On the Bilinear Hilbert Transform

Michael T. Lacey[1]

Abstract.

In joint work with C. Thiele, the author has shown that A. Calderón's bilinear Hilbert transform extends to a bounded operator on certain products of L^p spaces. This article illustrates the method of proof by giving a complete proof of an inequality which is slightly weaker than the original conjecture of Calderón concerning this transform.

1991 Mathematics Subject Classification: 42A50
Keywords and Phrases: hilbert transform, time–frequency analysis

1 Introduction

This note discusses a recently developed theory for the bilinear Hilbert transform, defined by

$$Hfg(x) := \lim_{\epsilon \to 0} \int_{|y|>\epsilon} f(x+y)g(x-y)\,\frac{dy}{y}.$$

A conjecture of A. Calderón concerned possible extensions of this transform to a bounded operator on products of L^p spaces. In this regard, note that the term dy/y is dimensionless, so that any inequalities satisfied by Hfg are those of Hölder's inequality. In collaboration with C. Thiele, the author has established

Theorem 1.1. *H extends to a bounded operator on $L^p \times L^q$ into L^r if*

$$1/p + 1/q = 1/r, \qquad 1 < p, q \leq \infty, \qquad 2/3 < r < \infty.$$

In particular, H maps $L^2 \times L^2$ into L^1, which is notable as the linear Hilbert transform does not preserve L^1. This was the form of Calderón's origonal conjecture dating from 1964.

The interest in the Theorem lies in the method of proof, as it can be seen as an outgrowth and rëexamination of a group of sophisticated and subtle techniques invented first by L. Carleson [1] and later by C. Fefferman [3]—those in the celebrated proof of the pointwise convergence of Fourier series. The central point is to

[1]The author has been supported by an NSF Grant, DMS—9706884.

exploit orthogonality in a situation which is highly sensitive to the temporal and frequency aspects of the functions under consideration.

To explain a significant part of the method of proof, we limit our discussion to a single instance of the Theorem above, one that is free of some of the technicalities of the general case treated in [4, 5, 6, 7]. The next section provides background for the rest of the paper. Then we present the geometric–combinatorial model of the bilinear Hilbert transform and prove that it maps $L^2 \times L^2$ into weak L^1.

I am grateful to L. Grafakos for taking the time to carefully read this manuscript and providing me with a long list of improvements.

2 Preliminaries

The Fourier transform is taken to be

$$\mathcal{F}f(\xi) = \hat{f}(\xi) := \int e^{-2\pi i x\xi} f(x)\, dx.$$

We refer to x as the time variable and ξ as the frequency variable. The inner product will be denoted by $\langle f, g\rangle = \int f\bar{g}\, dx$.

The linear Hilbert transform is given by the principal value of convolution with $1/y$.

$$Hf(x) := \lim_{\epsilon\to 0} \frac{1}{\pi} \int_{|y|>\epsilon} f(x-y)\, \frac{dy}{y}.$$

Context will distinguish linear and bilinear forms of the transform. We recall that

$$\mathcal{F}Hf(\xi) = \mathcal{F}f(\xi) \lim_{\epsilon\to 0} \frac{1}{\pi} \int_{\epsilon<|y|<1/\epsilon} e^{-2\pi i y\xi}\, \frac{dy}{y} = i\mathcal{F}f(\xi)\ \mathrm{sign}(\xi).$$

But $\mathcal{F}$ is unitary, so that H is bounded on L^2. This calculation also shows that the singularity of $1/y$ has the effect of distinguishing the origin in frequency—sign(ξ) has a jump discontinuity at 0.

Deeper properties of H are addressed by decomposing the non–local kernel $1/y$ into a sum of localized kernels. Fix a symmetric Schwartz function ψ which resolves the identity in that

$$\sum_{j=-\infty}^{\infty} \psi(2^j y) \equiv 1, \qquad y \neq 0. \tag{2.2}$$

Then, let $\psi_j(y) = y^{-1}\psi(2^j y)$, which is the same as $\psi_j(y) = 2^j\psi_1(2^j y)$. Each ψ_j gives rise to an operation

$$H_j f(x) = \int f(x-y)\psi_j(y)\, dy. \tag{2.3}$$

This is a local operation in that ψ_j is no longer a singular kernel. In fact, $|\psi_j(y)| \leq C2^j$, and it decays rapidly for $|y| > 2^{-j}$. Each H_j trivially maps L^p into itself for $1 \leq p \leq \infty$. We say that H_j as *scale* 2^j.

