesssim \sum_{T\in\mathcal{T}}\sum_{s\in T}|\langle f,\varphi_s\rangle|^2\sim\sigma^2\sum_{T\in\mathcal{T}}|\mathbf{top}(T)|.\end{aligned}$$

Now we look at the second term C. By Cauchy–Schwarz, we have

$$\begin{aligned}C &\le \sum_{T\in\mathcal{T}}\Big(\sum_{s\in T}|\langle f,\varphi_s\rangle|^2\Big)^{1/2}\Big(\sum_{s\in T}\Big|\sum_{T'\in\mathcal{T}'}\sum_{s'\in T'\colon |s'|<|s|}\langle\varphi_s,\varphi_{s'}\rangle\langle f,\varphi_{s'}\rangle\Big|^2\Big)^{1/2}\\ &\lesssim \sum_{T\in\mathcal{T}}\sigma\,|\mathbf{top}(T)|^{1/2}\,D(T)^{1/2},\end{aligned}$$

where

$$D(T)=\sum_{s\in T}\Big|\sum_{T'\in\mathcal{T}'}\sum_{s'\in T'\colon |s'|<|s|}\langle\varphi_s,\varphi_{s'}\rangle\langle f,\varphi_{s'}\rangle\Big|^2.$$

It remains to analyze $D(T)$ for a tree $T\in\mathcal{T}$. We claim that the set of tiles over which the inner sum ranges is actually independent of s. More specifically, define

$$\mathcal{A}=\Big\{s'\in\bigcup_{T'\neq T,T'\in\mathcal{T}}T'\colon \omega_{s,1}\cap\omega_{s',1}\neq\emptyset \text{ and } |s'|<|s| \text{ for some } s\in\mathcal{T}\Big\}.$$

Then we have:

Claim 12.1. *For each $s\in T$,*

$$\sum_{T'\in\mathcal{T}}\sum_{s'\in T'\colon |s'|<|s|}\langle\varphi_s,\varphi_{s'}\rangle\langle f,\varphi_{s'}\rangle=\sum_{s'\in\mathcal{A}}\langle\varphi_s,\varphi_{s'}\rangle\langle f,\varphi_{s'}\rangle.$$

Proof. It is clear from the definition of $\mathcal{A}$ that the summation on the left is over a set of tiles that is contained in $\mathcal{A}$. So suppose $s' \in \mathcal{A}$; by definition of $\mathcal{A}$, this gives us $\tilde{s} \in \mathcal{T}$ such that $|s'| < |\tilde{s}|$ and such that $\omega_{\tilde{s},1} \cap \omega_{s',1} \neq \emptyset$. This last condition guarantees that $\omega_{s',1} \supseteq \omega_T$. If $|s| \geq |\tilde{s}|$, then of course $|s| > |s'|$ and $\omega_{s,1} \cap \omega_{s',1} \neq \emptyset$, so that in fact the tile s' appears in the summation on the left-hand side of the claim. If $|s| < |\tilde{s}|$ and $|s| > |s'|$ then we are done as before. So assume $|s| \leq |s'| < |\tilde{s}|$. In this case $\omega_{s,1} \cap \omega_{s',1} = \emptyset$, which implies that $\langle \varphi_s, \varphi_{s'} \rangle = 0$, finishing the proof of the claim. □

Now for a collection of tiles $\mathcal{C}$, define

$$F(\mathcal{C}) = \sum_{t \in \mathcal{C}} \langle f, \varphi_t \rangle \, \varphi_t.$$

With this notation, we have

$$D(T) = \sum_{s \in T} |\langle \varphi_s, F(\mathcal{A}) \rangle|^2.$$

Before we proceed, we mention a key disjointness property of tiles in $\mathcal{A}$.

Claim 12.2. *Tiles in $\mathcal{A}$ are pairwise disjoint.*

Proof. Suppose $t, t' \in \mathcal{A}$. Then there are $s, s' \in \mathcal{T}$ such that $\omega_{t,2} \supseteq \omega_s \supseteq \omega_{\mathbf{top}(T)}$ and such that $\omega_{t',2} \supseteq \omega_{s'} \supseteq \omega_{\mathbf{top}(T')}$. Since $\omega_{t,2} \cap \omega_{t',2} \neq \emptyset$, we may assume without loss of generality that $\omega_{t,2} \subseteq \omega_{t',2}$, i.e., that $|t'| \leq |t|$. This means the tree T^* containing t was selected before the tree containing t'. Finally, note that t and t' cannot belong to the same 1-tree, since $\omega_{t,2} \subseteq \omega_{t',2}$. If $t \cap t' \neq \emptyset$, then in fact $t' \subseteq V(\mathbf{top}(T^*))$, and hence t' was included in the maximal tree $\widetilde{T^*}$ containing the 1-tree T^*; see the selection algorithm in Section 5 for construction of this tree $\widetilde{T^*}$. Hence the tiles in $\mathcal{A}$ are pairwise disjoint. □

We now introduce some more notation to sort the tiles in $\mathcal{A}$ according to how far they are from $\mathbf{top}(T)$. For $k > 1$, let $R_k = 2^k \mathbf{top}(T)$. Let $R_0 = \mathbf{top}(T)$. Then let

$$\mathcal{A}_k = \{s' \in \mathcal{A} \colon s' \subseteq R_k \text{ but } s' \not\subseteq R_{k-1}\}.$$

Now by the Minkowski inequality,

$$\Big(\sum_{s \in T} |\langle \varphi_s, F(\mathcal{A}) \rangle|^2 \Big)^{1/2} \leq \sum_k \Big(\sum_{s \in T} |\langle \varphi_s, F(\mathcal{A}_k) \rangle|^2 \Big)^{1/2}.$$

