Uniwersytet Wrocławski

Wydział Matematyki i Informatyki
Instytut Matematyczny

pl. Grunwaldzki 2/4
50-384 Wrocław
tel. +48 71 375 74 01, +48 71 328 07 17
fax +48 71 375 74 29
e-mail: dyrekcja@math.uni.wroc.pl
www.math.uni.wroc.pl

ssdnm

Time – frequency analysis and singular integrals

Vjekoslav Kovač

(University of Zagreb)

(Wrocław, 07 – 10.10.2014)

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Time-frequency analysis and singular integrals

Vjekoslav Kovač (University of Zagreb)

Mathematical Institute, University of Wrocław

a Ph.D. mini-course

October 6 – October 10, 2014

Abstract

Time-frequency analysis attempts to study individual functions and function systems by regarding them as objects that simultaneously "exist" both in the time domain and the frequency domain. One can even go so far as to consider the members of the system as being "morally supported" in certain subsets of the so-called "phase plane". This viewpoint can help get a better understanding of the classical systems such as the wavelet system or the Gabor system, but it also enables us to build a variety of "custom" systems such as wave packets.

Coming from a different background, singular integral operators are objects with subtle cancellation properties, which always have to be taken in consideration when proving any kind of estimates. If one wants to establish their boundedness by decomposing them using convenient function systems, then these systems have to respect the same groups of symmetries as the operator does and thus have to be chosen or built appropriately.

This methodology proved to be useful in general, but it particularly applies to multilinear integral operators, as they often posses more structural symmetries than their linear counterparts. This view at the wavelets was already used by R. R. Coifman and Y. F. Meyer in the study of linear and multilinear Calderón-Zygmund operators, while wave packet decompositions were applied by M. T. Lacey and C. M. Thiele to the more involved bilinear Hilbert transform and to the Carleson operator.

The goal of this course is to formalize and clarify the above connection between combinatorially/geometrically nontrivial time-frequency constructions and singular integral operators, which have become standard objects in harmonic analysis. The emphasis will be given on applying the former to prove L^p estimates for the latter. The course will be more inclined towards the prototypical examples of singular integral operators, both classical and the recent ones, than the most general classes of objects they belong to. Detailed proofs of boundedness will be presented for some of these examples, with remarks on how the same technique can be modified to handle the corresponding classes of operators.

Contents

Chapter 1

The phase plane and function systems

In this chapter we introduce (both heuristically and rigorously) the fundamental notion of the phase plane, also known as the time-frequency plane. Moreover, we give an overview of the most classical systems of functions, emphasizing the phase plane viewpoint. Several comments on how one could build more complicated custom systems are also given. Finally, we present methodologically very useful finite group models of the Fourier analysis.

1.1 The Fourier transform

The *Fourier transform* is one of the most important transformations in analysis. The reader has probably already met this concept in one form or another, but we review it for the completeness.

In the initial definition one takes an integrable function $f\colon \mathbb{R}^n \to \mathbb{C}$ and defines a new function $\mathcal{F}f = \hat{f}\colon \mathbb{R}^n \to \mathbb{C}$ by the formula

$$\hat{f}(\xi) := \int_{\mathbb{R}^n} f(x) e^{-2\pi i x\cdot\xi} dx, \quad \xi \in \mathbb{R}^n. \tag{1}$$

Here $x \cdot \xi$ denotes the standard scalar product of vectors $x, \xi \in \mathbb{R}^n$ and the integration is performed with respect to the Lebesgue measure on $\mathbb{R}^n$. When the dimension n equals 1, it is customary to think of f as a sort of "signal" evolving in time. In that case $\hat{f}(\xi)$ is simply an inner product of f with the pure exponential $e^{2\pi i x\xi}$ having "frequency" ξ and can thus be interpreted as a numerical contribution of that frequency in the composition of the overall signal. The main usefulness of the Fourier transform is in the fact that it switches the viewpoint, emphasizing completely different properties of the function f, while one can still interchange f and $\hat{f}$ easily.

For the purpose of inverting the above operation we also introduce the *inverse Fourier transform* $\mathcal{F}_{-1}f = \check{f}\colon \mathbb{R}^n \to \mathbb{C}$ by an almost identical formula

$$\check{f}(x) := \int_{\mathbb{R}^n} f(\xi)e^{2\pi i x\cdot\xi}d\xi = \hat{f}(-x), \quad x \in \mathbb{R}^n.$$

The well-known *Fourier inversion formula* states that

$$(\hat{f})\check{} = (\check{f})\hat{} = f \text{ a.e.} \tag{2}$$

whenever $f, \hat{f} \in \mathrm{L}^1(\mathbb{R}^n)$, thus justifying the term "inverse". By the *Riemann-Lebesgue lemma* the Fourier transform $f \mapsto \hat{f}$ is a well-defined and bounded linear operator $\mathcal{F}\colon \mathrm{L}^1(\mathbb{R}^n) \to \mathrm{C}_0(\mathbb{R}^n)$, where $\mathrm{C}_0(\mathbb{R}^n)$ denotes the space of continuous functions $g\colon \mathbb{R}^n \to \mathbb{C}$ vanishing at infinity, i.e. $\lim_{|x|\to\infty} |g(x)| = 0$. Yet another classical result enables the L^2-theory of the Fourier transform.

Theorem 1 (The Plancherel theorem)**.** *If* $f \in \mathrm{L}^1(\mathbb{R}^n) \cap \mathrm{L}^2(\mathbb{R}^n)$*, then* $\hat{f}, \check{f} \in \mathrm{L}^2(\mathbb{R}^n)$*. Furthermore,*

$$\mathcal{F}|_{\mathrm{L}^1(\mathbb{R}^n)\cap\mathrm{L}^2(\mathbb{R}^n)}\colon \mathrm{L}^1(\mathbb{R}^n) \cap \mathrm{L}^2(\mathbb{R}^n) \to \mathrm{L}^2(\mathbb{R}^n)$$

can be uniquely extended to a unitary isomorphism from $\mathrm{L}^2(\mathbb{R}^n)$ *to* $\mathrm{L}^2(\mathbb{R}^n)$ *(denoted again by* $\mathcal{F}$*) and the same holds for* $\mathcal{F}_{-1}|_{\mathrm{L}^1(\mathbb{R}^n)\cap\mathrm{L}^2(\mathbb{R}^n)}$*. The inverse operator (which is also the Hermitian adjoint) of*

$$\mathcal{F}\colon \mathrm{L}^2(\mathbb{R}^n) \to \mathrm{L}^2(\mathbb{R}^n)$$

is precisely

$$\mathcal{F}_{-1}\colon \mathrm{L}^2(\mathbb{R}^n) \to \mathrm{L}^2(\mathbb{R}^n).$$

Explicit consequences of Theorem 1 are

$$\langle \hat{f}, \hat{g}\rangle_{\mathrm{L}^2(\mathbb{R}^n)} = \langle f, g\rangle_{\mathrm{L}^2(\mathbb{R}^n)} \quad \text{and} \quad \|\hat{f}\|_{\mathrm{L}^2(\mathbb{R}^n)} = \|f\|_{\mathrm{L}^2(\mathbb{R}^n)}$$

for $f, g \in \mathrm{L}^2(\mathbb{R}^n)$, where $\langle f, g\rangle_{\mathrm{L}^2(\mathbb{R}^n)} := \int_{\mathbb{R}^n} f(x)\overline{g(x)}dx$ denotes the standard inner product. Detailed proofs of the above results can be found in many classical texts on real harmonic analysis, such as [11] and [28].

Using the so-called complex interpolation of L^p spaces it immediately follows that the Fourier transform is a contraction from $\mathrm{L}^p(\mathbb{R}^n)$ to $\mathrm{L}^q(\mathbb{R}^n)$, where $1 < p < 2$ and $q = \frac{p}{p-1}$ is its conjugated exponent. A significantly more difficult result gives the exact operator norm of this transformation.

Theorem 2 (The Babenko-Beckner inequality)**.** *Let p and q be conjugated exponents and $1 \le p \le 2$. For any $f \in \mathrm{L}^p(\mathbb{R}^n)$ one has*

$$\|\hat{f}\|_{\mathrm{L}^q(\mathbb{R}^n)} \le \Big(\frac{p^{1/p}}{q^{1/q}}\Big)^{n/2} \|f\|_{\mathrm{L}^p(\mathbb{R}^n)}.$$

The equality is attained for Gaussian functions.

Particular case of even integers q was first established by K. I. Babenko, while the actual theorem was first shown by W. Beckner [2]. We have stated it more as a curiosity than something we will need in the later discussion. By a rather general principle of Littlewood, one can show that the Fourier transform does not map $\mathrm{L}^p(\mathbb{R}^n)$ into $\mathrm{L}^q(\mathbb{R}^n)$ when $p > 2$.

There are many elegant and useful generalizations of the Fourier transform; only some will be mentioned (rather briefly) in these lectures. One can study functions on a locally compact abelian group $\mathbb{G}$ equipped with its Haar measure. The exponentials are then replaced by the so-called characters of $\mathbb{G}$ and the Fourier transform yields functions defined on the dual group $\hat{\mathbb{G}}$. All of the above results translate to that setting as well, see [13]. This approach unifies the Fourier transform on $\mathbb{R}^n$ with the theory of (multiple) Fourier series on the n-dimensional torus $\mathbb{T}^n$ and the various fast Fourier transforms on finite abelian groups. The study of Fourier analysis on nonabelian groups is significantly different and falls into the realm of representation theory. Another direction for generalizing the Fourier transform is to regard it as an operator on tempered distributions; the basics can be found in [11]. This is particularly useful in the study of linear partial differential equations. On the other hand, nonlinear PDEs require its nonlinear variants, called the scattering transforms.

* * *

Much more useful for us will be the symmetry properties of the Fourier transform, i.e. its relationship to the (mostly geometric) transformations of the ambient space $\mathbb{R}^n$. We begin by defining several basic operators on the Hilbert space $\mathrm{L}^2(\mathbb{R})$:

- the *translation* by $y \in \mathbb{R}^n$, defined as $(\mathrm{T}_y f)(x) := f(x-y)$, $x \in \mathbb{R}^n$,
- the *modulation* by $\eta \in \mathbb{R}^n$, defined as $(\mathrm{M}_\eta f)(x) := e^{2\pi i x\cdot\eta} f(x)$, $x \in \mathbb{R}^n$,
- the *dilation* by $r \in \mathbb{R} \setminus \{0\}$, defined as $(\mathrm{D}_r f)(x) := |r|^{-n/2} f(r^{-1}x)$, $x \in \mathbb{R}^n$.

If one takes an orthogonal transformation S on $\mathbb{R}^n$, then one can consider the *rotation* operator,

$$(\mathrm{R}_S f)(x) := f(S^{-1}x), \quad x \in \mathbb{R}^n.$$

We can also introduce a slightly less standard *quadratic modulation* by $r \in \mathbb{R}$ as

$$(\mathrm{Q}_r f)(x) := e^{\pi i r|x|^2} f(x), \quad x \in \mathbb{R}^n,$$

which will be mentioned a bit later in the particular case $n = 1$. All of the mentioned operators are unitary isometries of $\mathrm{L}^2(\mathbb{R}^n)$ with rather clear inverses.

Proposition 3. *One has*

$$\mathcal{F}\mathrm{T}_y = \mathrm{M}_{-y}\mathcal{F}, \quad \mathcal{F}\mathrm{M}_\eta = \mathrm{T}_\eta\mathcal{F}, \quad \mathcal{F}\mathrm{D}_r = \mathrm{D}_{r^{-1}}\mathcal{F}, \quad \mathcal{F}\mathrm{R}_S = \mathrm{R}_S\mathcal{F},$$

i.e.

$$(\mathrm{T}_y f)\hat{} = \mathrm{M}_{-y}\hat{f}, \quad (\mathrm{M}_\eta f)\hat{} = \mathrm{T}_\eta\hat{f}, \quad (\mathrm{D}_r f)\hat{} = \mathrm{D}_{r^{-1}}\hat{f}, \quad (\mathrm{R}_S f)\hat{} = \mathrm{R}_S\hat{f}$$

for any y, η, r, and S as before.

Proof. By Theorem 1 both sides of each of the four equalities are continuous linear operators on the space $\mathrm{L}^2(\mathbb{R})$, so it is enough to prove these identities on a dense subspace $\mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, on which the defining formula for the Fourier transform (1) applies.

$$(\mathrm{T}_y f)\hat{}(\xi) = \int_{\mathbb{R}^n} f(x-y)e^{-2\pi i x\cdot\xi}dx = \int_{\mathbb{R}^n} f(x)e^{-2\pi i(x+y)\cdot\xi}dx = \hat{f}(\xi)e^{-2\pi i y\cdot\xi} = (\mathrm{M}_{-y}\hat{f})(\xi)$$

$$(\mathrm{M}_\eta f)\hat{}(\xi) = \int_{\mathbb{R}^n} f(x)e^{2\pi i x\cdot\eta}e^{-2\pi i x\cdot\xi}dx = \hat{f}(\xi-\eta) = (\mathrm{T}_\eta\hat{f})(\xi)$$

$$(\mathrm{D}_r f)\hat{}(\xi) = \int_{\mathbb{R}^n} |r|^{-n/2} f(r^{-1}x)e^{-2\pi i x\cdot\xi}dx = \left[\begin{smallmatrix} z = r^{-1}x \\ dx = |r|^n dz \end{smallmatrix}\right] = |r|^{n/2}\hat{f}(a\xi) = (\mathrm{D}_{r^{-1}}\hat{f})(\xi)$$

$$(\mathrm{R}_S f)\hat{}(\xi) = \int_{\mathbb{R}^n} f(S^{-1}x)e^{-2\pi i x\cdot\xi}dx = \begin{bmatrix} z = S^{-1}x \\ dx = dz \\ Sz\cdot\xi = z\cdot S^{-1}\xi \end{bmatrix} = \hat{f}(S^{-1}\xi) = (\mathrm{R}_S\hat{f})(\xi)$$

□

Informally speaking, the Fourier transform interchanges translations and modulations, inverts the scale of dilations, and is invariant under rotations.

1.2 The uncertainty principle

It is fair to remark that the results in this section are not strictly logically needed in the rest of the chapter, but rather motivate all of the following material.

Rather heuristic ideas of W. Heisenberg in quantum mechanics were mathematically formalized by E. H. Kennard, H. Weyl, and H. P. Robertson. That way *the uncertainty principle* becomes a generic name for any result saying that f and $\hat{f}$ cannot be "well-localized" simultaneously, possibly under certain conditions imposed on f. Figuratively speaking, the signal cannot be too well localized both in time and frequency. Several results in this section and some results mentioned in Section 1.4 will provide rigorous formulations of this claim.

For simplicity we confine ourselves to one dimension. Here is a question that comes up naturally after the material presented in the previous section: Is there a nontrivial (meaning not a.e. equal to zero) function $\varphi \in \mathrm{L}^2(\mathbb{R})$ such that both φ and $\hat{\varphi}$ vanish outside a compact set? If such a function φ existed, it could be translated and modulated and it would make an ideal "building block" for generating perfectly localized systems of functions. Unfortunately, the answer to the above question is negative.

Theorem 4 (The qualitative uncertainty principle)**.** *If $\varphi \in \mathrm{L}^2(\mathbb{R})$ is a function such that φ and $\hat{\varphi}$ have compact supports, then $\varphi = 0$ a.e.*

Proof. Since a compactly supported L^2-function is also integrable, we known that $\varphi, \hat{\varphi} \in \mathrm{L}^1(\mathbb{R})$ and we can use the Fourier inversion formula. From (2) we see that f is a.e. equal to a continuous function $(\hat{f})\check{}$, so we can freely assume that f itself is a continuous function.

Define $F\colon \mathbb{C} \to \mathbb{C}$ by

$$F(\zeta) := \int_a^b \varphi(x) e^{-2\pi i x \zeta} dx,$$

where $[a, b]$ is any interval containing the support of φ. Clearly, F is a continuous function. In order to show that F is holomorphic we take an arbitrary closed piecewise C^1 contour C in the complex plane. By Cauchy's integral theorem and the fact that the exponential function is holomorphic we have $\oint_C e^{-2\pi i t \zeta} d\zeta = 0$. Interchanging integrals we get

$$\oint_C F(\zeta) d\zeta = \int_a^b \varphi(x) \Big(\oint_C e^{-2\pi i x \zeta} d\zeta \Big) dx = 0.$$

By Morera's theorem F must be an entire function.

Recall that F vanishes on a real half-line $[b,+\infty)$, so by the uniqueness theorem for holomorphic functions F must be equal to 0 on the whole $\mathbb{C}$. Since its restriction to the real line is precisely $\hat{\varphi}$, we have just showed $\hat{\varphi}=0$. Invoking the Fourier inversion formula (2) once again we conclude $\varphi=0$. □

The next reasonable thing to try is to investigate a quantitative limit to which φ and $\hat{\varphi}$ can be simultaneously localized.

Theorem 5 (The quantitative (Kennard's) uncertainty principle)**.** *For a function $\varphi\in \mathrm{L}^2(\mathbb{R})$ such that $\|\varphi\|_{\mathrm{L}^2}=1$ and for any $x_0,\xi_0\in\mathbb{R}$ one has*

$$\Big(\int_{\mathbb{R}}(x-x_0)^2|\varphi(x)|^2dx\Big)\Big(\int_{\mathbb{R}}(\xi-\xi_0)^2|\hat{\varphi}(\xi)|^2d\xi\Big)\geq\frac{1}{16\pi^2}.$$

We allow the left hand side to be infinite, in which case the inequality is trivial.

