Publl. Mat. **48** (2004), 251–307

CARLESON'S THEOREM: PROOF, COMPLEMENTS, VARIATIONS

Michael T. Lacey

Abstract

Carleson's Theorem from 1965 states that the partial Fourier sums of a square integrable function converge pointwise. We prove an equivalent statement on the real line, following the method developed by the author and C. Thiele. This theorem, and the proof presented, is at the center of an emerging theory which complements the statement and proof of Carleson's theorem. An outline of these variations is also given.

Contents

2000 *Mathematics Subject Classification.* 42A20, 42B20, 42B25.
Key words. Pointwise convergence, Fourier series, singular integrals, phase plane analysis.
This work has been supported by an NSF grant. These notes are based on a series of lectures given at the Erwin Schrödinger Institute, in Vienna, Austria. I am indebted to the Institute for the opportunity to present these lectures.

1. Introduction

L. Carleson's celebrated theorem of 1965 [**14**] asserts the pointwise convergence of the partial Fourier sums of square integrable functions. We give a proof of this fact, in particular the proof of Lacey and Thiele [**50**], as it can be presented in brief self contained manner, and a number of related results can be seen by variants of the same argument. We survey some of these variants, complements to Carleson's theorem, as well as open problems.

We are concerned with the Fourier transform on the real line, given by

$$\widehat{f}(\xi) = \int e^{-ix\xi} f(x)\, dx$$

for Schwartz functions f. For such functions, it is an important elementary fact that one has Fourier inversion,

$$f(x) = \lim_{N\to\infty} \frac{1}{2\pi} \int_{-N}^{N} \widehat{f}(\xi) e^{ix\xi}\, d\xi, \quad x \in \mathbb{R}, \tag{1.1}$$

the inversion holding for all Schwartz functions f. Indeed,

$$\frac{1}{2\pi} \int_{-N}^{N} \widehat{f}(\xi) e^{ix\xi}\, d\xi = D_N * f(x),$$

where $D_N(x) := \frac{\sin Nx}{\pi x}$ is the Dirchlet kernel.

The convergence in (1.1) for Schwartz functions follows from the classical facts

$$\int_{-\infty}^{\infty} D_N(x)\, dx = 1,$$

$$\lim_{N\to\infty} \int_{|x|\geq\epsilon} D_N(x)\, dx = 0, \quad \epsilon > 0.$$

L. Carleson's theorem asserts that (1.1) holds almost everywhere, for $f \in L^2(\mathbb{R})$. The form of the Dirchlet kernel already points out the essential difficulties in establishing this theorem. That part of the kernel that is convolution with $\frac{1}{x}$ corresponds to a singular integral. This can be done with the techniques associated to the Calderón Zygmund theory. In addition, one must establish some uniform control for the oscillatory term $\sin Nx$, which falls outside of what is commonly considered to be part of the Calderón Zygmund theory.

For technical reasons, we find it easier to consider the equivalent one sided inversion,

$$f(x) = \lim_{N\to\infty} \frac{1}{2\pi} \int_{-\infty}^{N} \widehat{f}(\xi) e^{ix\xi}\, d\xi. \tag{1.2}$$

Schwartz functions being dense in L^2, one need only show that the set of functions for which a.e. convergence holds is closed. The standard method for doing so is to consider the maximal function below, which we refer to as the Carleson operator

$$\mathcal{C}f(x) := \sup_N \left| \int_{-\infty}^{N} \widehat{f}(\xi) e^{ix\xi}\, d\xi \right|, \quad x \in \mathbb{R}. \tag{1.3}$$

There is a straight forward proposition.

Proposition 1.4. *Suppose that the Carleson operator satisfies*

$$|\{\mathcal{C}f(x) > \lambda\}| \lesssim \lambda^{-2} \|f\|_2^2, \quad f \in L^2(\mathbb{R}),\ \lambda > 0. \tag{1.5}$$

Then, the set of functions $f \in L^2(\mathbb{R})$ for which (1.2) *holds is closed and hence all of $L^2(\mathbb{R})$.*

Proof: For $f \in L^2(\mathbb{R})$, we should see that

$$L_f := \limsup_{N\to\infty} \left| f(x) - \frac{1}{2\pi} \int_{-\infty}^{N} \widehat{f}(\xi) e^{ix\xi}\, d\xi \right| = 0 \quad \text{a.e.}$$

To do so, we show that for all $\epsilon > 0$, $|\{L_f > \epsilon\}| \lesssim \epsilon$. We take g to be a smooth compactly supported function so that $\|f-g\|_2 \le \epsilon^{3/2}$. Now Fourier inversion holds for g, whence $L_f \le \mathcal{C}(f-g) + |f-g|$. Then, by the weak type inequality, (1.5), we have

$$|\{\mathcal{C}(f-g) > \epsilon\}| \lesssim \epsilon^{-2} \|f-g\|_2^2 \lesssim \epsilon. \qquad \square$$

This is a standard proposition, which holds in a general context, and serves as one of the prime motivations for considering maximal operators. Note in particular that we are not at this moment claiming that $\mathcal{C}$ is a bounded operator on L^2. Inequality (1.5) is the so called weak L^2 bound, and we shall utilize the form of this bound in a very particular way in the proof below.

It was one of L. Carleson's great achievements to invent a method to prove this weak type estimate.

Theorem 1.6. *The estimate* (1.5) *holds. As a consequence,* (1.2) *holds for all* $f \in L^2(\mathbb{R})$, *for almost every* $x \in \mathbb{R}$.

Carleson's original proof [**14**] was extended to L^p, $1 < p < \infty$, by Hunt. Also see [**67**]. Fefferman [**26**] gave an alternate proof that was influential by the explicit nature of it's "time frequency" analysis, of which we have more more to say below. We follow the proof of Lacey and Thiele [**50**]. More detailed comments on the history of the proof, and related results will come later.

The proof will have three stages, the first being an appropriate decomposition of the Carleson operator. The second being an introduction of three lemmas, which can be efficiently combined to give the proof of our theorem. The third being a proof of the lemmas.

We do not keep track of the value of generic absolute constants, instead using the notation $A \lesssim B$ iff $A \leq KB$ for some constant K. And $A \simeq B$ iff $A \lesssim B$ and $B \lesssim A$. The notation $\mathbf{1}_A$ denotes the indicator function of the set A. For an operator T, $\|T\|_p$ denotes the norm of T as an operator from L^p to itself.

2. Decomposition

The Fourier transform is a constant times a unitary operator on $L^2(\mathbb{R})$. In particular, we shall take the Plancherel's identity for granted.

Proposition 2.1. *For all* $f, g \in L^2(\mathbb{R})$,

$$\langle f, g \rangle = c \langle \widehat{f}, \widehat{g} \rangle$$

for appropriate constant $c = \frac{1}{2\pi}$.

The convolution of f and ψ is given by $f * \psi(x) = \int f(x-y)\psi(y)\,dy$. We shall also assume the following lemma.

Lemma 2.2. *If a bounded linear operator* T *on* $L^2(\mathbb{R})$ *commutes with translations, then* $Tf = \psi * f$, *where* ψ *is a distribution, which is to say a linear functional on Schwartz functions. In addition, the Fourier transform of* Tf *is given by*

$$\widehat{Tf} = \widehat{\psi}\widehat{f}.$$

Let us introduce the operators associated to translation, modulation and dilation on the real line.

$$\operatorname{Tr}_y f(x) := f(x-y), \tag{2.3}$$

$$\operatorname{Mod}_\xi f(x) := e^{i\xi x} f(x), \tag{2.4}$$

$$\operatorname{Dil}^p_\lambda f(x) := \lambda^{-1/p} f(x/\lambda), \quad 0 < p \leq \infty,\ \lambda > 0. \tag{2.5}$$

Note that the dilation operator preserves L^p norm. These operators are related through the Fourier transform, by

$$\widehat{\operatorname{Tr}_y} = \operatorname{Mod}_{-y}, \quad \widehat{\operatorname{Mod}_\xi} = \operatorname{Tr}_\xi, \quad \widehat{\operatorname{Dil}^2_\lambda} = \operatorname{Dil}^2_{1/\lambda}\,. \tag{2.6}$$

And we should also observe that the Carleson operator commutes with translation and dilation operators, while being invariant under modulation operators. For any $y, \xi \in \mathbb{R}$, and $\lambda > 0$,

$$\operatorname{Tr}_y \circ \mathcal{C} = \mathcal{C} \circ \operatorname{Tr}, \quad \operatorname{Dil}^2_\lambda \circ \mathcal{C} = \mathcal{C} \circ \operatorname{Dil}^2_\lambda, \quad \mathcal{C} \circ \operatorname{Mod}_\xi = \mathcal{C}.$$

Thus, our mode of analysis should exhibit the same invariance properties.

We have phrased the Carleson operator in terms of modulations of the operator $\mathrm{P}_- f(x) = \int_{-\infty}^0 \widehat{f}(\xi) e^{ix\xi}\, d\xi$, which is the Fourier projection on to negative frequencies. Specifically, since multiplication of f by an exponential is associated with a translation of $\widehat{f}$, we have

$$\mathcal{C} f = \sup_N |\mathrm{P}_-(e^{iN\cdot} f)|. \tag{2.7}$$

A characterization of the operator P_- will be useful to us.

Proposition 2.8. *Up to a constant multiple,* P_- *is the unique bounded operator on* $L^2(\mathbb{R})$ *which (a) commutes with translation (b) commutes with dilations (c) has as it's kernel precisely those functions with frequency support on the positive axis.*

Proof: Let T be a bounded operator on $L^2(\mathbb{R})$ which satisfies these three properties. Condition (a) implies that T is given by convolution with respect to a distribution. Such operators are equivalently characterized in frequency variables by $\widehat{Tf} = \tau \widehat{f}$ for some bounded function τ. Condition (b) then implies that $\tau(\xi) = \tau(\xi/|\xi|)$ for all $\xi \neq 0$. A function f is in the kernel of T iff $\widehat{f}$ is supported on the zero set of τ. Thus (c) implies that τ is identically 0 on the positive real axis, and non-zero on the negative axis. Thus, T must be a multiple of P_-. $\square$

We move towards the tool that will permit us to decompose Carleson operator, and take advantage of some combinatorics of the time-frequency plane. We let $\mathcal{D}$ be a choice of dyadic grids on the line. Of the different choices we can make, we take the grid to be one that is preserved under dilations by powers of 2. That is

$$\mathcal{D} = \{[j2^k, (j+1)2^k) : j, k \in \mathbb{Z}\}. \tag{2.9}$$

Thus, for each interval $I \in \mathcal{D}$ and $k \in \mathbb{Z}$, the interval $2^k I = \{2^k x : x \in I\}$ is also in $\mathcal{D}$.

A *tile* is a rectangle $s \in \mathcal{D} \times \mathcal{D}$ that has area one. We write a tile as $s = I \times \omega$, thinking of the first interval as a time interval and the second as frequency. The requirement of having area one is suggested by the uncertainty principle of the Fourier transform, or alternatively, our calculation of the Fourier transform of the dilation operators in (2.6). Let $\mathcal{T}$ denote the set of all tiles. While tiles all have area one, the ratio between the time and frequency coordinates is permitted to be arbitrary. See Figure 1 for a few possible choices of this ratio.

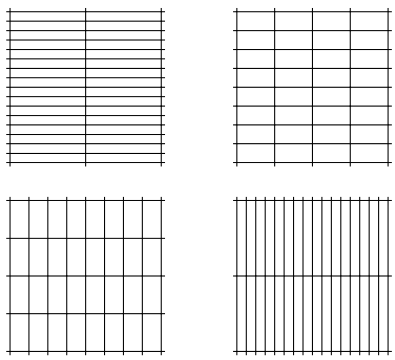

FIGURE 1. Four different aspects ratios for tiles. Each fixed ratio gives rise to a tiling for the time frequency plane.

Each dyadic interval is a union of its left and right halves, which are also dyadic. For an interval ω we denote these as ω_- and ω_+ respectively. We are in the habit of associating frequency intervals with the vertical axis. So ω_- will lie below ω_+. Associate to a tile $s = I_s \times \omega_s$ the rectangles $s_\pm = I_s \times \omega_{s\pm}$. These two rectangles play complementary roles in our proof.

Fix a Schwartz function φ with $\mathbf{1}_{[-1/9,1/9]} \le \widehat{\varphi} \le \mathbf{1}_{[-1/8,1/8]}$. Define a function associated to a tile s by

$$\varphi_s = \mathrm{Mod}_{c(\omega_{s-})} \mathrm{Tr}_{c(I_s)} \mathrm{Dil}^2_{|I_s|} \varphi \tag{2.10}$$

where $c(J)$ is the center of the interval J. Notice that φ_s has Fourier transform supported on ω_{s-}, and is highly localized in time variables around the interval I_s. That is, φ_s is essentially supported in the time-frequency plane on the rectangle $I_s \times \omega_{s-}$. Notice that the set of functions $\{\varphi_s : s \in \mathcal{T}\}$ has a set of invariances with respect to translation, modulation, and dilation that mimics those of the Carleson operator.

It is our purpose to devise a decomposition of the projection P_- in terms of the tiles just introduced. To this end, for a choice of $\xi \in \mathbb{R}$, let

$$\mathrm{Q}_\xi f = \sum_{s \in \mathcal{T}} \mathbf{1}_{\omega_{s+}}(\xi) \langle f, \varphi_s \rangle \varphi_s. \tag{2.11}$$

We should consider general values of ξ for the reason that the dyadic grid distinguishes certain points as being interior, or a boundary point, to an infinite chain of dyadic intervals. And moreover, for a given ξ, only certain tiles can contribute to the sum above, those tiles being determined by the expansion of ξ in a binary digits. See Figure 2. Let us list some relevant properties of these operators.

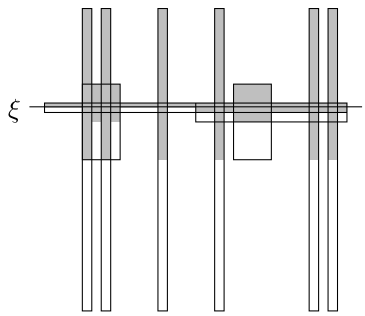

FIGURE 2. Some of the tiles that contribute to the sum for Q_ξ. The shaded areas are the tiles $I_s \times \omega_{s+}$.

Proposition 2.12. *For any ξ, the operator Q_ξ is a bounded operator on L^2, with bound independent of ξ. Its kernel contains those functions with Fourier transform supported on $[\xi,\infty)$, and it is positive semidefinite. Moreover, for each integer k,*

$$\mathrm{Q}_\xi = \mathrm{Dil}^2_{2^{-k}}\,\mathrm{Q}_{\xi 2^{-k}}\,\mathrm{Dil}^2_{2^k} \tag{2.13}$$

$$\mathrm{Q}_{\xi,k}\,\mathrm{Tr}_{2^k} = \mathrm{Tr}_{2^k}\,Q_{\xi,k}, \tag{2.14}$$

where $\mathrm{Q}_{\xi,k} = \sum_{\substack{s\in\mathcal{T}\\ |I_s|\le 2^k}} \mathbf{1}_{\omega_{s+}}(\xi)\langle f,\varphi_s\rangle\varphi_s$.

Proof: Let $\{\omega(n) : n \in \mathbb{Z}\}$ be the set of dyadic intervals for which $\xi \in \omega(n)_+$, listed in increasing order, thus $\cdots \subset \omega(n) \subset \omega(n+1) \subset \cdots$. Let $\mathcal{T}(n) = \{s \in \mathcal{T} : \omega_s = \omega(n)\}$, and

$$\mathrm{Q}_{(n)}\,f = \sum_{s\in\mathcal{T}(n)} \langle f,\varphi_s\rangle\varphi_s.$$

The intervals $\omega(n)_-$ are disjoint in n, and since φ_s has frequency support in ω_{s-}, it follows that the operators $\mathrm{Q}_{(n)}$ are orthogonal in n. The boundedness of Q_ξ reduces therefore to the uniform boundedness of $\mathrm{Q}_{(n)}$ in n.

Two operators $\mathrm{Q}_{(n)}$ and $\mathrm{Q}_{(n')}$ differ by composition with a modulation operator and a dilation operator that preserves L^2 norms. Thus, it suffices to consider the L^2 norm bound of a $\mathrm{Q}_{(n)}$ with $|I_s| = 1$ for all $s \in \mathcal{T}(n)$. Using the fact that φ_s is a rapidly decreasing function, we see that

$$|\langle\varphi_s,\varphi_{s'}\rangle| \lesssim \mathrm{dist}(I_s, I_{s'})^{-4}. \tag{2.15}$$

Now that the spatial length of the tiles is one, the tiles are separated by integral distances. Since $\sum_n n^{-4} < \infty$,

$$\begin{aligned}
\|\mathrm{Q}_{(n)}\,f\|_2^2 &= \sum_{s\in\mathcal{T}(n)}\sum_{s'\in\mathcal{T}(n)} \langle f,\varphi_s\rangle\langle\varphi_s,\varphi_{s'}\rangle\langle\varphi_{s'},f\rangle \\
&\lesssim \sup_{n\in\mathbb{Z}} \sum_{s\in\mathcal{T}(n)} |\langle f,\varphi_s\rangle\langle f,\varphi_{(I_s+n)\times\omega(n)}\rangle| \\
&\lesssim \sum_{s\in\mathcal{T}(n)} |\langle f,\varphi_s\rangle|^2.
\end{aligned}$$

The last inequality following by Cauchy-Schwarz. The last sum is easily controlled, by simply bringing in the absolute values. Since $|\langle f, \varphi_s\rangle|^2 \lesssim \int |f|^2 |\varphi_s|\, dx$

$$\begin{aligned}\sum_{s\in\mathcal{T}(n)} |\langle f, \varphi_s\rangle|^2 &\le \int |f|^2 \sum_{s\in\mathcal{T}(n)} |\varphi_s|\, dx \\ &\le \|f\|_2^2 \sup_x \sum_{s\in\mathcal{T}(n)} |\varphi_s(x)| \\ &\lesssim \|f\|_2^2.\end{aligned}$$

This completes the proof of the uniform boundedness of Q_ξ.