It remains to show

$$\sum_{s \in T} |\langle \varphi_s, F(\mathcal{A}_k) \rangle|^2 \lesssim 2^{-10k} \, \sigma \, |\mathbf{top}(T)|. \tag{12.1}$$

We will use the spatial localization of the tiles $s \in T$ to $\mathbf{top}(T)$ to obtain the desired decay in k. We have

$$\sum_{s \in T} |\langle \varphi_s, F(\mathcal{A}_k) \rangle|^2 \lesssim \sum_{s \in T} |\langle \varphi_s, \mathbf{1}_{R_{k-3}} F(\mathcal{A}_k) \rangle|^2 + \sum_{s \in T} |\langle \varphi_s, \mathbf{1}_{R_{k-3}^c} F(\mathcal{A}_k) \rangle|^2 = I_k + II_k.$$

First we estimate I_k. For $x \in R_{k-3}$ and $s \in \mathcal{A}_k$, we have

$$\varphi_s(x)\,\mathbf{1}_{R_{k-3}}(x) \lesssim 2^{-10k}\,\frac{1}{\sqrt{|s|}}\,\chi_s^{(\infty)}(x).$$

We now estimate $\|\mathbf{1}_{R_{k-3}}F(\mathcal{A}_k)\|_2$ by duality. We make one small observation as a preliminary:

Claim 12.3. *If M is the strong maximal operator, then*

$$\int \chi_s^{(\infty)}(x)\,g(x)\,dx \lesssim \int_s Mg(x)\,dx.$$

We remark that each $s \in \mathcal{A}$ essentially points in the direction of T, so the strong maximal operator is appropriate here.

Proof.

$$\int \chi_s^{(\infty)}(x)g(x)dx \lesssim |s| \sum_{k\geq 0} 2^{-3k}\frac{1}{|2^k s|}\int_{|2^k s|}|g| \lesssim |s| \inf_{x\in s} Mg(x) \lesssim \int_s Mg(x)dx. \quad \square$$

Consider a function $g \in L^2$, and remember that $|\langle f,\varphi_s\rangle| \lesssim \sigma\sqrt{|s|}$. Then using the claim above about disjointness of tiles $s \in \mathcal{A}_k$, we have

$$\begin{aligned}
\int F(\mathcal{A}_k)g\mathbf{1}_{R_{k-3}} &= \int \sum_{s\in\mathcal{A}_k}\langle f,\varphi_s\rangle\,\varphi_s(x)\,\mathbf{1}_{R_{k-3}}(x)\,g \lesssim \int \sum_{s\in\mathcal{A}_k} 2^{-10k}\,\sigma\,\chi_s^{(\infty)}(x)\,g \\
&\lesssim 2^{-10k}\sigma \sum_{s\in\mathcal{A}_k}\int_s Mg \leq 2^{-10k}\sigma \int_{\bigcup_{s\in\mathcal{A}_k} s} Mg \\
&\leq 2^{-10k}\sigma\,|R_k|^{1/2}\,\|g\|_2 \leq 2^{-10k}\,\sigma\,(2^{2k}\,|\mathbf{top}(T)|)^{1/2}\,\|g\|_2,
\end{aligned}$$

which implies that

$$I_k \lesssim \|\mathbf{1}_{R_{k-3}}F(\mathcal{A}_k)\|_2^2 \lesssim \sigma^2\,2^{-4k}\,|\mathbf{top}(T)|.$$

This proves (12.1) for I_k.

To estimate II_k, we need only estimate $\|F(\mathcal{A}_k)\|_2$ and apply Lemma 13.1. We do this just as above. Let g be such that $\|g\|_2 = 1$. Then

$$\int F(\mathcal{A}_k)\,g \leq \int \Big|\sum_{s\in\mathcal{A}_k}\langle f,\varphi_s\rangle\,\varphi_s\,g\Big| \lesssim \int \Big|\sum_{s\in\mathcal{A}_k}\sigma\,\chi_s^{(\infty)}\,g\Big| \lesssim \sigma\int_{\bigcup_{s\in\mathcal{A}_k} s} Mg \lesssim \sigma\,|R_k|^{1/2}.$$

So

$$\|F(\mathcal{A}_k)\|_2^2 \lesssim \sigma^2\,|\cup\mathcal{A}_k| \lesssim \sigma^2\,2^{2k}\,|\mathbf{top}(T)|.$$

Hence by Lemma 13.1,

$$II_k \lesssim 2^{-10k}\,\|F(\mathcal{A}_k)\|_2^2 \lesssim \sigma^2\,2^{-8k}\,|\mathbf{top}(T)|.$$

Summing in k proves $D(T) \lesssim \sigma^2|\mathbf{top}(T)|$, which finishes the proof.

13. Localized Bessel inequality

In this section we prove a Bessel inequality for 1-trees with functions supported away from the top of the tree. Specifically:

Lemma 13.1. *Let T be a 1-tree. For $k \geq 1$, let $R_k = 2^k \mathbf{top}(T)$. For $k \geq 1$, let $\Omega_k = R_k \setminus R_{k-1}$. Define $\Omega_0 = \mathbf{top}(T)$. Then for any $N > 0$,*

$$\sum_{s \in T} \left|\langle f\, \mathbf{1}_{\Omega_k}, \varphi_s \rangle\right|^2 \lesssim 2^{-Nk}\, \|f\, \mathbf{1}_{\Omega_k}\|_2^2.$$

Remark 13.2. For a classical one-dimensional tree, this can be proved by using the extreme spatial decay of the wave packets φ_s, $s \in T$, away from $\mathbf{top}(T)$. We use this in conjunction with orthogonality in the x-variable to handle interactions of the functions φ_s and $\varphi_{s'}$ horizontally close to the tree, where tail estimates do not improve for shorter tiles in the tree. This is the reason for the decomposition of Ω_k into $\mathbf{B}_k$ and $\mathbf{C}_k$ in the proof below.