Proof. It is enough to prove the inequality for Schwartz functions $\varphi\in\mathcal{S}(\mathbb{R})$, because the general case then follows by standard approximation arguments. Consider another function

$$\psi:=\mathrm{M}_{-\xi_0}\mathrm{T}_{-x_0}\varphi,\quad\text{i.e.}\quad\psi(x)=e^{-2\pi i x\xi_0}\varphi(x+x_0).$$

Proposition 3 gives

$$\hat{\psi}=\mathrm{T}_{-\xi_0}\mathrm{M}_{x_0}\hat{\varphi},\quad\text{i.e.}\quad\hat{\psi}(\xi)=e^{2\pi i x_0(\xi+\xi_0)}\hat{\varphi}(\xi+\xi_0).$$

Substituting $z=x-x_0$ and $\zeta=\xi-\xi_0$ the desired inequality becomes

$$\Big(\int_{\mathbb{R}}z^2|\psi(z)|^2dz\Big)\Big(\int_{\mathbb{R}}\zeta^2|\hat{\psi}(\zeta)|^2d\zeta\Big)\geq\frac{1}{16\pi^2}.$$

Introduce the two operators on Schwartz functions:

- the *position operator* X, defined as $(\mathrm{X}f)(x):=xf(x)$,
- the *momentum operator* D, defined as $(\mathrm{D}f)(x):=\frac{1}{2\pi i}f'(x)$.

They are interchanged by the Fourier transform:

$$(\mathrm{D}f)\hat{}=\mathrm{X}\hat{f},\quad\text{i.e.}\quad(\mathrm{D}f)\hat{}(\xi)=\xi\hat{f}(\xi),$$

because integration by parts gives

$$\begin{aligned}(\mathrm{D}f)\hat{\ }(\xi) &= \int_{\mathbb{R}} \frac{1}{2\pi i} f'(x) e^{-2\pi i x\xi} dx = -\int_{\mathbb{R}} \frac{1}{2\pi i} f(x) \Big(\frac{d}{dx} e^{-2\pi i x\xi}\Big) dx \\ &= \xi \int_{\mathbb{R}} f(x) e^{-2\pi i x\xi} dx = \xi \hat{f}(\xi).\end{aligned}$$

Observe that by the Plancherel identity we actually need to show

$$\|\mathrm{X}\psi\|_{\mathrm{L}^2(\mathbb{R})} \|\mathrm{D}\psi\|_{\mathrm{L}^2(\mathbb{R})} \geq \frac{1}{4\pi}. \tag{3}$$

Let us also observe that

$$\mathrm{DX} - \mathrm{XD} = \frac{1}{2\pi i}\mathrm{I},$$

which is just a concise way of writing

$$2\pi i(\mathrm{DX} - \mathrm{XD})f(x) = \frac{d}{dx}(xf(x)) - xf'(x) = f(x).$$

Moreover, X and D are clearly self-adjoint, in the sense that

$$\langle \mathrm{X}f, g\rangle_{\mathrm{L}^2(\mathbb{R})} = \langle f, \mathrm{X}g\rangle_{\mathrm{L}^2(\mathbb{R})} \quad \text{and} \quad \langle \mathrm{D}f, g\rangle_{\mathrm{L}^2(\mathbb{R})} = \langle f, \mathrm{D}g\rangle_{\mathrm{L}^2(\mathbb{R})}$$

for any $f, g \in \mathcal{S}(\mathbb{R})$. Let us expand out a nonnegative quadratic quantity

$$\|(\alpha\mathrm{X} + i\mathrm{D})\psi\|^2_{\mathrm{L}^2(\mathbb{R})}$$

for any $\alpha \in \mathbb{R}$.

$$\begin{aligned}\|(\alpha\mathrm{X} + i\mathrm{D})\psi\|^2_{\mathrm{L}^2(\mathbb{R})} &= \alpha^2 \langle \mathrm{X}\psi, \mathrm{X}\psi\rangle_{\mathrm{L}^2(\mathbb{R})} + \alpha i \langle \psi, (\mathrm{DX} - \mathrm{XD})\psi\rangle_{\mathrm{L}^2(\mathbb{R})} + \langle \mathrm{D}\psi, \mathrm{D}\psi\rangle_{\mathrm{L}^2(\mathbb{R})} \\ &= \alpha^2 \|\mathrm{X}\psi\|^2_{\mathrm{L}^2(\mathbb{R})} - \frac{\alpha}{2\pi}\|\psi\|^2_{\mathrm{L}^2(\mathbb{R})} + \|\mathrm{D}\psi\|^2_{\mathrm{L}^2(\mathbb{R})}.\end{aligned}$$

Since this quadratic function in α is nonnegative, we know that its discriminant has to be nonpositive, so finally

$$\tfrac{1}{4\pi^2}\|\mathrm{D}\psi\|^4_{\mathrm{L}^2(\mathbb{R})} - 4\|\mathrm{X}\psi\|^2_{\mathrm{L}^2(\mathbb{R})}\|\mathrm{D}\psi\|^2_{\mathrm{L}^2(\mathbb{R})} \leq 0,$$

which easily transforms into (3). □

Observe that condition $\|\varphi\|_{L^2} = 1$ can be restated as

$$\int_{\mathbb{R}} |\varphi(x)|^2 dx = \int_{\mathbb{R}} |\hat{\varphi}(\xi)|^2 d\xi = 1,$$

so $|\varphi|^2$ and $|\hat{\varphi}|^2$ are densities of two absolutely continuous probability distributions. Theorem 5 gives the strongest inequality when

$$x_0 = \int_{\mathbb{R}} x\, |\varphi(x)|^2\, dx, \quad \xi_0 = \int_{\mathbb{R}} \xi\, |\hat{\varphi}(\xi)|^2\, d\xi,$$

i.e. x_0 and ξ_0 are the expectations of these distributions, while the theorem itself gives the lower bound for the product of their variances:

$$\mathrm{Var}(|\varphi|^2)\mathrm{Var}(|\hat{\varphi}|^2) \geq \frac{1}{16\pi^2}.$$

Exercise 1. Prove that the inequality in Theorem 5 becomes an equality if and only if φ is a modulated Gaussian function of the form

$$\varphi(x) = c\, \sqrt[4]{2r}\, e^{-\pi r (x-x_0)^2} e^{2\pi i x \xi_0}$$

for some parameters $r > 0$ and $c \in \mathbb{C}$, $|c| = 1$.
Hint: For such an extremizer φ the function ψ from the above proof must satisfy an ordinary differential equation $\alpha x \psi(x) + \frac{1}{2\pi}\psi'(x) = 0$ for some $\alpha \in \mathbb{R}$.

Exercise 2. Prove the *operator* (Robertson's) version of *uncertainty principle*: If A and B are (possibly unbounded) Hermitian operators on a Hilbert space, then for any $\alpha, \beta \in \mathbb{R}$ and for any $v \in \mathrm{Domain}(AB) \cap \mathrm{Domain}(BA)$ one has

$$\|(A - \alpha I)v\|\, \|(B - \beta I)v\| \;\geq\; \tfrac{1}{2} \left| \big\langle (AB - BA)v, v \big\rangle \right|.$$

Observe that this claim implies Theorem 5.

Exercise 3. Prove the *entropic* (Hirschman's) version of *uncertainty principle*:

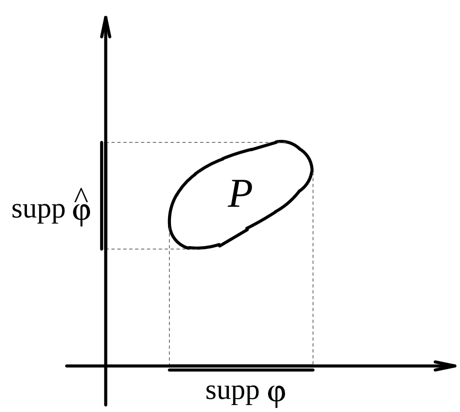
$$-\int_{\mathbb{R}} |f(x)|^2 \log |f(x)|^2 dx - \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \log |\hat{f}(\xi)|^2 d\xi \;\geq\; \log \frac{e}{2}.$$

The two terms on the left hand side are entropies of the probability distributions with densities $|\varphi|^2$ and $|\hat{\varphi}|^2$ respectively. It is possible to show that this result also implies Theorem 5.
Hint: Use Theorem 2 appropriately.

1.3 The phase plane portrait

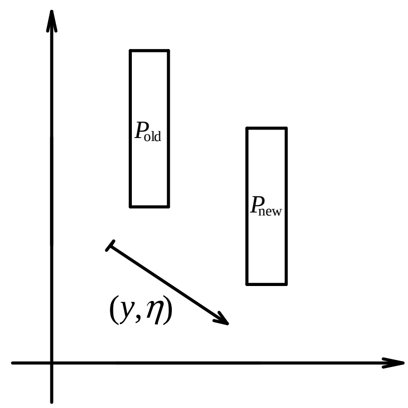
We will be working in the Cartesian product $\mathbb{R}^n \times \mathbb{R}^n$ and call it the *phase space*. The first n coordinates are thought of as the "time variables", while the last n coordinates can be named the "frequency variables". Alternative terminology coming from physics could be "position/momentum variables". We want to view the function $f \in \mathrm{L}^2(\mathbb{R}^n)$ as an object that "lives" in the phase space and reveals properties of both f and $\hat{f}$. A reader who knows about the Heisenberg groups might remark that the ambient space should rather be $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$, with the appropriate group structure. We intentionally neglect this subtle distinction since we will be more interested in the combinatorial/geometric than the algebraic structure of the phase plane. Throughout the course we will often be working in the particular case $n = 1$, when the ambient $\mathbb{R} \times \mathbb{R}$ becomes the *phase plane* or the *time-frequency plane* and the concepts can be graphically visualized.

Let us imagine for a moment that there exists a bounded set $P \subset \mathbb{R} \times \mathbb{R}$ in the time-frequency plane, where a carefully chosen function $\varphi \in \mathrm{L}^2(\mathbb{R})$ is concentrated. Even though we do not yet have a clear idea of what that set should be, let us call it the *phase plane support* of φ. Ideally, when we project P down to the horizontal axis, we should obtain the support of φ, while the projection of P to the vertical axis should be the support of $\hat{\varphi}$, as is depicted in the figure below.

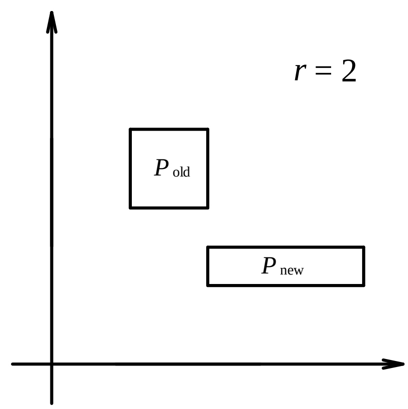


However, Theorem 4 prohibits the existence of any nonzero φ with such property, so let us rather proceed with a heuristics that φ and $\hat{\varphi}$ are only "roughly supported" (whatever that means) on the orthogonal projections of P to the axes.

If such a phase plane support exists, then it will nicely illustrate the actions of the operators T_y, M_η, and D_r on φ. Using Proposition 3 we see that the time-frequency support of $\mathrm{T}_y\varphi$ is simply obtained by shifting P to the right by length y, while the time-frequency support of $\mathrm{M}_\eta\varphi$ is simply obtained by shifting P upward by η. Thus, the composition $\mathrm{S}_{(y,\eta)} := \mathrm{M}_\eta \mathrm{T}_y$ deserves to be call the *time-frequency shift* by the vector (y, η).



On the other hand, the time-frequency support of $\mathrm{D}_r\varphi$ for some $r > 0$ is obtained by scaling P horizontally by factor r and vertically by factor r^{-1}, i.e. by applying the geometric transformation $(x, \xi) \mapsto (rx, r^{-1}\xi)$.



Can we make the concept of the phase plane support rigorous? The answer is affirmative and there are several possible ways of doing that.

* * *

The metaplectic representation. One can give up the notion of the actual phase space support as a set and rather concentrate on a class of geometric transformations on $\mathbb{R}^n$, which then correspond to unitary operators on the Hilbert space

$L^2(\mathbb{R}^n)$. This leads to the Segal-Shale-Weil representation of the metaplectic group. We can explain this construction in the simplest case $n = 1$.

Take an area and orientation preserving affine transformation $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, i.e. $A(x,\xi) = L(x,\xi) + (y,\eta)$, for some $L \in \mathrm{SL}(2,\mathbb{R})$, $y,\eta \in \mathbb{R}$. It is possible to define a unitary operator U_L on $L^2(\mathbb{R})$ satisfying the identity

$$U_L \mathrm{S}_{(a,b)} = c_{L,a,b} \mathrm{S}_{L(a,b)} U_L,$$

with some unimodular constant $c_{L,a,b} \in \mathbb{C}$, for any phase plane point $(a,b) \in \mathbb{R}^2$. In words, we do not immediately see the effect of the operator to the hypothetical time-frequency support of φ, but we rather note how it intertwines time-frequency shifts. This formulation is borrowed from [14]. Afterwards, it is natural to set

$$U_A := \mathrm{S}_{(y,\eta)} U_L.$$

The construction itself is not particularly complicated, as it is enough to define U_L for several transformations L that generate the whole group $\mathrm{SL}(2,\mathbb{R})$. We do this in the following table.

L	U_L
horizontal-vertical scaling $L(a,b) = (ra, r^{-1}b)$	dilation D_r
shear $L(a,b) = (a, b + ra)$	quadratic modulation Q_r
clockwise rotation by $\pi/2$ $L(a,b) = (b,-a)$	the Fourier transform $\mathcal{F}$

Let us prove that the table is indeed correct.

Proof.

$$(\mathrm{D}_r \mathrm{M}_b \mathrm{T}_a f)(x) = \tfrac{1}{\sqrt{r}} e^{2\pi i b x/r} f(\tfrac{x}{r} - a) = (\mathrm{M}_{b/r} \mathrm{T}_{ra} \mathrm{D}_r f)(x)$$

$$(\mathrm{Q}_r \mathrm{M}_b \mathrm{T}_a f)(x) = e^{\pi i (r x^2 + 2bx)} f(x-a) = e^{\pi i (r(x-a)^2 + 2(b+ra)x - ra^2)} f(x-a)$$

$$= e^{-\pi i r a^2} (\mathrm{M}_{b+ra} \mathrm{T}_a \mathrm{Q}_r f)(x)$$

$$\mathcal{F} \mathrm{M}_b \mathrm{T}_a = \mathrm{T}_b \mathcal{F} \mathrm{T}_a = \mathrm{T}_b \mathrm{M}_{-a} \mathcal{F} = e^{2\pi i a b} \mathrm{M}_{-a} \mathrm{T}_b \mathcal{F}$$

For the last equality we used Proposition 3. □

A very easy exercise is to show that shears and the rotation by $\frac{\pi}{2}$ already generate $\mathrm{SL}(2,\mathbb{R})$. Things are slightly different in higher dimensions as there one has to consider transformations that preserve the symplectic form rather than the volume.

Let us summarize that in this approach the time-frequency support is a relative notion rather than an absolute one. It does not make sense for a single function or for a family of unrelated functions. It only makes sense for a system of functions generated from a fixed function by applying some carefully chosen groups of unitary operators, such as translations, modulations, dilations, etc.

* * *

The Wigner transform. Another approach is not to consider the "phase plane portrait" as a set, but rather as a function on the time-frequency plane. For any Schwartz function $f \in \mathcal{S}(\mathbb{R})$ we define the *Wigner transform* of f as a two-dimensional function

$$(Vf)(x,\xi) := \int_{\mathbb{R}} e^{2\pi i t\xi} f(t+\tfrac{x}{2})\overline{f(t-\tfrac{x}{2})}dt.$$

Note that Vf it does not have to be nonnegative, unless we assume something on f. It is easy to show that the Wigner transform of the standard L^2 normalized Gaussian function $\varphi(x) = 2^{1/4}e^{-\pi x^2}$ is a two-dimensional Gaussian

$$(V\varphi)(x,\xi) = e^{-\frac{\pi}{2}(x^2+\xi^2)}.$$

One could easily compute Wigner transforms of translated and modulated Gaussians and observe that these are again 2D Gaussian functions, up to unimportant unimodular constants. However, the Gaussians rarely lead to satisfactory function systems for decompositions of operators in harmonic analysis. The following formula holds in the full generality.

Proposition 6. *For any $f,g \in \mathcal{S}(\mathbb{R})$ we have $\langle Vf, Vg\rangle_{\mathrm{L}^2(\mathbb{R}^2)} = |\langle f,g\rangle|^2_{\mathrm{L}^2(\mathbb{R})}$.*

Proof. Observe that $\xi \mapsto (Vf)(x,\xi)$ is precisely the inverse Fourier transform of the function $t \mapsto f(t+\frac{x}{2})\overline{f(t-\frac{x}{2})}$ for each fixed x. Applying the Plancherel formula gives

$$\begin{aligned}\langle Vf, Vg\rangle_{\mathrm{L}^2(\mathbb{R}^2)} &= \int_{\mathbb{R}} \langle (Vf)(x,\cdot), (Vg)(x,\cdot)\rangle_{\mathrm{L}^2(\mathbb{R})} dx \\ &= \int_{\mathbb{R}^2} f(t+\tfrac{x}{2})\overline{f(t-\tfrac{x}{2})g(t+\tfrac{x}{2})}g(t-\tfrac{x}{2})dtdx \\ &\quad \big[u = t+\tfrac{x}{2},\ u = t-\tfrac{x}{2},\ dudv = dtdx\big] \\ &= \int_{\mathbb{R}^2} f(u)\overline{f(v)g(u)}g(v)dudv = |\langle f, g\rangle|^2_{\mathrm{L}^2(\mathbb{R})}. \qquad \square\end{aligned}$$

Let us conclude that one does not need to have the phase plane portrait of φ localized to a certain set. It can stretch over the phase plane by only having most of the "mass" concentrated in a certain region, like the Gaussians do.