Since all of the functions φ_s that contribute to the definition of Q_ξ have frequency support below ξ, the conclusion abut the kernel of the operator is obvious. And that it is positive semidefinite, observe that

$$\langle \mathrm{Q}_\xi\, f, f\rangle = \sum_{\substack{s\in\mathcal{T}\\ \xi\in\omega_{s+}}} |\langle f, \varphi_s\rangle|^2 \ge 0. \tag{2.16}$$

In particular, $\langle \mathrm{Q}_\xi\, \varphi_s, \varphi_s\rangle \neq 0$ for $s \in \mathcal{T}(n)$.

To see (2.13) recall (2.6) and our specific choice of grids, (2.9). To see (2.14), observe that if $I \in \mathcal{D}$ has length at most 2^k, then $I + 2^k$ is also in $\mathcal{D}$. □

As the lemma makes clear, $\mathrm{Mod}_{-\xi}\, \mathrm{Q}_\xi\, \mathrm{Mod}_\xi$ serves as an approximation to P_-. A limiting procedure will recover P_- exactly. Consider

$$Q = \lim_{Y\to\infty} \int_{B(Y)} \mathrm{Dil}^2_{2^{-\lambda}}\, \mathrm{Tr}_{-y}\, \mathrm{Mod}_{-\xi}\, \mathrm{Q}_\xi\, \mathrm{Mod}_\xi\, \mathrm{Tr}_y\, \mathrm{Dil}^2_{2^\lambda}\, \mu(d\lambda, dy, d\xi). \tag{2.17}$$

Here, $B(Y)$ is the set $[1,2]\times[0,Y]\times[0,Y]$, and μ is normalized Lebesgue measure. Notice that the dilations are given in terms of 2^λ, so that in that parameter, we are performing an average with respect to the multiplicative Haar measure on $\mathbb{R}_+$.

Apply the right hand side to a Schwartz function f. It is easy to see that as $k \to -\infty$, the terms

$$\mathrm{Mod}_\xi\, \mathrm{Tr}_y\, \mathrm{Dil}^2_{2^\lambda}\, \mathrm{Q}_{\xi,k}\, f$$

tend to zero uniformly in the parameters ξ, y, and λ, with a rate that depends upon f. Here, $\mathrm{Q}_{\xi,k}$ is as in (2.14). Similarly, as $k \to \infty$, the terms

$$\mathrm{Mod}_\xi\, \mathrm{Tr}_y\, \mathrm{Dil}^2_{2^\lambda} (\mathrm{Q} - \mathrm{Q}_{\xi,k}) f$$

also tend to zero uniformly. Hence, the limit is seen to exist for all Schwartz functions. By Proposition 2.12, it follows that Q is a bounded operator on L^2. That Q is translation and dilation invariant follows from (2.14) and (2.13). Its kernel contains those functions with Fourier transform supported on $[0,\infty)$. Finally, if we verify that Q is not identically zero, we can conclude that it is a multiple of P_-. But, for e.g. $f=\operatorname{Mod}_{-1/8}\varphi$, it is easy to see that

$$\langle \mathrm{Q}_\xi \operatorname{Mod}_\xi \operatorname{Tr}_y \operatorname{Dil}^2_{2^\lambda} f, \operatorname{Mod}_\xi \operatorname{Tr}_y \operatorname{Dil}^2_{2^\lambda} f\rangle > 0$$

so that $Qf\neq 0$. Thus, Q is a multiple of P_-.

We can return to the Carleson operator. An important viewpoint emphasized by Fefferman's proof [**26**] is that we should linearize the supremum. That is we consider a measurable map $N\colon \mathbb{R}\mapsto\mathbb{R}$, which specifies the value of N at which the supremum in (1.3) occurs. Then, it suffices to bound the operator norm of the linear (not sublinear) operator

$$\mathrm{P}_- \operatorname{Mod}_{N(x)}.$$

Considering (2.17), we set

$$\mathcal{C}_N f(x) = \sum_{s\in\mathcal{T}} \mathbf{1}_{\omega_{s+}}(N(x))\langle f,\varphi_s\rangle\varphi_s(x).$$

Our main lemma is then

Lemma 2.18. *There is an absolute constant K so that for all measurable functions $N\colon\mathbb{R}\mapsto\mathbb{R}$, we have the weak type inequality*

$$|\{\mathcal{C}_N f>\lambda\}| \lesssim \lambda^{-2}\|f\|_2^2, \quad \lambda>0,\ f\in L^2(\mathbb{R}). \tag{2.19}$$

By the convexity of the weak L^2 norm, this theorem immediately implies the same estimate for $\mathrm{P}_-\operatorname{Mod}_{N(x)}$, and so proves Theorem 1.6. The proof of the lemma is obtained by combining the three estimates detailed in the next section.

2.1. Complements.

At the conclusion to the different sections of the proof, some complements to the ideas and techniques of the previous sections will be mentioned, but not proved. These items can be considered as exercises.

2.20. For Schwartz functions φ and ψ, set

$$\mathrm{A}\,f := \sum_{n\in\mathbb{Z}} \langle f, \operatorname{Tr}_n\varphi\rangle \operatorname{Tr}_n\psi$$

$$\mathrm{B}\,f := \int_0^1 \operatorname{Tr}_{-y}\mathrm{A}\operatorname{Tr}_y\,dy.$$

Then, B is a convolution operator, that is $\mathrm{B}\, f = \Psi * f$ for some function Ψ, which can be explicitly computed.

2.21. The identity operator is, up to a constant multiple, the unique bounded operator A on L^2 which commutes with all translation and modulation operators. That is $\mathrm{A}\colon L^2 \mapsto L^2$, and for all $y \in \mathbb{R}$ and $\xi \in \mathbb{R}$,

$$\mathrm{Tr}_y \mathrm{A} = \mathrm{A}\,\mathrm{Tr}_y, \quad \mathrm{Mod}_\xi \mathrm{A} = \mathrm{A}\,\mathrm{Mod}_\xi .$$

2.22. The operators

$$\mathrm{A}_j\, f := \sum_{\substack{s \in \mathcal{T} \\ |I_s| = 2^j}} \langle f, \varphi_s \rangle \varphi_s$$

are uniformly bounded operators on $L^2(\mathbb{R})$, assuming that φ is a Schwartz function.

2.23. Assuming that $\varphi \not\equiv 0$, the operator below is a non-zero multiple of the identity operator on L^2.

$$\int_0^{2^j} \int_0^{2^{-j}} \mathrm{Tr}_{-y} \mathrm{Mod}_{-\xi} \mathrm{A}_j \mathrm{Mod}_\xi \mathrm{Tr}_y \; d\xi \, dy.$$

3. The Central Lemmas

Observe that the weak type estimate of Lemma 2.18 is implied by

$$|\langle \mathcal{C}_N f, \mathbf{1}_E \rangle| \lesssim \|f\|_2 |E|^{1/2}, \tag{3.1}$$

for all functions f and sets E of finite measure. (In fact this inequality is equivalent to the weak L^2 bound.)

By linearity of $\mathcal{C}_N$, we may assume that $\|f\|_2 = 1$. By the invariance of $\mathcal{C}_N$ under dilations by a factor of 2 (with a change of measurable $N(x)$), we can assume that $1/2 < |E| \le 1$. Set

$$\phi_s = (\mathbf{1}_{\omega_{s+}} \circ N)\varphi_s. \tag{3.2}$$

We shall show that

$$\sum_{s \in \mathcal{T}} |\langle f, \varphi_s \rangle \langle \phi_s, \mathbf{1}_E \rangle| \lesssim 1. \tag{3.3}$$

To help keep the notation straight, note that φ_s is a smooth function, adapted to the tile. On the other hand, ϕ_s is the rough function paired with the indicator set $\mathbf{1}_{\omega_{s+}} \circ N$. From this point forward, the function f and the set E are fixed. We use data about these two objects to organize our proof.

As the sum above is over strictly positive quantities, we may consider all sums to be taken over some finite subset of tiles. Thus, there is never any question that the sums we treat are finite, and the iterative procedures we describe will all terminate. The estimates we obtain will be independent of the fact that the sum is formally over a set of finite tiles.

We need some concepts to phrase the proof. There is a natural partial order on tiles. Say that $s < s'$ iff $\omega_s \supset \omega_{s'}$ and $I_s \subset I_{s'}$. Note that the time variable of s is localized to that of s', and the frequency variable of s is similarly localized, up to the variability allowed by the uncertainty principle. Note that two tiles are incomparable with respect to the '$<$' partial order iff the tiles, as rectangles in the time frequency plane, do not intersect. A "maximal tile" will one that is maximal with respect to this partial order. See Figure 1.

We call a set of tiles $\mathbf{T} \subset \mathcal{S}$ a *tree* if there is a tile $I_{\mathbf{T}} \times \omega_{\mathbf{T}}$, called the *top of the tree,* such that for all $s \in \mathbf{T}$, $s < I_{\mathbf{T}} \times \omega_{\mathbf{T}}$. We note that the top is not uniquely defined. An important point is that a tree top specifies a location in time variable for the tiles in the tree, namely inside $I_{\mathbf{T}}$, and localizes the frequency variables, identifying $\omega_{\mathbf{T}}$ as a nominal origin.

We say that $\mathcal{S}$ *has count at most* A, and write

$$\operatorname{count}(\mathcal{S}) < A$$

iff $\mathcal{S}$ is a union $\bigcup_{\mathbf{T} \in \mathcal{T}} \mathbf{T}$, where each $\mathbf{T} \in \mathcal{T}$ is a tree, and

$$\sum_{\mathbf{T} \in \mathcal{T}} |I_{\mathbf{T}}| < A.$$

Fix $\chi(x) = (1+|x|)^{-\kappa}$, where κ is a large constant, whose exact value is unimportant to us. Define

$$\chi_I := \operatorname{Tr}_{c(I)} \operatorname{Dil}^1_{|I|} \chi, \tag{3.4}$$

$$\operatorname{dense}(s) := \sup_{s<s'} \int_{N^{-1}(\omega_{s'})} \chi_{I_{s'}}\, dx, \tag{3.5}$$

$$\operatorname{dense}(\mathcal{S}) := \sup_{s \in \mathcal{S}} \operatorname{dense}(s), \qquad \mathcal{S} \subset \mathcal{T}.$$

The first and most natural definition of a "density" of a tile, would be $|I_s|^{-1}|N^{-1}(\omega_{s+}) \cap I_s|$. But φ is supported on the whole real line, though does decay faster than any inverse of a polynomial. We refer to this as a "Schwartz tails problem". The definition of density as $\int_{N^{-1}(\omega_s)} \chi_{I_s}\, dx$, as it turns out, is still not adequate. That we should take the supremum over $s < s'$ only becomes evident in the proof of the "Tree Lemma" below.

The "Density Lemma" is

Lemma 3.6. *Any subset $\mathcal{S} \subset \mathcal{T}$ is a union of $\mathcal{S}_{\text{heavy}}$ and $\mathcal{S}_{\text{light}}$ for which*

$$\operatorname{dense}(\mathcal{S}_{\text{light}}) < \tfrac{1}{2} \operatorname{dense}(\mathcal{S}),$$

and the collection $\mathcal{S}_{\text{heavy}}$ satisfies

$$\operatorname{count}(\mathcal{S}_{\text{heavy}}) \lesssim \operatorname{dense}(\mathcal{S})^{-1}. \tag{3.7}$$

What is significant is that this relatively simple lemma admits a non-trivial variant intimately linked to the tree structure and orthogonality. We should refine the notion of a tree. Call a tree $\mathbf{T}$ with top $I_{\mathbf{T}} \times \omega_{\mathbf{T}}$ a $\pm$tree iff for each $s \in \mathbf{T}$, aside from the top, $I_{\mathbf{T}} \times \omega_{\mathbf{T}} \cap I_s \times \omega_{s\pm}$ is not empty. Any tree is a union of a $+$tree and a $-$tree. If $\mathbf{T}$ is a $+$tree, observe that the rectangles $\{I_s \times \omega_{s-} : s \in \mathbf{T}\}$ are disjoint. And, by the proof of Proposition 2.12, we see that

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

This motivates the definition

$$\operatorname{size}(\mathcal{S}) := \sup\{|I_{\mathbf{T}}|^{-1/2} \Delta(\mathbf{T}) : \mathbf{T} \subset \mathcal{S},\ \mathbf{T} \text{ is a } +\text{tree}\}. \tag{3.8}$$

The "Size Lemma" is

Lemma 3.9. *Any subset $\mathcal{S} \subset \mathcal{T}$ is a union of $\mathcal{S}_{\text{big}}$ and $\mathcal{S}_{\text{small}}$ for which*

$$\operatorname{size}(\mathcal{S}_{\text{small}}) < \tfrac{1}{2} \operatorname{size}(\mathcal{S}),$$

and the collection $\mathcal{S}_{\text{big}}$ satisfies

$$\operatorname{count}(\mathcal{S}_{\text{big}}) \lesssim \operatorname{size}(\mathcal{S})^{-2}. \tag{3.10}$$

Our final lemma relates trees, density and size. It is the "Tree Lemma".

Lemma 3.11. *For any tree $\mathbf{T}$*

$$\sum_{s \in \mathbf{T}} |\langle f, \varphi_s \rangle \langle \phi_s, \mathbf{1}_E \rangle| \lesssim |I_{\mathbf{T}}| \operatorname{size}(\mathbf{T}) \operatorname{dense}(\mathbf{T}). \tag{3.12}$$

The final elements of the proof are organized as follows. Certainly, $\operatorname{dense}(\mathcal{T}) < 2$ for κ sufficiently large. We take some finite subset $\mathcal{S}$ of $\mathcal{T}$, and so certainly $\operatorname{size}(\mathcal{S}) < \infty$. If $\operatorname{size}(\mathcal{S}) < 2$, we jump to the next stage of the proof. Otherwise, we iteratively apply Lemma 3.9 to obtain subcollections $\mathcal{S}_n \subset \mathcal{S}$, $n \geq 0$, for which

$$\operatorname{size}(\mathcal{S}_n) < 2^n, \quad n > 0, \tag{3.13}$$

and $\mathcal{S}_n$ satisfies

$$\operatorname{count}(\mathcal{S}_n) \lesssim 2^{-2n}. \tag{3.14}$$

We are left with a collection of tiles $\mathcal{S}' = \mathcal{S} - \bigcup_{n>0} \mathcal{S}_n$ which has both density and size at most 2.

Now, both Lemma 3.6 and Lemma 3.9 are set up for iterative application. And we should apply them so that the estimates in (3.7) and (3.10) are of the same order. (This means that we should have density about the square of the size.) As a consequence, we can achieve a decomposition of $\mathcal{S}$ into collections $\mathcal{S}_n$, $n \in \mathbb{Z}$, which satisfy (3.13), (3.14) and

$$\operatorname{dense}(\mathcal{S}_n) < \min(2, 2^{2n}). \tag{3.15}$$

Use the estimates (3.12)–(3.15). Write $\mathcal{S}_n$ as a union of trees $\mathbf{T} \in \widetilde{\mathbf{T}}_n$, this collection of trees satisfying the estimate of (3.14). We see that

$$\begin{aligned}
\sum_{s\in\mathcal{S}_n} |\langle f, \varphi_s\rangle\langle \phi_s, \mathbf{1}_E\rangle| &= \sum_{\mathbf{T}\in\widetilde{\mathbf{T}}_n} \sum_{s\in\mathbf{T}} |\langle f, \varphi_s\rangle\langle \phi_s, \mathbf{1}_E\rangle| \\
&\lesssim 2^n \min(2, 2^{2n}) \sum_{\mathbf{T}\in\widetilde{\mathbf{T}}_n} |I_{\mathbf{T}}| \qquad (3.16)\\
&\lesssim \min(2^{-n}, 2^n).
\end{aligned}$$

This is summable over $n \in \mathbb{Z}$ to an absolute constant, and so our proof (3.3) is complete, aside from the proofs of the three key lemmas.

3.1. Complements.

3.17. These two conditions are equivalent.

$$\sup_{\lambda>0} \lambda^{-2} |\{f > \lambda\}| \lesssim 1,$$

$$\int_E |f|\, dx \lesssim |E|^{1/2}, \qquad |E| < \infty.$$

3.18. Let A be an operator for which $\operatorname{Dil}^2_{2^k} \mathrm{A} = \mathrm{A} \operatorname{Dil}^2_{2^k}$ for all $k \in \mathbb{Z}$. Suppose that there is an absolute constant K so that for all functions $f \in L^2(\mathbb{R})$ of norm one,

$$|\{\mathrm{A}\, f > 1\}| \le K.$$

Then for all $\lambda > 0$,

$$|\{\mathrm{A}\, f > \lambda\}| \lesssim \lambda^{-2} \|f\|_2^2.$$

See [**22**].