Proof. For notational convenience, we will assume that the parallelogram $\mathbf{top}(T)$ is centered at the origin, has width 1, and has sides parallel to the coordinate axes. First note that

$$\sum_{s \in T} |\langle f\, \mathbf{1}_{\Omega_k}, \varphi_s \rangle|^2 = \sum_{s \in T} |\langle f\, \mathbf{1}_{\mathbf{B}_k}, \varphi_s \rangle|^2 + \sum_{s \in T} |\langle f\, \mathbf{1}_{\mathbf{C}_k}, \varphi_s \rangle|^2 =: B + C,$$

where

$$\mathbf{B}_k = \{(x, y) \in \Omega_k \colon |y| \geq 2^k\} \quad \text{and} \quad \mathbf{C}_k = \Omega_k \setminus \mathbf{B}_k.$$

To estimate B we will need to use orthogonality in the horizontal variable. To estimate C we will need only spatial decay, as in the one-dimensional case.

Note that, by Cauchy–Schwarz,

$$\begin{aligned} B^2 &= \int_{\mathbf{B}_k} f \sum_{s \in T} \langle f\, \mathbf{1}_{\mathbf{B}_k}, \varphi_s \rangle\, \varphi_s \\ &\leq \|f\, \mathbf{1}_{\Omega_k}\|_2 \Big(\sum_{s \in T} \sum_{s \in T'} \int_{|y| \geq 2^k} \int_{x \in \mathbb{R}} \langle f \mathbf{1}_{\mathbf{B}_k}, \varphi_s \rangle \langle f \mathbf{1}_{\mathbf{B}_k}, \varphi_{s'} \rangle \varphi_s(x, y) \varphi_{s'}(x, y)\, dx dy \Big)^{1/2}. \end{aligned}$$

Also note that if $|s| \neq |s'|$, then for every y, we have

$$\int_x \varphi_s(x, y) \varphi_{s'}(x, y) = 0.$$

This follows from the definition of the wave packets φ_s; specifically, note that $\pi_1(\mathrm{supp}(\hat{\varphi}_s)) \cap \pi_1(\mathrm{supp}(\hat{\varphi}_{s'})) = \emptyset$ whenever $\omega_{s,1} \cap \omega_{s,2} = \emptyset$, which happens whenever s and s' are in the same 1-tree and $|s| \neq |s'|$. By symmetry we may estimate $|\langle f\, \mathbf{1}_{\Omega_k}, \varphi_s \rangle \langle f\, \mathbf{1}_{\Omega_k}, \varphi_{s'} \rangle| \leq |\langle f\, \mathbf{1}_{\Omega_k}, \varphi_s \rangle|^2$, which gives us

$$\begin{aligned} \sum_{s \in T} \sum_{s' \in T} \int_{|y| \geq 2^k} \int_x \langle f\, \mathbf{1}_{B_k}, \varphi_s \rangle \langle f\, \mathbf{1}_{\Omega_k}, \varphi_{s'} \rangle\, \varphi_s(x, y)\, \varphi_{s'}(x, y) \\ \leq \sum_{s \in T} \sum_{s' \in T \colon |s| = |s'|} |\langle f\, \mathbf{1}_{B_k}\, \varphi_s \rangle|^2 \int_{|y| \geq 2^k} \int_x |\varphi_s|\, |\varphi_{s'}|. \end{aligned}$$

However, note that

$$\sum_{s'\in T:\,|s|=|s'|} \int_{|y|\geq 2^k} \int_x |\varphi_s|\,|\varphi_{s'}| \leq 2^{-Nk},$$

because the prototype φ is Schwartz, $s \in T$, and Ω_k is far away from $\mathbf{top}(T)$. Hence

$$B \lesssim 2^{-Nk/2}\, \|f\, \mathbf{1}_{\Omega_k}\|_2 \, \Big(\sum_{s\in T} |\langle f\, \mathbf{1}_{\Omega_k}\, \varphi_s\rangle|^2 \Big)^{1/2}.$$

We now estimate C. Define

$$T^j = \{s \in T\colon |s| = 2^{-j}|\mathbf{top}(T)|\}.$$

Note that if $s \in T^j$, then $|\langle f\mathbf{1}_{\mathbf{C}_k}, \varphi_s\rangle| \lesssim 2^{-Nk/2-50j}\, \|f\mathbf{1}_{\Omega_k}\|_2$ by Cauchy–Schwarz and the fact that $\|\varphi_s \mathbf{1}_{\mathbf{c}_k}\|_2 \lesssim 2^{-Nk/2-50j}$. This last claim follows from the fact that φ_s is highly localized on $\mathbf{top}(T)$, and because $\mathbf{C}_k$ is far away from $\mathbf{top}(T)$ horizontally. (Of course we could not make the same argument for B because we can do no better than $\|\varphi_s \mathbf{1}_{\mathbf{B}_k}\|_2 \lesssim 2^{-Nk}$ for $s \in T^j$; i.e., there is no decay in the parameter j.) This is already enough:

$$C \leq \sum_{j\geq 0} \sum_{s\in T^j} |\langle f\, \mathbf{1}_{\Omega_k}\, \varphi_s\rangle|^2 \lesssim 2^{-Nk/2}\, \|f\, \mathbf{1}_{\Omega_k}\|_2,$$

which finishes the proof of the lemma. □

14. Square function estimates

In this section we prove Lemma 10.2. The proof is similar to the standard proof of L^p boundedness for the analogous one-dimensional square function, with a few tweaks to handle the two-dimensionality. For notational convenience we will assume, without loss of generality, that the tree T has top that is axis parallel and centered at the origin. Proving the lemma with the spatial localization requires us to decompose Δ spatially as follows. For $k \geq 1$, define the set $\Omega_k = 2^k\mathbf{top}(T) \setminus 2^{k-1}\mathbf{top}(T)$. For $k = 0$, define $\Omega_k = \mathbf{top}(T)$. Now define

$$\Delta_k(f) = \Big(\sum_{s\in T} |\langle f, \mathbf{1}_{\Omega_k}\varphi_s\rangle|^2\, \frac{\mathbf{1}_s}{|s|} \Big)^{1/2}.$$