We refer the reader to the classical book [12] for more details on the previous two approaches.

* * *

A conventional compromise. The simplest way out is to index the function system we are interested in by certain subsets of $\mathbb{R} \times \mathbb{R}$, according to the very same heuristics as before, but postpone any justifications to the actual application, for example when we decompose a given operator using that function system. In most cases these sets are rectangles (possibly higher-dimensional) and we call them *tiles*. We keep the intuition that the sides of the rectangle are "morally" the time and frequency supports of the corresponding members of the system, but do not strive for such exact algebraic identities as earlier. When we really need to quantify what we have observed heuristically, we usually begin by showing an appropriate "almost orthogonality" statement for a family of functions corresponding to disjoint collection of tiles. This is how it is usually done in the applications of time-frequency analysis — when we get our hands dirty, stop contemplating and start proving estimates. A working example can be any research paper that uses wave-packet analysis; also see the introductory book [31].

One thing that was missing in the older literature was the combinatorial and geometric structure of function systems coming from the geometric relationship of their phase space supports, which is easier to discuss when we think of the phase plane supports as simple sets. Time-frequency analysis blossomed when an order was introduced to the set of tiles, depending on their mutual position. The idea goes back to the work of C. Fefferman [10] on the Carleson operator and is developed in a series of groundbreaking papers by M. Lacey and C. Thiele

[20],[21],[22]. The next chapter will give the reader a better understanding of this approach.

A nice treatment of phase plane analysis (somewhat influential for these notes) can also be found in [29].

* * *

Simplified models. The fourth alternative is to consider some "toy model" of the Fourier analysis in which the qualitative uncertainty principle fails, so it does not prevent us from having a "perfect" time-frequency localization. An example of such model will be presented in Section 1.5. One still has to confine themselves to defining the time-frequency support only for a very special system of functions.

Due to serious time limitations of this course we will not be able to present complete lengthy proofs of some of the famous results obtained by time-frequency analysis, such as the boundedness of the bilinear Hilbert transform. Thus, explaining the proof in an appropriate "toy model" turns out to be convenient. We will switch to an "alternative playground" at the key moment of the proof. That methodological trick was largely inaugurated by C. Thiele [30] and advocated by him and his collaborators. The reader should not consider this as a sort of cheating, but rather as simplifications in the exposition, because the actual proofs can be really technical.

Exercise 4.

(a) Compute $\mathcal{F}^2\varphi$, $\mathcal{F}^3\varphi$, and $\mathcal{F}^4\varphi$ in terms of a function φ. Relate the obtained result with the above interpretation of the Fourier transform as a phase plane rotation by $\pi/2$.

(b) If $A_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a phase plane rotation by an angle $0 < \theta < \pi/2$, try to find an explicit formula for the corresponding unitary operator U_{A_θ}, at least on a dense subspace of $\mathrm{L}^2(\mathbb{R})$. This operator is called the *fractional Fourier transform* and is denoted by $\mathcal{F}^\theta$.

1.4 Systems of functions

Three types of function systems, Gabor systems, wavelet systems, and wave packets, have become quite standard and have been applied many times over the years. More complicated ones, such as curvelets, ridgelets, edgelets, eyelets, composite dilation wavelets, chirplets, and polynomial phase wave packets have been introduced relatively recently, motivated mostly by applied math problems. Some of them also have nice applications in pure math problems; the papers [19], [23], and [24] are particularly enlightening examples. However, we will concentrate on the three standard systems in these lectures.

* * *

Gabor systems. Any $g \in \mathrm{L}^2(\mathbb{R})$ can generate a Gabor system $(g_{l,n})_{l,n\in\mathbb{Z}}$, which is of the form

$$g_{l,n}(x) := (\mathrm{M}_n \mathrm{T}_l g)(x) = e^{2\pi i n x} g(x-l),$$

i.e. it consists of integer time-frequency translates of g. Note that the operators M_n and T_l commute when $l, n \in \mathbb{Z}$ because $e^{2\pi i l n} = 1$. Thus, also

$$g_{l,n} = \mathrm{T}_l \mathrm{M}_n g.$$

It would be great to find a "nice" function g such that $(g_{l,n})_{l,n\in\mathbb{Z}}$ forms an orthonormal basis. The simplest example is $g = \mathbf{1}_{[0,1)}$, coming from the fact that the integer frequency exponentials form an orthonormal basis for $\mathrm{L}^2(\mathbb{T})$, where $\mathbb{T}$ is the torus $\mathbb{R}/\mathbb{Z} \equiv [0,1)$. However, $\hat{\mathbf{1}}_{[0,1)}$ is not a particularly well-localized function, in the sense that it is not even absolutely integrable. There is an addition to the quantitative uncertainty principle, called the *Balian-Low theorem*, which states that if g is a nonzero function such that $(g_{l,n})_{l,n\in\mathbb{Z}}$ forms an orthonormal basis for $\mathrm{L}^2(\mathbb{R})$, then

$$\int_{\mathbb{R}} x^2 |g(x)|^2 dx = +\infty \quad \text{or} \quad \int_{\mathbb{R}} \xi^2 |g(\xi)|^2 d\xi = +\infty.$$

Therefore, in some sense, we do not have a natural choice for g. (Note that the Gaussians are not even orthogonal, which rules them out immediately). Usually we give up the exact orthogonality requirement and just take a Schwartz function which is compactly supported in frequency.

Strange things are possible when l and n are not integers. For instance it is still an open problem if for any nonzero $g \in \mathrm{L}^2(\mathbb{R})$ or even just for any nonzero Schwartz function $g \in \mathcal{S}(\mathbb{R})$ and any finite set of points (a_i, b_i), $i = 1, 2, \ldots, N$ the system

$$\mathrm{M}_{b_i} \mathrm{T}_{a_i} g, \quad i = 1, 2, \ldots, N$$

has to be linearly independent. This formulation can be found in [14]; many partial results have been shown in the meantime.

* * *

Wavelet systems. A wavelet system is obtained by taking $\psi \in \mathrm{L}^2(\mathbb{R})$ and generating $(\psi_{j,l})_{j,l\in\mathbb{Z}}$, defined by

$$\psi_{j,l}(x) := (\mathrm{D}_{2^j}\mathrm{T}_l\psi)(x) = 2^{-j/2}\psi(2^{-j}x - l).$$

When $(\psi_{j,l})_{j,l\in\mathbb{Z}}$ forms an orthonormal basis for $\mathrm{L}^2(\mathbb{R})$ we say that ψ is an orthonormal wavelet. The simplest one is $\psi = \mathbf{1}_{[0,\frac{1}{2})} - \mathbf{1}_{[\frac{1}{2},1)}$, called the (dyadic) *Haar wavelet.* Note that $\psi_{j,l}$ is then supported on the interval $[2^j l, 2^j(l+1))$. In this case it is convenient to index the system by *dyadic intervals*

$$\mathcal{D} := \{[2^j l, 2^j(l+1)) : j, l \in \mathbb{Z}\},$$

rather than by pairs (j,l), so that it becomes $(\mathbf{h}_I)_{I\in\mathcal{D}}$, $\mathbf{h}_I = |I|^{-1/2}(\mathbf{1}_{I_{\mathrm{left}}} - \mathbf{1}_{I_{\mathrm{right}}})$.

A fundamental and nontrivial result of I. Daubechies (see [7]) is that for arbitrarily large positive integers k there exist compactly supported orthonormal wavelets of class C^k. A sort of uncertainty principle for wavelets is that there does not exist a compactly supported C^∞ orthonormal wavelet.

Good introductory texts on wavelets are [15] and [25].

* * *

Wave packets. Wave packet systems are generated from a single function φ by using all of the three groups: translations, modulations, and dilations. For instance, one such system is $(\varphi_{j,l,n})_{j,l,n\in\mathbb{Z}}$, where

$$\varphi_{j,l,n} := \mathrm{D}_{2^j}\mathrm{T}_l\mathrm{M}_n\varphi.$$

There are certainly too many functions in the system in order to form an orthonormal basis, so in the actual application one has to organize them into orthogonal (or almost orthogonal) parts. We continue discussing wave packets in the next section.

1.5 Dyadic models

Let us suppose that we want to define the phase plane support in the most literal way, at least in some idealized model, where the uncertainty principle fails. Somewhat surprisingly, we can achieve this if we are willing to abandon the usual group structure on $\mathbb{R}$ for a simpler one. This will lead to the replacement of the complex exponentials by the so-called Walsh functions. The idea is to build a model of Fourier analysis in which everything we desire holds literally and rigorously and in which we can gain intuition for a given problem and test our techniques before attempting to solve it in the classical (i.e. Euclidean) setting. One also has to be careful and choose the correct analogue of the problem, as incorrect interpretations can sometimes lead to unwanted oversimplifications.

The model we are about to introduce is a special case of the Fourier analysis on locally compact abelian groups, mentioned in Section 1.1, but there is no need to develop the whole general (and rather technical) theory for this particular purpose. For simplicity we will only present the dyadic case here, while trivial modifications are possible simply by replacing $\mathbb{Z}_2$ by $\mathbb{Z}_d$ for some integer $d \geq 2$. The latter are sometimes called the *Cantor group models*.

* * *

Recall that the torus $\mathbb{T} \equiv \mathbb{R}/\mathbb{Z}$ is a compact group and the measure on it is the Lebesgue measure coming from $\mathbb{T} \equiv [0,1)$. Denote $\mathbb{Z}_2 := \mathbb{Z}/2\mathbb{Z} \equiv \{0,1\}$ and consider the set

$$\mathbb{Z}_2^{\mathbb{N}} := \{(a_j)_{j\in\mathbb{N}} \,:\, (\forall j \in \mathbb{N})(a_j \in \mathbb{Z}_2)\}$$

with respect to the coordinate-wise addition. Define

$$\begin{aligned} &\Phi\colon \mathbb{Z}_2^{\mathbb{N}} \to [0,1], \quad && (a_j)_{j\in\mathbb{N}} \mapsto 0.a_1a_2a_3\ldots = \textstyle\sum_{j=1}^{\infty} a_j 2^{-j}, \\ &\Psi\colon [0,1] \to \mathbb{Z}_2^{\mathbb{N}}, \quad && t \mapsto (\lfloor 2^j t\rfloor \bmod 2)_{j\in\mathbb{N}} = (j\text{-th binary digit of } t)_{j\in\mathbb{N}}. \end{aligned}$$

Functions Φ and Ψ are Borel-measurable and $\Psi\circ\Phi = id_{\mathbb{Z}_2^{\mathbb{N}}}$ a.e., $\Phi\circ\Psi = id_{[0,1]}$ a.e. Besides that, the translation invariant measure on $\mathbb{Z}_2^{\mathbb{N}}$ is the image (the pushforward measure) of the Lebesgue measure λ on $[0,1]$ with respect to the function Ψ and the other way around by the function Φ. Hence, we can identify the probability measure spaces:

$$(\mathbb{Z}_2^{\mathbb{N}}, \mathcal{B}(\mathbb{Z}_2^{\mathbb{N}}), \lambda_{\mathbb{Z}_2^{\mathbb{N}}}) \equiv (\mathbb{T}, \mathcal{B}([0,1)), \lambda).$$

The only difference between these groups is in the binary operation, which in the former case is the binary addition "mod 2" without carrying over digits.

The *Walsh functions* [33] on $\mathbb{Z}_2^{\mathbb{N}}$ are the functions $(W_n)_{n\in\mathbb{N}_0}$ defined by

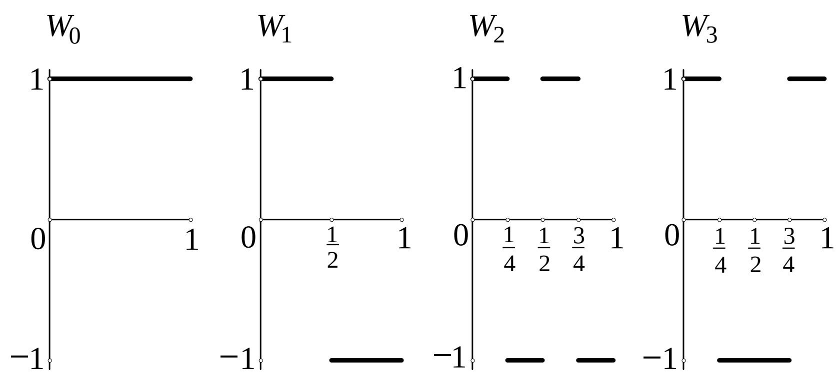
$$W_n((a_j)_{j\in\mathbb{N}}) := (-1)^{\sum_{j=1}^{k} b_j a_j},$$

where $n = \overline{b_k b_{k-1}\ldots b_2 b_1}$ is the binary representation of n. On the nonnegative integers $\mathbb{N}_0$ the natural operation $\oplus$ is the binary addition "mod 2" without carrying over digits once again, and it turns them into a group. Furthermore, $\{W_n : n \in \mathbb{N}_0\}$ is an orthonormal basis of the Hilbert space $\mathrm{L}^2(\mathbb{Z}_2^{\mathbb{N}}) \equiv \mathrm{L}^2([0,1))$.

Via the identification $\mathbb{Z}_2^{\mathbb{N}} \equiv [0,1)$ we can consider the Walsh functions on $[0,1)$ and then they satisfy the recurrence relations:

$$\begin{aligned} W_0(t) &= 1, \quad t \in [0,1), \\ W_{2n}(t) &= \begin{cases} W_n(2t), & t \in [0,\frac{1}{2}), \\ W_n(2t-1), & t \in [\frac{1}{2},1), \end{cases} \\ W_{2n+1}(t) &= \begin{cases} W_n(2t), & t \in [0,\frac{1}{2}), \\ -W_n(2t-1), & t \in [\frac{1}{2},1). \end{cases} \end{aligned}$$

Here are the graphs of the first several functions W_n.



In exactly the same way we introduce the dyadic analogue of the group $\mathbb{R}$. Actually, it will be more natural to define the group operation on $\mathbb{R}_+ = [0,+\infty)$. Consider the set

$$\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}} := \Big\{ (a_j)_{j\in\mathbb{Z}} \ : \ (\forall j \in \mathbb{Z})(a_j \in \mathbb{Z}_2), \quad (\exists j_0 \in \mathbb{Z})(\forall j \in \mathbb{Z})(j < j_0 \Rightarrow a_j = 0) \Big\}$$

of all double sides sequences of zeros and ones, such that from some place to the left they have only zeros. On $\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}$ we define the addition and multiplication as:

$$\begin{aligned} (a_j)_{j\in\mathbb{Z}} \oplus (b_j)_{j\in\mathbb{Z}} &= (a_j + b_j \bmod 2)_{j\in\mathbb{Z}} \\ (a_j)_{j\in\mathbb{Z}} \otimes (b_j)_{j\in\mathbb{Z}} &= (\sum_{k\in\mathbb{Z}} a_{j-k} b_k \bmod 2)_{j\in\mathbb{Z}} \end{aligned}$$

and then $(\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}, \oplus, \otimes)$ even becomes a field of characteristic 2.

This time define

$$\Phi\colon \mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}} \to [0,+\infty), \qquad (a_j)_{j\in\mathbb{N}} \mapsto \textstyle\sum_{j=-\infty}^{\infty} a_j 2^{-j},$$
$$\Psi\colon [0,+\infty) \to \mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}, \qquad t \mapsto (\lfloor 2^j t\rfloor \bmod 2)_{j\in\mathbb{Z}}.$$

The functions Φ and Ψ are Borel-measurable and $\Psi\circ\Phi = id_{\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}}$ a.e., $\Phi\circ\Psi = id_{[0,+\infty)}$ a.e. For that reason we can identify the measure spaces

$$(\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}, \mathcal{B}(\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}), \lambda_{\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}}) \equiv (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \lambda).$$

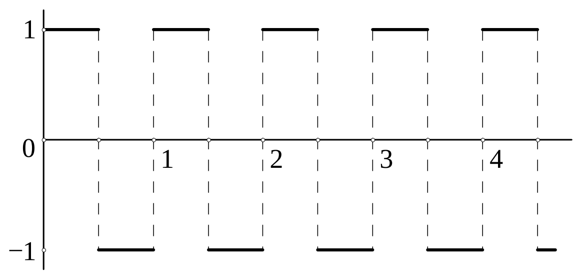
Denote: $E\colon \mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}} \to \mathbb{C}$, $E((c_j)_{j\in\mathbb{Z}}) = (-1)^{c_1}$. It is easy to verify that

$$E((a_j)_{j\in\mathbb{Z}} \oplus (b_j)_{j\in\mathbb{Z}}) = E((a_j)_{j\in\mathbb{Z}})E((b_j)_{j\in\mathbb{Z}}),$$

i.e. under the identification $\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}} \equiv \mathbb{R}_+$ we can write

$$E(x\oplus y) = E(x)E(y), \quad x,y\in[0,+\infty).$$

It is instructive to draw the graph of E and realize why it is sometimes called the *rectangular sine function.*



The Fourier transfrom on $\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}$ is called the *Walsh-Fourier transform* and it takes a function $f\in \mathrm{L}^1(\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}) \equiv \mathrm{L}^1(\mathbb{R}_+)$ and assigns to it $\hat{f}$ defined by:

$$\hat{f}(\xi) = \int_0^{+\infty} E(x\otimes\xi) f(x)\,dx, \qquad \xi\in\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}} \equiv \mathbb{R}_+.$$

The inverse Walsh-Fourier transform is the same object. Its basic properties are analogous to the ones from Section 1.1 and can be proved in the same way as before.