3.19. For any +tree $\mathbf{T}$,

$$\sum_{s\in\mathbf{T}} |\langle f, \varphi_s\rangle|^2 \lesssim \int |f|^2 \operatorname{Tr}_{c(I_{\mathbf{T}})} \operatorname{Dil}^\infty_{I_{\mathbf{T}}} \chi \, dx.$$

Moreover, one has the inequality

$$|I_{\mathbf{T}}|^{-1}\sum_{s\in\mathbf{T}}|\langle f,\varphi_s\rangle|^2 \lesssim \inf_{x\in\mathbf{T}} \mathrm{M}|f|^2(x).$$

Here, M is the maximal function,

$$\mathrm{M}\, f(x) = \sup_{t>0}(2t)^{-1}\int_{-t}^{t}|f(x-y)|\,dy.$$

4. The Density Lemma

Set $\delta = \mathrm{dense}(\mathcal{S})$. Suppose for the moment that density had the simpler definition

$$\mathsf{dense}(s) := \frac{|N^{-1}(\omega_s)\cap I_s|}{|I_s|}.$$

The collection $\mathcal{S}_{\text{heavy}}$ is to be a union of trees. So to select this collection, it suffices to select the tops of the trees in this set.

Select the tops of the trees, Tops as being those tiles s with $\mathsf{dense}(s)$ exceeding $\delta/2$, which are also maximal with respect to the partial order '$<$'. The tree associated to such a tile $s\in\mathsf{Tops}$ would just be all those tiles in $\mathcal{S}$ which are less than s. The tiles in Tops are pairwise incomparable with respect to the partial order '$<$', and so are pairwise disjoint rectangles in the time-frequency plane. And so the sets $N^{-1}(\omega_s)\cap I_s\subset E$ are pairwise disjoint, and each has measure at least $\frac{\delta}{2}|I_s|$. Hence the estimate below is immediate.

$$\sum_{s\in\mathsf{Tops}}|I_s|\lesssim \delta^{-1}. \tag{4.1}$$

The Schwartz tails problem prevents us from using this very simple estimate to prove this lemma, but in the present context, the Schwarz tails are a weak enemy at best. Let Tops be those $s\in\mathcal{S}$ which have $\mathrm{dense}(s)>\delta/2$ and are maximal with respect to '$<$'. It suffices to show (4.1). For an integer $k\geq 0$, and small constant c, let $\mathcal{S}_k$ be those $s\in\mathsf{Tops}$ for which

$$|2^kI_s\cap N^{-1}(\omega_s)|\geq c2^{2k}\delta|I_s|. \tag{4.2}$$

Every tile in Tops will be in some $\mathcal{S}_k$, with c sufficiently small, and so it suffices to show that

$$\sum_{s\in\mathcal{S}_k}|I_s|\lesssim 2^{-k}\delta^{-1}. \tag{4.3}$$

Fix k. Select from $\mathcal{S}_k$ a subset $\mathcal{S}'_k$ of tiles satisfying $\{2^k I_s \times \omega_s : s \in \mathcal{S}'_k\}$ are pairwise disjoint, and if $s \in \mathcal{S}_k$ and $s' \in \mathcal{S}'_k$ are tiles such that $2^k I_s \times \omega_s$ and $2^k I_{s'} \times \omega_{s'}$ intersect, then $|I_s| \le |I_{s'}|$. It is clearly possible to select such a subset. And since the tiles in $\mathcal{S}_k$ are incomparable with respect to '<', we can use (4.2) to estimate

$$\begin{aligned}\sum_{s \in \mathcal{S}_k} |I_s| &\le 2^{k+1} \sum_{s' \in \mathcal{S}'_k} |I_{s'}| \\ &\le \tfrac{2}{c} 2^{-k} \delta^{-1}.\end{aligned}$$

That is, we see that (4.3) holds, completing our proof.

4.1. Complements.

4.4. Let $\mathcal{S}$ be a set of tiles for which there is a constant K so that for all dyadic intervals J,

$$\sum_{\substack{s \in \mathcal{S} \\ I_s \subset J}} |I_s| \le K|J|.$$

Then for all $1 \le p < \infty$, and intervals J,

$$\Bigg\| \sum_{\substack{s \in \mathcal{S} \\ I_s \subset J}} \mathbf{1}_{I_s} \Bigg\|_p \lesssim K_p |J|^{1/p}.$$

In fact, $K_p \lesssim p$.

5. The Size Lemma

Set $\sigma = \operatorname{size}(\mathcal{S})$. We will need to construct a collection of trees $\mathbf{T} \in \widetilde{\mathbf{T}}_{\text{large}}$, with $\mathcal{S}_{\text{large}} = \bigcup_{\mathbf{T} \in \widetilde{\mathbf{T}}_{\text{large}}} \mathbf{T}$, and

$$\sum_{\mathbf{T} \in \widetilde{\mathbf{T}}_{\text{large}}} |I_{\mathbf{T}}| \lesssim \sigma^{-2}, \tag{5.1}$$

as required by (3.10).

The selection of trees $\mathbf{T} \in \widetilde{\mathbf{T}}_{\text{large}}$ will be done in conjunction with the construction of +trees $\mathbf{T}_+ \in \widetilde{\mathbf{T}}_{\text{large}+}$. This collection will play a critical role in the verification of (5.1).

The construction is recursive in nature. Initialize

$$\mathcal{S}^{\text{stock}} := \mathcal{S}, \quad \widetilde{\mathbf{T}}_{\text{large}} := \emptyset, \quad \widetilde{\mathbf{T}}_{\text{large}+} := \emptyset.$$

While size$(\mathcal{S}^{\rm stock}) > \sigma/2$, select a +tree $\mathbf{T}_+ \subset \mathcal{S}^{\rm stock}$ with

$$\Delta(\mathbf{T}_+) > \frac{\sigma}{2}|I_{\mathbf{T}_+}|. \tag{5.2}$$

In addition, the top of the tree $I_{\mathbf{T}_+} \times \omega_{\mathbf{T}_+}$ should be maximal with respect to the partial order '$<$' among all trees that satisfy (5.2). And $c(\omega_{\mathbf{T}_+})$ should be minimal, in the order of $\mathbb{R}$. Then, take $\mathbf{T}$ to be the maximal tree (without reference to sign) in $\mathcal{S}^{\rm stock}$ with top $I_{\mathbf{T}_+} \times \omega_{\mathbf{T}_+}$.

After this tree is chosen, update

$$\mathcal{S}^{\rm stock} := \mathcal{S}^{\rm stock} - \mathbf{T},$$

$$\widetilde{\mathbf{T}}_{\rm large} := \widetilde{\mathbf{T}}_{\rm large} \cup \mathbf{T},$$

$$\widetilde{\mathbf{T}}_{\rm large+} := \widetilde{\mathbf{T}}_{\rm large+} \cup \mathbf{T}_+.$$

Once the while loop finishes, set $\mathcal{S}_{\rm small} := \mathcal{S}^{\rm stock}$ and the recursive procedure stops.

It remains to verify (5.1). This is a orthogonality statement, but one that is just weaker than true orthogonality. Note that a particular enemy is the is the situation in which $\langle \varphi_s, \varphi_{s'} \rangle \neq 0$. When $\omega_s = \omega_{s'}$, this may happen, but as we saw in the proof of Proposition 2.12, this case may be handled by direct methods. Thus we are primarily concerned with the case that e.g. $\omega_{s-} \subset_{\neq} \omega_{s'-}$.

A central part of this argument is a bit of geometry of the time-frequency plane that is encoded in the construction of the +trees above. Suppose there are two trees $\mathbf{T} \neq \mathbf{T}' \in \widetilde{\mathbf{T}}_{\rm large+}$, and tiles $s \in \mathbf{T}$ and $s' \in \mathbf{T}'$ such that $\omega_{s-} \subset_{\neq} \omega_{s'-}$, then, it is the case that $I_{s'} \cap I_{\mathbf{T}} = \emptyset$. We refer to this property as 'strong disjointness'. It is a condition that is strictly stronger than just requiring that the sets in the time-frequency plane below are disjoint in $\mathbf{T}$.

$$\bigcup_{s \in \mathbf{T}} I_s \times \omega_{s-}, \quad \mathbf{T} \in \widetilde{\mathbf{T}}_{\rm large+}.$$

To see that strong disjointness holds, observe that $\omega_{\mathbf{T}} \subset \omega_s \subset_{\neq} \omega_{s'-}$. Thus $\omega_{\mathbf{T}'}$ lies above $\omega_{\mathbf{T}}$. That is, in our recursive procedure, $\mathbf{T}$ was constructed first. If it were the case that $I_{s'} \cap I_{\mathbf{T}} \neq \emptyset$, observe that one interval would have to be contained in the other. But tiles have area one, thus, it must be the case that $I_{s'} \subset I_{\mathbf{T}}$. That means that s' would have been in the tree (the one without sign) that was removed from $\mathcal{S}^{\rm stock}$ before $\mathbf{T}'$ was constructed. This is a contradiction which proves strong disjointness. See Figure 3.

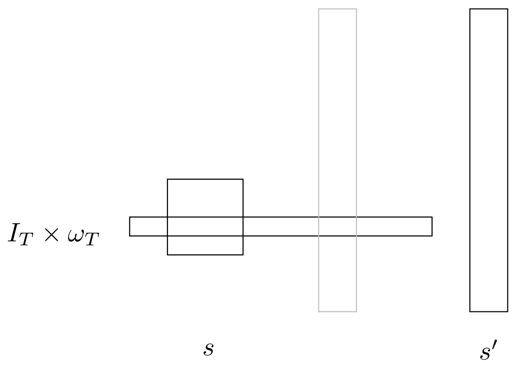

FIGURE 3. The proof of strongly disjoint trees. Note that the gray tile could be in the tree that was removed after the selection of the +-tree with the top indicated above.

We use this strong disjointness condition, and the selection criteria (5.2), to prove the bound (5.1). The method of proof is closely related to the so called T T* method. Set $\mathcal{S}' = \bigcup_{\mathbf{T}\in\widetilde{\mathbf{T}}_{\text{large}+}} \mathbf{T}$, and

$$F := \sum_{s\in\mathcal{S}'} \langle f, \varphi_s\rangle \varphi_s.$$

The operator $f \mapsto \langle f, \varphi_s\rangle\varphi_s$ is self-adjoint, so that

$$\begin{aligned}\sigma^2 \sum_{\mathbf{T}\in\widetilde{\mathbf{T}}_{\text{large}+}} |I_{\mathbf{T}}| &= \langle f, F\rangle \\ &\le \|f\|_2 \|F\|_2.\end{aligned}$$

And so, we should show that

$$\|F\|_2^2 \lesssim \sigma^2 \sum_{\mathbf{T}\in\widetilde{\mathbf{T}}_{\text{large}+}} |I_{\mathbf{T}}|. \tag{5.3}$$

This will complete the proof.

This last inequality is seen by expanding the square on the left hand side. In particular, the left hand side of (5.3) is at most the sum of the two terms

$$\sum_{\substack{s,s'\in\mathcal{S}'\\ \omega_s=\omega_{s'}}} \langle f,\varphi_s\rangle\langle\varphi_s,\varphi_{s'}\rangle\langle\varphi_{s'},f\rangle \tag{5.4}$$

$$2\sum_{\substack{s,s'\in\mathcal{S}'\\ \omega_s\subset_{\neq}\omega_{s'}}} |\langle f,\varphi_s\rangle\langle\varphi_s,\varphi_{s'}\rangle\langle\varphi_{s'},f\rangle|. \tag{5.5}$$

For the term (5.4), we have the obvious estimate on the inner product

$$|\langle\varphi_s,\varphi_{s'}\rangle| \lesssim \left(1+\frac{\operatorname{dist}(I_s,I_{s'})}{|I_s|}\right)^{-4}.$$

(Compare to (2.15).) Thus, by Cauchy-Schwarz,

$$(5.4) \lesssim \sum_{s\in\mathcal{S}'} |\langle f,\varphi_s\rangle|^2 \lesssim \sigma^2 \sum_{\mathbf{T}\in\widetilde{\mathbf{T}}_{\text{large}+}} |I_{\mathbf{T}}|.$$

For the term (5.5), we need only show that for each tree $\mathbf{T}$,

$$\sum_{s\in\mathbf{T}} \sum_{\substack{s'\in\mathcal{S}'\\ \omega_s\subset_{\neq}\omega_{s'}}} |\langle f,\varphi_s\rangle\langle\varphi_s,\varphi_{s'}\rangle\langle\varphi_{s'},f\rangle|^2 \lesssim \sigma^2|I_{\mathbf{T}}|. \tag{5.6}$$

Here, $\mathcal{S}(s) := \{s'\in\mathcal{S}'-\mathbf{T} : \omega_{s-}\subset_{\neq}\omega_{s'-}\}$. The implied constant should be independent of the tree $\mathbf{T}$.

Now, the strong disjointness condition enters in two ways. For $s\in\mathbf{T}$, and $s'\in\mathcal{S}(s)$, it is the case that $I_{s'}\cap I_{\mathbf{T}}=\emptyset$. But furthermore, for $s',s''\in\mathcal{S}(s)$, we have e.g. $\omega_{s-}\subset\omega_{s'-}\subset\omega_{s''-}$, so that $I_{s'}\cap I_{s''}$ is also empty.

At this point, rather clumsy estimates of (5.6) are in fact optimal. The definition of size gives us the bound

$$|\langle\varphi_{s'},f\rangle| \lesssim \sqrt{|I_{s'}|}\sigma.$$

And, since $\omega_s\subset\omega_{s'}$, we have $|I_s|\geq|I_{s'}|$, and I_s, and $I_{s'}$ are, in the typical situation, far apart. An estimation left to the reader gives

$$|\langle\varphi_s,\varphi_{s'}\rangle| \lesssim \sqrt{|I_{s'}||I_s|}\chi_{I_s}(c(I_{s'})). \tag{5.7}$$

Thus, we bound the left side of (5.6) by

$$\begin{aligned}\sum_{s\in\mathbf{T}}\sum_{s'\in\mathcal{S}(s)}|\langle f,\varphi_s\rangle\langle\varphi_s,\varphi_{s'}\rangle\langle\varphi_{s'},f\rangle| &\lesssim \sigma^2\sum_{s\in\mathbf{T}}|I_{s'}||I_s|\chi_{I_s}(c(I_{s'}))\\ &\lesssim \sigma^2\sum_{s\in\mathbf{T}}\int_{(I_\mathbf{T})^c}|I_s|\chi_{I_s}(x)\,dx \qquad (5.8)\\ &\lesssim \sigma^2|I_\mathbf{T}|\end{aligned}$$

as is easy to verify. This completes the proof of (5.6), and so finishes the proof of Lemma 3.9.

5.1. Complements.

5.9. Concerning the inequality (5.8), for any tree $\mathbf{T}$, we have

$$\sum_{s\in\mathbf{T}}\int_{I_\mathbf{T}^c}|I_s|\chi_{I_s}(x)\,dx\lesssim|I_\mathbf{T}|.$$

5.10. Let $\mathbf{T}$ be a +tree and set

$$F_\mathbf{T}=\sum_{s\in\mathbf{T}}\langle f,\varphi_s\rangle\varphi_s.$$

Then, the inequality below is true.

$$\begin{aligned}\|F_\mathbf{T}\|_2 &\simeq \left[\sum_{s\in\mathbf{T}}|\langle f,\varphi_s\rangle|^2\right]^{1/2}\\ &\lesssim \operatorname{size}(\mathbf{T})|I_\mathbf{T}|^{1/2}.\end{aligned}$$

5.11. With the notation above, assume that $0\in\omega_\mathbf{T}$. Then,

$$\begin{aligned}\operatorname{size}(\mathbf{T}) &\simeq \sup_J\left[|J|^{-1}\sum_{\substack{s\in\mathbf{T}\\ I_s\subset J}}|\langle f,\varphi_s\rangle|^2\right]^{1/2}\\ &\simeq \sup_J\left[|J|^{-1}\int_J\left|F_\mathbf{T}-|J|^{-1}\int_J F_\mathbf{T}\right|^2dx\right]^{1/2},\end{aligned}$$

where the supremum is over all intervals J. The last quantity is the *BMO* norm of $F_\mathbf{T}$.

5.12. It is an important heuristic that for a collection $\mathcal{S}$ of pairwise incomparable tiles, the functions $\{\varphi_s : s \in \mathcal{S}\}$ are nearly orthogonal. The heuristic permits a quantification in terms of the following weak type inequality. Let $\mathcal{S}$ be a collection of tiles that are pairwise incomparable with respect to '$<$'. Then for all $f \in L^2$ and all $\lambda > 0$,

$$\sum_{s \in \mathcal{S}_\lambda} |I_s| \lesssim \lambda^{-2} \|f\|_2^2,$$

where $\mathcal{S}_\lambda = \{s \in \mathcal{S} : |\langle f, \varphi_s \rangle| > \lambda \sqrt{|I_s|}\}$. Note that this in an inequality about the boundedness of a sublinear operator from $L^2(\mathbb{R})$ to $L^{2,\infty}(\mathbb{R}\times\mathcal{S})$. In the latter space, one uses counting measure on $\mathcal{S}$.

5.13. Another important heuristic is that the notion of "strong disjointness" for trees is as "pairwise incomparable" is for tiles. Let $\widetilde{\mathbf{T}}$ be a collection of strongly disjoint trees. Show that for all $f \in L^2$ and all $\lambda > 0$,

$$\sum_{\mathbf{T} \in \widetilde{\mathbf{T}}_\lambda} |I_{\mathbf{T}}| \lesssim \lambda^{-2} \|f\|_2^2,$$

where $\widetilde{\mathbf{T}}_\lambda = \{\mathbf{T} \in \widetilde{\mathbf{T}} : \Delta(\mathbf{T}) \geq \lambda |I_{\mathbf{T}}|\}$.

5.14. Let $\mathcal{S}$ be a collection of tiles that are pairwise incomparable with respect to '$<$'. Show that for all $2 < p < \infty$,

$$\left[\sum_{s \in \mathcal{S}} |\langle f, \varphi_s \rangle|^p \right]^{1/p} \lesssim \|f\|_p.$$

Notice that the form of this estimate at $p = \infty$ is obvious.