By Minkowski's inequality, we have

$$\Delta f(x) = \Big(\sum_{s\in T} \big|\big\langle f, \textstyle\sum_k \mathbf{1}_{\Omega_k}\varphi_s\big\rangle\big|^2\, \frac{\mathbf{1}_s}{|s|} \Big)^{1/2} \leq \sum_k \Delta_k\, f(x)$$

pointwise, so again by Minkowski's inequality we have

$$\|\Delta f\|_p \leq \sum_k \|\Delta_k f\|_p.$$

We will prove that for any N,

$$\|\Delta_k f\|_p \lesssim 2^{-Nk}\,\|\mathbf{1}_{\Omega_k} f\|_p. \tag{14.1}$$

With this, we can use Hölder's inequality to see that for any N, we have

$$\|\Delta f\|_p \lesssim \sum_k 2^{-Nk}\,\|\mathbf{1}_{\Omega_k} f\|_p \lesssim \Big(\sum_k 2^{-Nk}\int_{\Omega_k} |f|^p\Big)^{1/p} \lesssim \Big(\int |\beta_{N,T} f|^p\Big)^{1/p},$$

where $\beta_{N,T}$ is the function defined in the statement of Lemma 10.2, which finishes the proof of Lemma 10.2. It remains to prove (14.1). Note that Lemma 13.1 is exactly this when $p = 2$. By interpolation, it is enough to prove the following weak-type estimate:

$$|\{\Delta_k f > \lambda\}| \lesssim 2^{2k}\,\frac{\|f\|_1}{\lambda^1}.$$

By dividing the function f into $\lesssim 2^{2k}$ pieces, we may assume the support of f is contained in a translate of $\mathbf{top}(T)$. With this assumption, it is enough to prove for such f that

$$|\{\Delta_k f > \lambda\}| \lesssim \frac{\|f\|_1}{\lambda^1}.$$

Our argument proceeds more or less by the usual path of Calderón–Zygmund decomposition.

Denote by R_k the rectangle with the same center and length as R but 2^k times the height. Let $\widetilde{\mathcal{B}}$ be the collection of maximal rectangles of width w taken from the collection such that

$$\frac{1}{|R_k|}\int_{R_k} |f| > 2^{5k}\,\lambda,$$

and for each $R \in \widetilde{\mathcal{B}}$, let $R' = \pi_1(R) \times \pi_2(C\,\mathbf{top}(T))$. Then let $\mathcal{B} = \{R' \colon R \in \widetilde{\mathcal{B}}\}$. We can see already that $\sum_{R\in\widetilde{\mathcal{B}}} |R| \le \sum_{R\in\mathcal{B}} |R| \lesssim \|f\|_1/\lambda$. This follows from the weak (1,1) inequality for the Hardy–Littlewood maximal function, which holds for rectangles of fixed width: if we write, for $k \ge 0$,

$$\widetilde{\mathcal{B}}_k = \Big\{R \in \widetilde{\mathcal{B}} \colon \frac{1}{|R_k|}\int_{R_k} |f| > 2^{5k}\lambda\Big\},$$

then we have

$$\sum_{R\in\widetilde{\mathcal{B}}} |R| \lesssim \sum_{k\ge 0} 2^k\,\frac{\|f\|_1}{2^{5k}\lambda} \lesssim \frac{\|f\|_1}{\lambda}.$$

For each $(x, y) \in R$, let

$$b(x, y) = f(x, y) - \frac{1}{|\pi_1(R)|}\int_{\pi_1(R)} f(z, y)\,dz.$$

Note that by definition we have that, for each $y \in \pi_2(\mathbf{top}(T))$,

$$\int_{\pi_1(R)} b(x, y)\,dx = 0.$$

We also have the following helpful fact:

Claim 14.1. *For each $y \in \pi_2(C\,\mathbf{top}(T))$, we have*

$$\frac{1}{|\pi_1(R)|}\int_{\pi_1(R)} f(z,y)\,dz \le C\,\lambda.$$

Proof of Claim. Note that $\widehat{f}$ is supported in the annulus of width $1/w$. Let k be a function such that $\widehat{k}(\xi) = 1$ for $\xi \in [-4w, 4w]$. Then

$$f(x,y) = \int f(x,w)\,k(y-w)\,dw,$$

so

$$\frac{1}{|\pi_1(R)|}\int_{\pi_1(R)} |f(z,y)|\,dz = \frac{1}{|\pi_1(R)|}\int_{\pi_1(R)} \Big|\int f(z,w)k(y-w)\,dw\Big|\,dz.$$

Because k rapidly decays away from a rectangle of height w, if we denote by R_j the rectangle with same center and length as R but 2^j times the height, then

$$\frac{1}{|\pi_1(R)|}\int_{\pi_1(R)} \Big|\int f(z,w)\,k(y-w)\,dw\Big|\,dz$$
$$\lesssim \frac{1}{|\pi_1(R)|}\int_{\pi_1(R)} \sum_{j\ge 1} \frac{1}{2^j}\int_{2^{j-1}}^{2^j} f(z,w)\,2^{-10j}\,dw\,dz \le \lambda,$$

where the last inequality is by assumption on R. □

With this claim, we define

$$g(x,y) = f(x,y) \quad \text{for } (x,y) \notin \bigcup_{R\in\mathcal{B}} R$$

and

$$g(x,y) = \frac{1}{|\pi_1(R)|}\int_{\pi_1(R)} f(z,y)\,dz \quad \text{for } (x,y) \in R \in \mathcal{B}.$$