Proposition 7. *The Walsh-Fourier transform extends by density to* $\mathrm{L}^2(\mathbb{R}_+)$*. Take* $f \in \mathrm{L}^2(\mathbb{R}_+)$*,* $y, \eta \in \mathbb{R}_+$*,* $j \in \mathbb{Z}$*. Define the folowing transformations*

$$\begin{aligned} \textit{translations:} \quad & (\mathrm{T}_y f)(x) := f(x \oplus y), \\ \textit{modulations:} \quad & (\mathrm{M}_\eta f)(x) := E(x \otimes \eta) f(x), \\ \textit{dilations:} \quad & (\mathrm{D}_{2^j} f)(x) := 2^{-j/2} f(2^{-j} x). \end{aligned}$$

Then

$$(\mathrm{T}_y f)\hat{} = \mathrm{M}_y \hat{f}, \quad (\mathrm{M}_\eta f)\hat{} = \mathrm{T}_\eta \hat{f}, \quad (\mathrm{D}_{2^j} f)\hat{} = \mathrm{D}_{2^{-j}} \hat{f}.$$

The phase space is now $\mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}} \times \mathbb{Z}_2^{\mathbb{Z},\,\mathrm{fin}}$, which as a measure space can be identified with $(\mathbb{R}_+)^2 = [0, +\infty)^2$, i.e. the first quadrant. We call it the *Walsh phase plane*. The function $\mathbf{1}_{[0,1)}$ is its own Walsh-Fourier transform and it serves as an analogue of the Gaussian $e^{-\pi x^2}$ for the Fourier transform on $\mathbb{R}$. We see that this time the perfect localization is actually possible!

* * *

Let us now concentrate on the geometry of the Walsh phase plane. The *Walsh wave packet* system is $(w_{j,l,n})_{j \in \mathbb{Z},\, l,n \in \mathbb{N}_0}$, where we apply the three transformation groups (translations, modulations, dilations) to the function $\mathbf{1}_{[0,1)}$, i.e.

$$w_{j,l,n} := \mathrm{D}_{2^j} \mathrm{T}_l \mathrm{M}_n \mathbf{1}_{[0,1)}, \quad j, l, n \in \mathbb{Z}, \ l, n \geq 0,$$

which explicitly reads

$$w_{j,l,n}(x) = 2^{-\frac{j}{2}}\, \widetilde{W}_n(2^{-j} x - l), \quad x \in \mathbb{R}_+. \tag{4}$$

Here $\widetilde{W}_n$ denotes the extension of W_n by zero outside of the interval $[0, 1)$. Using the basic properties of the Walsh-Fourier transform we easily obtain

$$\hat{w}_{j,l,n} := \mathrm{D}_{2^{-j}} \mathrm{T}_n \mathrm{M}_l \mathbf{1}_{[0,1)} = w_{-j,n,l}.$$

Hence,

$w_{j,l,n}$ is supported on $[2^j l, 2^j (l + 1))$,

$\hat{w}_{j,l,n}$ is supported on $[2^{-j} n, 2^{-j}(n + 1))$, so it is natural to assign to a function $w_{j,l,n}$ a "tile" in the phase plane,

$$P_{j,l,n} := [2^j l, 2^j(l+1)) \times [2^{-j}n, 2^{-j}(n+1)).$$

In the following text a *tile* will be any rectangle in the phase plane $(\mathbb{R}_+)^2$ whose sides are dyadic intervals and whose area equals 1. A tile $P_{j,l,n}$ can be called the *phase plane support* of $w_{j,l,n}$. The correspondence between the set of all tiles and the Walsh wave packet system will be written as $P \mapsto w_P$. Let us also agree that in this case the time interval of P will be denoted I_P, while the frequency interval will be denoted Ω_P, i.e. $P = I_P \times \Omega_P$.

Observe that the functions w_P are normalized in L^2. Their L^∞ normalizations will also be convenient in the next chapter and we denote them by $\tilde{w}_P$, i.e.

$$\tilde{w}_P(x) = \tilde{w}_{j,l,n}(x) = \widetilde{W}_n(2^{-j}x - l),$$

so that

$$w_P = |I_P|^{-1/2}\tilde{w}_P.$$

Lemma 8. *If P and P' are disjoint tiles, then the functions $w_P, w_{P'} \in \mathrm{L}^2(\mathbb{R}_+)$ are mutually orthogonal.*

Proof. Observe that either time intervals $I_P, I_{P'}$ are disjoint (when the statement is obvious), or frequency intervals $\Omega_P, \Omega_{P'}$ are disjoint (when the statement follows by applying the Plancherel theorem). □

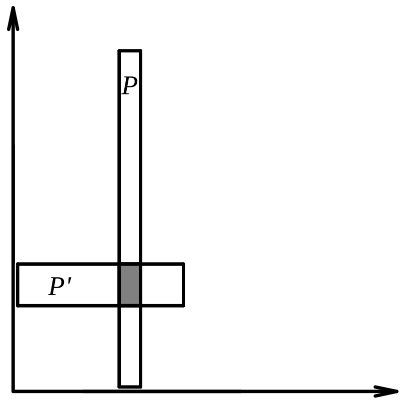
The following lemma is a dyadic analogue of Proposition 6.

Lemma 9. *For any two tiles P and P' we have $|\langle w_P, w_{P'}\rangle|^2 = |P \cap P'|$.*

Proof. Because of the previous lemma we can assume that $P \cap P' \neq \emptyset$.

$$\begin{aligned} P &= P_{j,l,n} = [2^j l, 2^j(l+1)) \times [2^{-j}n, 2^{-j}(n+1)) \\ P' &= P_{j',l',n'} = [2^{j'} l', 2^{j'}(l'+1)) \times [2^{-j'}n', 2^{-j'}(n'+1)) \end{aligned}$$

Without loss of generality $j \leq j'$.



Note that $I_P \cap I_{P'} \neq \emptyset$ implies $I_P \subseteq I_{P'}$, which gives

$$2^{j'-j} l' \leq l \leq 2^{j'-j} l' + 2^{j'-j} - 1. \tag{5}$$

Also, from $\Omega_P \cap \Omega_{P'} \neq \emptyset$ we get $\Omega_P \supseteq \Omega_{P'}$, which gives

$$2^{j'-j} n \leq n' \leq 2^{j'-j} n + 2^{j'-j} - 1. \tag{6}$$

Applying formula (4) we compute:

$$\langle w_P, w_{P'} \rangle = \int_{\mathbb{R}_+} w_P \overline{w_{P'}} = 2^{-\frac{j+j'}{2}} \int_{[2^j l, 2^j (l+1))} \widetilde{W}_n(2^{-j}x - l) \widetilde{W}_{n'}(2^{-j'}x - l')\, dx.$$

We are going to show that the function under the last integral sign is constantly equal to either -1 or 1 on the whole interval $[2^j l, 2^j(l+1))$. This will mean that

$$\langle w_P, w_{P'} \rangle = \pm 2^{-\frac{j+j'}{2}} 2^j = \pm 2^{-\frac{j'-j}{2}},$$

while

$$\lambda(P \cap P') = |I_P| \cdot |\Omega_{P'}| = 2^j 2^{-j'} = 2^{-(j'-j)}$$

and the proof will be completed.

Using the recurrence relations for the Walsh functions we can write

$$\widetilde{W}_{2k}(t) = \widetilde{W}_k(2t) + \widetilde{W}_k(2t-1), \tag{7}$$

$$\widetilde{W}_{2k+1}(t) = \widetilde{W}_k(2t) - \widetilde{W}_k(2t-1). \tag{8}$$

By repeated applications of (7) and (8) because of (6) we get

$$\widetilde{W}_{n'}(t) = \sum_{m=0}^{2^{j'-j}-1} \pm \widetilde{W}_n(2^{j'-j} t - m)$$

for some choice of $2^{j'-j}$ plus or minus signs. Substituting $t = 2^{-j'}x - l'$ and multiplying by $\widetilde{W}_n(2^{-j}x - l)$ we get

$$\widetilde{W}_n(2^{-j}x - l)\widetilde{W}_{n'}(2^{-j'}x - l') = \sum_{m=0}^{2^{j'-j}-1} \pm\widetilde{W}_n(2^{-j}x - l)\widetilde{W}_n(2^{-j}x - (2^{j'-j}l' + m)).$$

Recall condition (5), which claims that $l = 2^{j'-j}l' + m$ holds for precisely one $0 \leq m \leq 2^{j'-j} - 1$. Therefore, exactly one summand in the above sum equals

$$\pm\widetilde{W}_n(2^{-j}x - l)^2 = \pm 1 \quad \text{for } x \in [2^j l, 2^j(l+1)),$$

while the others are equal to 0. Finally,

$$\widetilde{W}_n(2^{-j}x - l)\widetilde{W}_{n'}(2^{-j'}x - l') = \pm 1 \quad \text{for } x \in [2^j l, 2^j(l+1)). \qquad \square$$

The following theorem enables the correspondence between "pavable" subsets of the phase plane $(\mathbb{R}_+)^2$ and closed subspaces of $\mathrm{L}^2(\mathbb{R}_+)$.

Theorem 10. *If $\mathcal{P}$ and $\mathcal{P}'$ are two collections of pairwise disjoint tiles such that $\bigcup_{P\in\mathcal{P}} P = \bigcup_{P'\in\mathcal{P}'} P'$, then*

$$\overline{[\{w_P : P \in \mathcal{P}\}]} = \overline{[\{w_{P'} : P' \in \mathcal{P}'\}]},$$

where $[\cdot]$ and $\overline{\cdot}$ denote the linear span and the closure in $\mathrm{L}^2(\mathbb{R}_+)$ respectively.

Proof. Take some tile $P \in \mathcal{P}$ and use the fact that $P \cap P'$, $P' \in \mathcal{P}'$ constitutes a countable partition of P.

$$\sum_{P'\in\mathcal{P}'} |\langle w_P, w_{P'}\rangle|^2 = (\text{Lemma 9}) = \sum_{P'\in\mathcal{P}'} \lambda(P \cap P') = \lambda(P) = 1.$$

By Lemma 8 the set $\{w_{P'} : P' \in \mathcal{P}'\}$ must be an orthonormal basis of the subspace $\overline{[\{w_{P'} : P' \in \mathcal{P}'\}]}$, so the square of the norm of the orthogonal projection of w_P onto that subspace must be

$$\Big\| \sum_{P'\in\mathcal{P}'} \langle w_P, w_{P'}\rangle w_{P'} \Big\|^2 = \sum_{P'\in\mathcal{P}'} |\langle w_P, w_{P'}\rangle|^2 = 1,$$

while (by the Pythagorean theorem) the square of the distance from w_P to that subspace must be

$$\Big\| w_P - \sum_{P'\in\mathcal{P}'} \langle w_P, w_{P'}\rangle w_{P'} \Big\|^2 = \|w_P\|^2 - \Big\| \sum_{P'\in\mathcal{P}'} \langle w_P, w_{P'}\rangle w_{P'} \Big\|^2 = 1 - 1 = 0.$$

We conclude $w_P \in \overline{[\{w_{P'} : P' \in \mathcal{P}'\}]}$ and, since $P \in \mathcal{P}$ was arbitrary, we have just shown $\overline{[\{w_P : P \in \mathcal{P}\}]} \subseteq \overline{[\{w_{P'} : P' \in \mathcal{P}'\}]}$. The other inclusion follows by symmetry. □

Therefore, to each set $S \subseteq (\mathbb{R}_+)^2$ that is a union of some family $\mathcal{P}$ of pairwise disjoint tiles we can assign a closed subspace of $\mathrm{L}^2(\mathbb{R}_+)$ given by the formula

$$V_S := \overline{[\{w_P : P \in \mathcal{P}\}]}$$

and the orthogonal projection Π_S in $\mathrm{L}^2(\mathbb{R}_+))$ onto that subspace acting by the formula

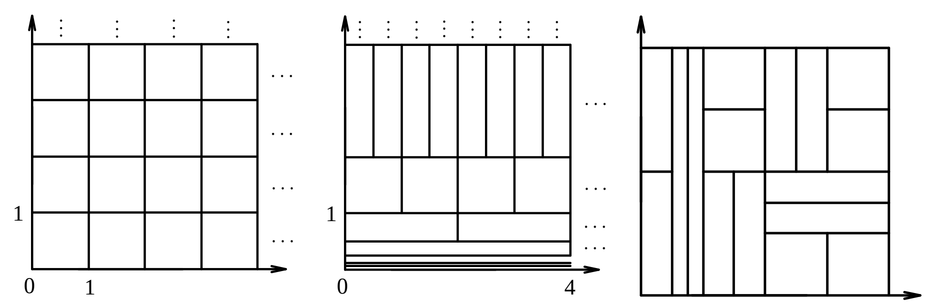
$$\Pi_S f := \sum_{P \in \mathcal{P}} \langle f, w_P \rangle w_P.$$

Because of Theorem 10 the definitions of V_S and Π_S do not depend on the actual tiling of S.

The mapping $S \mapsto V_S$ obviously has properties:

$$\begin{aligned} S \subseteq S' \quad &\Rightarrow \quad V_S \subseteq V_{S'}, \\ S \cap S' = \emptyset \quad &\Rightarrow \quad V_S \perp V_{S'}. \end{aligned}$$

It is interesting to note that each tiling of the whole phase plane $(\mathbb{R}_+)^2$ clearly gives one orthonormal basis for $\mathrm{L}^2(\mathbb{R}_+)$. The following figures show that there are many possible tilings.



Three sources of materials on dyadic harmonic analysis that complement each other well are [1], [27], and [30].

Chapter 2

Linear and multilinear singular integrals

In this chapter we finally use ideas from the previous one to successfully decompose and bound integral operators. The material is chosen to present some basic techniques in the field, but also not to overwhelm the beginner with the technicalities.

2.1 Symmetries of an operator

A *linear singular integral* is typically an operator of the form

$$(Tf)(x) := \text{p.v.} \int_{\mathbb{R}^n} K(x,y)f(y)dy.$$

The kernel K has to be controllably singular close to the "diagonal"

$$\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : x = y\}$$

and sometimes its changes of sign have to guarantee subtle cancellation properties. For instance, $K(x,y) = \frac{x_1-y_1}{|x-y|^{n+1}}$ is the so-called *Riesz kernel*. The letters "p.v." denote that the integral has to be understood in the principal value sense, which in this case means

$$\lim_{\varepsilon \searrow 0} \int_{\{y\in\mathbb{R}^n : |x-y|>\varepsilon\}} K(x,y)f(y)dy.$$

A rather broad and very useful class are the *Calderón-Zygmund operators*, which we do not discuss here.

One is typically interested in the estimates that a given operator satisfies and the most basic ones are the L^p estimates, for instance

$$\|Tf\|_{\mathrm{L}^p(\mathbb{R}^n)} \lesssim_p \|f\|_{\mathrm{L}^p(\mathbb{R}^n)}.$$

Here we write $A \lesssim_p B$ if the two quantities A and B satisfy $A \leq C_p B$ with some constant depending on p. The constant is made implicit as its actual value is often unimportant. The advantage of this notation is that we can change the constant from line to line in lengthy computations. The people who contributed most to the early developments of the theory of singular integral operators are Alberto P. Calderón, Antoni Zygmund, Elias M. Stein, and Guido L. Weiss.

Multilinear singular integral operators can also appear naturally and some motivating examples will be presented later in the course. One possible general scheme would be

$$T(f_1, f_2, \ldots, f_k)(x) := \text{p.v.} \int_{\Omega} K(x, y_1, y_2, \ldots, y_k) f_1(y_1) f_2(y_2) \cdots f_k(y_k) \\ d\sigma(y_1, y_2, \ldots y_k),$$

where Ω is a higher-dimensional plane in the Euclidean space in which $(y_1, \ldots, y_k)$ lives and σ is the translation-invariant measure on Ω (which coincides with the Hausdorff measure here). Therefore, we want to allow the possibilities when the variables $y_1, y_2, \ldots, y_k$ of the functions $f_1, f_2, \ldots, f_k$ are not necessarily independent, but rather related by certain linear constraints. This time "p.v." means that we "cut out" the region where K is singular and then let this region shrink. Several bilinear examples are

$$T(f,g)(x) := \text{p.v.} \int_{\mathbb{R}} f(x-t) g(x+t) \frac{dt}{t},$$
$$T(f,g)(x,y) := \text{p.v.} \int_{\mathbb{R}} f(x+t, y) g(x, y+t) \frac{dt}{t},$$
$$T(f,g)(x,y) := \text{p.v.} \int_{\mathbb{R}^2} f(x+s, y) g(x, y+t) \frac{s ds dt}{(s^2+t^2)^{3/2}}.$$

We are primarily interested in the L^p estimates again:

$$\|T(f_1, f_2, \ldots, f_k)\|_{\mathrm{L}^p(\mathbb{R}^n)} \lesssim_{p, p_1, \ldots, p_k} \prod_{j=1}^{k} \|f_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^{n_j})}. \tag{1}$$

Some people who contributed to bringing up multilinear singular integrals as an active area of research and who invented some of the most important tools are Ronald R. Coifman, Yves F. Meyer, Michael T. Lacey, and Christoph M. Thiele.

Note that everything will be "flat" in this course and we do not investigate the effects of curvature. Otherwise, singular integrals on manifolds are also an interesting and active area of study.

It is usually more convenient to convert multilinear operators into multilinear forms by dualizing them with an extra function.

$$\Lambda(f_1, f_2, \dots, f_k, f_{k+1}) := \int_{\mathbb{R}^n} T(f_1, f_2, \dots, f_k)(x) f_{k+1}(x) dx.$$

The desired estimate becomes

$$\Lambda(f_1, f_2, \dots, f_k, f_{k+1}) \lesssim_{p_1,\dots,p_k,p_{k+1}} \prod_{j=1}^{k+1} \|f_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^{n_j})}, \tag{2}$$

where p_{k+1} is the conjugated exponent of p and we write n_{k+1} for n. Inequalities (1) and (2) are equivalent as long as $p \geq 1$. Having $k+1$ functions instead of k of them is usually not a big conceptual difference, but the main advantage is that the symmetries of the operator might manifest themselves better. Those symmetries can also dictate the range of exponents in estimates (1) and (2), as we present on the following two examples.