5.15. Let $\widetilde{\mathbf{T}}$ be a collection of strongly disjoint trees. Then for all $2 < p < \infty$,

$$\sum_{\mathbf{T} \in \widetilde{\mathbf{T}}} \Delta(\mathbf{T})^p \lesssim \|f\|_p.$$

5.16. The L^p estimates of the previous two complements can in some instances be improved. For each integer k,

$$\left\| \left[\sum_{\substack{s \in \mathcal{T} \\ |I_s| = 2^k}} \frac{|\langle f, \varphi_s \rangle|^2}{|I_s|} \mathbf{1}_{I_s} \right]^{1/2} \right\|_p \lesssim \|f\|_p, \quad 2 < p < \infty.$$

This can be seen by showing that

$$\sum_{\substack{s\in\mathcal{T}\\|I_s|=2^k}} \frac{|\langle f,\varphi_s\rangle|^2}{|I_s|}\mathbf{1}_{I_s} \lesssim (\mathrm{Dil}^1_{2^k}\,\chi)*|f|^2.$$

6. The Tree Lemma

We begin with some remarks about the maximal function, and a particular form of the same that we shall use at a critical point of this proof. Consider the maximal function

$$\mathrm{M}\, f = \sup_{I\in\mathcal{D}} \mathbf{1}_I |\langle f,\chi_I\rangle|.$$

It is well known that this is bounded on L^2. A proof follows. Consider a linearized version of the supremum. To each $I\in\mathcal{D}$, associate a set $E(I)\subset I$, and require that the sets $\{E(I) : I\in\mathcal{D}\}$ be pairwise disjoint. (Thus, for fixed f, $E(I)$ is that subset of I on which the supremum above is equal to $|\langle f,\chi_I\rangle|$.) Define

$$\mathrm{A}\, f = \sum_{I\in\mathcal{D}} \mathbf{1}_{E(I)}\langle f,\chi_I\rangle.$$

We show that $\|\mathrm{A}\|_2$ is bounded by a constant, independent of the choice of the sets $E(I)$.

The method is that of $\mathrm{T}\,\mathrm{T}^*$. Note that for positive f

$$\begin{aligned}\mathrm{A}\,\mathrm{A}^* f &\le 2\sum_{I\in\mathcal{D}}\sum_{|J|\le|I|} \mathbf{1}_{E(I)}\langle\chi_I,\chi_J\rangle\langle\mathbf{1}_{E(J)},f\rangle\\ &\lesssim \sum_{I\in\mathcal{D}} \mathbf{1}_{E(I)}\langle f,\chi_I\rangle.\end{aligned}$$

It follows that

$$\begin{aligned}\|\mathrm{A}^*\|_2^2 &= \sup_{\|f\|_2=1}\langle \mathrm{A}^* f,\mathrm{A}^* f\rangle\\ &= \sup_{\|f\|_2=1}\langle f,\mathrm{A}\,\mathrm{A}^* f\rangle\\ &\lesssim \|\mathrm{A}\|_2\end{aligned}$$

and so $\|\mathrm{A}\|_2\lesssim 1$, as claimed.

We shall have recourse to not only this, but a particular refinement. Let $\mathcal{J}$ be a partition of $\mathbb{R}$ into dyadic intervals. To each $J \in \mathcal{J}$, associate a subset $G(J) \subset J$, with $|G(J)| \leq \delta|J|$, where $0 < \delta < 1$ is fixed. Consider

$$\mathrm{M}_\delta\, f := \sum_{J \in \mathcal{J}} \mathbf{1}_{G(J)} \sup_{I \supset J} |\langle f, \chi_I \rangle|. \tag{6.1}$$

Then $\|\mathrm{M}_\delta\|_2 \lesssim \sqrt{\delta}$. The proof is

$$\begin{aligned}\int |\mathrm{M}_\delta\, f|^2\, dx &= \sum_{J \in \mathcal{J}} |G(J)| \sup_{I \supset J} |\langle f, \chi_I \rangle| \\ &\leq \delta \sum_{J \in \mathcal{J}} |J| \sup_{I \supset J} |\langle f, \chi_I \rangle| \\ &\leq \delta \int |\mathrm{M}\, f|^2\, dx \\ &\lesssim \delta \|f\|_2^2.\end{aligned}$$

We begin the main line of the argument. Let $\delta = \mathrm{dense}(\mathbf{T})$, and $\sigma = \mathrm{size}(\mathbf{T})$. Make a choice of signs $\varepsilon_s \in \{\pm 1\}$ such that

$$\sum_{s \in \mathbf{T}} |\langle f, \varphi_s \rangle \langle \phi_s, \mathbf{1}_E \rangle| = \int_E \sum_{s \in \mathbf{T}} \varepsilon_s \langle f, \varphi_s \rangle \phi_s\, dx.$$

By the "Schwartz tails", the integral above is supported on the whole real line. Let $\mathcal{J}$ be a partition of $\mathbb{R}$ consisting of the maximal dyadic intervals J such that $3J$ does not contain any I_s for $s \in \mathbf{T}$. It is helpful to observe that for such J if $|J| \leq |I_{\mathbf{T}}|$, then $J \subset 3I_{\mathbf{T}}$. And if $|J| \geq |I_{\mathbf{T}}|$, then $\mathrm{dist}(J, I_{\mathbf{T}}) \gtrsim |J|$. The integral above is at most the sum over $J \in \mathcal{J}$ of the two terms below.

$$\sum_{\substack{s \in \mathbf{T} \\ |I_s| \leq |J|}} |\langle f, \varphi_s \rangle| \int_{J \cap E} |\phi_s|\, dx \tag{6.2}$$

$$\int_{J \cap E} \Bigg| \sum_{\substack{s \in \mathbf{T} \\ |I_s| > |J|}} \varepsilon_s \langle f, \varphi_s \rangle \phi_s \Bigg|\, dx. \tag{6.3}$$

Notice that for the second sum to be non-zero, we must have $J \subset 3I_{\mathbf{T}}$.

The first term (6.2) is controlled by an appeal to the “Schwartz tails”. Fix an integer $n \geq 0$, and only consider those $s \in \mathbf{T}$ for which $|I_s| = 2^{-n}|J|$. Now, the distance of I_s to J is at least $\gtrsim |J|$. And,

$$|\langle f, \varphi_s\rangle| \int_{J\cap E} |\phi_s|\, dx \leq \sigma\delta(|I_s|^{-1}\operatorname{dist}(I_s, J))^{-10}|I_s|.$$

The $I_s \subset I_{\mathbf{T}}$, so that summing this over $|I_s| = 2^{-n}|J|$ will give us

$$\sigma\delta 2^{-n}\min\bigl(|J|, |I_{\mathbf{T}}|(\operatorname{dist}(J, I_{\mathbf{T}})|I_{\mathbf{T}}|^{-1})^{-10}\bigr).$$

This is summed over $n \geq 0$ and $J \in \mathcal{J}$ to bound (6.2) by $\lesssim \sigma\delta|I_{\mathbf{T}}|$, as required.

Critical to the control of (6.3) is the following observation. Let

$$G(J) = J \cap \bigcup_{\substack{s\in\mathbf{T}\\ |I_s|\geq|J|}} N^{-1}(\omega_{s+}). \tag{6.4}$$

Then $|G(J)| \lesssim \delta|J|$. To see this, let J' be the next larger dyadic interval that contains J. Then $3J'$ must contain some $I_{s'}$, for $s' \in \mathbf{T}$. Let s'' be that tile with $I_{s'} \subset I_{s''}$, $|I_{s''}| = |J|$, and $\omega_{\mathbf{T}} \subset \omega_{s''}$. Then, $s' < s''$, and by the definition of density,

$$\int_{E\cap N^{-1}(\omega_{s''})} \chi_{I_{s''}}\, dx \leq \delta.$$

But, for each s as in (6.4), we have $\omega_s \subset \omega_{s''}$, so that $G(J) \subset N^{-1}(\omega_{s''})$. Our claim follows.

Suppose that $\mathbf{T}$ is a $-$tree. That means that the tiles $\{I_s \times \omega_{s+} : s \in \mathbf{T}\}$ are disjoint. We use an estimation absent of any cancellation effects. Then, the bound for (6.3) is no more than

$$|G(J)| \Biggl\| \sum_{\substack{s\in\mathbf{T}\\ |I_s|\geq|J|}} |\langle f, \varphi_s\rangle\phi_s| \Biggr\|_\infty \lesssim \delta\sigma|J|.$$

This is summed over $J \subset 3I_{\mathbf{T}}$ to get the desired bound.

Suppose that $\mathbf{T}$ is a +tree. (This is the interesting case.) Then, the tiles $\{I_s \times \omega_{s-} : s \in \mathbf{T}\}$ are pairwise disjoint, and we set

$$F = \operatorname{Mod}_{-c(\omega_{\mathbf{T}})} \sum_{s \in \mathbf{T}} \varepsilon_s \langle f, \varphi_s \rangle \varphi_s.$$

Here it is useful to us that we only use the "smooth" functions φ_s in the definition of this function. Note that $\|F\|_2 \lesssim \sigma\sqrt{|I_{\mathbf{T}}|}$, which is a consequence of the definition of size and Proposition 2.12. Set $\tau(x) = \sup\{|I_s| : s \in \mathbf{T},\, N(x) \in \omega_{s+}\}$, and observe that for each J, and $x \in J$,

$$\sum_{\substack{s \in \mathbf{T} \\ |I_s| \geq |J|}} \varepsilon_s \langle f, \varphi_s \rangle \phi_s(x) = \sum_{\substack{s \in \mathbf{T} \\ \tau(x) \geq |I_s| \geq |J|}} \varepsilon_s \langle f, \varphi_s \rangle \varphi_s(x).$$

This is so since all of the intervals ω_{s+} must contain $\omega_{\mathbf{T}}$, and if $N(x) \in \omega_{s+}$, then it must also be in every other $\omega_{s'+}$ that is larger. What is significant here is that on the right we have a truncation of the sum that defines F.

This last sum can be dominated by a maximal function. For any $\tau > 0$ and $J \in \mathcal{J}$, let

$$F_{\tau,J} = \operatorname{Mod}_{-c(\omega_{\mathbf{T}})} \sum_{\substack{s \in \mathbf{T} \\ \tau \geq |I_s| \geq |J|}} \varepsilon_s \langle f, \varphi_s \rangle \varphi_s.$$

This function has Fourier support in the interval $\left[-\frac{7}{8}|J|^{-1}, -\frac{1}{4}\tau^{-1}\right]$. In particular, recalling how we defined φ, we can choose $\frac{1}{16} < a, b < \frac{1}{4}$ so that

$$F_{\tau,J} = (\operatorname{Dil}^1_{a|J|} \varphi - \operatorname{Dil}^1_{b\tau} \varphi) * F.$$

We conclude that for $x \in J$,

$$|F_{\tau(x),J}(x)| \lesssim \mathrm{M}_\delta\, F(x),$$

where M_δ is defined as in (6.1).

The conclusion of this proof is now at hand. We have

$$\begin{aligned}
\sum_{\substack{J \in \mathcal{J} \\ |J| \leq 3|I_{\mathbf{T}}|}} \int_{G(J)} |F_{\tau(x),J}|\, dx &\lesssim \int_{\bigcup_{|J| \leq 3|I_{\mathbf{T}}|} G(J)} \mathrm{M}_\delta\, F\, dx \\
&\lesssim \Bigg| \bigcup_{|J| \leq 3|I_{\mathbf{T}}|} G(J) \Bigg|^{1/2} \|\mathrm{M}_\delta\, F\|_2 \\
&\lesssim \delta\sqrt{|I_{\mathbf{T}}|}\|F\|_2 \\
&\lesssim \sigma\delta|I_{\mathbf{T}}|.
\end{aligned}$$

6.1. Complements.

6.5. The estimate below is somewhat cruder than the one just obtained, and therefore easier to obtain. For all trees $\mathbf{T}$,

$$\Big\|\sum_{s\in\mathbf{T}}\langle f,\varphi_s\rangle\phi_s\Big\|_2 \lesssim \Big[\sum_{s\in\mathbf{T}}|\langle f,\varphi_s\rangle|^2\Big]^{1/2}$$

$$\lesssim \operatorname{size}(\mathbf{T})|I_{\mathbf{T}}|^{1/2}.$$

6.6. The maximal function M_δ in (6.1) admits the bounds

$$\|\mathrm{M}_\delta\|_p \lesssim \delta^{1/p}, \quad 1<p<\infty.$$

This depends upon the fact that the maximal function itself maps L^p into itself, for $1<p<\infty$.

6.7. For any tree $\mathbf{T}$,

$$\Big\|\sum_{s\in\mathbf{T}}\langle g,\varphi_s\rangle\phi_s\Big\|_p \lesssim \delta^{1/p}\|g\|_p, \quad 1<p<\infty.$$

6.8. For a +tree $\mathbf{T}$,

$$\Big\|\sum_{s\in\mathbf{T}}\langle f,\varphi_s\rangle\varphi_s\Big\|_p \lesssim \Big\|\Big[\sum_{s\in\mathbf{T}}\frac{|\langle f,\varphi_s\rangle|^2}{|I_s|}\mathbf{1}_{I_s}\Big]^{1/2}\Big\|_p, \quad 1<p<\infty.$$

Conclude that

$$\Big\|\sum_{s\in\mathbf{T}}\langle f,\varphi_s\rangle\varphi_s\Big\|_p \lesssim \operatorname{size}(\mathbf{T})|I_{\mathbf{T}}|^{1/p}, \quad 1<p<\infty.$$

7. Carleson's Theorem on L^p, $1<p\neq 2<\infty$

We outline a proof that the Carleson maximal operator maps L^p into itself for all $1<p<\infty$. The key point is that we should obtain a distributional estimate for the model operator.

Proposition 7.1. *For $1<p<\infty$, there is an absolute constant K_p so that for all sets $E\subset\mathbb{R}$ of finite measure and measurable functions N, we have*

$$|\{|\mathcal{C}_N\mathbf{1}_E|>\lambda\}| \le K_p^p\lambda^{-p}|E|. \tag{7.2}$$

Interpolation provides the L^p inequalities. We shall in fact prove that for all sets E and F, there is a set $F' \subset F$ of measure $|F'| \ge \frac{1}{2}|F|$,

$$|\langle \mathcal{C}\mathbf{1}_E, \mathbf{1}_{F'}\rangle| \lesssim \min(|E|,|F|)\left(1+\left|\log \tfrac{|E|}{|F|}\right|\right). \tag{7.3}$$

It is a routine matter to see that this estimate implies that

$$|\{|\mathcal{C}\mathbf{1}_E| > \lambda\}| \lesssim |E| \begin{cases} \lambda|\log \lambda| & \text{if } 0<\lambda<1/2 \\ e^{-c\lambda} & \text{otherwise.} \end{cases} \tag{7.4}$$

Here, c is an absolute constant. This distributional inequality is in fact the best that is known about the Carleson operator. See Sjölin [**67**] for the Walsh case, and [**68**] for the Fourier case. A more recent publication proving the same point is Arias de Reyna [**2**]. Both authors present a proof along the lines of Carleson. We follow the weak type inequality approach of Muscalu, Tao, and Thiele [**56**]. The relevance of this approach to the Carleson theorem was demonstrated by Grafakos, Tao, and Terwilleger [**30**].

We shall find it necessary to appeal to some deeper properties of the Calderón Zygmund theory, and in particular a weak L^1 bound for the maximal function, but also the bound in (7.11) below.

In proving (7.3), we can rely upon invariance under dilations, up to a change in the measurable $N(x)$, to assume that $1/2 < |E| \le 1$. As we already know the weak L^2 estimate, (7.3) is obvious for $\frac{1}{3} < |F| < 3$. The argument then splits into two cases, that of $|F| < \frac{1}{3}$ or $|F| \ge 3$.

Note that our measurable function $N(x)$ is defined on the set F. It is clear that our Density Lemma, Lemma 3.6, continues to hold in this context, with the change that the measure of F should be added to the right hand side of (3.7).

7.1. The case of $|F| < \frac{1}{3}$.

In this case, we will take $F' = F$. Recall that $\mathcal{T}$ denotes the set of all tiles. Clearly, $\operatorname{size}(\mathcal{T}) \lesssim 1$. We repeat the argument of (3.13)–(3.16). Here, we should keep in mind that we want to balance out the estimate for the count$(\cdot)$ function, and that we have a better upper bound on the count function coming from the Density Lemma. Thus, $\mathcal{T}$ is a union of collections $\mathcal{S}_n$, for $n \ge 0$, so that

$$\operatorname{dense}(\mathcal{S}_n) \lesssim 2^{-2n}, \tag{7.5}$$

$$\operatorname{size}(\mathcal{S}_n) \lesssim \min(1, 2^{-n}|F|^{-1/2}), \tag{7.6}$$

$$\operatorname{count}(\mathcal{S}_n) \lesssim 2^{2n}|F|. \tag{7.7}$$

Then by the calculation of (3.16), we have

$$\sum_{s\in\mathcal{S}_n} |\langle \mathbf{1}_E, \varphi_s\rangle\langle\phi_s, \mathbf{1}_{F'}\rangle| \lesssim \operatorname{dense}(\mathcal{S}_n)\operatorname{size}(\mathcal{S}_n)2^{2n}|F|$$

$$\lesssim \min(|F|, |F|^{1/2}2^{-n}).$$

The sum of this terms over $n \geq 0$ is no more than

$$\lesssim |F| \cdot |\log|F||$$

which is as required.