Carleson's Theorem can now be recalled in a form specific to our purposes. To prove the pointwise convergence of Fourier series, the maximum partial Fourier sums must be controlled. On the real line, this is equivalent to providing a control on the maximal operator

$$\mathcal{C}f(x) := \sup_{\lambda} |H[e^{2\pi i\lambda\cdot} f](x)|.$$

Carleson proved that $\mathcal{C}$ maps L^2 into itself. See [1].

The aspect of $\mathcal{C}$ that distinguishes it from other operators of harmonic analysis is its invariance under conjugation of f by exponentials. Conjugation being dual to translation, we see that $\mathcal{C}$ has no distinguished point in frequency, which is a feature shared with the bilinear Hilbert transform. Denote, for the moment, $f^\lambda(x) = e^{2\pi i\lambda x} f(x)$. Then one readily sees that $Hf^\lambda g^\lambda(x) = e^{4\pi i\lambda x} Hfg(x)$. That is, Hfg commutes with conjugation, and so again it has no distinguished points in frequency. This suggests that the analysis of the two operators should be intimately related.

It also suggests that such an analysis cannot distinguish points in the frequency variable. We follow such a path. Each scale of the $1/y$ kernel requires a separate decomposition of f, g that is sensitive to both temporal and frequency aspects of the functions. It is convienent to introduce this with elements of the geometry of the phase plane.

Let φ be a Schwartz function on L^2 norm one, with Fourier transform supported in $[-\frac{1}{2}, \frac{1}{2}]$, so that

$$|\hat{\varphi}(\xi)|^2 + |\hat{\varphi}(\xi + \tfrac{1}{2})|^2 \equiv \text{constant} \qquad -\tfrac{1}{2} \le \xi < \tfrac{1}{2}.$$

Let $s = I_s \times \omega_s$ be a rectangle of area 1 and set

$$\varphi_s(x) := e^{2\pi i c(\omega_s)x} |I_s|^{-1/2} \varphi\Big(\frac{s - c(I_s)}{|I_s|}\Big), \tag{2.4}$$

where $c(J)$ denotes the center of the interval J. With this definition, I_s plays the role of the temporal variable and ω_s of the frequency. The initial set of rectangles we are interested in are

$$\mathbf{S}_1 := \{(n, n+1) \times (m/2, m/2+1) \mid m, n \in \mathbb{Z}\},$$

which have scale 1. To make our considerations independent of scale, for each integer j, let $\mathbf{S}_j$ be the rectangles which are images of those in $\mathbf{S}_1$ under the area preserving dilation $(x, \xi) \to (2^{-j}x, 2^j\xi)$. And set $I_j f = \frac{1}{2}\sum_{s\in\mathbf{S}_j} \langle f, \varphi_s\rangle \varphi_s$. Each I_j is just a change of scale applied to I_1. And I_1 is in fact the identity on L^2. See [2, Section 3.4.4] for a full discussion of this fact.

Now, for the transform, take a particular scale of the transform, as given in (2.3). We have

$$H_j fg = I_j[\, H_j(I_j f, I_j g)\,] = \sum_{s1,s2,s3\in\mathbf{S}_j} \langle f, \varphi_{s1}\rangle \langle g, \varphi_{s2}\rangle \langle H_j \varphi_{s1}\varphi_{s2}, \varphi_{s3}\rangle. \tag{2.5}$$

The point underlying this expression is that the triple sum over $\mathbf{S}_j$ diagonalizes, and the "diagonal terms," collected over 2^j, have useful combinatorial structures.

The significant aspect of the diagonalization takes place on the Fourier side, which is made explicit by choosing the kernel ψ in (2.2) so that $\mathcal{F}\psi$ is supported on $(-10,-4] \cup [4,10)$. Recall that φ is supported on $[-\frac{1}{2},\frac{1}{2}]$ and therefore φ_s is supported on ω_s.