* * *

Consider a trilinear form on 2D functions with determinantal kernel,

$$\Lambda(f, g, h) := \text{p.v.} \int_{\mathbb{R}^6} \frac{f(x_1, x_2) g(y_1, y_2) h(z_1, z_2)}{\begin{vmatrix} 1 & 1 & 1 \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix}} dx_1 dx_2 dy_1 dy_2 dz_1 dz_2.$$

Recall that the denominator equals 0 if and only if the three points (x_1, x_2), (y_1, y_2), and (z_1, z_2) lie on the same line in $\mathbb{R}^2$. Suppose that we have an estimate

$$|\Lambda(f, g, h)| \lesssim_{p,q,r} \|f\|_{\mathrm{L}^p(\mathbb{R}^2)} \|g\|_{\mathrm{L}^q(\mathbb{R}^2)} \|h\|_{\mathrm{L}^r(\mathbb{R}^2)} \tag{3}$$

for some exponents p, q, r. The same estimate must remain satisfied if we replace f, g, h by the dilates $\mathrm{D}_r f, \mathrm{D}_r g, \mathrm{D}_r h$ for any $r > 0$. On the one hand,

$$\begin{aligned} &\Lambda(\mathrm{D}_r f, \mathrm{D}_r g, \mathrm{D}_r h) \\ &= \text{p.v.} \int_{\mathbb{R}^6} \frac{r^{-3} f(r^{-1}x_1, r^{-1}x_2) g(r^{-1}y_1, r^{-1}y_2) h(r^{-1}z_1, r^{-1}z_2)}{\begin{vmatrix} 1 & 1 & 1 \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix}} dx_1 dx_2 dy_1 dy_2 dz_1 dz_2 \end{aligned}$$

$$\begin{aligned}
&\quad [x'_i = r^{-1}x_i,\ y'_i = r^{-1}y_i,\ z'_i = r^{-1}z_i] \\
&= \text{p.v.} \int_{\mathbb{R}^6} \frac{r^{-3} f(x'_1, x'_2) g(y'_1, y'_2) h(z'_1, z'_2)}{r^2 \left| \begin{smallmatrix} 1 & 1 & 1 \\ x'_1 & y'_1 & z'_1 \\ x'_2 & y'_2 & z'_2 \end{smallmatrix} \right|} r^6 dx'_1 dx'_2 dy'_1 dy'_2 dz'_1 dz'_2 \\
&= r\,\Lambda(f, g, h).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\|\mathrm{D}_r f\|_{\mathrm{L}^p(\mathbb{R}^2)} &= \Big(\int_{\mathbb{R}^2} |r^{-1} f(r^{-1}x_1, r^{-1}x_2)|^p dx_1 dx_2 \Big)^{1/p} \\
&\quad [x'_i = r^{-1}x_i] \\
&= \Big(\int_{\mathbb{R}^2} r^{-p} |f(x'_1, x'_2)|^p r^2 dx'_1 dx'_2 \Big)^{1/p} = r^{\frac{2}{p}-1} \|f\|_{\mathrm{L}^p(\mathbb{R}^2)},
\end{aligned}$$

so

$$\|\mathrm{D}_r f\|_{\mathrm{L}^p(\mathbb{R}^2)} \|\mathrm{D}_r g\|_{\mathrm{L}^q(\mathbb{R}^2)} \|\mathrm{D}_r h\|_{\mathrm{L}^r(\mathbb{R}^2)} = r^{2(1/p+1/q+1/r)-3} \|f\|_{\mathrm{L}^p(\mathbb{R}^2)} \|g\|_{\mathrm{L}^q(\mathbb{R}^2)} \|h\|_{\mathrm{L}^r(\mathbb{R}^2)}.$$

Applying (3) to $\mathrm{D}_r f, \mathrm{D}_r g, \mathrm{D}_r h$ in the places of f, g, h, we obtain

$$|\Lambda(f, g, h)| \lesssim_{p,q,r} r^{2(1/p+1/q+1/r)-4} \|f\|_{\mathrm{L}^p(\mathbb{R}^2)} \|g\|_{\mathrm{L}^q(\mathbb{R}^2)} \|h\|_{\mathrm{L}^r(\mathbb{R}^2)}.$$

Letting $r \to 0$ and $r \to +\infty$ we conclude that the necessary condition for having the desired estimate is

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 2.$$

Actually, Λ does satisfy many such L^p estimates, as is shown in [32].

* * *

Let us try another example,

$$T(f, g)(x, y) := \text{p.v.} \int_{\mathbb{R}^2} f(x - s, y - t) g(x + s, y + t) \frac{ds}{s} \frac{dt}{t},$$

i.e.

$$\Lambda(f, g, h) := \text{p.v.} \int_{\mathbb{R}^4} f(x - s, y - t) g(x + s, y + t) h(x, y) \frac{ds}{s} \frac{dt}{t} dx dy.$$

This time we have

$$\begin{aligned}
&\Lambda(\mathrm{D}_r f, \mathrm{D}_r g, \mathrm{D}_r h) \\
&= \text{p.v.} \int_{\mathbb{R}^4} r^{-3} f\big(r^{-1}(x-s), r^{-1}(y-t)\big) g\big(r^{-1}(x+s), r^{-1}(y+t)\big) \\
&\qquad\qquad\qquad\qquad h\big(r^{-1}x, r^{-1}y\big) \frac{ds}{s}\frac{dt}{t} dx dy \\
&\quad \big[x' = r^{-1}x,\ y' = r^{-1}y,\ s' = r^{-1}s,\ t' = r^{-1}t\big] \\
&= \text{p.v.} \int_{\mathbb{R}^4} r^{-3} f(x'-s', y'-t') g(x'+s', y'+t') h(x', y') \frac{ds'}{s'}\frac{dt'}{t'} r^2 dx' dy' \\
&= r^{-1}\Lambda(f, g, h).
\end{aligned}$$

Applying (3) to $\mathrm{D}_r f, \mathrm{D}_r g, \mathrm{D}_r h$ gives

$$|\Lambda(f,g,h)| \lesssim_{p,q,r} r^{2(1/p+1/q+1/r)-2} \|f\|_{\mathrm{L}^p(\mathbb{R}^2)} \|g\|_{\mathrm{L}^q(\mathbb{R}^2)} \|h\|_{\mathrm{L}^r(\mathbb{R}^2)}.$$

Thus, the necessary condition for the estimate is

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1.$$

However, one has to be aware that there is no guarantee that such estimates are actually true. Indeed, it is known that this trilinear form satisfies no L^p estimates at all; the counterexample can be found in [26].

Exercise 5. Prove that if

$$\Lambda(f,g,h) := \text{p.v.} \int_{\mathbb{R}^3} f(x,y) g(y,z) h(z,x) \frac{dx dy dz}{x+y+z}$$

satisfies estimate (3) for some exponents p, q, r, then we must have $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$. No estimates have been established so far for this multilinear form. It is only known that the estimates fail unless $1 < p, q, r < \infty$.

2.2 The Hilbert transform

It is quite likely that the reader has already met the *Hilbert transform*. It is defined for $f \in \mathrm{C}_c^1(\mathbb{R})$ as

$$(Hf)(x) := \text{p.v.} \int_{\mathbb{R}} f(x-t) \frac{dt}{t}.$$

The requirement that f is C^1 and compactly supported is required in order for the limit

$$\lim_{\varepsilon\to 0}\int_{\{t:|t|\geq\varepsilon\}} f(x-t)\frac{dt}{t} = \int_{\{t:|t|\leq 1\}} \frac{f(x-t)-f(x)}{t}dt + \int_{\{t:|t|\geq 1\}} f(x-t)\frac{dt}{t}.$$

to exist. Once we know that H is bounded on some space $\mathrm{L}^p(\mathbb{R})$, $1<p<\infty$, then we can exend it uniquely by continuity, but for now the dense subspace $\mathrm{C}_c^1(\mathbb{R})$ is fine as its domain.

It is not difficult to show the formula

$$(Hf)\widehat{\ }(\xi) = -i\,\mathrm{sgn}\,\xi\,\hat{f}(\xi).$$

Another way to state it is to say that the Fourier transform of the tempered distribution p.v.$\frac{1}{\pi t}$ equals the function $-i\,\mathrm{sgn}\,\xi$, or that the Hilbert transform is a Fourier multiplier with symbol $-i\,\mathrm{sgn}\,\xi$. Combining with the Plancherel theorem we see that H is an isometry with respect to the L^2 norm,

$$\|Hf\|_{\mathrm{L}^2(\mathbb{R})} = \|(Hf)\widehat{\ }\|_{\mathrm{L}^2(\mathbb{R})} = \|\hat{f}\|_{\mathrm{L}^2(\mathbb{R})} = \|f\|_{\mathrm{L}^2(\mathbb{R})},$$

so it is a unitary operator on $\mathrm{L}^2(\mathbb{R})$. A slightly harder task is to prove boundedness of H on $\mathrm{L}^p(\mathbb{R})$ for each $1<p<\infty$. This is a typical application of the Littlewood-Paley theory.

Observe that H commutes with translations and dilations. Indeed, for $f\in \mathrm{C}_c^1(\mathbb{R})$ we have

$$\begin{aligned}(H\mathrm{T}_y f)(x) &= \mathrm{p.v.}\int_{\mathbb{R}} f(x-y-t)\frac{dt}{t} = (Hf)(x-y) = (\mathrm{T}_y Hf)(x),\\ (H\mathrm{D}_r f)(x) &= \mathrm{p.v.}\int_{\mathbb{R}} r^{-1/2} f(r^{-1}(x-t))\frac{dt}{t} = [s=r^{-1}t,\ dt = r ds]\\ &= \mathrm{p.v.}\int_{\mathbb{R}} r^{-1/2} f(r^{-1}x-s)\frac{ds}{s} = r^{-1/2}(Hf)(r^{-1}x) = (\mathrm{D}_r Hf)(x).\end{aligned}$$

Take some orthonormal wavelet system $(\psi_{j,k})_{j,k\in\mathbb{Z}}$, $\psi_{j,k} = \mathrm{D}_{2^{-j}}\mathrm{T}_k\psi$. From the previous property we see that $(H\psi_{j,k})_{j,k\in\mathbb{Z}}$ is again a wavelet system. Moreover by unitarity of H this system must also be an orthonormal basis for $\mathrm{L}^2(\mathbb{R})$, so it is actually another orthonormal wavelet system. Denote $\vartheta_{j,k} := H\psi_{j,k}$. Expanding an L^2 function as $f=\sum_{j,k\in\mathbb{Z}}\langle f,\psi_{j,k}\rangle\psi_{j,k}$ we obtain the presentation

$$\langle Hf, g\rangle_{\mathrm{L}^2(\mathbb{R})} = \sum_{j,k\in\mathbb{Z}}\langle f,\psi_{j,k}\rangle\langle H\psi_{j,k}, g\rangle = \sum_{j,k\in\mathbb{Z}}\langle f,\psi_{j,k}\rangle\overline{\langle g,\vartheta_{j,k}\rangle},$$

i.e.

$$\langle Hf, g\rangle_{\mathrm{L}^2(\mathbb{R})} = \sum_{I\in\mathcal{D}} \langle f, \psi_I\rangle \overline{\langle g, \vartheta_I\rangle}.$$

There is a serious problem with this representation: If ψ is for instance chosen from $\mathrm{C}^1_c(\mathbb{R})$, then $\vartheta = H\psi$ does not have to be nice at all! Usually it is a better idea to decompose an operator in a single "nice" wavelet basis and then H would prove to be "almost diagonal" in the sense that the coefficients $\langle H\psi_I, \psi_J\rangle$ decay very rapidly when the intervals I and J are either distant or have very different lengths. More importantly, the same property holds for higher-dimensional Calderón-Zygmund operators that satisfy $T(\mathbf{1}) = 0 = T^*(\mathbf{1})$.

Instead of presenting the proof of boundedness of H on $\mathrm{L}^p(\mathbb{R})$, let us rather replace ψ_I and ϑ_I by the Haar wavelet $\mathbf{h}_I$ and give a simple proof of the bound

$$\sum_{I\in\mathcal{D}} |\langle f, \mathbf{h}_I\rangle\langle g, \mathbf{h}_I\rangle| \lesssim_{p,q} \|f\|_{\mathrm{L}^p(\mathbb{R})}\|g\|_{\mathrm{L}^q(\mathbb{R})} \tag{4}$$

for conjugated exponents $1 < p, q < \infty$.

We need the following well-known result.

Proposition 11 (Boundedness of the square function)**.** *Define the dyadic square function by*

$$\mathrm{S}f := \Big(\sum_{I\in\mathcal{D}} |I|^{-1}|\langle f, \mathbf{h}_I\rangle|^2 \mathbf{1}_I\Big)^{\frac{1}{2}}.$$

Then $\|\mathrm{S}f\|_{\mathrm{L}^p(\mathbb{R})} \lesssim_p \|f\|_{\mathrm{L}^p(\mathbb{R})}$ *for any* $1 < p < \infty$.

Turning back to (4) we rewrite the left hand side as

$$\int_{\mathbb{R}} \sum_{I\in\mathcal{D}} |I|^{-1/2}|\langle f, \mathbf{h}_I\rangle|\, \mathbf{1}_I \; |I|^{-1/2}|\langle g, \mathbf{h}_I\rangle|\, \mathbf{1}_I,$$

which is by the Cauchy-Schwarz inequality in I at most

$$\int_{\mathbb{R}} (\mathrm{S}f)\,(\mathrm{S}g) \le \|\mathrm{S}f\|_{\mathrm{L}^p(\mathbb{R})}\|\mathrm{S}g\|_{\mathrm{L}^q(\mathbb{R})} \lesssim_{p,q} \|f\|_{\mathrm{L}^p(\mathbb{R})}\|g\|_{\mathrm{L}^q(\mathbb{R})}.$$

Exercise 6. For any $R > 0$ let

$$(S_R f)(x) := \int_{-R}^{R} \hat{f}(\xi) e^{2\pi i x\xi} d\xi$$

denote the truncated Fourier integrals. Show the formula

$$S_R = \tfrac{i}{2}\big(\mathrm{M}_{-R} H \mathrm{M}_R - \mathrm{M}_R H \mathrm{M}_{-R}\big).$$

From this conclude that the operators $(S_R)_{R>0}$ are uniformly bounded on $\mathrm{L}^p(\mathbb{R})$, $1 < p < \infty$ and then that for each $f \in \mathrm{L}^p(\mathbb{R})$ one has $\lim_{R\to+\infty} S_R f = f$ in the L^p norm.
Remark: An analogous statement holds for the Fourier series on the torus $\mathbb{T}$. This is how M. Riesz proved that partial Fourier sums of a function $f \in \mathrm{L}^p(\mathbb{T})$, $p > 1$ converge in the L^p norm.

2.3 The bilinear Hilbert transform

The *bilinear Hilbert transform* is defined as

$$B(f,g)(x) := \text{p.v.} \int_{\mathbb{R}} f(x-t)g(x+t)\frac{dt}{t}.$$

It was introduced by A. Calderón [3] regarding the conjecture on boundedness of the Cauchy integral along Lipschitz curves,

$$(C_{\downarrow} f)(z) := \lim_{\varepsilon\searrow 0} \frac{1}{2\pi i} \int_{\Gamma} \frac{f(\zeta)}{\zeta - (z + i\varepsilon)} d\zeta,$$

which was later established by “softer” techniques [4],[6], but we do not discuss them here.

Once again, B rather obviously commutes with translations and dilations. Let us dualize it with the third function in order to reveal yet another symmetry,

$$\Lambda(f,g,h) := \int_{\mathbb{R}} \text{p.v.} \int_{\mathbb{R}} f(x-t)g(x+t)h(x)\frac{dt}{t} dx.$$

It is the modulation symmetry. Namely, for any $\eta \in \mathbb{R}$ we have

$$\begin{aligned}&\Lambda\big(\mathrm{M}_\eta f, \mathrm{M}_\eta g, \mathrm{M}_{-2\eta} h\big)\\ &= \int_{\mathbb{R}} \text{p.v.} \int_{\mathbb{R}} e^{2\pi i(x-t)\eta} f(x-t) e^{2\pi i(x+t)\eta} g(x+t) e^{-4\pi i x\eta} h(x)\frac{dt}{t} dx = \Lambda(f,g,h).\end{aligned}$$

A general class of objects of which the BHT is a prominent representative is called the *modulation invariant forms*.

We would like to prove the bound

$$|\Lambda(f,g,h)| \lesssim_{p,q,r} \|f\|_{\mathrm{L}^p(\mathbb{R})}\|g\|_{\mathrm{L}^q(\mathbb{R})}\|h\|_{\mathrm{L}^r(\mathbb{R})} \tag{5}$$

for any exponents $2 < p,q,r < \infty$ such that $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$. The range of exponents for which the estimate holds is actually larger, but this is the simplest and chronologically the first established case.