Note that by the claim we have $g(x,y) \lesssim \lambda$ for $(x,y) \in R$. Further, for almost every $(x,y) \notin \bigcup_{R\in\mathcal{B}} R$ such that $g(x,y) = f(x,y) \gg \lambda$, there exists a horizontal line segment L through (x,y) such that $\frac{1}{|L|}\int_L f \gg \lambda$, which implies there is a rectangle of width w containing (x,y) on which the average of f is larger than λ, contradicting our assumption that $(x,y) \notin \bigcup_{R\in\mathcal{B}} R$. Hence $g \lesssim \lambda$ almost everywhere.

To see the purpose of including the rectangles $5CR'$ in the exceptional set (rather than a small dilate of R itself), consider a rectangle R north of the tree T, and a mean zero function h supported on R. Analysis of $\int_{(5CR)^c} \Delta h$ is a bit more complicated than in the one-dimensional case because the collection $\{\varphi_s\}_{s\in T}$ has no orthogonality in the vertical direction. However by excluding R', we need only consider small tiles s supported away from the vertical translate of $5CR$, allowing us to take advantage of the spatial decay (in the horizontal variable) of the functions φ_s.

With this modification, the proof now proceeds as expected: use the fact that $|g| \lesssim \lambda$, together with the L^2 estimate on Δ, to see that

$$|\{\Delta_k g > \lambda\}| \lesssim \frac{\int |g|^2}{\lambda^2} \lesssim \frac{\|f\|_1}{\lambda}.$$

Additionally, by the Chebyshev and triangle inequalities, together with the sublinearity of Δ_k, we have

$$\Big|\Big\{x \notin E \colon \Delta_k\big(\sum_R b_R\big) > \lambda\Big\}\Big| \leq \frac{1}{\lambda} \sum_R \int_{(5CR')^c} |\Delta_k(b_R)|.$$

To finish the proof we show that for each $R \in \mathcal{B}$, we have

$$\int_{(5CR')^c} |\Delta_k(b_R)| \lesssim \int |b_R|, \tag{14.2}$$

which will give us that

$$\Big|\Big\{x \notin E \colon \Delta\big(\sum_R b_R\big) > \lambda\Big\}\Big| \lesssim \frac{1}{\lambda} \sum_R \int |b_R| \lesssim \sum_R |R| \lesssim \frac{\|f\|_1}{\lambda}.$$

Once again, to prove (14.2), we essentially follow the one-dimensional argument, dealing with a few extra nuisances along the way. A reader having trouble seeing through the technicalities should note that all of the computations below are essentially the same as in the one-dimensional case. The problem is understanding why the present situation is essentially the same as the one-dimensional case. More specifically, to prove (14.2), it is convenient to make a few simplifying (and valid) assumptions. For each parallelogram $s \in T$ define

$$\tilde{s} = \pi_1(s) \times C\pi_2(\mathbf{top}(T)).$$

Since $s \subseteq \tilde{s}$, it is clear that if we define

$$\tilde{\Delta}_k f = \Big(\sum_{s \in T} |\langle f\, \mathbf{1}_{\Omega_k}, \varphi_s\rangle|^2 \,\frac{\mathbf{1}_{\tilde{s}}}{|s|}\Big)^{1/2}, \tag{14.3}$$

then $\Delta_k f \leq \tilde{\Delta}_k f$ pointwise. For each $s \in T$, we know that $\pi_1(\tilde{s})$ is contained in the union of two dyadic intervals $\tilde{s}_L$ and $\tilde{s}_R$ each of size $\lesssim \pi_1(\tilde{s})$. Further, because the set of tiles of a given size and orientation partition $\mathbb{R}^2$ (i.e., for each $\omega \in \mathcal{D}$, we have $\bigcup_{R \in \mathcal{U}_\omega} R = \mathbb{R}^2$; see the definitions in Section 3), and because $|\pi_1(s)| \geq |\pi_2(s)|$ we know that for any dyadic interval I, there are $\lesssim 1$ tiles $s \in T$ such that $I = \pi_1(\tilde{s_L})$ or $I = \pi_1(\tilde{s_R})$. All of this allows us to assume (possibly after dividing T into ~ 1 pieces) that the tiles s are parameterized by dyadic intervals, and that for each $x \in C\,\mathbf{top}(T)$, and each dyadic interval I, there is at most one $s \in T$ such that $x \in \tilde{s}$ and $\pi_1(\tilde{s}) = I$.

To prove (14.2), we split the sum inside Δf into two pieces, one over tiles whose vertical projection is smaller than the length of R, and the other over tiles whose vertical projection is larger than the length of R. We begin by controlling the sum over smaller tiles. Note that the dominant term in both cases comes from tiles such that $|\pi_1(s)| \sim |\pi_1(R)|$. In the integral below, we need only consider $x \in C\mathbf{top}(T)$ such that $\pi_1(x) \notin \pi_1(5CR)$. This allows us to prove the desired estimate using spatial decay alone. Further, since $\mathbf{1}_{\tilde{s}}(x)$ is constant on vertical segments projecting to $\pi_2(C\mathbf{top}(T))$, we have

$$\begin{aligned}\int_{x\in K\mathbf{top}(T)\cap(5CR')^c} &\Big(\sum_{|\pi_1(s)|\leq|\pi_1(R)|} |\langle b_R,\varphi_s\rangle|^2\,\frac{\mathbf{1}_s}{|s|}\Big)^{1/2}\\ &\lesssim \int_{x\in K\mathbf{top}(T)\cap(5CR')^c}\Big(\frac{\|b_R\|_1^2}{|R|^2}\Big(\frac{|x-c(R)|}{|\pi_1(R)|}\Big)^{-10}\Big)^{1/2}\\ &\lesssim \frac{\|b_R\|_1}{|\pi_1(R)|}\int_{t\in\mathbb{R}:\ |t|\geq 5|\pi_1(R)|}\frac{1}{|t/|\pi_1(R)||^5}\,dt \lesssim \|b_R\|_1.\end{aligned}$$

