

Graduate Harmonic Analysis

Spring 2026

Notes by M. Iliopoulou

Contents

Chapter 1. The Fourier transform: Where can we define it?	5
1.1. The Fourier transform on L^p , for $p \in (1, 2)$ - the Hausdorff-Young inequality	6
1.2. What happens for $p > 2$? Negative results via Khinchin's inequality	8
Chapter 2. An interlude: The inverse Fourier transform	14
Chapter 3. Do partial Fourier integrals converge?	16
3.1. An initial reduction	16
3.2. Pointwise convergence	17
3.3. L^p -convergence	18
3.4. Fourier multipliers: the basics	20
3.5. Back to S_1 : Fefferman's ball multiplier theorem	22
3.6. The uncertainty principle	23
3.7. The proof of Fefferman's ball multiplier theorem	34
Chapter 4. The Kakeya conjectures	53
4.1. The Kakeya set conjecture	53
4.2. The Kakeya maximal operator conjecture	57
4.3. Kakeya max $\implies$ Kakeya	59
4.4. The Kakeya maximal operator conjecture holds in $\mathbb{R}^2$.	60
4.5. The Kakeya problem in finite fields	62
4.6. The polynomial method: the basic idea	65
4.7. The proof of the finite field Kakeya conjecture	65
Chapter 5. The Fourier restriction conjecture	73
5.1. The Fourier restriction problem: the basics	73
5.2. A dual formulation via the Fourier extension operator	76
5.3. The Fourier extension operator in PDE	82
5.4. The Fourier restriction conjecture	83
5.5. Restriction implies Kakeya	91
5.6. The Stein-Tomas restriction theorem	95
5.7. The Fourier restriction conjecture holds in $\mathbb{R}^2$.	103
5.8. The wave packet decomposition	111
5.9. The polynomial method in Fourier restriction theory: a motivation	115
Chapter 6. Polynomial partitioning	117
6.1. The advantage	117
6.2. Statement and proof	117
6.3. The numerology of the cellular case	121
6.4. Dealing with the cellular case via induction	125
6.5. The joints problem	128
6.6. Counting joints using Dvir's vanishing lemma (Quilodrán's proof)	129
6.7. Counting joints using polynomial partitioning	132
6.8. Counting joints with multiplicities in $\mathbb{R}^3$ (using polynomial partitioning)	135

CHAPTER 1

The Fourier transform: Where can we define it?

Let $n \in \mathbb{N}$. We know that the spaces $L^p(\mathbb{R}^n)$, for $1 \leq p < +\infty$, are all distinct. However, *all* of them contain the space of Schwartz functions $\mathcal{S}(\mathbb{R}^n)$, which is dense in every one of these spaces $(L^p(\mathbb{R}^n), \|\cdot\|_p)$. Recall that

$\mathcal{S}(\mathbb{R}^n) := \{f \in C^\infty(\mathbb{R}^n) : \text{the function } x^{\vec{k}} D^{\vec{m}} f(x) : \mathbb{R}^n \rightarrow \mathbb{C} \text{ is bounded, for all } \vec{k}, \vec{m} \in \mathbb{Z}_{\geq 0}^n\}$.

The Fourier transform on $L^1(\mathbb{R}^n)$. The Fourier transform $\hat{\cdot}$ is first defined on $L^1(\mathbb{R}^n)$ as follows: For every $f \in L^1(\mathbb{R}^n)$, we define $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{C}$ by

$$(1) \quad \hat{f}(\xi) := \int e^{-2\pi i \langle \xi, x \rangle} f(x) dx, \quad \text{for all } x \in \mathbb{R}^n.$$

Observe that the above integral makes sense exactly because $f \in L^1(\mathbb{R}^n)$. Indeed, the integral exists (in $\mathbb{C}$) as the function $e^{-2\pi i \langle \xi, \cdot \rangle} f$ is (Lebesgue-)measurable and $\int |e^{-2\pi i \langle \xi, \cdot \rangle} f| = \int |f| < +\infty$. Therefore, $L^1(\mathbb{R}^n)$ is the correct space for the Fourier transform to be defined on pointwise.