Then the final inner product in (2.5) is determined in part by the Fourier support of $H_j\varphi_{s1}\varphi_{s2}$. We claim that this function is zero unless $\pm(\omega_{s1}-\omega_{s2})\cap[4\cdot 2^j, 10\cdot 2^j) \neq 0$, and then the Fourier transform is supported in $\omega_{s1}+\omega_{s2}$. That is, the three frequency intervals ω_{sk} depend in fact on only one parameter. A formal calculation verifies the claim. Expand φ_{s1} in a dual variable τ_1 and φ_{s2} in τ_2, then

$$H_j\varphi_{s1}\varphi_{s2}(x) = \iint e^{2\pi i(\tau_1+\tau_2)x}\widehat{\varphi_{s1}}(\tau_1)\widehat{\varphi_{s2}}(\tau_2)\widehat{\psi}(2^{-j}(\tau_1-\tau_2))\, d\tau_1 d\tau_2.$$

In this expression, recall that $\mathcal{F}\varphi_s$ is supported on ω_s. Note that $\tau_1+\tau_2$ is the frequency variable for the left hand side, and then our claims follow.

The second diagonalization is of a more trivial nature. Recall that φ is a Schwartz function, so that φ_s is highly localized around I_s. In addition, ψ has rapid decay away from the origin. From these considerations, it follows that for all $s1, s2, s3$,

$$|\langle H_j\varphi_{s1}\varphi_{s2},\varphi_{s3}\rangle| \leq C_N|I_{s1}|^{-1/2}[1+|I_{s1}|^{-1}\max_{j\neq k}\operatorname{dist}(I_{sj},I_{sk})]^{-N}, \quad N\geq 1.$$

A final diagonalization procedure, which we do not present here, reduces the sum over j in (2.5) to a sum of "model sums" as defined in the next section. This is worth doing because the model sums can be analysized using only natural geometric–combinatorial considerations: There are no *ad hoc* features of the argument for the model sums.

3 Model Sums

In this section we state and prove a theorem which is general enough to prove the most important single inequality for the bilinear Hilbert transform. But this will require some definitions—we give them and illustrate with special cases which contain all of the difficulty of the general case.

A collection of intervals $\mathcal{G}$ is a *grid* if for all $I,I'\in\mathcal{G}$, we have $I\cap I'$ equal to $\emptyset$, I or I', and $I\subset_{\neq} I'$ implies $2|I|\leq|I'|$. The special cases of interest are collections of dyadic and triadic intervals.

A collection $\mathbf{S}$ of rectangles $s=R_s\times\rho_s$ are called *tiles* if for all $s,s'\in\mathbf{S}$ we have $|s|\leq 4|s'|$, and in addition, $\{R_s \mid s\in\mathbf{S}\}$ and $\{\rho_s \mid s\in\mathbf{S}\}$ are grids. For an example consider triadic rectangles $s=R_s\times\rho_s$ of area one. Each ρ_s is a union of three triadic intervals of equal length, $\rho_{s1},\rho_{s2},\rho_{s3}$. Then all of the rectangles $R_s\times\rho_s, R_s\times\rho_{sj}$, $1\leq j\leq k$ form a collection of tiles.

This is most relevant, because of the connection to the decomposition in (2.5), and the diagonalization that was discussed there. For that reason, we make a further definition along these lines: A collection of tiles $\mathbf{S}$ are called *tri–tiles* if to each $s=R_s\times\rho_s\in\mathbf{S}$ there are three tiles $sj=R_s\times\rho_{sj}$, $1\le j\le 3$, with these properties. For all $s\in\mathbf{S}$ and all j, (a) $\rho_{sj}\subset\rho_s$, (b) $\xi_1<\xi_2<\xi_3$, for all $\xi_j\in\rho_{sj}$, and (c) $\{s,s1,s2,s3\mid s\in\mathbf{S}\}$ is a collection of tiles.

The tri–tiles will describe the location of functions in the time–frequency plane. A collection of functions $\{\phi_{sj}\mid s\in\mathbf{S},\ 1\le j\le 3\}$ are *adapted* to a collection of tri–tiles $\mathbf{S}$ if for all $s\in\mathbf{S}$ and all j, $\|\phi_{sj}\|_2\le 1$, there is an affine map α_j on $\mathbb{R}$ so that $\mathcal{F}\phi_{sj}$ is suported on $\alpha_j(\rho_{sj})$, and we have

$$|\phi_{sj}(x)|\le\frac{C_0}{\sqrt{|R_s|}}\Big(1+\frac{|x-c(I_s)|}{|R_s|}\Big)^{-10}, \tag{3.6}$$

$$\langle\phi_{sj},\phi_{s'j}\rangle=0\quad\text{if } I_s\ne I_{s'},\ \omega_{sj}=\omega_{s'j}. \tag{3.7}$$

One should compare these conditions to the definition of φ_s in (2.4).