We emphasize that the integral in the last line is one-dimensional. It remains to control the sum over the tiles with vertical projection larger than $|\pi_1(R)|$. This requires using the mean-zero-along-horizontal-line-segments property of the function b_R. Note that for any smooth function h, we have

$$\begin{aligned}\langle b_R,h\rangle &= \int_{y\in\pi_2(R)}\int_{x\in\pi_1(R)} b_R(x,y)h(x,y)dxdy\\ &\leq \int_{y\in\pi_2(R)}\int_{x\in\pi_1(R)} |b_R(x,y)||h(x,y)-h(c_{\pi_1(R)},y)|dxdy.\end{aligned}$$

Our goal is to apply this to the wave packets φ_s. Specifically, we will show

Claim 14.2.

$$|\langle b_R,\varphi_s\rangle| \lesssim \|b_R\|_1\,\frac{1}{|s|^{1/2}}\,\frac{|\pi_1(R)|}{|\pi_1(s)|}\,\min\Big(1,\Big(\frac{|x-c(R)|}{|\pi_1(s)|}\Big)^{-10}\Big)$$

Proof. We must deal with a small technicality here: the tiles s need not be precisely axis parallel, but fortunately they are close. Precisely, we have that the vertical component (when using the coordinate frame of s) of $(x,y)-(c_{\pi_1(R)},y)$ is less than $w|\pi_1(R)|/|\pi_1(s)|$. Of course we have the horizontal component (when using the coordinate frame of s) of $(x,y)-(c_R,y)$ is less than $|\pi_1(R)|$. Further, we know that

$$\begin{aligned}D_1\varphi_s(x,y) &\leq \frac{1}{\sqrt{|s|}}\,\frac{1}{|\pi_1(s)|}\,\frac{\partial}{\partial x}\,\varphi\Big(\frac{x}{|\pi_1(s)|},\frac{y}{w}\Big)\\ D_2\varphi_s(x,y) &\leq \frac{1}{\sqrt{|s|}}\,\frac{1}{w}\,\frac{\partial}{\partial y}\,\varphi\Big(\frac{x}{|\pi_1(s)|},\frac{y}{w}\Big).\end{aligned}$$

Hence

$$\begin{aligned}|\varphi_s(x,y)-\varphi_s(c(\pi_1(R)),y)| &\lesssim \frac{w\,|\pi_1(R)|}{|\pi_1(s)|}\,\frac{1}{\sqrt{|s|}}\,\frac{1}{w}\,\frac{\partial}{\partial y}\,\varphi\Big(\frac{x}{|\pi_1(s)|},\frac{y}{w}\Big)\\ &\quad + |\pi_1(R)|\,\frac{1}{\sqrt{|s|}}\,\frac{1}{|\pi_1(s)|}\,\frac{\partial}{\partial x}\,\varphi\Big(\frac{x}{|\pi_1(s)|},\frac{y}{w}\Big).\square\end{aligned}$$

The claim yields, writing $\Gamma = K\mathbf{top}(T) \cap (5CR')^c$,

$$\int_\Gamma \Big(\sum_{|\pi_1(s)|>|\pi_1(R)|} |\langle b_R, \varphi_s\rangle|^2 \frac{\mathbf{1}_{\tilde{s}}(x)}{|s|}\Big)^{1/2} dx$$

$$\lesssim \int_\Gamma \|b_R\|_1 \, |\pi_1(R)| \Bigg(\sum_{|\pi_1(s)|>|\pi_1(R)|} \Bigg(\frac{\min\Big(1, \Big(\frac{|x-c(R)|}{|\pi_1(s)|}\Big)^{-10}\Big)}{|\pi_1(s)| \, |s|^{1/2}}\Bigg)^2 \frac{\mathbf{1}_{\tilde{s}}(x)}{|s|}\Bigg)^{1/2} dx$$

$$\lesssim \|b_R\|_1.$$

This completes the proof of (14.2) and thus the proof of Lemma 10.2.

15. BMO type estimates for the square function

In this section we prove Lemma 10.3. As in the previous section, we consider the related operator $\tilde{\Delta}$. See (14.3) for the definition, as well as the discussion immediately following the definition for several simplifying assumptions that we make. To prove the lemma, we prove the following key claim. Here, and in the rest of the proof, we write $\sigma = \mathbf{size}(T)$; note that we also have

$$\sigma \sim \Big(\frac{1}{|\mathbf{top}(T)|} \sum_{s\in T} |\langle f, \varphi_s\rangle|^2\Big)^{1/2}.$$

As in the last section, we consider a slightly modified version of Δ: define

$$\tilde{\Delta} f = \Big(\sum_{s\in T} |\langle f, \varphi_s\rangle|^2 \, \frac{\mathbf{1}_{\tilde{s}}}{|s|}\Big)^{1/2}$$

where the rectangles $\tilde{s}$ are defined immediately above (14.3).