It is not too hard to show that, for every $f \in L^1(\mathbb{R}^n)$, it holds that $\hat{f} \in C_0(\mathbb{R}^n)$; that is, $\hat{f}$ is continuous, and $\hat{f}(\xi) \rightarrow 0$ as $|\xi| \rightarrow +\infty$. There is no guarantee, however, that $\hat{f} \in L^1(\mathbb{R}^n)$. The simpler fact that $\hat{f} \in L^\infty(\mathbb{R}^n)$ is trivial, as

$$(2) \quad \|\hat{f}\|_\infty \leq \|f\|_1 \quad \text{for every } f \in L^1(\mathbb{R}^n),$$

by the triangle inequality: for every $\xi \in \mathbb{R}^n$,

$$|\hat{f}(\xi)| = \left| \int e^{-2\pi i \langle \xi, x \rangle} f(x) dx \right| \leq \int |f(x)| dx = \|f\|_1.$$

Note that (2) means that $\hat{\cdot} : L^1(\mathbb{R}^n) \rightarrow L^\infty(\mathbb{R}^n)$ is a bounded linear operator.

A great property of the Fourier transform is that we in general can recover f from $\hat{f}$ via the Fourier inversion formula: for every $f \in L^1(\mathbb{R}^n)$ such that $\hat{f} \in L^1$,

$$f(x) = \hat{\hat{f}}(-x) = \int e^{2\pi i \langle x, \xi \rangle} \hat{f}(\xi) d\xi, \quad \text{for a.e. } x \in \mathbb{R}^n.$$

Specifically, the inversion formula above holds at every Lebesgue point x of f , and therefore at every $x \in \mathbb{R}^n$ where f is continuous. In particular $f \in \mathcal{S}(\mathbb{R}^n)$, then the inversion formula holds at every $x \in \mathbb{R}^n$.

The inversion formula is so useful that it is worth wondering whether we can extend the definition of the Fourier transform to other spaces (different from $L^1(\mathbb{R}^n)$), in a way that still allows us to recover f from $\hat{f}$. This will gradually lead us to the discussion of the Hausdorff-Young inequality, Fefferman's ball multiplier theorem and Bochner-Riesz means.

The Fourier transform on $L^2(\mathbb{R}^n)$. For general $f \in L^2(\mathbb{R}^n) \setminus L^1(\mathbb{R}^n)$, the integral in (1) *does not make sense*. However, the Fourier transform has already been defined on the

space $\mathcal{S}(\mathbb{R}^n)$ of Schwartz functions, which is dense in $(L^2(\mathbb{R}^n), \|\cdot\|_2)$. Using this density, we will be able to extend its definition to the whole of $L^2(\mathbb{R}^n)$. The useful fact here is that

$$\widehat{\cdot} : (\mathcal{S}(\mathbb{R}^n), \|\cdot\|_2) \rightarrow (\mathcal{S}(\mathbb{R}^n), \|\cdot\|_2),$$

is an *isometry*¹, i.e.

$$\|\widehat{f}\|_2 = \|f\|_2 \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n).$$

Thanks to this, we will manage to linearly extend the Fourier transform from $\mathcal{S}(\mathbb{R}^n)$ to the bigger space $L^2(\mathbb{R}^n)$, so that it remains an isometry, i.e. so that

$$(3) \quad \|\widehat{f}\|_2 = \|f\|_2 \quad \text{for all } f \in L^2(\mathbb{R}^n)$$

(known as Plancherel's theorem).

[In fact, we will have that $\widehat{\cdot} : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ is a 1-1, onto, bounded linear operator (an isometry), and that the inversion formula still holds: for every $f \in L^2(\mathbb{R}^n)$,

$$f(x) = \widehat{\widehat{f}}(-x) \quad \text{for a.e. } x \in \mathbb{R}^n.$$

This time however it is not obtained via an integral, but via density.]

—> So, how do we extend $\widehat{\cdot}$ from $\mathcal{S}(\mathbb{R}^n)$ to the whole of $L^2(\mathbb{R}^n)$?

Let $f \in L^2(\mathbb{R}^n)$. There exists a sequence $(g_k)_{k \in \mathbb{N}}$ of Schwartz functions with

$$\|g_k - f\|_2 \rightarrow 0 \quad \text{as } k \rightarrow +\infty \quad (\text{density}).$$

This implies that $(g_k)_{k \in \mathbb{N}}$ is a Cauchy sequence in $(L^2(\mathbb{R}^n), \|\cdot\|_2)$, and by Plancherel's theorem (3) $(\widehat{g_k})_{k \in \mathbb{N}}$ is also Cauchy in $(L^2(\mathbb{R}^n), \|\cdot\|_2)$.

Since $(L^2(\mathbb{R}^n), \|\cdot\|_2)$ is a Banach space, it follows that $(\widehat{g_k})_{k \in \mathbb{N}}$ converges in $(L^2(\mathbb{R}^n), \|\cdot\|_2)$; that is, there exists some $\widetilde{f} \in L^2(\mathbb{R}^n)$ with

$$\|\widehat{g_k} - \widetilde{f}\|_2 \rightarrow 0 \quad \text{as } k \rightarrow +\infty.$$

We define $\widehat{f} := \widetilde{f}$. Note that this is a function defined a.e.