The discrete combinatorial model of the bilinear Hilbert transform is

$$\mathcal{H}^{\mathbf{S}}f_1f_2(x)=\sum_{s\in\mathbf{S}}|R_s|^{-1/2}\phi_{s3}(x)\prod_{j=1}^{2}\langle f_j,\phi_{sj}\rangle.$$

Here, $\mathbf{S}$ is a set of tri–tiles and the ϕ_{sj} are adapted to $\mathbf{S}$. Compare this sum to (2.5).

These operators obey the same inequalities as in Theorem 1.1. But for this exposition, we restrict our attention to that case which follows from purely L^2 arguments, that is

Lemma 3.8. *The operator $\mathcal{H}^{\mathbf{S}}$ maps $L^2\times L^2$ into weak L^1. The norm of the operator depends only on the constant C_0 in (3.6).*

3.1 The Key Lemma

We state and prove the key Lemma in the proof of Lemma 3.8. But first we shall have to delineate the structures with which the Lemma must be stated. The tri–tiles admit a partial order. Thus we write $s<s'$ if $\rho_s\supset\rho_{s'}$ and $I_{s'}\supset I_s$. We note that s and s' intersect as rectangles if and only if they are comparable under '$<$.'

We say that a collection of tri–tiles $\mathbf{T}\subset\mathbf{S}$ is a *tree with top t* if for every $s\in\mathbf{T}$, $s<t$. (The top need not be in the tree and tops are not unique.) It is easy to see that this partial order does not admit a cycle, so our use of the phrase tree conforms to common usage. We note that any collection $\mathbf{S}$ of tri–tiles is a union of trees. Simply let $\mathbf{S}^*$ denote those elements of $\mathbf{S}$ that are maximal under '$<$,' and for each $t\in\mathbf{S}^*$, let $\mathbf{T}_t$ be the maximal tree in $\mathbf{S}$ with top t. $\mathbf{S}$ is the union of the $\mathbf{T}_t$, $t\in\mathbf{S}^*$.

We refine the notion of a tree by saying that $\mathbf{T}$ is a j–tree if $\mathbf{T}$ is a tree with top t and $\rho_{sj}\cap\rho_t=\emptyset$ for all $s\ne t\in\mathbf{T}$. Under this condition, the functions $\{\phi_{sj}\mid s\in\mathbf{T}\}$ are orthogonal. If on the other hand we were to assume that $\rho_{sk}\cap\rho_t\ne\emptyset$ for all $s\in\mathbf{T}$, it follows from the grid structure that $\mathbf{T}$ is a j–tree for $j\ne k$. Now, for $s\ne t\in\mathbf{T}$, at most one of the three intervals ρ_{sk} can intersect ρ_t, so an arbitrary tree is a union of at most three subtrees which are j–trees for two choices of j.

Trees have important analytic properties. The time intervals in a tree refine that of the top. The frequency intervals increase in length, but only at the rate dictated by uncertainty. And all the frequency intervals contain that of the top, hence the trees have localized the frequency variables in the transform. For a k–tree set

$$\Delta(T,k) := \Big[|R_t|^{-1}\sum_{s\in\mathbf{T}}|\langle f_k,\phi_{sk}\rangle|^2\Big]^{1/2},$$

Here, we specifically include the case of $k=3$, as well as $k=1,2$, for ultimately we will form the inner product of $\mathcal{H}f_1f_2$ with a well–chosen third function f_3. For an arbitrary collection of tiles $\mathbf{S}$, we set

$$k\text{–size}(\mathbf{S}) = \sup\{\Delta(\mathbf{T},k) : \mathbf{T}\subset\mathbf{S} \text{ is a } k\text{–tree.}\}.$$

We note that a singleton $\mathbf{T}=\{t\}$ is a k–tree for all k. So k–size$(\mathbf{S})$ dominates the terms $|R_s|^{-1/2}|\langle f_k,\phi_{sk}\rangle|$ for all $s\in\mathbf{S}$.