Equivalently, one can view B as a bilinear multiplier, i.e.

$$B(f,g)(x) := \int_{\mathbb{R}^2} \big(-\pi i \operatorname{sgn}(\xi-\eta)\big) e^{2\pi i x(\xi+\eta)} \hat{f}(\xi)\hat{g}(\eta)d\xi d\eta.$$

Let us also comment that the pointwise product is a trivial example of a bilinear multiplier,

$$f(x)g(x) := \int_{\mathbb{R}^2} e^{2\pi i x(\xi+\eta)} \hat{f}(\xi)\hat{g}(\eta)d\xi d\eta,$$

and it trivially satisfies bounds (5). Therefore, by considering a linear combination of these, it is enough to bound the multiplier associated with $\mathbf{1}_{(0,+\infty)}$,

$$T(f,g)(x) := \int_{\mathbb{R}^2} \mathbf{1}_{(0,+\infty)}(\xi-\eta) e^{2\pi i x(\xi+\eta)} \hat{f}(\xi)\hat{g}(\eta)d\xi d\eta.$$

Note that the symbol of the multiplier is singular along the whole line $\xi=\eta$, which might be a heuristic explanation why it is more difficult then the linear Hilbert transform.

The first natural step is to perform a smooth decomposition of the symbol

$$\mathbf{1}_{(0,+\infty)}(\tau) = \sum_{j\in\mathbb{Z}} \theta(3^{-j}\tau),$$

where θ is a C^∞ function compactly supported in the open interval $(0,+\infty)$. If we set $\psi=\check{\theta}$, then we can write

$$\check{\mathbf{1}}_{(0,+\infty)}(t) = \sum_{j\in\mathbb{Z}} 3^j\psi(3^j t).$$

The effect is that the corresponding trilinear form decomposes into

$$\widetilde{\Lambda}(f,g,h) = \sum_{j\in\mathbb{Z}} \int_{\mathbb{R}^2} f(x-t)g(x+t)h(x)3^j\psi(3^j t)dtdx.$$

We further perform the smooth partition of unity in order to localize in the x variable,

$$\mathbf{1}_{\mathbb{R}}(x) = \sum_{k\in\mathbb{Z}} \varphi(x-k),$$

for some C^{∞} compactly supported function φ. This finally leads to

$$\widetilde{\Lambda}(f,g,h) = \sum_{j,k\in\mathbb{Z}} \int_{\mathbb{R}^2} f(x-t)g(x+t)h(x)\varphi(x-k)3^j\psi(3^j t)dtdx.$$

Observe that $\varphi(x-k)$ is a function with its time support around k, while $\psi(3^j t)$ is a function with its frequency support around 3^j. At this point switching to a toy model will be methodologically convenient.

2.4 Triadic model of the BHT

Motivated by the previous decomposition we define the triadic model of $\widetilde{\Lambda}$ as

$$\widetilde{\Lambda}_3(f,g,h) = \sum_{j\in\mathbb{Z}} \sum_{\substack{I \text{ triadic interval} \\ |I|=3^{-j}}} \Big| \int_{\mathbb{R}^2} f(x\ominus t)g(x\oplus t)h(x)\mathbf{1}_I(x)3^j\tilde{\mathbf{h}}_{[0,3^{-j})}(t)dtdx \Big|.$$

Here $\oplus$ and $\ominus$ denote the operations in the $\mathbb{Z}_3$ Cantor group model for $\mathbb{R}_+$, i.e. we are adding real numbers in base 3 without carrying over digits. We choose to work in characteristic 3 (instead of 2) in order for $x\ominus t$, $x\oplus t$, x to be mutually different when $t\neq 0$. The Haar functions $\tilde{\mathbf{h}}_I$ are now L^{∞} normalized, so

$$\tilde{\mathbf{h}}_I = \mathbf{1}_{I_0} + \omega\mathbf{1}_{I_1} + \omega^2\mathbf{1}_{I_2},$$

where I_0, I_1, I_2 are the thirds of I and $\omega = e^{2\pi i/3}$. Indeed $\tilde{\mathbf{h}}_I$ is just the L^{∞} normalized first Walsh function $\tilde{w}_{I,1}$, but in the characteristics 3. The reader should not feel uncomfortable in this setting, as everything from Section 1.5 applies again.

Inserting absolute values in $\widetilde{\Lambda}_3$ is important, as otherwise the form telescopes to the pointwise product, which is trivially bounded. A metaphysical reason is that on the positive frequency half-axis we see no difference between p.v.$\frac{1}{\pi t}$ and δ_0. Only after being broken into a sequence of scales, the finite characteristic model becomes faithful.

Let us substitute $y = x\ominus t$, so that

$$t = x\ominus y \qquad \text{and} \qquad x\oplus t = 2x\ominus y = \ominus x\ominus y,$$

since we are working in characteristic 3. Using

$$\mathbf{1}_I(x)\tilde{\mathbf{h}}_{[0,3^{-j})}(t) = \mathbf{1}_I(x)\mathbf{1}_I(y)\tilde{\mathbf{h}}_{[0,3^{-j})}(x \ominus y) = \tilde{\mathbf{h}}_I(x)\tilde{\mathbf{h}}_I(\ominus y) = \tilde{\mathbf{h}}_I(x)\overline{\tilde{\mathbf{h}}_I(y)}$$

the form becomes

$$\widetilde{\Lambda}_3(f,g,h) = \sum_{j\in\mathbb{Z}} 3^j \sum_{\substack{I \text{ triadic} \\ |I|=3^{-j}}} \Big| \int_{\mathbb{R}^2} h(x)f(y)g(\ominus x \ominus y)\tilde{\mathbf{h}}_I(x)\overline{\tilde{\mathbf{h}}_I(y)}dxdy\Big|.$$

Decompose the function g into the triadic Walsh-Fourier series,

$$g(z) = |I|^{-1}\sum_{n=0}^{\infty}\langle g, \tilde{w}_{I,n}\rangle \tilde{w}_{I,n}(z),$$

i.e.

$$g(\ominus x \ominus y) = |I|^{-1}\sum_{n=0}^{\infty}\langle g, \tilde{w}_{I,n}\rangle \overline{\tilde{w}_{I,n}(x)\tilde{w}_{I,n}(y)},$$

where $\tilde{w}_{I,n}$ are now the L^∞ normalized Walsh functions, while we keep the notation $w_{I,n}$ for the L^2 normalized ones. Plugging in and using

$$\tilde{w}_{I,m}\tilde{w}_{I,n} = \tilde{w}_{I,m\oplus n}$$

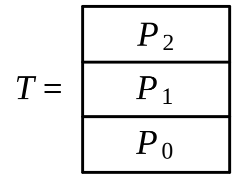
finally gives

$$\begin{aligned}\widetilde{\Lambda}_3(f,g,h) &= \sum_{j\in\mathbb{Z}} 3^{2j} \sum_{\substack{I \text{ triadic} \\ |I|=3^{-j}}} \Big| \sum_{n=0}^{\infty}\int_{\mathbb{R}^2} h(x)f(y)\langle g, \tilde{w}_{I,n}\rangle \overline{\tilde{w}_{I,n\ominus 1}(x)\tilde{w}_{I,n\oplus 1}(y)}dxdy\Big| \\ &\leq \sum_{I \text{ triadic interval}}\sum_{n=0}^{\infty} |I|^{-2}\big|\langle f, \tilde{w}_{I,n\oplus 1}\rangle\langle g, \tilde{w}_{I,n}\rangle\langle h, \tilde{w}_{I,n\ominus 1}\rangle\big| \\ &\leq \sum_{I \text{ triadic interval}}\sum_{n=0}^{\infty} |I|^{-1/2}\big|\langle f, w_{I,n\oplus 1}\rangle\langle g, w_{I,n}\rangle\langle h, w_{I,n\ominus 1}\rangle\big|.\end{aligned}$$

The right hand side can now be written as three mutually similar sums of the form

$$\Lambda(f,g,h) = \sum_{T \text{ tritile}} |I_T|^{-1/2}\big|\langle f, w_{P_0}\rangle\langle g, w_{P_1}\rangle\langle h, w_{P_2}\rangle\big|.$$

The sum is taken over all tritiles $T = I_T \times \Omega_T$ vertically divided into three tiles P_0, P_1, P_2.



Theorem 12. *The estimate*

$$|\Lambda(f,g,h)| \;\lesssim_{p,q,r} \|f\|_{\mathrm{L}^p(\mathbb{R})}\|g\|_{\mathrm{L}^q(\mathbb{R})}\|h\|_{\mathrm{L}^r(\mathbb{R})} \tag{6}$$

holds for $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1,\;\; 2<p,q,r<\infty.$

In all that follows, a tritile will be a rectangle of area 3 whose sides are triadic intervals. A fundamental property of tritiles we will use in the proof is that if the lower thirds P_0 of some collection of tritiles all intersect, then their middle thirds P_1 are mutually disjoint and the same holds for their upper thirds P_2. Two analogous properties, for intersecting middle or upper thirds, are equally obvious.

The rest of the section is dedicated to the proof of the above theorem. It closely follows [30], with only a few details written as in [17]. Before we do anything, let us observe that by the usual limiting arguments we can assume that f,g,h are bounded compactly supported functions and also that is it enough to consider only tritiles T such that $|I_T|\geq 3^{-N}$ for some "large" fixed positive integer N. The bound we prove will not depend on N so we will be able to take $N\to\infty$. The advantage of these restrictions is that they make all of the following arguments finite.

We can define partial order on the set of all tritiles T by

$$T\preceq T' \quad \text{if and only if} \quad I_T\subseteq I_{T'} \text{ and } \Omega_T\supseteq\Omega_{T'}.$$

Observe that tritiles T and T' are comparable if and only if $T\cap T'\neq\emptyset$. A collection of tritiles $\mathcal{C}$ is *convex* if for any three tritiles T,T',T''

$$(T\preceq T'\preceq T'')\,\&\,(T,T''\in\mathcal{C}) \;\Rightarrow\; T'\in\mathcal{C}.$$

Lemma 13. *If $\mathcal{C}$ is any finite convex collection of tritiles, then the union of $\mathcal{C}$ (as a subset of $(\mathbb{R}_+)^2$) can be decomposed into a collection $\mathcal{D}$ of mutually disjoint tiles. In particular, the orthogonal projection $\Pi_{\cup\mathcal{C}}=\Pi_{\cup\mathcal{D}}$ makes sense. Moreover, the collection $\mathcal{D}$ can be chosen in a way that it "preserves" minimal tritiles in $\mathcal{C}$; more precisely, if T is any minimal tritile in $\mathcal{C}$, then T decomposes horizontally into three disjoint tiles from $\mathcal{D}$.*

Proof of Lemma 13. We can prove the statement by the induction on the total number of tritiles in $\mathcal{C}$. It clearly holds when $\mathcal{C}$ is either empty or consists of a single tritile. Suppose that we are given a nonempty finite convex collection $\mathcal{C}$, take a maximal tritile $T \in \mathcal{C}$, and suppose that it is divided horizontally into tiles R_0, R_1, R_2. For some i consider a tritile Q_i which can be divided vertically into R_i and two other tiles.

- If no tritiles from $\mathcal{C} \setminus \{T\}$ intersect R_i, then R_i can be chosen for the output collection.
- If there exists a tritile $T' \in \mathcal{C} \setminus \{T\}$ intersecting R_i, then $T' \preceq Q_i \preceq T$. By the convexity of $\mathcal{C}$ we must have $Q_i \in \mathcal{C}$. Therefore, $\mathcal{C} \setminus \{T\}$ covers R_i, so this tile can be discarded for now.

Since T was maximal, the collection $\mathcal{C} \setminus \{T\}$ is still convex and the induction hypothesis applies. □

Lemma 14. *For any tritile T divided vertically into tiles P_0, P_1, P_2 we have*

$$\tfrac{1}{3}\|\Pi_T f\|_{\mathrm{L}^\infty} \le |I_T|^{-1/2} \max_{1\le j\le 3} |\langle f, w_{P_j}\rangle| \le \|\Pi_T f\|_{\mathrm{L}^\infty}.$$

Proof of Lemma 14. Observe that by the recurrence relations for Walsh functions we easily get

$$\begin{aligned} \|\Pi_T f\|_{\mathrm{L}^\infty} &= \|\langle f, w_{P_0}\rangle w_{P_0} + \langle f, w_{P_1}\rangle w_{P_1} + \langle f, w_{P_2}\rangle w_{P_2}\|_{\mathrm{L}^\infty} \\ &= |I_T|^{-1/2} \max\big\{\big|\langle f, w_{P_0}\rangle + \langle f, w_{P_1}\rangle + \langle f, w_{P_2}\rangle\big|, \\ &\qquad\qquad \big|\langle f, w_{P_0}\rangle + \langle f, w_{P_1}\rangle\omega + \langle f, w_{P_2}\rangle\omega^2\big|, \\ &\qquad\qquad \big|\langle f, w_{P_0}\rangle + \langle f, w_{P_1}\rangle\omega^2 + \langle f, w_{P_2}\rangle\omega\big|\big\}. \end{aligned}$$

The lemma is now obvious by the triangle inequality. □

Let the *density* of a tritile T with respect to a function f be defined as

$$\delta(T, f) := \sup_{T' \succeq T} \|\Pi_{T'} f\|_{\mathrm{L}^\infty}.$$

Observe that $\delta(T, f)$ decays to 0 as $|I_T|$ grows, simply because f is bounded and compactly supported. Also, the tritiles with $\delta(T, f) = 0$ can be discarded from Λ.

We introduce the collections of tritiles that have comparable density. For any $k \in \mathbb{Z}$ we define

$$\mathcal{P}_k^f := \big\{T \;:\; 2^k \le \delta(T, f) < 2^{k+1}\big\},$$

and let $\mathcal{M}_k^f$ denote the family of maximal tritiles in $\mathcal{P}_k^f$. Collections $\mathcal{P}_k^g$, $\mathcal{M}_k^g$, $\mathcal{P}_k^h$, $\mathcal{M}_k^h$ are defined analogously. Furthermore, for any triple of integers k_1, k_2, k_3 we set

$$\mathcal{P}_{k_1,k_2,k_3} := \mathcal{P}_{k_1}^f \cap \mathcal{P}_{k_2}^g \cap \mathcal{P}_{k_3}^h \,,$$

and let $\mathcal{M}_{k_1,k_2,k_3}$ denote the family of maximal tritiles in $\mathcal{P}_{k_1,k_2,k_3}$. Let us enumerate the tritiles from $\mathcal{M}_{k_1,k_2,k_3}$ as $Q_1, Q_2, \ldots$. For each i we can consider the subcollection of $\mathcal{P}_{k_1,k_2,k_3}$ consisting only of tritiles that are dominated by Q_i, i.e.

$$\mathcal{T}_i := \{T \in \mathcal{P}_{k_1,k_2,k_3} : T \preceq Q_i\}.$$

Some of the tritiles might belong to more than one such collection, but this is allowed. Note that $\mathcal{T}_i$ is a finite convex collection of tritiles with Q_i as its greatest element with respect to $\preceq$. Convexity follows simply from the fact that the density $\delta(\cdot, f)$ is monotonically decreasing with respect to the partial order $\preceq$. We can say that each $\mathcal{T}_i$ is a *convex tree* having Q_i as its root.

For each tree of tritiles $\mathcal{T}$ we introduce the form $\Lambda_{\mathcal{T}}(f, g, h)$, defined in exactly the same way as Λ, but summing over tritiles $T \in \mathcal{T}$ only. If $\mathcal{T}$ is any tree of tritiles from $\mathcal{P}_{k_1,k_2,k_3}$ obtained by applying the previous procedure, the key estimate we need to show is the so-called *single tree estimate*:

$$|\Lambda_{\mathcal{T}}(f, g, h)| \lesssim 2^{k_1+k_2+k_3} |I_{\mathcal{T}}|, \tag{7}$$

where $I_{\mathcal{T}}$ is the frequency interval of the root of $\mathcal{T}$.

In order to prove it we denote the root of $\mathcal{T}$ simply by Q and further split $\mathcal{T}$ into three subcollections all having Q in common (but otherwise disjoint),

$$\begin{aligned}
\mathcal{T}_0 &:= \{T \in \mathcal{T} : \Omega_{P_0} \supseteq \Omega_Q\}, \\
\mathcal{T}_1 &:= \{T \in \mathcal{T} : \Omega_{P_1} \supseteq \Omega_Q\}, \\
\mathcal{T}_2 &:= \{T \in \mathcal{T} : \Omega_{P_2} \supseteq \Omega_Q\}.
\end{aligned}$$

Observe that $\mathcal{T}_i$ need **not** be convex. Without loss of generality let us concentrate on $\mathcal{T}_0$. For tritiles $T \in \mathcal{T}_0$ the middle thirds P_1 are mutually disjoint tiles and the upper thirds P_2 are also mutually disjoint. Consequently, the corresponding wave packets w_{P_1} are mutually orthogonal and the same is true for w_{P_2}. Also recall that by Lemma 14 and the construction

$$|I_T|^{-1/2} |\langle f, w_{P_0} \rangle| \leq \|\Pi_T f\|_{\mathrm{L}^\infty(\mathbb{R})} \leq \delta(T, f) \leq 2^{k_1+1}.$$

Therefore we estimate

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_0} |I_T|^{-1/2}\big|\langle f, w_{P_0}\rangle\langle g, w_{P_1}\rangle\langle h, w_{P_2}\rangle\big| \\
&\lesssim \sum_{T\in\mathcal{T}_0} 2^{k_1}\big|\langle g, w_{P_1}\rangle\langle h, w_{P_2}\rangle\big| \\
&\leq 2^{k_1}\Big(\sum_{T\in\mathcal{T}_0} |\langle g, w_{P_1}\rangle|^2\Big)^{1/2}\Big(\sum_{T\in\mathcal{T}_0} |\langle h, w_{P_2}\rangle|^2\Big)^{1/2} \\
&\leq 2^{k_1}\|\Pi_{\cup\mathcal{T}} g\|_{\mathrm{L}^2(\mathbb{R})}\|\Pi_{\cup\mathcal{T}} h\|_{\mathrm{L}^2(\mathbb{R})}.
\end{aligned}$$

Here the orthogonal projection $\Pi_{\cup\mathcal{T}}$ corresponding to the area of the whole tree $\cup\mathcal{T} = \bigcup_{T\in\mathcal{T}} T$ appears. We could have introduced it because of Lemma 13 and the fact that $\mathcal{T}$ is convex. Observe that $\Pi_{\cup\mathcal{T}} g$ is supported on $I_{\mathcal{T}}$ and that we have $\|\Pi_{\cup\mathcal{T}} g\|_{\mathrm{L}^\infty(\mathbb{R})} \lesssim 2^{k_2}$. To see this we take $x \in I_{\mathcal{T}}$ and the minimal $T_{\min} \in \mathcal{T}$ containing x. By Lemma 13 again we can write

$$\Pi_{\cup\mathcal{T}} g = \Pi_{T_{\min}} g + \Pi_{(\cup\mathcal{T})\setminus T_{\min}} g$$

and observe that the two functions on the right hand side have disjoint time supports, so

$$(\Pi_{\cup\mathcal{T}} g)(x) = (\Pi_{T_{\min}} g)(x) \quad \text{and} \quad |(\Pi_{T_{\min}} g)(x)| \leq \delta(T_{\min}, g) \leq 2^{k_2+1}.$$

Finally,

$$2^{k_1}\|\Pi_{\cup\mathcal{T}}(g)\|_{\mathrm{L}^2(\mathbb{R})}\|\Pi_{\cup\mathcal{T}}(h)\|_{\mathrm{L}^2(\mathbb{R})} \leq 2^{k_1}2^{k_2}|I_{\mathcal{T}}|^{1/2}2^{k_3}|I_{\mathcal{T}}|^{1/2},$$

which is what we needed.