7.2. The case of $|F| \geq 3$.

This case corresponds to the analysis of the Carleson operator on L^p for $1 < p < 2$. We shall have need of a more delicate weak type inequality below to complete this proof. To define the set F', let

$$\Omega = \{\mathrm{M}\,\mathbf{1}_E > C_1|F|^{-1}\}.$$

By the weak L^1 inequality for the maximal function, for an absolute choice of C_1, we have $|\Omega| < \frac{1}{2}|F|$. And we take $F' = F \cap \Omega^c$. The inner product in (7.3) is less than the sum of

$$\sum_{\substack{s\in\mathcal{T}\\ I_s\subset\Omega}} |\langle f, \varphi_s\rangle\langle\phi_s, \mathbf{1}_{F'}\rangle| \tag{7.8}$$

$$\sum_{\substack{s\in\mathcal{T}\\ I_s\not\subset\Omega}} |\langle f, \varphi_s\rangle\langle\phi_s, \mathbf{1}_{F'}\rangle|. \tag{7.9}$$

These sums are handled separately.

For (7.8), observe that φ_s is essentially supported inside of Ω while ϕ_s is essentially not supported there. Thus, we should rely upon Schwartz tails to handle this term. Let $J \subset \Omega$ be an interval such that $2^k J \subset \Omega$ but $2^{k+1}J \not\subset \Omega$. We observe two inequalities for such an interval, which are stated using the function χ_J, as defined in (3.4). The first is that

$$\int_{F'} \chi_J\, dx \leq \int_{(2^k J)^c} \chi_J\, dx \lesssim 2^{-(\kappa-1)k}.$$

Here, κ is a large constant in the definition of χ. Also, we have

$$\begin{aligned}\int_E \chi_J\,dx &\lesssim 2^k \int_E \chi_{2^{k+1}J}\,dx \\ &\lesssim 2^k \inf_{x\in 2^{k+1}J} \mathrm{M}\,\mathbf{1}_E(x) \\ &\lesssim 2^k |F|^{-1}.\end{aligned}$$

The last line follows as some point in $2^{k+1}J$ must be in Ω.

Observe that among all tiles s with $I_s = J$, there is exactly one tile s with $N(x) \in \omega_{s+}$. Hence

$$\begin{aligned}\sum_{\substack{s\in\mathcal{T}\\ I_s=J}} |\langle f,\varphi_s\rangle\langle\phi_s,\mathbf{1}_{F'}\rangle| &\lesssim |J| \sum_{\substack{s\in\mathcal{T}\\ I_s=J}} \langle \mathbf{1}_E,\chi_J\rangle\langle\mathbf{1}_{F'},\chi_J\rangle \\ &\lesssim 2^{-(\kappa-2)k}|F|^{-1}|J|.\end{aligned}$$

Recall that k is associated to how deeply J is embedded in Ω, and that Ω has measure at most $\lesssim |F|$. Hence the right hand side above can be summed over $J \subset \Omega$ to see that

$$(7.8) \lesssim 1,$$

which is better than desired.

We turn to the second estimate. Set $\mathcal{T}_{\text{out}} := \{s \in \mathcal{T} : I_s \not\subset \Omega\}$, which is the collection of tiles summed over in (7.9). The essential aspect of the definition of Ω is this lemma.

Lemma 7.10.

$$\operatorname{size}(\mathcal{T}_{\text{out}}) \lesssim |F|^{-1}.$$

Assuming the lemma, we turn to the line of argument (3.13)–(3.16). The collection $\mathcal{T}_{\text{out}}$ can be decomposed into collections $\mathcal{S}_n$, for $n \geq 0$, for which (7.5) and (7.7) holds, and in addition

$$\operatorname{size}(\mathcal{S}_n) \lesssim \min(|F|^{-1}, |F|^{-1/2}2^{-n}).$$

Then by the calculation of (3.16), we have

$$\sum_{s\in\mathcal{S}_n} |\langle \mathbf{1}_E,\varphi_s\rangle\langle\phi_s,\mathbf{1}_{F'}\rangle| \lesssim \min(1, |F|^{1/2}2^{-n}),$$

making the sum over $n \geq 0$ no more than $\log|F|$, as required. This completes the proof of the (7.2).

Proof: This is a consequence of the particular structure of a +tree $\mathbf{T}$, and the fact for $s \in \mathbf{T}$, the distance of $\operatorname{supp}(\widehat{\varphi_s})$ to $\omega_{\mathbf{T}}$ is approximately $|\omega_s|$. The Calderón Zygmund theory applies, and shows that for any choice of signs $\varepsilon_s \in \{\pm 1\}$, for $s \in \mathbf{T}$,

$$\left| \left\{ \sum_{s \in \mathbf{T}} \varepsilon_s \langle f, \varphi_s \rangle \varphi_s > \lambda \right\} \right| \lesssim \lambda^{-1} \| f | I_{\mathbf{T}} | \chi_{I_{\mathbf{T}}} \|_1, \quad \lambda > 0. \tag{7.11}$$

We apply this inequality for trees $\mathbf{T} \in \mathcal{T}_{\text{out}}$, and $f = \mathbf{1}_E$. By taking the average over all choices of signs, we can conclude a distributional estimate on the square functions

$$\Delta_{\mathbf{T}} := \left[\sum_{s \in \mathbf{T}} \frac{|\langle \mathbf{1}_E, \varphi_s \rangle|^2}{|I_s|} \mathbf{1}_{I_s} \right]^{1/2}. \tag{7.12}$$

Namely, that for each +tree $\mathbf{T} \subset \mathcal{T}_{\text{out}}$,

$$|\{\Delta_{\mathbf{T}} > \lambda\}| \lesssim \lambda^{-1} |I_{\mathbf{T}}| |F|^{-1}. \tag{7.13}$$

As this inequality applies to all subtrees of $\mathbf{T}$, it can be strengthened. (This is a reflection of the John Nirenberg inequality.) Fix the +tree $\mathbf{T} \subset \mathcal{T}_{\text{out}}$. We wish to conclude that

$$|I_{\mathbf{T}}|^{-1} \int_{I_{\mathbf{T}}} \Delta_{\mathbf{T}}^2 \, dx \lesssim |F|^{-1}. \tag{7.14}$$

For a subset $\mathbf{T}' \subset \mathbf{T}$, let

$$\operatorname{sh}(\mathbf{T}') := \bigcup_{s \in \mathbf{T}'} I_s$$

be the *shadow of* $\mathbf{T}'$. A shadow is not necessarily an interval. Define $\Delta_{\mathbf{T}'}$ as in (7.12). And finally set

$$G(\lambda) = \sup_{\mathbf{T}' \subset \mathbf{T}} |F| |\operatorname{sh}(\mathbf{T}')|^{-1} |\{\Delta_{\mathbf{T}'} > \lambda\}|. \tag{7.15}$$

Notice that (7.13) implies that $G(\lambda) \lesssim \lambda^{-1}$, for $\lambda > 0$. If we show that $G(\lambda) \lesssim \lambda^{-4}$, for $\lambda > 1$, we can conclude (7.14). In fact we can show that $G(\lambda)$ decays at an exponential squared rate, which is the optimal estimate.

Observe that (7.14), implies that we have

$$\frac{|\langle \mathbf{1}_E, \varphi_s \rangle|}{\sqrt{|I|}} \leq \lambda_0 < \infty.$$

Thus, the square functions we are considering $\Delta_{\mathbf{T}}$ can only take incremental steps of a strictly bounded size.

For any $\lambda \ge \sqrt{2}\lambda_0$, let us bound $G(\sqrt{2}\lambda)$. Fix $\mathbf{T}'$ achieving the supremum in the definition of $G(\sqrt{2}\lambda)$. Consider a somewhat smaller threshold, namely $\{\Delta_{\mathbf{T}'} > \lambda\}$. In order to proceed, consider a function $\tau\colon \operatorname{sh}\mathbf{T}' \mapsto \mathbb{R}_+$ such that

$$\sum_{\substack{s\in\mathbf{T}\\ |I_s|\ge\tau(x)}} \frac{|\langle f,\varphi_s\rangle|^2}{|I_s|}\mathbf{1}_{I_s}(x) \ge \lambda^2.$$

In addition require that $\tau(x)$ is the smallest such function satisfying this condition. It is the case that the sum above can be no more than $\lambda^2+\lambda_0^2$.

Take $\mathbf{T}'' \subset \mathbf{T}'$ to be the tree

$$\mathbf{T}'' := \{s \in \mathbf{T}' : |I_s| \le \tau(x),\ x \in I_s\}.$$

The point of these definitions is that

$$\Delta_{\mathbf{T}'}(x) \ge \sqrt{2}\lambda \quad \text{implies} \quad \Delta_{\mathbf{T}''}(x) \ge \sqrt{\lambda^2-\lambda_0^2}.$$

Therefore,

$$\begin{aligned}
|F|^{-1}\operatorname{sh}(\mathbf{T}')G(\sqrt{2}\lambda) &= |\{\Delta_{\mathbf{T}'} > \sqrt{2}\lambda\}| \\
&\le |\{\Delta_{\mathbf{T}''} > \sqrt{\lambda^2-\lambda_0^2}\}| \\
&\le |F|^{-1}\operatorname{sh}(\mathbf{T}'')G(\sqrt{\lambda^2-\lambda_0^2}) \\
&\le |F|^{-1}\operatorname{sh}(\mathbf{T}')G(\lambda)G(\sqrt{\lambda^2-\lambda_0^2}).
\end{aligned}$$

We conclude that $G(\sqrt{2}\lambda) \le G(\sqrt{\lambda^2-\lambda_0^2})^2$.

To conclude, we should in addition require that λ_0 is so large that $G(\kappa\lambda_0) \le \frac{1}{2}$, where

$$\kappa := \prod_{k=1}^{\infty}\sqrt{1-2^{-k}}.$$

An induction argument will then show that

$$G(2^{k/2}\lambda_0) \le G(\kappa\lambda_0)^{2^k} \le 2^{-2^k}, \quad k \ge 0, \tag{7.16}$$

which is the claimed exponential decay. Our proof of the lemma is done. □

7.3. Complements.

7.17. In the inequality (7.2), one can show that the constants K_p on the right hand side obey

$$K_p \lesssim \frac{p^2}{p-1}.$$

7.18. If it is the case that for some $0 < \alpha < 1$, we have the inequality

$$\sup_{\mathbf{T}' \subset \mathbf{T}} |\mathrm{sh}(\mathbf{T}')|^{-1/\alpha} \|\Delta_{\mathbf{T}'}\|_\alpha < \infty$$

then, the stronger estimate below holds.

$$\|\Delta_{\mathbf{T}}\|_1 \lesssim |\mathrm{sh}(\mathbf{T})|.$$

7.19. In (7.2), we assert the restricted weak type inequality for $1 < p < \infty$. The weak type estimate for $2 < p < \infty$ is in fact directly available. That is, for $2 < p < \infty$, and $f \in L^p$ of norm one,

$$|\{\mathcal{C}_N f > \lambda\}| \lesssim \lambda^{-p}, \quad \lambda > 0.$$

The key point is to take advantage of the fact that f is locally square integrable. A very brief sketch of the argument follows. (1) It suffices to prove the inequality above for $\lambda = 1$. (2) Define $\Omega = \{\mathrm{M}|f|^2 > 1\}$, and show that $|\Omega| \lesssim 1$. (3) Define sums as in (7.8) and (7.9), and control each term separately. One will need to replace the Size Lemma as stated with 5.15.

8. Remarks

8.1. After Carleson [**14**] proved his theorem, Hunt [**32**] extended the argument to L^p, for $1 < p < \infty$. A similar extension, in the Walsh Paley case, was done by Billiard [**9**], in the case of L^2, and Sjölin [**67**], for all $1 < p < \infty$. The Carleson theorem has equivalent formulations on the groups $\mathbb{R}$, $\mathbb{T}$, and $\mathbb{Z}$. The last case, of the integers, was explicitly discussed by Máté [**55**]. This paper was overlooked until recently.

8.2. Fefferman [**26**] devised an alternate proof, which proved to be influential through it's use of methods of analysis that used both time and frequency information in an operator theoretic fashion. The proof of Lacey and Thiele [**50**] presented here borrows several features of that proof. The notion of tiles, and the partial order on tiles is due to Fefferman [**26**]. Likewise, the Density Lemma and the Tree Lemma, and the proof of the same, have clear antecedents in this paper.

8.3. Those familiar with the Littlewood Paley theory know that it is very useful in decoupling the scales of operators, like those of the Hilbert transform. The Carleson operator is, however, not one in which scales can be decoupled. This is another source of the interest in this theorem.

8.4. We present the proof of the Carleson theorem on the real line due to the presence of the dilation structure.

8.5. We choose to express the Carleson operator in terms of the projection P_-. This operator is a linear combination of the identity operator and the Hilbert transform given by

$$\mathrm{H}\, f(x) := \lim_{\epsilon \to 0} \int_{\epsilon < |y| < 1/\epsilon} f(x-y) \frac{dy}{y}.$$

Hence, an alternate form of the Carleson operator is

$$\sup_N |\mathrm{H} \circ \mathrm{Mod}_N\, f| = \sup_N \left| \int e^{iNy} f(x-y) \frac{dy}{y} \right|. \tag{8.6}$$

This form is suggestive of other questions related to the Carleson Theorem, a point we rely upon below.

8.7. Despite the fact that Carleson's operator maps L^2 into itself, all three known proofs of Carleson's theorem establish the weak type bound on L^2. The strong type bound must be deduced by interpolation. On the other hand, the weak type bound is a known consequence of the pointwise convergence of Fourier series. This was observed by Calderón, as indicated by a footnote in [**84**], and is a corollary of a general observation of Stein [**70**].

8.8. Hunt and Young [**33**] have established a weighted estimate for the Carleson operator. Namely for a weight w in the class A^p, the Carleson operator maps $L^p(w)$ into itself, for $1 < p < \infty$. The method of proof utilizes the known Carleson bound, and distribution inequalities for the Hilbert transform.

8.9. The Proposition 2.8 has a well known antecedent in a characterization of (a constant times) the Hilbert transform as the unique operator A such that A is bounded operator on L^2 that commutes with dilation, is invariant under dilations, A^2 is the identity, but is not itself the identity. See [**71**].

8.10. The inequality (3.3) eschews all additional cancellations. It shows that all the necessary cancellation properties are already encoded in the decomposition of the operator. In addition the combinatorial model of the Carleson operator is in fact unconditionally convergent in $s \in \mathcal{T}$.

This turns out to be extremely useful fact in the course of the proof: one is free to group the tiles in anyway that one likes.

8.11. The Size Lemma should be compared to Rubio de Francia's extension of the classical Littlewood Paley inequality [**66**]. Also see the author's recent survey of this theorem [**44**].

8.12. A +tree $\mathbf{T}$ is a familiar object. Aside from a modulation by $c(\omega_{\mathbf{T}})$, it shares most of the properties associated with sums of wavelets. In particular, if $0 \in \omega_{\mathbf{T}}$, note that

$$\operatorname{size}(\mathbf{T}) \simeq \Big\| \sum_{s \in \mathbf{T}} \langle f, \varphi_s \rangle \varphi_s \Big\|_{BMO}$$

where the last norm is the *BMO* norm.

8.13. The key instance of the Tree Lemma is that of a +tree. This case corresponds to a particular maximal function applied to a function associated to the tree. It is this point at which the supremum of Carleson's theorem is controlled by a much tamer supremum: The one in the ordinary maximal function.

8.14. The statement and proof of the size lemma, Lemma 3.9, replaces the initial arguments of this type that are in [**47**]. This argument has proven to be very flexible in it's application. And, in some instances it produces sharp estimates, as explained by Barrionuevo and Lacey [**8**].

8.15. The set of functions $\mathcal{T}_k := \{s \in \mathcal{T} : |I_s| = 2^k\}$ is an example of a Gabor basis. For appropriate choice of φ, the operator

$$\mathrm{A}_k f := \sum_{\substack{s \in \mathcal{T} \\ |I_s| = 2^k}} \langle f, \varphi_s \rangle \varphi_s$$

is in fact the identity operator. See the survey of Daubechies [**25**].

9. Complements and Extensions

9.1. Equivalent formulations of Carleson's theorem.

The Fourier transform has a formulation on each of the Euclidean groups $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{T}$. Carleson's original proof worked on $\mathbb{T}$. Fefferman's proof translates very easily to $\mathbb{R}$. Mátè [**55**] extended Carleson's proof to $\mathbb{Z}$. Each of the statements of the theorem can be stated in terms of a maximal Fourier multiplier theorem, and we have stated it as such in this paper. Inequalities for such operators can be transferred between these three Euclidean groups, and was done so by Auscher and Carro [**7**].

9.2. Fourier series near L^1, Part 1.

The point of issue here is the determination of that integrability class which guarantees the pointwise convergence of Fourier series. The natural setting for these questions is the unit circle $\mathbb{T} = [0,1]$, and the partial Fourier sums

$$S_N f(x) = \sum_{|n| \le N} \widehat{f}(n) e^{2\pi i n x}, \quad \widehat{f}(n) = \int_0^1 f(x) e^{-inx}\, dx.$$

In the positive direction, one seeks the "smallest" function ψ such that if $\int_{\mathbb{T}} \psi(f)\, dx < \infty$, then the Fourier series of f converge pointwise.

Antonov [**1**] has found the best result to date,

Theorem 9.1. *For all functions $f \in L(\log L)(\log\log\log L)(\mathbb{T})$, the partial Fourier series of f converges pointwise to f.*

This extends the result of Sjölin [**67**], [**68**], who had the result above, but with a double log where there is a triple log above. Arias de Reyna [**2**], [**3**] has noted an extension of this theorem, in that one can define a rearrangement invariant Banach space B, so that pointwise convergence holds for all $f \in B$, and B contains $L(\log L)(\log\log\log L)$.