It should be noted, for we will relie upon this fact latter, that for any collection of tri–tiles $\mathbf{S}$, we have k–size$(\mathbf{S})\le K\|f_k\|_\infty$.

Now, fix $k=1,2$ or 3 and let $\mathbf{T}\subset\mathbf{S}$ be a k–tree for $j\ne k$ and let t denote the top of the tree. Then, by applying Cauchy–Schwarz,

$$|\langle\mathcal{H}^{\mathbf{T}}f_1f_2,f_3\rangle| \le \sum_{s\in\mathbf{T}}\frac{|\langle f_j,\phi_{sj}\rangle|}{\sqrt{|R_s|}}\prod_{j\ne k}|\langle f_k,\phi_{sk}\rangle| \le |R_t|\prod_{j=1}^{3}k\text{–size}(\mathbf{S}). \tag{3.9}$$

This is the central estimate on a tree.

Lemma 3.10. *Let $k=1,2$ or 3. Let f_k be a Schwartz function and $\mathbf{S}$ any collection of tri–tiles. Then $\mathbf{S}$ is a union of $\mathbf{S}_1$ and $\mathbf{S}_2$ which have these two properties. Let $\mathbf{S}_1^*$ denote the elements in $\mathbf{S}_1$ that are maximal under $<$. We have*

$$\sum_{t\in\mathbf{S}_1^*}|R_t| \le C_1 k\text{–size}(\mathbf{S})^{-2}\|f_k\|_2^2, \tag{3.11}$$

$$k\text{–size}(\mathbf{S}_2) \le \tfrac12[k\text{–size}(\mathbf{S})]. \tag{3.12}$$

The constant C_1 depends only on C_0 in (3.6).

Note that the second collection $\mathbf{S}_2$ is better in that it has smaller size. But we have controlled the number of trees that are in $\mathbf{S}_1$ the first collection—this is the critical condition, the one that orthogonality gives us.

For the proof of the Lemma it suffices to consider the case of k–size$(\mathbf{S})=1$. We construct $\mathbf{S}_1$ in two similar steps, with the construction being motivated by the particulars of the issues of orthogonality with which we conclude this argument.

It is required to distinguish between two types of k–trees. A k–tree $\mathbf{T}$ with top t will be called a *left tree* if for all $s\ne t\in\mathbf{T}$, ρ_{sk} lies to the left of ρ_t. Note that for s, s' in a left tree $\mathbf{T}$, the relation $s < s'$ implies that ρ_{sk} lies to the left of $\rho_{s'k}$. A *right tree* has a corresponding definition.

We construct $\mathbf{S}_{1\ell} \subset \mathbf{S}$ as a union of trees $\tilde{\mathbf{T}}_l$ with tops $t(l)$, for integers $l \geq 1$. Each $\tilde{\mathbf{T}}_l$ will be associated to a left tree $\mathbf{T}_l$. The construction is inductive. We take $\mathbf{T}_1 \subset \mathbf{S}^1$ to be a left tree which satisfies several conditions. First, $\Delta(\mathbf{T}_1, k) \geq 1/4$. Second, $\mathbf{T}_1$ is maximal with respect to inclusion. Third, the top $t(1)$ is to be maximal with respect to '$<$.' Finally, $\rho_{t(1)}$ is to be left most—that is $\min\{\xi \mid \xi \in \rho_{t(1)}\}$ is minimal among the maximal tops. After selecting $\mathbf{T}_1$, we take $\tilde{\mathbf{T}}_1$ to be the maximal tree with top $t(1)$ in $\mathbf{S}$. We remove $\tilde{\mathbf{T}}_1$ from $\mathbf{S}$ and repeat this procedure until there is no left tree $\mathbf{T}_l \subset \mathbf{S}$ meeting these criteria. The union of the $\tilde{\mathbf{T}}_l$ is then $\mathbf{S}_{1\ell}$.