It is easy to see that $\widehat{f}$ is well-defined, in the sense that using a different sequence of Schwartz functions converging to f will not make any difference, and that for $\widehat{f} \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ (a dense subspace of $L^2(\mathbb{R}^n)$, the two so far definitions agree.

The following question is therefore natural.

Q1: On which other L^p spaces can we define the Fourier transform?

1.1. The Fourier transform on L^p , for $p \in (1, 2)$ - the Hausdorff-Young inequality

It is easy to see that, for any $p \in (1, 2)$, $L^p \subseteq L^1 + L^2$. Indeed, for every $f \in L^p$, we can write

$$f = f \cdot \chi_{\{|f| \geq 1\}} + f \cdot \chi_{\{|f| < 1\}},$$

¹It is also onto $\mathcal{S}(\mathbb{R}^n)$.

where $f \cdot \chi_{\{|f| \geq 1\}} \in L^1$, as

$$\int |f \cdot \chi_{\{|f| \geq 1\}}| = \int_{\{|f| \geq 1\}} |f| \stackrel{p>1}{\leq} \int_{\{|f| \geq 1\}} |f|^p \leq \int |f|^p < +\infty,$$

and $f \cdot \chi_{\{|f| < 1\}} \in L^2$, as

$$\int |f \cdot \chi_{\{|f| < 1\}}|^2 = \int_{\{|f| < 1\}} |f|^2 = \int_{\{|f| < 1\}} |f|^p |f|^{2-p} \stackrel{p<2}{\leq} \int_{\{|f| < 1\}} |f|^p \leq \int |f|^p < +\infty.$$

Therefore, the Fourier transform can be defined on L^p , as follows.

DEFINITION 1.1.1. *Let $p \in (1, 2)$. For every $f \in L^p$, we define*

$$\widehat{f} := \widehat{f}_1 + \widehat{f}_2,$$

where $f_1 := f \cdot \chi_{\{|f| \geq 1\}} \in L^1$ and $f_2 := f \cdot \chi_{\{|f| < 1\}} \in L^2$.

Now it makes sense to ask what properties the Fourier transform of an L^p function has - for instance, does it live in some L^q , for appropriate q ? Even better, is the Fourier transform a bounded operator from L^p to some L^q ? The answer is yes, and is given by the Hausdorff-Young inequality.

THEOREM 1.1.2. (The Hausdorff-Young inequality) *For every $p \in [1, 2]$, it holds that*

$$(4) \quad \|\widehat{f}\|_{p'} \leq \|f\|_p \quad \text{for all } f \in L^p.$$

Note that we know (4) for $p = 1$ and $p = 2$. The statement for intermediate p follows immediately by the Riesz-Thorin interpolation theorem, which we state without proof.

THEOREM 1.1.3. (Riesz-Thorin interpolation) *Let $1 \leq p_1 < p_2 \leq +\infty$, and T be a linear operator defined on $L^{p_1} + L^{p_2}$, satisfying, for some $q_1, q_2 \in [1, +\infty]$, that*

$$T : L^{p_1} \rightarrow L^{q_1} \text{ is bounded, with } \|Tf\|_{q_1} \leq A_1 \|f\|_{p_1} \text{ for all } f \in L^{p_1},$$

and

$$T : L^{p_2} \rightarrow L^{q_2} \text{ is bounded, with } \|Tf\|_{q_2} \leq A_2 \|f\|_{p_2} \text{ for all } f \in L^{p_2}.$$

For any $p \in [p_1, p_2]$, write

$$\frac{1}{p} = \frac{\theta}{p_1} + \frac{1-\theta}{p_2}$$

(for the unique $\theta \in [0, 1]$ that works). Then, for the q given by

$$\frac{1}{q} = \frac{\theta}{q_1} + \frac{1-\theta}{q_2},$$

T is bounded from L^p to L^q , and in particular

$$\|Tf\|_q \leq A_1^\theta A_2^{1-\theta} \|f\|_p \text{ for all } f \in L^p.$$

Proof of Theorem 1.1.2. We know that the linear operator $\widehat{\cdot}$ satisfies

$$\|\widehat{f}\|_\infty \leq \|f\|_1 \quad \text{for all } f \in L^1$$

and

$$\|\widehat{f}\|_2 = \|f\|_2 \quad \text{for all } f \in L^2.$$

Therefore, we can apply the Riesz-Thorin interpolation theorem for $(p_1, q_1) = (1, \infty)$, $(p_2, q_2) = (2, 2)$ and $A_1 = A_2 = 1$. In particular, let $p \in (1, 2)$. There exists (unique) $\theta \in (0, 1)$ for which