The method of proof takes as it's starting point, the distributional estimate of (7.4). One seeks to "extrapolate" these inequalities to the setting of the theorem above and Antonov nicely exploits the explicit nature of the kernels involved in this maximal operator. Also see the work of Sjölin and Soria [**69**] who demonstrate that Antonov's approach extends to other maximal operator questions.

9.3. Fourier series near L^1, Part 2.

In the negative direction, Kolmogorov's fundamental example [**36**], [**37**] of an integrable function with pointwise divergent Fourier series admits a strengthening to the following statement, as obtained by Körner [**40**].

Theorem 9.2. *For all $\psi(x) = o(\log\log x)$, there is a function $f : [0, 2\pi] \to \mathbb{R}$ with divergent Fourier series, and $\int |f|\psi(f)\, dx < \infty$.*

The underlying method of proof was, in some essential way, unsurpassed until quite recently, when Konyagin [**38**], [**39**] proved

Theorem 9.3. *The previous theorem holds assuming only*

$$\psi(x) = o\left(\sqrt{\frac{\log x}{\log\log x}}\right). \tag{9.4}$$

There is a related question, on the growth of partial sums of Fourier series of integrable functions. Hardy [**31**] showed that for integrable functions f, one has $S_n f = o(\log n)$ a.e., and asked if this is the best possible estimate. This question is still open, with the best result from below following from Konyagin's example. With ψ as in (9.4), there is an $f \in L^1(\mathbb{T})$ with

$$\limsup_{n\to\infty} \frac{S_n f}{\psi(n)} = \infty \quad \text{for all } x \in \mathbb{T}.$$

Let δ_t denote the Dirac point mass at $t \in \mathbb{T}$. The method of proof is to construct measures

$$\mu = K^{-1} \sum_{k=1}^{K} \delta_{t_k}$$

a set $E \subset \mathbb{T}$ with measure at least $1/4$, and choices of integers N for which

$$\sup_{n<N} |S_n \mu(x)| \geq \psi(N), \quad x \in E.$$

Kolmogorov's example consists of uniformly distributed point masses, whereas Konyagin's example consists of point masses that have a distribution reminiscent of a Cantor set.

9.3.1. Probabilistic series.

It is of interest from the point of view of probability and ergodic theory, to consider the version of the Hilbert transform and Carleson theorem that arises from the integers. Here, we consider the probabilistic versions. Let X_k be independent and identically distributed copies of a mean zero random variable X. The question is if the sum

$$\sum_{k=1}^{\infty} \frac{X_k}{k}$$

converges a.s. Without additional assumption on the distribution of X, a necessary and sufficient condition is that $\mathbb{E}X \log(2 + |X|) < \infty$. One direction of this is in [**53**]. If however X is assumed to be symmetric, integrability is necessary and sufficient. This addresses the issue of the Hilbert transform.

Carleson's theorem, in this language, concerns the convergence of the series

$$Y(t) := \sum_{k=1}^{\infty} \frac{X_k}{k} e^{2\pi i k t} \quad \text{for all } t \in \mathbb{T}.$$

The role of the quantifiers should be emphasized. Convergence holds for all $t \in \mathbb{T}$, on a set of full probability. Given this, abstract results on 0–1 Laws assure us that if the series converges for all t, off of a single set of probability zero, then the limiting function is continuous with probability one. The paper of Talagrand [**79**] gives necessary and sufficient conditions for the convergence of this series.

Theorem 9.5. *Let X_k be independent identically distributed copies of a mean zero symmetric random variable X. $X \in L(\log\log L)$ iff the series $Y(t)$ converges to a continuous function on $\mathbb{T}$ almost surely.*

The assumption of symmetry should be added to the statement of the theorem in [**79**]. Cuzick and Lai [**24**] provide an example of a non symmetric mean zero $X \in L(\log\log L)$ for which the series $Y(t)$ is divergent. This series is a borderline series in that it just falls out of the scope of the powerful theory of Marcus and Pisier [**54**] on random Fourier series.

My thanks to several people who provided me with some references for this section. They are James Campbell, Ciprian Demeter, Michael Lin, and Anthony Quas.

9.4. The Wiener-Wintner question.

A formulation of Carleson's operator on $\mathbb{Z}$ is

$$C_{\mathbb{Z}} f(j) := \sup_{\tau} \sup_{N} \left| \sum_{0<|k|<N} f(j-k) \frac{e^{i\tau k}}{k} \right| .$$

See [**55**]. Unaware of this work which followed soon after Carleson, Campbell and Petersen [**12**] considered this operator on ℓ^2, with equivalence in ℓ^p established by Assani and Petersen [**5**], [**12**]. Also see Assani, Petersen and White [**6**], for these and other equivalences. The latter authors had additional motivations from dynamical systems, which we turn to now.

Calderón [**11**] observed that inequalities for operators on $\mathbb{Z}$ which commute with translation can be transferred to discrete dynamical systems. Let (X, μ) be a probability space, and $T\colon X \to X$ a map which preserves μ measure. Thus, $\mu(T^{-1}A) = \mu(A)$ for all measurable $A \subset X$. A Carleson operator on (X, μ, T) is

$$\mathrm{C_{mps}}\, f(x) := \sup_{\tau} \sup_{N} \left| \sum_{0<|k|<N} f(T^k x) \frac{e^{i\tau k}}{k} \right| .$$

And it is a consequence of Calderón's observation and Carleson's theorem that this operator is bounded on $L^2(X)$.

There is however a curious point that distinguishes this case from the other settings of Euclidean groups. It is the case that one has pointwise convergence of

$$\lim_{N\to\infty} \sum_{0<|k|<N} f(T^k x)\frac{e^{i\tau k}}{k} \quad \text{exists for all } \tau$$

holding for almost every $x \in X$? The boundedness of the maximal function C_{mps} shows that this would hold on a closed set in $L^2(X)$. The missing ingredient is the dense class for which the convergence above holds. Unlike the setting of Euclidean groups, there is no natural dense class.

This conjecture was posed by Campbell and Petersen [**12**].

Conjecture 9.6. *For all measure preserving systems (X, μ, T), and all $f \in L^2(X)$, we have the following:*

$$\mu\left\{\lim_{N\to\infty} \sum_{0<|k|<N} f(T^k x)\frac{e^{i\tau k}}{k} \quad \textit{exists for all } \tau\right\} = 1.$$

A theorem of Wiener and Wintner [**83**] provides a classical motivation of this question. This theorem concerns the same phenomena, but with the discrete Hilbert transform replaced by the averages.

Theorem 9.7. *For all measure preserving systems (X, μ, T), and all $f \in L^2(X)$, we have the following:*

$$\mu\left\{\lim_{N\to\infty} N^{-1}\sum_{k=0}^{N-1} f(T^k x)e^{i\tau k} \quad \textit{exists for all } \tau\right\} = 1.$$

This theorem admits a simple proof. And note that this theorem trivially supplies a dense class in all L^p spaces, $1 \le p < \infty$.

The Wiener-Wintner theorem has several interesting variants, for which one can phrase related questions by replacing averages by Hilbert transforms. As far as is known to us, none of these questions is answered.

An attractive theorem proved by Lesigne [**51**], [**52**] is

Theorem 9.8. *For any measure preserving system (X, μ, T) and all integrable functions f, there is a subset $X_f \subset X$ of full measure so that for all $x \in X_f$, all polynomials p, and all 1-periodic functions ϕ, the limit below exists:*

$$\lim_{N\to\infty} N^{-1}\sum_{n=1}^{N} \phi(p(n)) f(T^n x).$$

Extending this theorem to the Hilbert transform would be an extraordinary accomplishment, whereas if one replaced the discrete dynamical system by flows, it could be that the corresponding result for the Hilbert transform might be within reach.

In connection to this, Arkhipov and Oskolkov [**4**] have proved the following theorem.

Theorem 9.9. *For all integers d,*

$$\sup_{\deg(p)=d} \left| \sum_{n\neq 0} \frac{e^{ip(n)}}{n} \right| < \infty,$$

with the supremum formed over all polynomials of degree d.

This is a far more subtle fact than the continuous analog stated in (9.10). Arkhipov and Oskolkov use the Hardy Littlewood Circle method of exponential sums, with the refinements of Vinogradov. See [**4**], [**62**], [**63**].

By Plancherel, this theorem shows that for a polynomial p which maps the integers to the integers, the operators on the integers given by

$$T_p f(j) = \sum_{n\neq 0} f(x - p(n))$$

is a bounded operator on $\ell^2(\mathbb{Z})$.

Stein and Wainger have established ℓ^2 mapping properties for certain Radon transforms [**76**], [**77**].

9.5. E. M. Stein's maximal function.

A prominent theme of the research of Stein and Wainger concerns oscillatory integrals, with polynomial phases. It turns out to be of interest to determine what characteristics of the polynomial govern allied analytic quantities. In many instances, this characteristic is just the the degree of the polynomial. For instance, the following is a corollary to a theorem of Stein and Wainger from 1970 [**75**]. Namely, that one has a bound

$$\sup_{\deg(P)=d} \left| \int e^{ip(y)} \frac{dy}{y} \right| \lesssim 1, \quad d = 1, 2, \ldots \tag{9.10}$$

A conjecture of Stein's concerns an extension of Carleson's maximal operator to one in which one forms a supremum over all polynomial choices of phase with a fixed degree. Thus,

Conjecture 9.11. *For each integer d, the maximal function below maps L^p into itself for $1 < p < \infty$.*

$$\mathcal{C}_d f(x) = \sup_{\deg(P)=d} \left| \int e^{ip(y)} f(x-y) \frac{dy}{y} \right|.$$

Note that the case of $d = 1$ corresponds to Carleson's theorem. Let us set $\mathcal{C}'_d$ to be the maximal operator above, but with the the restriction that the polynomials p do not have a linear term. It is useful to make this distinction, as it is the linear terms that are intertwined with the Fourier transform.

Stein [**73**] considered the purely quadratic terms, and showed that $\mathcal{C}'_2$ maps L^p into itself for all $1 < p < \infty$. The essence of the matter is the bound on L^2, and there his argument is a variant on the method of $\mathrm{T\,T}^*$, emphasizing a frequency decomposition of the operator. Stein and Wainger [**78**] have proved that $\mathcal{C}'_d$ is bounded on all L^p's, for all $d \geq 2$. Again the L^2 case is decisive and the argument is an application of the $\mathrm{T\,T}^*$ method, but with a spatial decomposition of the operator.

Let us comment in a little more detail about how these results are proved. If, for the moment, one consider a fixed polynomial $P(y)$, and the oscillatory integral

$$T_P f(x) := \int e^{iP(y)} f(x-y) \frac{dy}{y}. \tag{9.12}$$

One may utilize the scale invariance of the the Hilbert transform kernel to change variables. With the correct change of variables, one may assume that the polynomial $P(y) = \sum_{j=1}^d a_j y^j$ satisfies $\sum |a_j| = 1$. Then, it is evident that for $|y| < 1$, say, that the integral above is well approximated by a truncation of the Hilbert transform. Thus, it is those scales of the operator larger than 1 that must be controlled.

It is a consequence of the van der Corput estimates that some additional decay can be obtained from these terms. In particular, one has this estimate. To set notation, in the one dimensional case only, set

$$P_{\vec{a}}(x) = a_d x^d + \cdots + a_1 x, \quad \vec{a} = (a_d, \ldots, a_1).$$

Lemma 9.13. *Let χ be a smooth bump function. Then we have the estimate*

$$\left\| \widehat{\mathrm{e}^{iP_{\vec{a}}(y)}\, \chi(y)}(\xi) \right\|_\infty \lesssim (1 + \|\vec{a}\|_1)^{-1/d}.$$

In particular, by the Plancherel identity, we have the estimate

$$\left\| [\mathrm{e}^{iP_{\vec{a}}(y)}\, \chi(y)] * f(x) \right\|_2 \lesssim (1 + \|\vec{a}\|_1)^{-1/d} \|f\|_2. \tag{9.14}$$

Notice that these estimates are better than the trivial ones. And that the second estimate can be interpolated to obtain a range of L^p inequalities for which one has decay, with a rate that depends upon the degree and the L^p space in question.

In a discussion of the extensions of this principle in for example [**73**], [**78**], one establishes appropriate extensions of this last lemma, always seeking some additional decay that arises from the polynomial. For instance, in [**78**], Stein and Wainger prove a far reaching extension of this principle.

Lemma 9.15. *There is a constant $\delta > 0$, depending only on the degree d, so that we have the estimate $\|S_\lambda\|_2 \lesssim (1+\lambda)^{-\delta}$, for all $\lambda > 0$, where*

$$\mathrm{S}_\lambda f(x) := \sup_{\substack{\|\vec{a}\|_1 \geq \lambda \\ a_1 = 0}} \sup_{t>0} |[\mathrm{Dil}_t^1 \, \mathrm{e}^{iP_{\vec{a}}(\cdot)} \, \chi] * f|.$$

It is essential in this supremum be formed over polynomials $P_{\vec{a}}$ which do not have a linear term.

Ionescu has pointed out that this lemma is not true with the linear term included, even in the case of second degree polynomials. The example, which we will see again below, begins by taking a function $f(x)$, and replacing it by the function $g(x) = \mathrm{e}^{i\lambda x^2} f(x)$. Then, in the supremum defining S_λ above, take the dilation parameter to be $t = 1$, and the polynomial to be $P(y) = -y^2 + 2xy$. Note that as we are taking a supremum, we can in particular take a polynomial that depends upon x.

In this example, the modulation of f by "chirp" is then canceled out by the choice of P. There is no decay in the estimate. This estimate is special to the case of the second power, so it is natural to guess that it plays a distinguished role in these considerations. This also points out an error in the author's paper [**43**]. (The error enters in specifically at the equation (2.9). The phase plane analysis of that paper might yet find some use.) At this point, the resolution of Stein's conjecture is not settled. And it appears that a positive bound of the operator $\mathcal{C}_2$ will in particular require a novel phase plane analysis with quadratic phase. This should be compared to the notion of *degeneracy* in Section 9.8 below.

9.6. Fourier series in two dimensions.

In this section we extend the Fourier transform to functions of the plane

$$\widehat{f}(\xi) = \int f(x) e^{ix\cdot\xi} \, dx$$

where $\xi = (\xi_1, \xi_2)$, and $x = (x_1, x_2)$. The possible extensions of Carleson's theorem to the two dimensional setting are numerous. The state of our knowledge is not so great.

9.6.1. Fourier series in two dimensions, Part I.

Consider the pointwise convergence of the Fourier sums in the plane given by

$$\int_{tP} \widehat{f}(\xi) e^{ix\cdot\xi}\, dx.$$

Here, P is polygon with finitely many sides in the plane, with the origin in it's interior. Proving the pointwise convergence of these averages is controlled by the maximal function

$$\mathcal{C}_P f := \sup_{t>0} \left| \int_{tP} \widehat{f}(\xi) e^{ix\cdot\xi}\, dx \right|.$$

Fefferman [**28**] has observed that this maximal operator can be controlled by a sum of operators which are equivalent to the Carleson operator.

For simplicity, we just assume that the polygon is the unit square. And let

$$Qf = \int_0^\infty \int_{-\xi_1}^{\xi_1} \widehat{f}(\xi_1, \xi_2) e^{i(x_1\xi_1 + x_2\xi_2)}\, d\xi_2\, d\xi_1.$$

This is the Fourier projection of f onto the sector swept out by the right hand side of the square. Notice that $\mathcal{C}_P \circ Q$ is the one dimensional Carleson operator applied in the first coordinate.

Thus, $\mathcal{C}_P$ is dominated by a sum of terms which are obtained from the one dimensional Carleson operator, and so $\mathcal{C}_P$ maps L^p into itself for $1 < p < \infty$. This argument works for any polygon with a finite number of sides. While we have stressed the two dimensional aspect of this argument, it also works in any dimension.

Nevertheless, it is of some interest to consider maximal operators of the form

$$\sup_{\xi\in\mathbb{R}^2} \left| \int f(x-y) e^{i\xi\cdot y} K(y)\, dy \right|$$

where K is a Calderón Zygmund kernel. This is the question addressed by Sjölin [**28**], and more recently by Sjölin and Prestini [**64**] and Grafakos, Tao and Terwilleger [**30**].

9.6.2. Fourier series in two dimensions, Part II.

What other methods can be used to sum Fourier series in the plane? One method that comes to mind is over arbitrary rectangles. That is, one considers the maximal operator

$$\mathcal{R}f(x) := \sup_{\omega} \left| \int_{\omega} \widehat{f}(\xi) e^{i\xi \cdot x} \, d\xi \right| .$$

The supremum is formed over arbitrary rectangles ω with center at the origin. Fefferman [**27**] has shown however that this is a badly behaved operator.

Proposition 9.16. *There is a bounded, compactly supported function f for which $\mathcal{R}f = \infty$ on a set of positive measure.*

This maximal operator has an alternate formulation, see (8.6), as

$$\sup_{\alpha,\beta} \left| \iint f(x - x', y - y') e^{i(\alpha x' + \beta y')} \frac{dx'}{x'} \frac{dy'}{y'} \right| . \tag{9.17}$$

The example of Fefferman is a sum of terms $f_\lambda(x, y) = e^{i\lambda xy} \chi(x, y)$, where $\lambda > 3$, and χ is a smooth bump function satisfying e.g. $\mathbf{1}_{[-1,1]^2} \leq \chi \leq \mathbf{1}_{[-2,2]}$. The key observation in the construction of the example is

Lemma 9.18. *We have the estimate*

$$\mathcal{R}f_\lambda(x, y) \gtrsim \log \lambda$$

for $(x, y) \in \left[-\frac{1}{2}, \frac{1}{2}\right]^2$.