Under this construction, it is obvious that $\Delta(\mathbf{T}, k) < 1/4$ for all left trees $\mathbf{T} \subset \mathbf{S} - \mathbf{S}_{1\ell}$. Moreover—and this is the essential combinatorial observation—the $\mathbf{T}_l$ satisfy this disjointness condition.

$$\text{If } s \in \mathbf{T}_l,\ s' \in \mathbf{T}_{l'} \text{ and } \rho_{sk} \subset_{\neq} \rho_{s'k} \text{ then } R_{t(l)} \cap R_{s'} = \emptyset. \tag{3.13}$$

Indeed, the grid structure implies that $\rho_{t(l)} \subset \rho_s \subset \rho_{sk'}$, and so $\rho_{t(l)}$ lies to the left of $\rho_{t(l')}$. Thus, the tree $\mathbf{T}_l$ was constructed first. But then $s' \notin \tilde{\mathbf{T}}_l$, so that we must have $s' \not< t(l)$, which is to say $R_{t(l)} \cap R_{s'} = \emptyset$.

We verify that $\mathbf{S}_{1\ell}$ satisfies (3.11) momentarily.

We finally construct $\mathbf{S}_{1r}$ as a union of right trees in $\mathbf{S} - \mathbf{S}_{1\ell}$, with the obvious changes in the argument above. We conclude that this collection satisfies (3.11) as well. Then $\mathbf{S}_1$ is the union $\mathbf{S}_{1\ell} \cup \mathbf{S}_{1r}$ and $\mathbf{S}_2 := \mathbf{S} - \mathbf{S}_1$. For any left or right tree $\mathbf{T} \subset \mathbf{S}_2$, we have $\Delta(\mathbf{T}, k) < 1/4$ so that (3.12) follows.

We have still to establish (3.11), for the collections $\mathbf{S}_{1\ell}$ and $\mathbf{S}_{1r}$ defined above. This is the point that orthogonality enters the proof, and in fact we need only consider $\mathbf{S}_{1\ell}$. For the purposes of this argument, we suppose that $\mathbf{S} = \bigcup_l \mathbf{T}_l$, as constructed above. The properties of the $\mathbf{T}_l$ that we need are

$$\Delta(\mathbf{T}_l, k) \geq 1/4 \quad \text{for all } l,$$
$$\frac{|\langle f_k, \phi_{sk} \rangle|}{\sqrt{|R_s|}} \leq 1 \quad \text{for all } s \in \mathbf{S},$$

and the trees satisfy the disjointness condition (3.13). We demonstrate the inequality

$$\sum_{l=1}^{\infty} |R_{t(l)}| \leq K \|f_k\|_2^2. \tag{3.14}$$

For the proof of (3.14), it suffices to assume that $\mathbf{S}$ is finite, so that the number below is finite.

$$B := \Big\| \sum_{s \in \mathbf{S}} \langle f_k, \phi_{sk} \rangle \phi_{sk} \Big\|_2.$$

We show that $B \leq K\|f_k\|_2$, which proves the Lemma as the next inequality shows.

$$\begin{aligned}\sum_{l=1}^{\infty}|R_{t(l)}| &\leq \sum_{s\in\mathbf{S}}|\langle f_k,\phi_{sk}\rangle|^2\\ &= \Big\langle f_k, \sum_{s\in\mathbf{S}}\langle f_k,\phi_{sk}\rangle\phi_{sk}\Big\rangle\\ &\leq \|f_k\|_2 B.\end{aligned}$$

Note that we are exploiting the self–dual nature of the problem.

Now, we expand B^2, to get a diagonal term $\mathcal{D}$ and off–diagonal term $\mathcal{O}$. The diagonal term is $\mathcal{D} = \sum_{s\in\mathbf{S}}|\langle f_k,\phi_{sk}\rangle|^2$, which is no more than $B\|f_k\|_2$ as we have just seen. The off–diagonal term is $\mathcal{O} := 2\sum_{s\in\mathbf{S}}|\langle f_k,\phi_{sk}\rangle|\mathcal{O}(s)$, where

$$\mathcal{O}(s) := \sum_{s'\in\mathbf{S}(s)} |\langle\phi_{sk},\phi_{s'k}\rangle||\langle f_k,\phi_{s'k}\rangle|,$$

and we use the notation $\mathbf{S}(s) := \{s'\in\mathbf{S} \mid \rho_{sk} \subset_{\neq} \rho_{s'k}\}$. This is justified by (3.7) and the fact that the Fourier transform of ϕ_{sk} is supported on an affine image of ρ_{sk}. Hence the inner product of ϕ_{sk} and $\phi_{s'k}$ is zero unless ρ_{sk} and $\rho_{s'k}$ intersect. But then we can assume that one interval is contained in another due to the grid structure.