$$\frac{1}{p} = \frac{\theta}{1} + \frac{1-\theta}{2} = \frac{1}{2} + \frac{\theta}{2},$$

specifically $\frac{\theta}{2} = \frac{1}{p} - \frac{1}{2}$. For this θ , define q via

$$\frac{1}{q} = \frac{\theta}{\infty} + \frac{1-\theta}{2} = \frac{1}{2} - \frac{\theta}{2} = \frac{1}{2} - \left(\frac{1}{p} - \frac{1}{2}\right) = 1 - \frac{1}{p} = \frac{1}{p'},$$

that is, $q = p'$. By the Riesz-Thorin interpolation theorem,

$$\|\hat{f}\|_{p'} \leq 1^\theta \cdot 1^{1-\theta} \|f\|_p = \|f\|_p \quad \text{for all } f \in L^p.$$

□

1.2. What happens for $p > 2$? Negative results via Khinchin's inequality

Given any $p \in [1, +\infty)$, an idea that would help us linearly extend the Fourier transform from $\mathcal{S}(\mathbb{R}^n)$ to the whole of L^p is the following. (Note that, by the Hausdorff-Young inequality, the idea works for $p \in [1, 2]$. However, we write it down for general $p \in [1, +\infty)$, as the strategy makes sense independently of the range of p .)

Idea: We will definitely be able to linearly extend the Fourier transform from $\mathcal{S}(\mathbb{R}^n)$ to L^p , for some $1 \leq p < +\infty$, if the operator

$$\hat{\cdot} : (\mathcal{S}(\mathbb{R}^n), \|\cdot\|_p) \rightarrow (\mathcal{S}(\mathbb{R}^n), \|\cdot\|_q)$$

is bounded, for some $q \in [1, +\infty]$; that is, if there exists $q \in [1, +\infty]$ such that

$$(5) \quad \|\hat{f}\|_q \leq C_{p,q} \|f\|_p \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n)$$

(where $C_{p,q} > 0$ is a constant depending only on p, q and n). [By the Hausdorff-Young inequality, we already know this to be true for $p \in [1, 2]$, in which case $q = p'$ works.]

Note that the specific value of q is not important here. All that matters is that, just like in the case of L^2 , density of $\mathcal{S}(\mathbb{R}^n)$ in L^p combined with (5) allows us to linearly extend the Fourier transform from $\mathcal{S}(\mathbb{R}^n)$ to the whole of L^p . Moreover, (5) will be automatically extended to the whole of L^p : we will have that

$$\|\hat{f}\|_q \leq C_{p,q} \|f\|_p \quad \text{for all } f \in L^p.$$

PROOF. Suppose that (5) holds. Let $f \in L^p$. There exists a sequence $(g_k)_{k \in \mathbb{N}}$ of Schwartz functions with

$$\|g_k - f\|_p \rightarrow 0 \quad \text{as } k \rightarrow +\infty \quad (\text{density}).$$

This implies that $(g_k)_{k \in \mathbb{N}}$ is a Cauchy sequence in $(L^p, \|\cdot\|_p)$, and by (5), $(\hat{g}_k)_{k \in \mathbb{N}}$ is also Cauchy in $(L^q, \|\cdot\|_q)$.

Since $(L^q, \|\cdot\|_q)$ is a Banach space, it follows that $(\hat{g}_k)_{k \in \mathbb{N}}$ converges in $(L^q, \|\cdot\|_q)$; that is, there exists some $\tilde{f} \in L^q$ with

$$\|\hat{g}_k - \tilde{f}\|_q \rightarrow 0 \quad \text{as } k \rightarrow +\infty.$$

We define $\hat{f} := \tilde{f}$ (a function defined a.e.). It is now easy to see that $\hat{\cdot}$ is a well-defined, bounded linear operator from L^p to L^q . Indeed, by the above, for every $f \in L^p$ we have

$$\|\hat{f}\|_q = \lim_{k \rightarrow +\infty} \|\hat{g}_k\|_q \stackrel{(5)}{\leq} C_{p,q} \lim_{k \rightarrow +\infty} \|g_k\|_p = C_{p,q} \|f\|_p.$$

□

Given this idea, we can ask a more precise question than **Q1**:

Q2: For which p does there exist $q \in [1, +\infty]$ such that

$$\|\hat{f}\|_q \leq C_{p,q} \|f\|_p \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n)?$$

We have already answered this question for $p \in [1, 2]$; for each such p , there exists q for which the above holds ($q = p'$). We will show however that, **when $p > 2$, there is no q for which the above holds. So, for $p > 2$, the above strategy of extending the Fourier transform to the whole of L^p does not work** (and in general it is unclear what to do).