Proof: In the supremum over α and β in the definition of $\mathcal{R}$, let $\alpha = \lambda y$ and $\beta = \lambda x$, and consider

$$\begin{aligned} R(x, y) &= \int \left[\int f_\lambda(x - x', y - y') e^{i\lambda(x'y + xy')} \frac{dx'}{x'} \right] \frac{dy'}{y'} \\ &= e^{i\lambda xy} \int \left[\int e^{i\lambda x'y'} \chi(x - x', y - y') \frac{dx'}{x'} \right] \frac{dy'}{y'} . \end{aligned}$$

The inside integral in the brackets admits these two estimates for all $(x, y) \in \left[-\frac{1}{2}, \frac{1}{2}\right]^2$.

$$\begin{aligned} I(x, y, y') &= \left[\int e^{i\lambda x'y'} \chi(x - x', y - y') \frac{dx'}{x'} \right] \\ &= \begin{cases} c \operatorname{sign}(\lambda y') + O(\lambda |y'|)^{-1} \\ O(1 + \lambda |y'|) \end{cases} \end{aligned}$$

c is a non-zero constant. Both of these estimates are well-known.

Then, we should estimate

$$R(x,y) = \int_{|y-y'|<1/\lambda} I(x,y,y') \frac{dy'}{y'} + \int_{|y-y'|>1/\lambda} I(x,y,y') \frac{dy'}{y'}.$$

The first term on the right is no more than $O(1)$, and the second term is $\gtrsim \log 1/\lambda$. $\square$

This example shows that there are bounded functions f for which

$$\sup_{|\alpha|,|\beta|<N} \left| \iint f(x-x', y-y') e^{i(\alpha x' + \beta y')} \frac{dx'}{x'} \frac{dy'}{y'} \right| \gtrsim \log N.$$

It might be of interest to know if this estimate is best possible.

The integrals in (9.17) are singular integrals in the product setting. There is as of yet no positive results relating to Carleson's theorem in a product setting.

9.6.3. Fourier series in two dimensions, Part III.

The exponential $e^{i\xi \cdot x}$ is an eigenfunction of $-\Delta$, the positive Laplacian, with eigenvalue $|\xi|^2$. It would be appropriate to sum Fourier series according to this quantity. This concerns the operator of Fourier restriction to the unit disc

$$Tf := \int_{|\xi|<1} \widehat{f}(\xi) e^{i\xi \cdot x} \, d\xi.$$

It is a famous result of Fefferman [**29**] that T is bounded on $L^2(\mathbb{R}^2)$ iff $p = 2$. The fundamental reason for this is the existence of the Besicovitch set, a set contained in a large ball, that contains a unit line segment in each direction, but has Lebesgue measure one. The relevance of this set is indicated by the observation that the restriction of T to very small disc placed on the boundary of the disc is well approximated by a projection onto a half space. Such a projection is a one dimensional Fourier projection performed in the normal direction to the disc. And the normal directions can point in arbitrary directions. An extension of Fefferman's argument shows that the Fourier restriction to any smooth set with a curved boundary can only be a bounded operator on L^2.

Nevertheless, the question of summability in the plane remains open. Namely,

Question 9.19. *Is it the case that the maximal operator below maps $L^2(\mathbb{R}^2)$ into weak $L^2(\mathbb{R}^2)$?*

$$\sup_{r>0} \left| \int_{|\xi|<r} \widehat{f}(\xi) e^{i\xi \cdot x} \, d\xi \right|.$$

In the known proofs of Carleson's theorem, the truncations of singular integrals plays a distinguished role. In the proof we have presented, this is seen in the Tree Lemma, also cf. 8.13. In view of this, it interesting to suppose that if this conjecture is true, what could play a role similar to the Tree Lemma. It appears to be this.

Question 9.20. *Is it the case that the maximal operator below maps $L^2(\mathbb{R}^2)$ into weak $L^2(\mathbb{R}^2)$?*

$$\sup_{k\in\mathbb{N}} \left| \int_{|\xi|<1+2^{-k}} \widehat{f}(\xi) e^{i\xi\cdot x}\, d\xi \right| .$$

It is conceivable that a positive answer here could lead to a proof of spherical convergence of Fourier series.

9.6.4. Fourier series in two dimensions, Part IV.

In order to bridge the gaps between Parts I and III, the following question comes to mind. Is there a polygon with infinitely many sides which one could sum Fourier series with respect to?

Mockenhoupt pointed out to the author that there is a natural first choice for P. It is a polygon P_{lac} which in the first quadrant has vertices at the points $e^{\pi i 2^{-k}}$ for $k \in \mathbb{N}$. Call this the *lacunary sided polygon.*

It is a fact due to Córdoba and Fefferman [**23**] that the lacunary sided polygon is a bounded L^p multiplier, for all $1 < p < \infty$. That is the operator below maps L^p into itself for $1 < p < \infty$.

$$\mathrm{T}_{\mathrm{lac}}\, f(x) = \int_{P_{\mathrm{lac}}} \widehat{f}(\xi) e^{i\xi\cdot x}\, d\xi.$$

This fact is in turn linked to the boundedness of the maximal function in a lacunary set of directions:

$$\mathrm{M}_{\mathrm{lac}}\, f(x) = \sup_{k\in\mathbb{N}} \sup_{t>0} (2t)^{-1} \int_{-t}^{t} |f(x - u(1, 2^{-k}))|\, du.$$

Note that this is a one dimensional maximal function computed in a set of directions in the plane that, in a strong sense, is zero dimensional. Let us state as a conjecture.

Conjecture 9.21. *For $2 \le p < \infty$, the maximal function below maps $L^p(\mathbb{R}^2)$ into itself.*

$$\sup_{t>0} \left| \int_{tP_{\mathrm{lac}}} \widehat{f}(\xi) e^{i\xi\cdot x}\, d\xi \right| .$$

Even a restricted version of this conjecture remains quite challenging. In analogy to Question 9.20

Conjecture 9.22. *For $2 \le p < \infty$, the maximal function below maps $L^p(\mathbb{R}^2)$ into itself.*

$$\sup_{k\in\mathbb{N}} \left| \int_{(1+2^{-k})P_{\mathrm{lac}}} \widehat{f}(\xi) e^{i\xi\cdot x}\, d\xi \right| .$$

Another question, with a somewhat more quantitative focus, considers uniform polygons with N sides, but then seek norm bounds on these two maximal operators, on L^2 say, which grow logarithmically in N. We do not have a good conjecture as to the correct order of growth of these constants. If one could prove that the bounds where independent of N, then the spherical summation conjecture would be a consequence.

9.7. The bilinear Hilbert transforms.

The bilinear Hilbert transforms are given by

$$\mathrm{H}_\alpha(f,g)(x) := \int f(x-\alpha y) g(x-y) \frac{dy}{y}, \quad \alpha \in \mathbb{R},$$

with the convention that $\mathrm{H}_0(f,g) = f\,\mathrm{H}\,g$, and $\mathrm{H}_\infty(f,g) = (\mathrm{H}\,f)g$. A third degenerate value is $\alpha = 1$.

These transforms commute with appropriate joint translations of f and g, and dilations of f and g. They are related to Carleson's theorem through the observation that for $\alpha \notin \{0,1,\infty\}$, H_α enjoys an invariance property with respect to modulation. Namely,

$$\mathrm{H}_\alpha(\mathrm{Mod}_\beta f, \mathrm{Mod}_{-\alpha\beta}\, g) = \mathrm{Mod}_{\beta-\alpha\beta}\, \mathrm{H}_\alpha(f,g).$$

That is, the bilinear Hilbert transforms share the essential characteristics of the Carleson operator.

It was the study of these transforms that lead Lacey and Thiele to the proof of Carleson's theorem presented here. The bilinear Hilbert transforms are themselves interesting objects, with surprising properties. Indeed, it is natural to ask what L^p mapping properties are enjoyed by these transforms. Note that in the integral, the term dy/y is dimensionless, so that the L^p mapping properties should be those of Hölder's inequality. Thus, H_α should map $L^2 \times L^2$ into L^1. Note that this is false in the degenerate case of $\alpha = 1$, as the Hilbert transform does not preserve L^1. Nevertheless, this was conjectured by Calderón in the non-degenerate cases.

See [**46**]–[**49**] for a proof of this theorem.

Theorem 9.23. *For $1 < p, q \le \infty$, if $0 < 1/r = 1/p + 1/q < 3/2$, and $\alpha \notin \{0,1,\infty\}$, then*

$$\|\mathrm{H}_\alpha(f,g)\|_r \lesssim \|f\|_p \|g\|_q.$$

We should mention that in a certain sense the proof of this theorem is easier than that for Carleson's operator. The proof outlined in [**42**] contains the notions of tiles, trees, and size. But the estimate that corresponds to the tree lemma is a triviality. The reason for this gain in simplicity is that there is no need for a mechanism to control a supremum.

The subject of multilinear operators with modulation invariance has inspired a large number of results, and is worthy of survey on it alone. We refer the reader to Thiele's article [**82**] for a survey of recent activity in this area.

9.8. Multilinear oscillatory integrals.

Consider the bilinear oscillatory integral

$$\mathrm{B}_2(f_1, f_2)(x) := \int f_1(x-y) f_2(x+y) \frac{\mathrm{e}^{2iy^2}}{y}\, dy.$$

This is seen to be a disguised form of a bilinear Hilbert transform. Setting $g_j(x) := \mathrm{e}^{ix^2} f_j(x)$, one sees that

$$\mathrm{B}_2(g_1, g_2)(x) := \mathrm{e}^{2ix^2} \int f_1(x-y) f_2(x+y) \frac{dy}{y}.$$

(Compare this to Ionescu's example mentioned at the end of Section 9.5.) As it turns out, for a polynomial of any other degree, the integral above is bounded. The proof demonstrates a multilinear variant of the van der Corput type inequality, of which Lemma 9.13 is just one example.

More generally, Christ, Li, Tao and Thiele [**20**] define multilinear functionals

$$\Lambda_\lambda(f_1, f_2, \ldots, f_n) = \int_{\mathbb{R}^m} e^{i\lambda P(x)} \prod_{j=1}^{n} f_j(\pi_j(x))\eta(x)\, dx \tag{9.24}$$

where $\lambda \in \mathbb{R}$ is a parameter, $P\colon \mathbb{R}^m \mapsto \mathbb{R}$ is a real-valued polynomial, $m \geq 2$, and $\eta \in C_0^1(\mathbb{R}^m)$ is compactly supported. Each π_j denotes the orthogonal projection from $\mathbb{R}^m$ to a linear subspace $V_j \subset \mathbb{R}^m$ of any dimension $\kappa \leq m-1$, and $f_j\colon V_j \to \mathbb{C}$ is always assumed to be locally integrable with respect to Lebesgue measure on V_j.

Notice that by taking $n = 3$, and taking projections $\pi_j\colon \mathbb{R}^2 \mapsto V_j$ where

$$\begin{aligned} V_1 &= \{(x,x) : x \in \mathbb{R}\}, \\ V_2 &= \{(x,-x) : x \in \mathbb{R}\}, \\ V_3 &= \{(x,0) : x \in \mathbb{R}\}, \end{aligned} \tag{9.25}$$

we can recover for instance a bilinear Hilbert transform.

And they say that a polynomial P has *a power decay property* if there is a $\delta > 0$, so that for all $f_j \in L^\infty(V_j)$, we have the estimate

$$|\Lambda(f_1, \dots, f_n)| \lesssim (1 + |\lambda|)^{-\delta} \prod_{j=1}^{n} \|f_j\|_\infty.$$

From this estimate, a range of power decay estimates hold in all relevant products of L^p spaces. This should be compared to Lemma 9.13 and in particular (9.14) below.

Clearly, there are obstructions to a power decay property, and this obstruction can be formalized in a definition. A polynomial P is said to be *degenerate (relative to* $\{V_j\}$*)* if there exist polynomials $p_j : V_j \to \mathbb{R}$ such that $P = \sum_{j=1}^{n} p_j \circ \pi_j$. Otherwise P is nondegenerate. In the case $n = 0$, where the collection of subspaces $\{V_j\}$ is empty, P is considered to be nondegenerate if and only if it is nonconstant. And in the example (9.25), we see that $P(y) = 2x^2 + 2y^2 = (x + y)^2 + (x - y)^2$ is degenerate.

It is natural to conjecture that non degeneracy is sufficient for a power decay property. This is verified in a wide range of special cases in the paper by Christ, Li, Tao, and Thiele [**20**], by a range of interesting techniques. It is of interest to determine if the natural conjecture here is indeed correct.

9.9. Hilbert transform on smooth families of lines.

This question has its beginnings in the Besicovitch set, which we already mentioned in connection to spherical summation of Fourier series. One may construct Besicovitch sets with these properties. For choices of $0 < \epsilon, \alpha < 1$, there is a Besicovitch set K in the square $[0, 4]^2$ say, for which K has measure at most ϵ, and there is a function $g : \mathbb{R}^2 \to \mathbb{T}$, so that for a set of x's in $[0, 4]^2$ of measure $\gtrsim 1$, $K \cap \{x + tv(x) : t \in \mathbb{R}\}$ contains a line segment of length one, and v is Hölder continuous of order α.

One can ask if the Hölder continuity condition is sharp. A beautiful formulation of a conjecture in this direction is attributed to Zygmund.

Conjecture 9.26. *Let $v : \mathbb{R}^2 \to \mathbb{T}$ be Hölder continuous (of order 1). Then for all square integrable functions f,*

$$f(x) = \lim_{t \to 0} (2t)^{-1} \int_{-t}^{t} f(x - uv(x))\, du \quad a.e\ (x).$$

This is a differentiablity question, on a choice of lines specified by v. The only stipulation is that v is Hölder continuous. This is only known under more stringent conditions on v, such as analytic due to Stein [**72**], or real analytic due to Bourgain [**10**]. There is a partial result due to Katz [**35**] (also see [**34**]) that demonstrates at worst "log log" blowup assuming the Hölder continuity of v. The question is open, even if one assumes that $v \in C^{1000}$.

The difficulty in this problem arises from those points at which the gradient of v is degenerate; assumptions such as analyticity certainly control such degeneracies.

Stein [**72**] posed the Hilbert transform variant, namely defining

$$H_v f(x) := \text{p.v.} \int_{-1}^{1} f(x - yv(x)) \frac{dy}{y},$$

is it the case that there is a constant c so that if $\|v\|_{\text{Höl}} < c$, then H_v maps $L^2(\mathbb{R}^2)$ into itself. A curious fact about this question is that this inequality, if known, implies Carleson's theorem for one dimensional Fourier series.

To see this, observe that the symbol for the transform is $\psi(\xi \cdot v(x))$, where ψ is the Fourier transform of $y^{-1}\mathbf{1}_{\{|y|<1\}}$. Suppose the vector field is of the form $v(x) = (1, \nu(x_1))$ where we need only assume that ν is Hölder continuous of norm 1 say, and consider the trace of the symbol on the line $\xi_2 = -N$. Then, the symbol is $\psi((\xi_1, N) \cdot (1, \nu(x_1)) = \psi(\xi_1 - N\nu(x_1))$. We conclude that this symbol defines a bounded linear operator on $L^2(\mathbb{R})$, with bound that is independent of N. That is, for *any* Lipschitz function $\nu(x_1)$, and *any* $N > 1$ the symbol $\psi(\xi_1 - N\nu(x_1))$ is the symbol of a bounded linear operator on $L^2(\mathbb{R})$. By varying N and ν, we may replace $N\nu(x_1)$ by an arbitrary measurable function. This is the substance of Carleson's theorem.

But the implication is entirely one way: A positive answer to the family of lines question seems to require techniques quite a bit more sophisticated than those that imply Carleson's theorem. Recently Lacey and Li [**45**] have been able to obtain a partial answer, assuming only that the vector field has $1 + \epsilon$ derivatives.

Theorem 9.27. *Assume that $v \in C^{1+\epsilon}$ for some $\epsilon > 0$. Then the operator H_v is bounded on $L^2(\mathbb{R}^2)$. The norm of the operator is at most*

$$\|H_v\|_2 \lesssim [1 + \log_+ \|v\|_{C^{1+\epsilon}}]^2.$$

9.10. Schrödinger operators, scattering transform.

There is a beautiful line of investigation relating Schrödinger equations in one dimension to aspects of the Fourier transform, and in particular, Carleson's theorem. There is a further connection to scattering transforms and nonlinear Fourier analysis. All in all, these topics are extremely broad, with several different sets of motivations, and a long list of contributors.

We concentrate on a succinct way to see the connection to Carleson's theorem, an observation made explicitly by Christ and Kiselev [**15**], [**16**], also see [**65**]. The basic object is a time independent Schrödinger operator on the real line,

$$H = -\frac{d^2}{dx^2} + V$$

where V is an appropriate potential on the real line. The idea is that if V is small, in some specific senses, then the spectrum of H should resemble that of $-\frac{d^2}{dx^2}$. In particular eigenfunctions should be perturbations of the exponentials.

Standard examples show that one should seek to show that for almost all λ, the eigenfunctions of energy λ, that is the solutions to

$$(H - \lambda^2 I)$$

are bounded perturbations of $e^{\pm i\lambda x}$.