Note that by Cauchy–Schwarz, $\mathcal{O} \leq 2B^{1/2}[\sum_{s\in\mathbf{S}}\mathcal{O}(s)^2]^{1/2}$ and we claim that for each tree $\mathbf{T}_l$

$$\sum_{s\in\mathbf{T}_l}\mathcal{O}(s)^2 \leq K|R_{t(l)}|. \tag{3.15}$$

This will complete the proof of the Lemma, for we will then have $B^2 \leq KB\|f\|_2$.

To see the claim, fix a tree top $t(l)$ and consider $s\in\mathbf{T}_l$. We observe that (3.6) implies that

$$|\langle\phi_{sk},\phi_{s'k}\rangle| \leq K\frac{\sqrt{|R_{s'}|}}{\sqrt{|R_s|}}(1+\operatorname{dist}(R_s,R_{s'})|R_s|^{-1})^{-5}.$$

A detailed proof of the estimate is left to the reader, but note that the right hand side is only slightly bigger than $\|\phi_{s'k}\|_{L^1(R_{s'})}\|\phi_{sk}\|_{L^\infty(R_{s'})}$. But also note that (3.13) implies that every $s'\in\mathbf{S}(s)$ must satisfy $R_{t(l)}\cap R_{s'} = \emptyset$. Hence,

$$\begin{aligned}\mathcal{O}(s) &\leq K|R_s|^{-1/2}\sum_{s'\in\mathbf{S}(s)}(1+\operatorname{dist}(R_s,R_{s'})|R_s|^{-1})^{-5}|R_s|\\ &\leq K|R_s|^{1/2}\int_{R^c_{t(l)}}(1+\operatorname{dist}(R_s,x)|R_s|^{-1})^{-5}\frac{dx}{|R_s|}\\ &\leq K|R_s|^{1/2}(1+\operatorname{dist}(R_s,R^c_{t(l)})|R_s|^{-1})^{-4}.\end{aligned}$$

But the tree structure imposes restrictions on the intervals R_s: They are in relation to $R_{t(l)}$ as the dyadic intervals $[j2^{-n},(j+1)2^{-n}]$, $0\leq j<2^n$ are to $[0,1]$. Thus, one sees that (3.15) holds. The proof is done.

3.2 Application of the Key Lemma

The Key Lemma, together with some considerations of a more familiar nature, give Lemma 3.8. It suffices to prove the inequality

$$|\{x \mid \mathcal{H}^{\mathbf{S}} f_1 f_2(x) \geq 2\}| \leq K, \tag{3.16}$$

for all f_1, f_2 of L^2 norm one and any collection of tri–tiles $\mathbf{S}$. K depends only on C_0 in (3.6). This is due to an invariance of the model operators under dilation. As this is a commonplace reduction, we omit the easy argument to pass from the inequality above to Lemma 3.8.

As f_1 and f_2 are Schwartz functions, $\mathbf{S}$ has finite 1 and 2 size, say no more than 2^{-n_0}. If $n_0 \geq 0$, there is nothing for us to do at this point. So assume that $n_0 < 0$. We iteratively apply Lemma 3.10 in the following fashion. Apply Lemma 3.10 for those $k = 1, 2$ for which k–size$(\mathbf{S}) \geq 2^{-n_0-1}$. We see that $\mathbf{S} = \mathbf{S}_{n_0} \cup \mathbf{S}^{n_0+1}$ where

$$\sum_{t \in \mathbf{S}^*_{n_0}} |R_t| \leq C 2^{2n_0} \quad \text{and} \quad k\text{–size}(\mathbf{S}^{n_0+1}) \leq 2^{-n_0-1}, \ k = 1, 2.$$

The point that distinguishes the case $n_0 < 0$ is that the sum of the tops of the trees is less than a constant.

Continuing this procedure, we may write $\mathbf{S} = \bigcup_{n=-\infty}^{0} \mathbf{S}_n$, where k–size$(\mathbf{S}_n) \leq 2^{-n}$ for all $-\infty < n \leq 0$ and $k = 1, 2$. And for n strictly less than zero,

$$\sum_{t \in \mathbf{S}^*_n} |R_t| \leq C 2^{2n}, \quad n < 0.$$

There is no such control for $\mathbf{S}_0$, and we shall return to this collection momentarily.