* * *

We decompose Λ into a sum of $\Lambda_{\mathcal{T}}$ over all $k_1, k_2, k_3 \in \mathbb{Z}$ and all trees $\mathcal{T}$ with roots from $\mathcal{M}_{k_1,k_2,k_3}$. In order to finish the proof of (6) using (7), it remains to show

$$\sum_{k_1,k_2,k_3\in\mathbb{Z}} 2^{k_1+k_2+k_3} \sum_{Q\in\mathcal{M}_{k_1,k_2,k_3}} |I_Q| \ \lesssim_{p,q,r} \|f\|_{\mathrm{L}^p}\|g\|_{\mathrm{L}^q}\|h\|_{\mathrm{L}^r}. \tag{8}$$

The trick is in the following observation. Fix any triple $k_1, k_2, k_3 \in \mathbb{Z}$ and some $Q \in \mathcal{M}^f_{k_1}$. Consider the collection of all $Q' \in \mathcal{M}_{k_1,k_2,k_3}$ that are $\preceq Q$. They are incomparable by maximality, but their frequency intervals $\Omega_{Q'}$ contain Ω_Q. Therefore their time intervals $I_{Q'}$ must be disjoint and are also contained in I_Q, so

$$\sum_{\substack{Q'\in\mathcal{M}_{k_1,k_2,k_3}\\ Q'\preceq Q}} |I_{Q'}| \ \leq \ |I_Q|.$$

We can now sum over all $Q \in \mathcal{M}_{k_1}^f$ and observe that each $Q' \in \mathcal{M}_{k_1,k_2,k_3}$ appears at least once on the left hand side, since it is certainly dominated by some maximal element in the bigger collection. Therefore,

$$\sum_{Q' \in \mathcal{M}_{k_1,k_2,k_3}} |I_{Q'}| \leq \sum_{Q \in \mathcal{M}_{k_1}^f} |I_Q|.$$

The same is true with $\mathcal{M}_{k_2}^g$ and $\mathcal{M}_{k_3}^h$. Thus, it suffices to prove

$$\sum_{k_1,k_2,k_3 \in \mathbb{Z}} 2^{k_1+k_2+k_3} \min\Big(\sum_{Q \in \mathcal{M}_{k_1}^f} |I_Q|, \sum_{Q \in \mathcal{M}_{k_2}^g} |I_Q|, \sum_{Q \in \mathcal{M}_{k_3}^h} |I_Q| \Big) \lesssim_{p,q,r} \|f\|_{\mathrm{L}^p} \|g\|_{\mathrm{L}^q} \|h\|_{\mathrm{L}^r}\,. \tag{9}$$

Lemma 15. *For $2 < p < \infty$ we have*

$$\sum_{k \in \mathbb{Z}} 2^{pk} \sum_{Q \in \mathcal{M}_k^f} |I_Q| \ \lesssim_p \ \|f\|_{\mathrm{L}^p}^p.$$

Proof of Lemma 15. Consider the following version of the maximal function,

$$\mathrm{M}_2 f \ := \sup_{I \text{ triadic interval}} \Big(\frac{1}{|I|} \int_I |f|^2 \Big)^{1/2} \mathbf{1}_I\,.$$

Clearly, $\mathrm{M}_2 f$ is bounded by the square root of the uncentered maximal function applied to $|f|^2$, which yields that M_2 is bounded on $\mathrm{L}^p(\mathbb{R})$ for any $2 < p < \infty$.

For each $Q \in \mathcal{M}_k^f$ and any tritile $\tilde{Q}$ such that $\tilde{Q} \succ Q$ we know that $\tilde{Q} \notin \mathcal{P}_k^f$ (by the maximality of Q), i.e. $\|\Pi_{\tilde{Q}} f\|_{\mathrm{L}^\infty} < 2^k$, so

$$\|\Pi_Q f\|_{\mathrm{L}^\infty} = \delta(Q, f) \geq 2^k.$$

Furthermore, if we decompose Q vertically into tiles P_0, P_1, P_2, then by Lemma 14 we can choose one of them, denoted by P_Q, such that

$$|I_Q|^{-1/2} |\langle f, w_{P_Q} \rangle| \geq \frac{1}{3} 2^k > 2^{k-2}. \tag{10}$$

Therefore, for $x \in I_Q$ we have

$$(\mathrm{M}_2 f)(x) \geq \Big(\frac{1}{|I_Q|} \int_{I_Q} |f|^2 \Big)^{1/2} \geq \frac{1}{|I_Q|} \int_{I_Q} |f| \geq \frac{|\langle f, \tilde{w}_{P_Q} \rangle|}{|I_Q|} = \frac{|\langle f, w_{P_Q} \rangle|}{|I_Q|^{1/2}} > 2^{k-2},$$

which gives us

$$I_Q \subseteq A_k := \{\mathrm{M}_2 f > 2^{k-2}\}.$$

However, we have not yet used orthogonality arguments. Recall that the tritiles in $\mathcal{M}_k^f$ are mutually disjoint (by maximality), so $\{P_Q : Q \in \mathcal{M}_k^f\}$ is a collection of disjoint tiles. By (10) and orthogonality of the wave packets w_{P_Q} we have

$$\sum_{Q\in\mathcal{M}_k^f} |I_Q| \lesssim 2^{-2k} \sum_{Q\in\mathcal{M}_k^f} |\langle f, w_{P_Q}\rangle|^2 = 2^{-2k} \sum_{Q\in\mathcal{M}_k^f} |\langle f\mathbf{1}_{A_k}, w_{P_Q}\rangle|^2 \le 2^{-2k}\|f\|_{\mathrm{L}^2(A_k)}^2.$$

Consider the collection of all maximal triadic intervals J contained in A_k; they are disjoint and cover A_k completely. For any such J take $\tilde{J}$ to be the unique three times larger triadic interval containing J, so by the maximality we know

$$\Big(\frac{1}{|\tilde{J}|}\int_{\tilde{J}} |f|^2\Big)^{1/2} \le 2^{k-2},$$

which implies

$$\int_J |f|^2 \lesssim 2^{2k}|J|.$$

Summing over all such J we obtain

$$\|f\|_{\mathrm{L}^2(A_k)}^2 \lesssim 2^{2k}|A_k|.$$

Finally,

$$\begin{aligned}\sum_{k\in\mathbb{Z}} 2^{pk} \sum_{Q\in\mathcal{M}_k^f} |I_Q| &\lesssim \sum_{k\in\mathbb{Z}} 2^{(p-2)k}\|f\|_{\mathrm{L}^2(A_k)}^2 \lesssim \sum_{k\in\mathbb{Z}} 2^{pk}|A_k| \\ &= \sum_{k\in\mathbb{Z}} 2^{pk}|\{\mathrm{M}_2 f > 2^{k-2}\}| \sim_p \|\mathrm{M}_2 f\|_{\mathrm{L}^p}^p \lesssim_p \|f\|_{\mathrm{L}^p}^p,\end{aligned}$$

because M_2 is bounded on $\mathrm{L}^p(\mathbb{R})$. This completes the proof of the lemma. □

In order to bound (9), we split it into three parts, depending on which of the numbers

$$\frac{2^{pk_1}}{\|f\|_{\mathrm{L}^p}^p},\ \frac{2^{qk_2}}{\|g\|_{\mathrm{L}^q}^q},\ \frac{2^{rk_3}}{\|h\|_{\mathrm{L}^r}^r}$$

is the largest. By obvious symmetry it is enough to show how to bound the part of the sum when $\frac{2^{pk_1}}{\|f\|_{\mathrm{L}^p}^p} \ge \frac{2^{qk_2}}{\|g\|_{\mathrm{L}^q}^q}$ and $\frac{2^{pk_1}}{\|f\|_{\mathrm{L}^p}^p} \ge \frac{2^{rk_3}}{\|h\|_{\mathrm{L}^r}^r}$. Write

$$\sum_{k_1\in\mathbb{Z}} \frac{2^{pk_1}}{\|f\|_{\mathrm{L}^p}^p}\Big(\sum_{Q\in\mathcal{M}_{k_1}^F} |I_Q|\Big) \sum_{k_2,k_3} \bigg(\frac{2^{qk_2}/\|g\|_{\mathrm{L}^q}^q}{2^{pk_1}/\|f\|_{\mathrm{L}^p}^p}\bigg)^{\frac{1}{q}} \bigg(\frac{2^{rk_3}/\|h\|_{\mathrm{L}^r}^r}{2^{pk_1}/\|f\|_{\mathrm{L}^p}^p}\bigg)^{\frac{1}{r}} \lesssim_{p,q,r} 1\,,$$

sum the two convergent geometric series, and finally use Lemma 15.

2.5 The Carleson operator

Suppose that we want to recover $f \in \mathrm{L}^2(\mathbb{R})$ from its Fourier transform. There are many ways of doing that, but the most natural would probably be

$$\lim_{R\to+\infty} \int_{-R}^{R} \hat{f}(\xi) e^{2\pi i x\xi} d\xi = f(x) \qquad \text{for a.e. } x \in \mathbb{R}. \tag{11}$$

Note that we are not allowed to integrate $\hat{f}(\xi)e^{2\pi i x\xi}$ over the real line immediately, as $\hat{f}$ is an L^2 function and might not be in $\mathrm{L}^1(\mathbb{R})$. This statement turned out to be extremely difficult to prove and is today known as the *Carleson theorem* [5]. The reader might have heard about this problem for the Fourier series on the torus $\mathbb{T}$, but these two are essentially equivalent via the so-called transference principle between $\mathbb{R}$ and $\mathbb{T}$. The pointwise convergence is clear on a dense subclass of functions, say when f is a Schwartz function, so it is enough to bound the maximal partial integral:

$$\Big\| \sup_{R>0} \Big| \int_{-R}^{R} \hat{f}(\xi) e^{2\pi i x\xi} d\xi \Big| \Big\|_{\mathrm{L}^2_x(\mathbb{R})} \lesssim \|f\|_{\mathrm{L}^2(\mathbb{R})}.$$

Even the weak L^2 norm on the left hand side would suffice, but we will not introduce these.

One of the previous exercises was to present the partial Fourier integrals

$$(S_R f)(x) := \int_{-R}^{R} \hat{f}(\xi) e^{2\pi i x\xi} d\xi$$

as

$$S_R = \tfrac{i}{2}\big(\mathrm{M}_{-R} H \mathrm{M}_R - \mathrm{M}_R H \mathrm{M}_{-R}\big),$$

where H is the Hilbert transform. Therefore we actually need to bound the maximally modulated Hilbert transform, named the *Carleson operator*,

$$(Cf)(x) := \sup_{R\in\mathbb{R}} \Big| \mathrm{p.v.} \int_{\mathbb{R}} f(x-t) e^{iRt} \frac{dt}{t} \Big|.$$

What Carleson actually did in 1966 was that he proved the weak L^2 bound for C, but even more is true.

Theorem 16 (The Carleson-Hunt theorem, [5],[16])**.** *The estimate*

$$\|Cf\|_{\mathrm{L}^p(\mathbb{R})} \lesssim_p \|f\|_{\mathrm{L}^p(\mathbb{R})}$$

holds for any $1 < p < \infty$*.*

This theorem actually implies (11) for any $f \in \mathrm{L}^p(\mathbb{R})$, $1 < p < \infty$. Alternatively, we can integrate only over $\mathbb{T} \equiv [-\frac{1}{2}, \frac{1}{2})$.

* * *

The proof of boundedness of the Carleson operator also uses time-frequency analysis, quite similarly as it was done for the BHT, but it is too lengthy for this course. We will rather explain a beautiful observation of C. Demeter and C. Thiele how it can actually be encoded into a single multilinear singular integral form! Then in Section 2.6 we will prove a boundedness result on a dyadic version of that form, using much simpler time-frequency arguments.

The first step is to use "the worst choice function" $N\colon \mathbb{R} \to \mathbb{R}$, which realizes the supremum in Cf. Therefore we actually need bounds for the linearized operator

$$(C_N f)(x) := \text{p.v.} \int_{\mathbb{T}} f(x-t) e^{iN(x)t} \frac{dt}{t} \tag{12}$$

independently of the function N. Note that N is only measurable.

Demeter and Thiele [9] studied two-dimensional variants of the bilinear Hilbert transform and one of its instances is

$$\Lambda(F_1, F_2, F_3) := \int_{\mathbb{T}^4} F_1(x-s, y-t) F_2(x-t, y) F_3(x, y) K(s,t) dx dy ds dt,$$

where K denotes a 2D Calderón-Zygmund kernel. For instance, one can take

$$K(s,t) = \sum_{k=0}^{\infty} \varphi_k(s) \psi_k(t),$$

where $\varphi_k(s) = 2^k \varphi(2^k s)$, $\psi_k(t) = 2^k \psi(2^k t)$, φ is a C^∞ positive function supported in $[-\frac{1}{4}, \frac{1}{4}]$ with integral 1, and ψ is a C^∞ function such that $\sum_{k=0}^{\infty} \psi_k(t) = \frac{1}{t}$ for $t \in [-\frac{1}{4}, \frac{1}{4}] \setminus \{0\}$. Let us also take F_1, F_2, F_3 to be of the form

$$F_1(x,y) = f(y), \quad F_2(x,y) = e^{-ixN(y)} g(y), \quad F_3(x,y) = e^{ixN(y)} h(y)$$

for some one-dimensional functions f, g, h. After this substitution $\Lambda(F_1, F_2, F_3)$ becomes

$$\begin{aligned} &\int_{\mathbb{T}^4} f(y-t) e^{-i(x-t)N(y)} g(y) e^{ixN(y)} h(y) \frac{1}{t} dx dy ds dt \\ &= \int_{\mathbb{T}^2} f(y-t) e^{itN(y)} g(y) h(y) dt dy = \int_{\mathbb{T}} (C_N f)\, g\, h. \end{aligned}$$

If we have an estimate for Λ,

$$|\Lambda(F_1, F_2, F_3)| \lesssim_{p,q,r} \|F_1\|_{\mathrm{L}^p(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^q(\mathbb{R}^2)} \|F_3\|_{\mathrm{L}^r(\mathbb{R}^2)} \tag{13}$$

for some triple of exponents $1 < p, q, r < \infty$, $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, then it immediately implies

$$\Big| \int_{\mathbb{R}} (C_N f)\, g\, h \Big| \lesssim_{p,q,r} \|f\|_{\mathrm{L}^p(\mathbb{R})} \|g\|_{\mathrm{L}^q(\mathbb{R})} \|h\|_{\mathrm{L}^r(\mathbb{R})},$$

i.e. Theorem 16 holds for such p. Actually, Demeter and Thiele showed (13) under an additional constraint $p, q, r > 2$, which consequently implies (11) for $2 < p < \infty$ and thus only slightly misses the L^2 case. Actually, we will prove bound (13) with $p = 2$, $q = r = 4$ in the next section, but for a dyadic model of Λ, which does not quite imply the Carleson theorem, but is still interesting.

2.6 Yet another trilinear form

This section attempts to incorporate one rather powerful technique, known as the *method of Bellman functions*, into the time-frequency analysis. The proofs obtained this way are often conceptually simple and elegant. A slight disadvantage is that the technique primarily works in the Cantor group models and then one additionally has to transfer the result to the Euclidean case, which sometimes might not work. As a rule of thumb, the proofs using wavelet systems alone can be transferred easily, while for the ones using wave packets we need to rewrite the proof from scratch. However, very often the intuition is gained by considering these "toy models". The material in this section is mostly taken from [18].

We turn back to the trilinear form related to the Carleson operator from the previous chapter. In the dyadic model the kernel K will be replaced by

$$K_{\mathrm{W}}(s,t) := \sum_{k=0}^{\infty} 2^{2k} \mathbf{1}_{[0,2^{-k})}(s) \tilde{\mathbf{h}}_{[0,2^{-k})}(t).$$

Here $\mathbf{1}_I$ denotes the characteristic function of an interval I, while $\tilde{\mathbf{h}}_I$ is the L^∞ normalized dyadic Haar wavelet. Demeter considers the following trilinear form in [8]:

$$\Lambda(F_1, F_2, F_3) := \sum_{k=0}^{\infty} 2^{2k} \int_{[0,1)^4} F_1(x \oplus s, y \oplus t)\, F_2(x \oplus t, y)\, F_3(x, y) \\ \mathbf{1}_{[0,2^{-k})}(s)\, \tilde{\mathbf{h}}_{[0,2^{-k})}(t)\, dx dy ds dt.$$

Here we prove a single L^p estimate for Λ.