Claim 15.1.

$$\big|\{\tilde{\Delta} f > \sigma n\}\big| \lesssim 2^{-n^2} \big|\{\tilde{\Delta} f > \sigma\}\big|.$$

(*Of course we do not need the full exponential-squared decay, but we do have it.*)

With the claim, we are almost done:

$$\begin{aligned}
\|\tilde{\Delta} f\|_2^2 &\lesssim \int_{\{\tilde{\Delta} f \le \sigma\}} (\tilde{\Delta} f)^2 + \sum_n \sum_{n=1}^{\infty} (\sigma n)^2 |\{\tilde{\Delta} f > n\sigma\}| \\
&\lesssim \int_{\{\tilde{\Delta} f \le \sigma\}} (\tilde{\Delta} f)^2 + \sum_n \sum_{n=1}^{\infty} (\sigma n)^2 |2^{-n^2} \{\tilde{\Delta} f > \sigma\}| \\
&\lesssim \int_{\{\tilde{\Delta} f \le \sigma\}} (\tilde{\Delta} f)^2 + \sigma^2 |\{\tilde{\Delta} f > \sigma\}| \\
&\lesssim \sigma \int_{\{\tilde{\Delta} f \le \sigma\}} \tilde{\Delta} f + \sigma \int_{\{\tilde{\Delta} f > \sigma\}} \tilde{\Delta} f = \sigma \int \tilde{\Delta} f.
\end{aligned}$$

With this, we see that

$$\sigma^2|\mathbf{top}(T)| \sim \|\tilde{\Delta} f\|_2^2 \lesssim \sigma \int \tilde{\Delta} f,$$

which proves that

$$\|\tilde{\Delta} f\|_2 \sim \sigma|\mathbf{top}(T)|^{1/2} \lesssim \frac{1}{|\mathbf{top}(T)|^{1/2}} \int \tilde{\Delta} f,$$

which is what we need. It remains to prove the claim.

Proof of Claim 15.1. Of course to prove the claim it is enough to show that

$$|\{\tilde{\Delta} f > \sqrt{n}\sigma\}| \lesssim 2^{-n}\, |\{\tilde{\Delta} f > \sigma\}|,$$

and this is equivalent to showing

$$|\{(\tilde{\Delta} f)^2 > n\sigma^2\}| \lesssim 2^{-n} |\{(\tilde{\Delta} f)^2 > \sigma^2\}|,$$

which can be shown in a rather straightforward manner following the proof of the John–Nirenberg inequality. Recall that for each dyadic I we have an associated tile in T, which we call $s(I)$. For notational convenience, define for intervals I and K

$$a_{I,K}(x) = \sum_{I \subseteq J \subseteq K} |\langle f, \varphi_{s(J)}\rangle|^2 \, \frac{\mathbf{1}_s(x)}{|s(J)|}.$$

We first note that for any K, if I is a maximal interval on which

$$a_{I,K} > m\sigma^2,$$

then we know

$$a_{I,K} < (m+2)\sigma^2,$$

since

$$|\langle f, \varphi_{s(I)}\rangle|^2 \frac{1}{|s(I)|} \leq \sigma^2.$$

We begin by defining a collection of intervals $\mathcal{I}_0$:

$$\mathcal{I}_0 = \{\, \text{maximal dyadic } I \colon a_{I,\pi_1(C\mathbf{top}(T))} > 100\sigma^2\}.$$

Then having defined $\mathcal{I}_{n-1}$, define for any $K \in \mathcal{I}_{n-1}$,

$$\mathcal{I}_n(K) = \{\, \text{maximal dyadic } I \colon a_{I,K} > 100\sigma^2\}$$
$$\mathcal{I}_n = \bigcup_{K \in \mathcal{I}_{n-1}} \mathcal{I}_n(K).$$

We remark that for any K,

$$\bigcup_{I \in \mathcal{I}_n(K)} |I| \leq \frac{1}{2}\, |K|.$$

To see this we only need to use the Chebyshev inequality, and the estimate on **size**(T):

$$\Big|\bigcup_{I\in\mathcal{I}_n(K)} I\Big| \le \frac{1}{10\sigma^2}\int a_{I,K} \le \frac{1}{10\sigma^2}\sum_{J\subseteq K} |\langle f,\varphi_{s(J)}\rangle|^2 \le \frac{1}{10}|K|,$$

where the last inequality is due to the estimate on **size**(T). Similarly,

$$\Big|\bigcup_{I\in\mathcal{I}_0} I\Big| \le \frac{1}{2}\big|\pi_1(C\mathbf{top}(T))\big|.$$

Putting together all K in $\mathcal{I}_{n-1}$ gives us that

$$\bigcup_{I\in\mathcal{I}_n} |I| \le \frac{1}{2}\bigcup_{I\in\mathcal{I}_{n-1}} |I|,$$

and iterating this gives us that

$$\bigcup_{I\in\mathcal{I}_n} |I| \le 2^{-n}\bigcup_{I\in\mathcal{I}_0} |I|,$$

which proves Claim 15.1 since

$$(\tilde{\Delta} f)^2(x) \lesssim n\sigma^2$$

for x such that $\pi_1(x)\notin\bigcup_{I\in\mathcal{I}_n} I$. □

16. Appendix. The case $p > 2$

In this appendix we briefly discuss the proof of Theorem 2.1 for $p > 2$, which is essentially the proof in [6].

Following the tree decomposition of Section 5 and the remarks in Section 6, we need to show

$$\sum_\delta\sum_\sigma\sum_{T\in\mathcal{T}_{\delta,\sigma}} \delta\sigma\,|\mathbf{top}(T)| \lesssim |F|^{1/p}\,|E|^{1-1/p}.$$

This time we care most about p close to ∞. We may assume $|E|\le|F|$ because if $|E|>|F|$ then we may apply the previous arguments for the case $p\le 2$. We emphasize here that there is no circularity. Both the argument in this section (in which we assume $|E|\le|F|$) and the argument in the bulk of the paper (in which we assume $|E|\ge|F|$) work when $p=2$. Hence the $p=2$ case of the estimate in (3.3) is established for arbitrary E and F. This allows us to assume $|E|\le|F|$ in this section, where $p\ge 2$, and allows us to assume $|E|\ge|F|$ in the earlier part of the paper, where $p\le 2$.