Seeking such an eigenfunction, one can formally write

$$u(x) = e^{i\lambda x} + \frac{1}{i\lambda}\int_x^\infty \sin(\lambda(x-y))V(y)u(y)\,dy.$$

Iterating this formula, again formally, one has

$$u(x) = e^{i\lambda x}$$

$$+\frac{1}{i\lambda}\int_x^\infty \sin(\lambda(x-y))V(y)e^{i\lambda y}\,dy \tag{9.28}$$

$$+\frac{1}{(i\lambda)^2}\iint_{x\le y_1\le y_2} \sin(x-y_1)\sin(y_1-y_2)V(y_1)V(y_2)u(y_2)\,dy_1\,dy_2. \tag{9.29}$$

Observe that (9.28) no longer contains u, and is a linear combination of

$$e^{i\lambda x}\int_x^\infty e^{2i\lambda y}V(y)\,dy \tag{9.30}$$

$$e^{i\lambda x}\int_x^\infty V(y)\,dy. \tag{9.31}$$

One seeks estimates of these in the mixed norm space of say, $L^2_\lambda L^\infty_x$. From such estimates, one deduces that for almost all λ, there is an eigenfunction with is a perturbation of $e^{i\lambda x}$.

Concerning (9.30), notice that if $V \in L^2$, we can, by Plancherel, regard V as $\widehat{f}$, for some $f \in L^2$. The desired estimate is then a consequence of Carleson's theorem. This is indicative of the distinguished role that L^2 plays in this subject. Also of the intertwining of the roles of frequency and time that occur in the subject.

Concerning (9.31), unless $V \in L^1$, there is no reasonable interpretation that can be placed on this term. In practice, a different approach than the one given here must be adopted.

If one continues the expansion in (9.29), one gets a bilinear operator with features that resemble both the Carleson operator, and the bilinear Hilbert transform. See the papers by Muscalu, Tao, and Thiele [**57**], [**59**]–[**61**].

We refer the reader to these papers by Christ and Kiselev [**15**]–[**18**]. For a survey of this subject, see [**19**]. The reader should also consult the ongoing investigations of Muscalu, Tao, and Thiele [**58**]. This paper begins with an interesting summary of the perspective of the nonlinear Fourier transform.

References

[1] N. Yu. Antonov, Convergence of Fourier series, in: *"Proceedings of the XX Workshop on Function Theory"* (Moscow, 1995), *East J. Approx.* **2(2)** (1996), 187–196.

[2] J. Arias de Reyna, *"Pointwise convergence of Fourier series"*, Lecture Notes in Mathematics **1785**, Springer-Verlag, Berlin, 2002.

[3] J. Arias de Reyna, Pointwise convergence of Fourier series, *J. London Math. Soc. (2)* **65(1)** (2002), 139–153.

[4] G. I. Arkhipov and K. I. Oskolkov, A special trigonometric series and its applications, (Russian), *Mat. Sb. (N.S.)* **134(176)**, no. 2 (1987), 147–157, 287; translation in *Math. USSR-Sb.* **62(1)** (1989), 145–155.

[5] I. Assani and K. Petersen, The helical transform as a connection between ergodic theory and harmonic analysis, *Trans. Amer. Math. Soc.* **331(1)** (1992), 131–142.

[6] I. Assani, K. Petersen and H. White, Some connections between ergodic theory and harmonic analysis, in: *"Almost everywhere convergence, II"* (Evanston, IL, 1989), Academic Press, Boston, MA, 1991, pp. 17–40.

[7] P. Auscher and M. J. Carro, On relations between operators on $\mathbf{R}^N$, $\mathbf{T}^N$ and $\mathbf{Z}^N$, *Studia Math.* **101(2)** (1992), 165–182.

[8] J. Barrionuevo and M. T. Lacey, A weak-type orthogonality principle, *Proc. Amer. Math. Soc.* **131(6)** (2003), 1763–1769 (electronic).

[9] P. BILLARD, Sur la convergence presque partout des séries de Fourier-Walsh des fonctions de l'espace $L^2(0,1)$, *Studia Math.* **28** (1966/1967), 363–388.

[10] J. BOURGAIN, A remark on the maximal function associated to an analytic vector field, in: *"Analysis at Urbana, Vol. I"* (Urbana, IL, 1986–1987), London Math. Soc. Lecture Note Ser. **137**, Cambridge Univ. Press, Cambridge, 1989, pp. 111–132.

[11] A.-P. CALDERÓN, Ergodic theory and translation-invariant operators, *Proc. Nat. Acad. Sci. U.S.A.* **59** (1968), 349–353.

[12] J. CAMPBELL AND K. PETERSEN, The spectral measure and Hilbert transform of a measure-preserving transformation, *Trans. Amer. Math. Soc.* **313(1)** (1989), 121–129.

[13] A. CARBERY, A. SEEGER, S. WAINGER AND J. WRIGHT, Classes of singular integral operators along variable lines, *J. Geom. Anal.* **9(4)** (1999), 583–605.

[14] L. CARLESON, On convergence and growth of partial sumas of Fourier series, *Acta Math.* **116** (1966), 135–157.

[15] M. CHRIST AND A. KISELEV, WKB asymptotic behavior of almost all generalized eigenfunctions for one-dimensional Schrödinger operators with slowly decaying potentials, *J. Funct. Anal.* **179(2)** (2001), 426–447.

[16] M. CHRIST AND A. KISELEV, Scattering and wave operators for one-dimensional Schrödinger operators with slowly decaying nonsmooth potentials, *Geom. Funct. Anal.* **12(6)** (2002), 1174–1234.

[17] M. CHRIST AND A. KISELEV, WKB and spectral analysis of one-dimensional Schrödinger operators with slowly varying potentials, *Comm. Math. Phys.* **218(2)** (2001), 245–262.

[18] M. CHRIST AND A. KISELEV, Maximal functions associated to filtrations, *J. Funct. Anal.* **179(2)** (2001), 409–425.

[19] M. CHRIST AND A. KISELEV, One-dimensional Schrödinger operators with slowly decaying potentials: spectra and asymptotics (or: Baby Fourier analysis meets toy quantum mechanics), Lecture Notes, University of Arkansas Spring Lecture Series (2002).

[20] M. CHRIST, X. LI, T. TAO AND C. M. THIELE, On multilinear oscillatory integrals, nonsingular and singular, Arxiv, math.CA/0311039 (2003).

[21] G. COHEN AND M. LIN, Laws of large numbers with rates and the one-sided ergodic Hilbert transform, *Illinois J. Math.* **47(4)** (2003), 997–1031.

[22] R. R. COIFMAN, Remarks on weak type inequalities for operators commuting with translations, *Bull. Amer. Math. Soc.* **74** (1968), 710–714.

[23] A. CÓRDOBA AND R. FEFFERMAN, On the equivalence between the boundedness of certain classes of maximal and multiplier operators in Fourier analysis, *Proc. Nat. Acad. Sci. U.S.A.* **74(2)** (1977), 423–425.

[24] J. CUZICK AND TZE LEUNG LAI, On random Fourier series, *Trans. Amer. Math. Soc.* **261(1)** (1980), 53–80.

[25] I. DAUBECHIES, *"Ten lectures on wavelets"*, CBMS-NSF Regional Conference Series in Applied Mathematics **61**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[26] C. FEFFERMAN, Pointwise convergence of Fourier series, *Ann. of Math. (2)* **98** (1973), 551–571.

[27] C. FEFFERMAN, On the divergence of multiple Fourier series, *Bull. Amer. Math. Soc.* **77** (1971), 191–195.

[28] C. FEFFERMAN, On the convergence of multiple Fourier series, *Bull. Amer. Math. Soc.* **77** (1971), 744–745.

[29] C. FEFFERMAN, The multiplier problem for the ball, *Ann. of Math. (2)* **94** (1971), 330–336.

[30] L. GRAFAKOS, T. TAO AND E. TERWILLEGER, L^p bounds for a maximal dyadic sum operator, *Math. Z.* **246(1–2)** (2004), 321–337.

[31] G. H. HARDY, On the summability of Fourier's series, *J. London Math. Soc.* **12** (1913), 365–372.

[32] R. A. HUNT, On the convergence of Fourier series, in: *"1968 Orthogonal Expansions and their Continuous Analogues"* (Proc. Conf., Edwardsville, Ill., 1967), Southern Illinois Univ. Press, Carbondale, Ill., pp. 235–255.

[33] R. A. HUNT AND WO SANG YOUNG, A weighted norm inequality for Fourier series, *Bull. Amer. Math. Soc.* **80** (1974), 274–277.

[34] N. H. KATZ, Maximal operators over arbitrary sets of directions, *Duke Math. J.* **97(1)** (1999), 67–79.

[35] N. H. KATZ, A partial result on Lipschitz differentiation, in: *"Harmonic analysis at Mount Holyoke"*, (South Hadley, MA, 2001), Contemp. Math. **320**, Amer. Math. Soc., Providence, RI, 2003, pp. 217–224.

[36] A. N. KOLMOGOROV, Une série de Fourier-Lebesque divergente presque partout, *Fund. Math.* **4** (1923), 324–328.

[37] A. N. KOLMOGOROV, Une série de Fourier-Lebesque divergente presque partout, *C. R. Acad. Sci. Paris Sér. I Math.* **183** (1926), 1327–1329.

[38] S. V. KONYAGIN, On the divergence everywhere of trigonometric Fourier series, (Russian), *Mat. Sb.* **191(1)** (2000), 103–126; translation in *Sb. Math.* **191(1–2)** (2000), 97–120.

[39] S. V. KONYAGIN, On divergence of trigonometric Fourier series everywhere, *C. R. Acad. Sci. Paris Sér. I Math.* **329(8)** (1999), 693–697.

[40] T. W. KÖRNER, Everywhere divergent Fourier series, *Colloq. Math.* **45(1)** (1981), 103–118 (1982).

[41] M. T. LACEY, Transferring the Carleson-Hunt theorem in the setting of Orlicz spaces, in: *"Interaction between functional analysis, harmonic analysis, and probability"* (Columbia, MO, 1994), Lecture Notes in Pure and Appl. Math. **175**, Dekker, New York, 1996, pp. 307–314.

[42] M. T. LACEY, On the bilinear Hilbert transform, in: *"Proceedings of the International Congress of Mathematicians, Vol. II"* (Berlin, 1998), *Doc. Math.* **Extra Vol. II** (1998), 647–656 (electronic).

[43] M. T. LACEY, Carleson's theorem with quadratic phase functions, *Studia Math.* **153(3)** (2002), 249–267.

[44] M. T. LACEY, Issues related to Rubio de Francia's Littlewood-Paley Inequailty: A Survey, math.CA/0306417 (2003).

[45] M. T. LACEY AND X. LI, Maximal theorems for directional Hilbert transform on the plane, Preprint.

[46] M. T. LACEY AND C. M. THIELE, L^p estimates for the bilinear Hilbert transform, *Proc. Nat. Acad. Sci. U.S.A.* **94(1)** (1997), 33–35.

[47] M. T. LACEY AND C. M. THIELE, L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$, *Ann. of Math. (2)* **146(3)** (1997), 693–724.

[48] M. T. LACEY AND C. M. THIELE, On Calderón's conjecture for the bilinear Hilbert transform, *Proc. Natl. Acad. Sci. USA* **95(9)** (1998), 4828–4830 (electronic).

[49] M. T. LACEY AND C. M. THIELE, On Calderón's conjecture, *Ann. of Math. (2)* **149(2)** (1999), 475–496.

[50] M. T. LACEY AND C. M. THIELE, A proof of boundedness of the Carleson operator, *Math. Res. Lett.* **7(4)** (2000), 361–370.

[51] E. LESIGNE, Spectre quasi-discret et théorème ergodique de Wiener-Wintner pour les polynômes, *Ergodic Theory Dynam. Systems* **13(4)** (1993), 767–784.

[52] E. LESIGNE, Un théorème de disjonction de systèmes dynamiques et une généralisation du théorème ergodique de Wiener-Wintner, *Ergodic Theory Dynam. Systems* **10(3)** (1990), 513–521.

[53] J. Marcinkiewicz and A. Zygmund, Sur les fonctions indépendentes, *Fund. Math.* **29** (1937), 60–90.

[54] M. B. Marcus and G. Pisier, Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes, *Acta Math.* **152(3–4)** (1984), 245–301.

[55] A. Máté, Convergence of Fourier series of square integrable functions, (Hungarian, with English summary), *Mat. Lapok* **18** (1967), 195–242.

[56] C. Muscalu, T. Tao and C. M. Thiele, Multi-linear operators given by singular multipliers, *J. Amer. Math. Soc.* **15(2)** (2002), 469–496 (electronic).

[57] C. Muscalu, T. Tao and C. M. Thiele, A discrete model for the bi-Carleson operator, *Geom. Funct. Anal.* **12(6)** (2002), 1324–1364.

[58] C. Muscalu, T. Tao and C. M. Thiele, A Carleson theorem for a Cantor group model of the scattering transform, *Nonlinearity* **16(1)** (2003), 219–246.

[59] C. Muscalu, T. Tao and C. M. Thiele, L^p estimates for the "biest" I. The Walsh model, math.CA/0102083.

[60] C. Muscalu, T. Tao and C. M. Thiele, L^p estimates for the "biest" II. The Fourier model, math.CA/0102084.

[61] C. Muscalu, T. Tao and C. M. Thiele, A counterexample to a multilinear endpoint question of Christ and Kiselev, *Math. Res. Lett.* **10(2–3)** (2003), 237–246.

[62] K. I. Oskolkov, Schrödinger equation and oscillatory Hilbert transforms of second degree, *J. Fourier Anal. Appl.* **4(3)** (1998), 341–356.

[63] K. I. Oskolkov, A class of I. M. Vinogradov's series and its applications in harmonic analysis, in: *"Progress in approximation theory"* (Tampa, FL, 1990), Springer Ser. Comput. Math. **19**, Springer, New York, 1992, pp. 353–402.

[64] E. Prestini and P. Sjölin, A Littlewood-Paley inequality for the Carleson operator, *J. Fourier Anal. Appl.* **6(5)** (2000), 457–466.

[65] C. Remling, The absolutely continuous spectrum of one-dimensional Schrödinger operators with decaying potentials, *Comm. Math. Phys.* **193(1)** (1998), 151–170.

[66] J. L. Rubio de Francia, A Littlewood-Paley inequality for arbitrary intervals, *Rev. Mat. Iberoamericana* **1(2)** (1985), 1–14.

[67] P. Sjölin, An inequality of Paley and convergence a.e. of Walsh-Fourier series, *Ark. Mat.* **7** (1969), 551–570 (1969).

[68] P. Sjölin, Convergence almost everywhere of certain singular integrals and multiple Fourier series, *Ark. Mat.* **9** (1971), 65–90.

[69] P. SJÖLIN AND F. SORIA, Remarks on a theorem by N. Yu. Antonov, *Studia Math.* **158(1)** (2003), 79–97.
[70] E. M. STEIN, On limits of seqences of operators, *Ann. of Math. (2)* **74** (1961), 140–170.
[71] E. M. STEIN, *"Singular integrals and differentiability properties of functions"*, Princeton Mathematical Series **30**, Princeton University Press, Princeton, N.J., 1970.
[72] E. M. STEIN, Problems in harmonic analysis related to curvature and oscillatory integrals, in: *"Proceedings of the International Congress of Mathematicians, Vol. 1, 2"* (Berkeley, Calif., 1986), Amer. Math. Soc., Providence, RI, 1987, pp. 196–221.
[73] E. M. STEIN, Oscillatory integrals related to Radon-like transforms, in: *"Proceedings of the Conference in Honor of Jean-Pierre Kahane"* (Orsay, 1993), *J. Fourier Anal. Appl.* **Special Issue** (1995), 535–551.
[74] E. M. STEIN, *"Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals"*, Princeton Mathematical Series **43**, Monographs in Harmonic Analysis III, Princeton University Press, Princeton, NJ, 1993.
[75] E. M. STEIN AND S. WAINGER, The estimation of an integral arising in multiplier transformations, *Studia Math.* **35** (1970), 101–104.
[76] E. M. STEIN AND S. WAINGER, Discrete analogues in harmonic analysis. I. l^2 estimates for singular Radon transforms, *Amer. J. Math.* **121(6)** (1999), 1291–1336.
[77] E. M. STEIN AND S. WAINGER, Discrete analogues in harmonic analysis. II. Fractional integration, *J. Anal. Math.* **80** (2000), 335–355.
[78] E. M. STEIN AND S. WAINGER, Oscillatory integrals related to Carleson's theorem, *Math. Res. Lett.* **8(5–6)** (2001), 789–800.
[79] M. TALAGRAND, A borderline random Fourier series, *Ann. Probab.* **23(2)** (1995), 776–785.
[80] C. M. THIELE, Time-frequency analysis in the discrete phase plane, in: *"Topics in analysis and its applications"*, World Sci. Publishing, River Edge, NJ, 2000, pp. 99–152.
[81] C. M. THIELE, The quartile operator and pointwise convergence of Walsh series, *Trans. Amer. Math. Soc.* **352(12)** (2000), 5745–5766.
[82] C. M. THIELE, Singular integrals meet modulation invariance, in: *"Proceedings of the International Congress of Mathematicians, Vol. II"* (Beijing, 2002), Higher Ed. Press, Beijing, 2002, pp. 721–732.

[83] N. WIENER AND A. WINTNER, On the ergodic dynamics of almost periodic systems, *Amer. J. Math.* **63** (1941), 794–824.

[84] A. ZYGMUND, *"Trigonometric series. Vol. I, II"*, Third edition, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2002.

School of Mathematics
Georgia Institute of Technology
Atlanta GA 30332
USA
E-mail address: lacey@math.gatech.edu
http://www.math.gatech.edu/~lacey

Primera versió rebuda el 30 d'octubre de 2003,
darrera versió rebuda el 18 de juny de 2004.