Theorem 17. $|\Lambda(F_1,F_2,F_3)| \lesssim \|F_1\|_{\mathrm{L}^2([0,1)^2)}\|F_2\|_{\mathrm{L}^4([0,1)^2)}\|F_3\|_{\mathrm{L}^4([0,1)^2)}$.

We perform the time-frequency decomposition of the form. Let $\mathcal{D}_+$ denote the family of dyadic intervals in $[0,+\infty)$. For each $I \in \mathcal{D}_+$ its left half will be denoted I_0 and its right half will be denoted I_1. A *dyadic step function* on $\mathbb{R}$ will simply be a finite linear combination of characteristic functions of intervals from $\mathcal{D}_+$. Similarly, a dyadic step function on $\mathbb{R}^2$ means a finite linear combination of characteristic functions of dyadic rectangles, i.e. rectangles with sides in $\mathcal{D}_+$.

Let us introduce the notation

$$[f(x)]_{x,I\times\Omega} = [f(x)]_{x,I,n} := \frac{1}{|I|}\int_I f(x)\tilde{w}_{I,n}(x)dx$$

for a locally integrable function f and for

$$I = \big[2^{-k}\ell, 2^{-k}(\ell+1)\big), \quad \Omega = \big[2^k n, 2^k(n+1)\big),$$

$k,\ell,n \in \mathbb{Z},\ \ell,n \geq 0$. In particular, $[f(x)]_{x,I,0}$ is the ordinary average of f on I. We want to emphasize notationally in which variable we are averaging, because we will be dealing with expressions in several variables.

Fix three real-valued dyadic step functions F_1,F_2,F_3 supported on $[0,1)^2$. Take a positive integer M that is large enough so that F_1,F_2,F_3 are constant on dyadic squares with sides of length 2^{-M}. No estimates will depend on M and we only use it to keep the arguments finite. Finally, since the estimate we are proving is homogenous, we can normalize the functions by

$$\|F_1\|_{\mathrm{L}^2} = 1, \quad \|F_2\|_{\mathrm{L}^4} = \|F_3\|_{\mathrm{L}^4} = 1.$$

We begin to decompose Λ by breaking the integrals over $[0,1)$ in x and y into integrals over dyadic intervals of length 2^{-k}. Then we change $\mathbf{h}_{[0,2^{-k})}$ to $\tilde{w}_{[0,2^{-k}),1}$:

$$\Lambda(F_1,F_2,F_3) = \sum_{k=0}^{\infty} 2^{2k} \sum_{\substack{I,J\in\mathcal{D}_+,\ I,J\subseteq[0,1)\\ |I|=|J|=2^{-k}}} \int_{\mathbb{R}^4} F_1(x\oplus s, y\oplus t)\, F_2(x\oplus t, y)\, F_3(x,y)$$
$$\mathbf{1}_I(x)\mathbf{1}_J(y)\mathbf{1}_{[0,2^{-k})}(s)\tilde{w}_{[0,2^{-k}),1}(t)\, dxdydsdt.$$

We substitute $x_1 = x\oplus s,\ x_2 = x\oplus t,\ x_3 = x$, so that $s = x_1\oplus x_3$ and $t = x_2\oplus x_3$. Note that translation $s\mapsto x\oplus s$ preserves the Lebesgue measure and that it also preserves dyadic intervals of length 2^{-k}. Using these facts we get

$$\Lambda(F_1,F_2,F_3) = \sum_{k=0}^{M-1} 2^{2k} \sum_{\substack{I,J\in\mathcal{D}_+,\ I,J\subseteq[0,1)\\ |I|=|J|=2^{-k}}} \int_{\mathbb{R}^4} F_1(x_1, y\oplus x_2\oplus x_3)F_2(x_2,y)F_3(x_3,y)$$
$$\tilde{w}_{I,0}(x_1)\tilde{w}_{I,1}(x_2)\tilde{w}_{I,1}(x_3)\tilde{w}_{J,0}(y)\, dx_1dx_2dx_3dy.$$

Decomposing into the Walsh-Fourier series in the second variable we get

$$F_1(x_1, y \oplus x_2 \oplus x_3) = \sum_{n=0}^{2^M|J|-1} [F_1(x_1,y_1)]_{y_1,J,n}\, \tilde{w}_{J,n}(y \oplus x_2 \oplus x_3)$$

for $x_1, x_2, x_3 \in I$, $y \in J$. Here we recall that $[F_1(x_1,y_1)]_{y_1,J,n}$ vanishes when $n \geq 2^M|J|$, because of our assumptions on F_1, F_2, F_3. The character properties give

$$\begin{aligned}&\tilde{w}_{J,n}(y \oplus x_2 \oplus x_3)\tilde{w}_{I,0}(x_1)\tilde{w}_{I,1}(x_2)\tilde{w}_{I,1}(x_3)\tilde{w}_{J,0}(y)\\ &= \tilde{w}_{I,0}(x_1)\tilde{w}_{I,n\oplus 1}(x_2)\tilde{w}_{I,n\oplus 1}(x_3)\tilde{w}_{J,n}(y),\end{aligned}$$

and since $|I \times J| = 2^{-2k}$, the form finally becomes

$$\begin{aligned}\Lambda(F_1,F_2,F_3) = \sum_{\substack{I,J\in\mathcal{D}_+\\ I,J\subseteq[0,1)\\ |I|=|J|\geq 2^{-M+1}}} \sum_{0\leq n<2^M|I|} &|I\times J| \big[\,[F_1(x_1,y_1)]_{x_1,I,0}\big]_{y_1,J,n}\\ &\big[\,[F_2(x_2,y)]_{x_2,I,n\oplus 1}\,[F_3(x_3,y)]_{x_3,I,n\oplus 1}\big]_{y,J,n}. \quad (14)\end{aligned}$$

* * *

Let us now construct a time-frequency Bellman function. A rectangular cuboid $I \times J \times \Omega$ formed by $I, J, \Omega \in \mathcal{D}_+$ will be called a *tile* if $|I| = |J| = |\Omega|^{-1}$ and a *multitile* if $|I| = |J| = 2|\Omega|^{-1}$. We will only be interested in tiles and multitiles contained in $[0,1)^2 \times [0,2^M)$. Therefore we define the collections:

$$\begin{aligned}&\mathcal{T}_k := \big\{ I\times J\times\Omega \ \text{ is a tile } : \ I,J\subseteq[0,1),\ \Omega\subseteq[0,2^M),\ |I|=2^{-k}\big\},\\ &\mathcal{T} := \textstyle\bigcup_{k=0}^{M}\mathcal{T}_k,\\ &\mathcal{M}_k := \big\{ I\times J\times\Omega \ \text{ is a multitile } : \ I,J\subseteq[0,1),\ \Omega\subseteq[0,2^M),\ |I|=2^{-k}\big\},\\ &\mathcal{M} := \textstyle\bigcup_{k=0}^{M-1}\mathcal{M}_k.\end{aligned}$$

A multitile $P = I \times J \times \Omega$ can be divided "horizontally" into four tiles

$$P_{0,0} := I_0\times J_0\times\Omega,\quad P_{0,1} := I_0\times J_1\times\Omega,\quad P_{1,0} := I_1\times J_0\times\Omega,\quad P_{1,1} := I_1\times J_1\times\Omega$$

and "vertically" into two tiles

$$P^0 := I\times J\times\Omega_0,\quad P^1 := I\times J\times\Omega_1.$$

For any function on tiles $\mathcal{B}\colon \mathcal{T}\to\mathbb{R}$ we define its *first-order difference* as a function on multitiles $\Box\mathcal{B}\colon \mathcal{M}\to\mathbb{R}$ given by the formula

$$(\Box\mathcal{B})(P) := \frac{1}{4}\sum_{\alpha,\beta\in\{0,1\}}\mathcal{B}(P_{\alpha,\beta}) - \sum_{\gamma\in\{0,1\}}\mathcal{B}(P^{\gamma}).$$

Note that (14) can now be rewritten as

$$\Lambda = \sum_{I\times J\times\Omega\in\mathcal{M}} |I\times J|\ \mathcal{A}(I\times J\times\Omega), \tag{15}$$

where $\mathcal{A}\colon \mathcal{M}\to\mathbb{R}$ is given by

$$\mathcal{A}(I\times J\times\Omega) := \sum_{\gamma\in\{0,1\}} \big[\, [F_1(x,y)]_{x,I,0}\,\big]_{y,J\times\Omega_\gamma} \big[\, [F_2(x,y)]_{x,I\times\Omega_{\gamma\oplus 1}}\, [F_3(x,y)]_{x,I\times\Omega_{\gamma\oplus 1}}\,\big]_{y,J\times\Omega_\gamma}.$$

This follows from the fact that n and $n\oplus 1$ are two consecutive nonnegative integers and the smaller one is even, so the union of

$$I\times J\times\big[\tfrac{n}{|I|},\tfrac{n+1}{|I|}\big) \quad\text{and}\quad I\times J\times\big[\tfrac{n\oplus 1}{|I|},\tfrac{(n\oplus 1)+1}{|I|}\big)$$

constitutes a multitile from $\mathcal{M}$.

We will be writing $I\times J\times\Omega$ for a generic multitile. In order to control $\mathcal{A}$ we need to introduce several relevant expressions. Define $\mathcal{B}_i\colon \mathcal{T}\to\mathbb{R},\ i=1,2,3,4,5$ by

$$\begin{aligned}
\mathcal{B}_1(I\times J\times\Omega) &:= \big[\, [F_1(x,y)]_{x,I,0}\,\big]_{y,J\times\Omega}\big[\, [F_2(x,y)]_{x,I\times\Omega}\, [F_3(x,y)]_{x,I\times\Omega}\,\big]_{y,J\times\Omega},\\
\mathcal{B}_2(I\times J\times\Omega) &:= \big[\, [F_1(x,y)]_{x,I,0}\,\big]^2_{y,J\times\Omega},\\
\mathcal{B}_3(I\times J\times\Omega) &:= \big[\, [F_2(x,y)]_{x,I\times\Omega}\, [F_3(x,y)]_{x,I\times\Omega}\,\big]^2_{y,J\times\Omega},\\
\mathcal{B}_4(I\times J\times\Omega) &:= \big[\, [F_2(x,y)]^2_{x,I\times\Omega}\,\big]^2_{y,J,0},\\
\mathcal{B}_5(I\times J\times\Omega) &:= \big[\, [F_3(x,y)]^2_{x,I\times\Omega}\,\big]^2_{y,J,0}.
\end{aligned}$$

We remark that the scope of each averaging variable (i.e. x or y) is only inside the corresponding bracket. Finally, denote

$$\mathcal{B}_- := \mathcal{B}_1 - \mathcal{B}_2 - \tfrac{1}{2}\mathcal{B}_3 - \tfrac{1}{2}\mathcal{B}_4 - \tfrac{1}{2}\mathcal{B}_5 \quad\text{and}\quad \mathcal{B}_+ := \mathcal{B}_1 + \mathcal{B}_2 + \tfrac{1}{2}\mathcal{B}_3 + \tfrac{1}{2}\mathcal{B}_4 + \tfrac{1}{2}\mathcal{B}_5.$$

The key local estimate is contained in the following lemma. We omit the proof as it is purely computational and the reader can find the details in [18].

Lemma 18. *For every $I \times J \times \Omega \in \mathcal{M}$ we have*

$$\Box\mathcal{B}_-(I\times J\times\Omega) \;\leq\; \mathcal{A}(I\times J\times\Omega) \;\leq\; \Box\mathcal{B}_+(I\times J\times\Omega).$$

Now we complete the proof of Theorem 17. We introduce two auxiliary expressions

$$\Xi_k^+ := \sum_{I\times J\times\Omega\in\mathcal{T}_k} |I\times J|\;\mathcal{B}_+(I\times J\times\Omega), \quad \Xi_k^- := \sum_{I\times J\times\Omega\in\mathcal{T}_k} |I\times J|\;\mathcal{B}_-(I\times J\times\Omega)$$

for $k = 0, 1, 2, \ldots, M$. By definition of the operator $\Box$ we have

$$\Xi_{k+1}^\pm - \Xi_k^\pm = \sum_{I\times J\times\Omega\in\mathcal{M}_k} |I\times J|\;(\Box\mathcal{B}_\pm)(I\times J\times\Omega).$$

Summing for $k = 0, 1, \ldots, M-1$, using (15), and applying Lemma 18 we conclude

$$\Xi_M^- - \Xi_0^- \leq \Lambda \leq \Xi_M^+ - \Xi_0^+.$$

However, it is easy to control "single scale quantities" $\Xi_M^\pm$ and $\Xi_0^\pm$,

$$|\Xi_M^\pm| \lesssim 1, \quad |\Xi_0^\pm| \lesssim 1,$$

which proves the estimate

$$|\Lambda(F_1, F_2, F_3)| \lesssim 1$$

and finally establishes Theorem 17.

* * *   * * *   * * *

We tried to present some of the techniques from time-frequency analysis used for bounding singular integrals. After learning the main ideas in the Cantor group "toy model" the reader should have less difficulties going through the proofs in the classical setting, given in the papers [20],[21],[22].

Bibliography

[1] P. Auscher, S. Hofmann, C. Muscalu, T. Tao, C. Thiele, *Carleson measures, trees, extrapolation, and $T(b)$ theorems*, Publications Matematiques Barcelona **46** (2002), 257–325.

[2] W. Beckner, *Inequalities in Fourier analysis*, Ann. of Math. (2) **102** (1975), 159–182.

[3] A. P. Calderón, *Commutators of Singular Integral Operators*, Proc. Nat. Acad. Sci. U.S.A. **53** (1965), 1092–1099.

[4] A. P. Calderón, *Cauchy Integrals on Lipschitz Curves and Related Operators*, Proc. Nat. Acad. Sci. U.S.A. **74** (1977), 1324–1327.

[5] L. Carleson, *On Convergence and Growth of Partial Sums of Fourier Series*, Acta Math. **116** (1966), 135–157.

[6] R. R. Coifman, P. W. Jones, S. Semmes, *Two elementary proofs of the* L^2 *boundedness of Cauchy integrals on Lipschitz curves*, Journal of the AMS **2** (1989), 553–564.

[7] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Series, Appl. Math., Society for Industrial and Applied Mathematics, 1992.

[8] C. Demeter, *A guide to Carleson's Theorem*, lecture notes.

[9] C. Demeter, C. M. Thiele, *On the two-dimensional bilinear Hilbert transform*, Amer. J. Math. **132** (2010), 201–256.

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

[11] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, John Wiley and Sons, 1984.

[12] G. B. Folland, *Harmonic Analysis in Phase Space*, Princeton University Press, 1989.

[13] G. B. Folland, *A Course in Abstract Harmonic Analysis*, Studies in Advanced Mathematics, CRC Press, 1995.

[14] C. Heil, J. Ramanathan, P. Topiwala, *Linear independence of time-frequency translates*, Proc. Amer. Math. Soc. **124** (1996), 2787–2795.

[15] E. Hernández, G. Weiss, *A First Course on Wavelets*, Studies in Advanced Mathematics, CRC Press, 1996.

[16] R. A. Hunt, *On the Convergence of Fourier Series*, Orthogonal Expansions and their Continuous Analogues, Proc. Conf. Edwardsville (1967), 235–255.

[17] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam. **28** (2012), 1143–1164.

[18] V. Kovač, *On a trilinear form related to the Carleson theorem*, J. Math. Anal. Appl. 405 (2013), no. 1, 220–226.

[19] M. Lacey, X. Li, *On a conjecture of E. M. Stein on the Hilbert transform on vector fields*, Mem. Amer. Math. Soc. **205** (2010).

[20] M. Lacey, C. Thiele, *L^p Estimates on the Bilinear Hilbert Transform for* $2 < p < \infty$, Ann. of Math. (2) **146** (1997), 693–724.

[21] M. Lacey, C. Thiele, *On Calderón's conjecture*, Ann. of Math. (2) **149** (1999), 475–496.

[22] M. Lacey, C. Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. **7** (2000), 361–370.

[23] V. Lie, *The (weak-L^2) boundedness of the quadratic Carleson operator*, Geom. Funct. Anal. **19** (2009), 457–497.

[24] V. Lie, *The Polynomial Carleson Operator*, preprint.

[25] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, 1992.

[26] C. Muscalu, J. Pipher, T. Tao, C. Thiele, Bi-parameter paraproducts, Acta Math. **193** (2004), 269–296.

[27] M. C. Pereyra, *Lecture Notes on Dyadic Harmonic Analysis*, Contemporary Mathematics, 289, AMS, 2001, pp. 1–61.

[28] E. M. Stein, G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, 1971.

[29] T. Tao, *Harmonic analysis in the phase plane*, lecture notes, 2001.

[30] C. Thiele, *Time-frequency Analysis in the Discrete Phase Plane*, Ph.D. Dissertation, Yale University, 1995.

[31] C. Thiele, *Wave Packet Analysis*, CBMS Regional Conference Series in Mathematics, AMS, 2006.

[32] S. I. Valdimarsson, *A multilinear generalisation of the Hilbert transform and fractional integration*, Rev. Mat. Iberoam. **28** (2012), 25–55.

[33] J. L. Walsh, *A closed set of normal orthogonal functions*, Amer. J. Math. **45** (1923), 5–24.