By Estimates 5.3 and 5.4 it suffices to prove

$$\sum_\delta\sum_\sigma\sum_{T\in\mathcal{T}_{\delta,\sigma}} \delta\sigma\min\Big(\frac{|E|}{\delta},\frac{|F|}{\sigma^2}\Big) \lesssim |F|^{1/p}\,|E|^{1-1/p} \tag{16.1}$$

for $p\ge 2$. The following simple estimate will be helpful:

Claim 16.1. *For any δ, we have*

$$\sum_{\sigma} \delta\sigma \min\left(\frac{|E|}{\delta}, \frac{|F|}{\sigma^2}\right) \lesssim \sqrt{\delta|E||F|}.$$

Proof. We need only observe that the two terms in the minimum are equal when $\sigma = \sqrt{\delta|F|/|E|}$ and split the sum over σ accordingly. □

We split the sum (16.1) in δ into two pieces, with the dividing line being $\delta = |E|/|F|$. For smaller δ, we use Claim 16.1 above:

$$\sum_{\delta \le |E|/|F|} \sum_{\sigma} \delta\sigma \min\left(\frac{|E|}{\delta}, \frac{|F|}{\sigma^2}\right) \lesssim \sum_{\delta \le |E|/|F|} \sqrt{\delta|E||F|} \lesssim |E|.$$

For larger δ, we use the estimate $\mathbf{size} \lesssim 1$:

Claim 16.2. *If the function in the definition of* $\mathbf{size}(T)$ *is called* f*, then*

$$\mathbf{size}(T) \lesssim \|f\|_\infty.$$

Of course we are using $f = \mathbf{1}_F$, which proves that here $\mathbf{size}(T) \lesssim 1$.

Proof. For $k \ge 1$, define

$$\begin{aligned}\Omega_0 &= \mathbf{top}(T)\\ \Omega_k &= 2^k\,\mathbf{top}(T) \setminus 2^{k-1}\,\mathbf{top}(T).\end{aligned}$$

We need only note that for any 1-tree T, by Lemma 13.1,

$$\begin{aligned}\left(\sum_{s\in T} |\langle f, \varphi_s\rangle|^2\right)^{1/2} &\le \sum_k \left(\sum_{s\in T} |\langle \mathbf{1}_{\Omega_k} f, \varphi_s\rangle|^2\right)^{1/2}\\ &\lesssim \sum_k 2^{-Nk}\, \|\mathbf{1}_{\Omega_k} f\|_2^2 \lesssim \|f\|_\infty^2\, |\mathbf{top}(T)|\end{aligned}$$

since $|\Omega_k| \lesssim 2^{2k}\,|\mathbf{top}(T)|$. This proves the claim. □

Hence

$$\sum_{\delta \ge |E|/|F|} \sum_{\sigma \le 1} \delta\sigma \frac{|E|}{\delta} \lesssim |E| \log \frac{|F|}{|E|}.$$

Combining these two estimates proves (16.1) since $|E| \le |F|$.

References

[1] Bateman, M.: Kakeya sets and directional maximal operators in the plane. *Duke Math. J.* **147** (2009), no. 1, 55–77.

[2] Bateman, M.: L^p estimates for maximal averages along one-variable vector fields in $\mathbb{R}^2$. *Proc. Amer. Math. Soc.* **137** (2009), 955–963.

[3] Bateman, M.: Maximal averages along a planar vector field depending on one variable. *Trans. Amer. Math. Soc.* **365** (2013), no. 8, 4063–4079.

[4] Bateman, M. and Thiele, C.: L^p estimates for the Hilbert transform along a one variable vector field. ArXiv: 1109.6396 [math.CA]. To appear in *Anal. PDE.*

[5] Katz, N. H.: Maximal operators over arbitrary sets of direction. *Duke Math. J.* **97** (1999), no. 1, 67–79.

[6] Lacey, M. and Li, X.: Maximal theorems for the directional Hilbert transform on the plane. *Trans. Amer. Math. Soc.* **358** (2006), 4099–4117.

[7] Lacey, M. and Li, X.: On a conjecture of E. M. Stein on the Hilbert transform on vector fields. *Mem. Amer. Math. Soc.* **205** (2010), no. 965.

[8] Lacey, M. and Thiele, C.: A proof of boundedness of the Carleson operator. *Math. Res. Lett.* **7** (2000), no. 4, 361–370.

[9] Nagel, A., Stein, E. M., and Wainger, S.: Differentiation in lacunary directions. *Proc. Nat. Acad. Sci. USA* **75** (1978), 1060–1062

[10] Sjögren, P. and Sjölin, P.: Littlewood–Paley decompositions and Fourier multipliers with singularities on certain sets. *Ann. Inst. Fourier (Grenoble)* **31** (1981), no. 1, 157–175.

[11] Stein, E. and Street, B.: Multi-parameter singular Radon transforms. *Math. Res. Lett.* **18** (2011), no. 2, 257–277.

[12] Strömberg, J.-O.: Maximal functions for rectangles with given directions. Thesis, Mittag-Leffler Inst., Djursholm, Sweden.

Received December 2, 2011; revised February 20, 2013.

Michael Bateman: UCLA Mathematics Department, Box 951555, Los Angeles, CA 90095-1555, USA;

and

Department of Pure Mathematics and Mathematical Statistics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK.
E-mail: bateman@math.ucla.edu, m.bateman@dpmms.cam.ac.uk

The author was partially supported by NSF grant DMS-0902490.