Now for a fixed tree $\mathbf{T}$ with top t in $\mathbf{S}_n$, away from the interval $2R_t$, we have

$$\begin{aligned} |\mathcal{H}^{\mathbf{T}} f_1 f_2(x)| &\leq \sup_{s \in \mathbf{T}} \prod_{1}^{2} \frac{|\langle f_j, \phi_{sj} \rangle|}{\sqrt{|R_s|}} \sum_{s \in \mathbf{T}} \sqrt{|R_s|} |\phi_{s3}(x)| \\ &\leq K 2^{-2n} \Big(1 + \frac{\operatorname{dist}(x, R_t)}{|R_t|}\Big)^{-10}, \quad \text{if } x \notin 2R_t, \end{aligned}$$

as (3.6) and the tree structure easily imply. Thus, if we set $E_n = \bigcup_{t \in \mathbf{S}^*_n} 2^{-n} R_t$, then we have $|E_n| \leq K 2^n$. Thus, the union of these sets has bounded measure, so we need only estimate $\mathcal{H}^{\mathbf{S}_n}$ off of the set E_n, but then

$$\|\mathcal{H}^{\mathbf{S}_n} f_1 f_2\|_{L^1(E_n^c)} \leq K 2^{3n}.$$

Consequently, the collection $\mathbf{S}^0 = \bigcup_{n=-\infty}^{-1} \mathbf{S}_n$ satisfies (3.16).

Therefore, we can assume that the 1 and 2 size of $\mathbf{S}$ is at most 1. This is the case in which the Key Lemma is decisive. Set $G := \{x \mid \mathcal{H}^{\mathbf{S}} f_1 f_2(x) > 1/2\}$, which is a set of finite measure as we take f_1 and f_2 to be Schwartz functions. We can assume that $|G| > K'$ for otherwise there is nothing to prove. Then take f_3 to be a Schwartz functions approximating $|G|^{-1/2} \mathbb{I}_G$ so well that

$$|G|^{1/2} \leq 2 |\langle \mathcal{H}^{\mathbf{S}} f_1 f_2, f_3 \rangle|,$$

but in addition $|f_3(x)| \le 2|G|^{-1/2}$ for all x. It follows that the 3–size($\mathbf{S}$) can be assumed to be at most one, provided K' is large enough.

By applying Lemma 3.10 inductively, we can write $\mathbf{S}$ as $\bigcup_{n=0}^{\infty} \mathbf{S}_n$ where each $\mathbf{S}_n$ has k–size at most 2^{-n}, for $1 \le k \le 3$, and

$$\sum_{t \in \mathbf{S}_n^*} |R_t| \le C 2^{2n}.$$

But then by (3.9) it follows that

$$|\langle \mathcal{H}^{\mathbf{S}_n} f_1 f_2, f_3 \rangle| \le K 2^{-3n} \sum_{t \in \mathbf{S}_n^*} |R_t| \le K 2^{-n}.$$

And this is summable over $n > 0$, from which we conclude that $|G|^{1/2} \le K$. The proof is done.

References

[1] L. Carleson. "On convergence and growth of partial sums of Fourier series." *Acta Math.* 116 (1966) pp. 135-157.

[2] I. Daubechies. *Ten Lectures on Wavelets.* CBMS–NSF regional conference series in applied mathematics. SIAM (1992) Philadelphia.

[3] C. Fefferman. "Pointwise convergence of Fourier series." *Ann. Math.* 98 (1973) 551–571.

[4] M.T. Lacey, C.M. Thiele. "Bounds for the bilinear Hilbert transform on L^p." *Proc. Nat. Acad. Sci.* 94 (1997) 33—35.

[5] M.T. Lacey, C.M. Thiele. "On Calderòn's Conjecture." To appear in *Ann. Math.*

[6] M.T. Lacey, C.M. Thiele. "On Calderòn's Conjecture for the bilinear Hilbert transform." *Proc. Nat. Acad. Sci.* 95 (1998) 4828–4830.

[7] M.T. Lacey, C.M. Thiele. "L^p Bounds for the bilinear Hilbert transform, $p > 2$." Ann. Math. 146 (1997) 693—724.

Michael T. Lacey
School of Mathematics
Georgia Institute of Technology
Atlanta GA 30332
USA
email: `lacey@math.gatech.edu`
web: `www.math.gatech.edu/~lacey`