In our effort to answer **Q2**, it would be helpful to know, given some $p \in [1, +\infty)$, what are the possible candidates q for which (5) could hold. This is also useful for $p \in [1, 2]$, as it would tell us whether the Fourier transform of the arbitrary L^p function could perhaps be in L^q , for other q than just p' .

It turns out that only $q = p'$ may work.

PROPOSITION 1.2.1. *Let $p \in [1, +\infty)$, and suppose that, for some $q \in [1, +\infty]$ and a constant $C > 0$, we have that*

$$(6) \quad \|\hat{f}\|_q \leq C \|f\|_p \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n).$$

Then, $q = p'$.

PROOF. The proof is a classical scaling argument. Fix some non-zero $f \in \mathcal{S}(\mathbb{R}^n)$. Then, $\|f\|_p \neq 0$ and $\|\hat{f}\|_q \neq 0$.

For all $\lambda > 0$, we define $f_\lambda : \mathbb{R}^n \rightarrow \mathbb{C}$ by

$$f_\lambda(x) := f\left(\frac{x}{\lambda}\right), \quad \text{for all } x \in \mathbb{R}^n.$$

Observe that this simply rescales the domain of f . Moreover, we know that $\hat{f}_\lambda(\xi) = \lambda^n \hat{f}(\lambda\xi)$, for all $\xi \in \mathbb{R}^n$.

Now, we calculate $\|f_\lambda\|_p$ and $\|\hat{f}_\lambda\|_q$. We will then plug them into (6), to see what conditions the freedom of λ imposes on q . Indeed,

$$\|f_\lambda\|_p = \left(\int \left| f\left(\frac{x}{\lambda}\right) \right|^p dx \right)^{1/p} \xrightarrow{\frac{x}{\lambda}=y, dx=\lambda^n dy} \left(\int |f(y)|^p \lambda^n dy \right)^{1/p} = \lambda^{n/p} \|f\|_p$$

and

$$\|\hat{f}_\lambda\|_q = \left(\int |\lambda^n \hat{f}(\lambda\xi)|^q d\xi \right)^{1/q} \xrightarrow{\lambda\xi=u, d\xi=\frac{du}{\lambda^n}} \lambda^n \left(\int |\hat{f}(u)|^q \cdot \frac{du}{\lambda^n} \right)^{1/q} = \lambda^{n-\frac{n}{q}} \|\hat{f}\|_q.$$

Therefore, assuming that (6) holds, it follows that, for this fixed f ,

$$\lambda^{n-\frac{n}{q}} \|\hat{f}\|_q \leq C \cdot \lambda^{n/p} \|f\|_p \quad \text{for all } \lambda > 0,$$

and thus that

$$(7) \quad \lambda^{n-\frac{n}{q}-\frac{n}{p}} \leq C',$$

where $C' > 0$ is a constant (depending on f , but independent of λ).

As $\lambda \rightarrow +\infty$, (7) is possible only if $n - \frac{n}{q} - \frac{n}{p} \leq 0$.

As $\lambda \rightarrow 0$, (7) is possible only if $n - \frac{n}{q} - \frac{n}{p} \geq 0$.

Therefore, for (7) can only hold if $n - \frac{n}{q} - \frac{n}{p} = 0$, i.e. if $\frac{1}{q} + \frac{1}{p} = 1$.

□

Therefore, for $p \in [1, 2]$, the story is complete: the Fourier transform is bounded from L^p to $L^{p'}$ alone. And we can reformulate **Q2** as follows.

Q3: For which $p > 2$ do we have that $\|\hat{f}\|_{p'} \leq C_p \|f\|_p$ for all $f \in \mathcal{S}(\mathbb{R}^n)$?

The answer is: **for none**. So, we have not yet discovered a way to define the Fourier transform on L^p , for $p > 2$.

PROPOSITION 1.2.2. (Boundedness failure for $p > 2$) Let $p > 2$. Then, the Fourier transform

$$\hat{\cdot}: (\mathcal{S}(\mathbb{R}^n), \|\cdot\|_p) \rightarrow (\mathcal{S}(\mathbb{R}^n), \|\cdot\|_{p'})$$

is not a bounded operator. In other words, there exists **no constant** $C > 0$ with the property that

$$\|\hat{f}\|_{p'} \leq C \|f\|_p \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n).$$

We will prove this probabilistically, via Khinchin's inequality (which we now state and discuss, but do not prove).

PROPOSITION 1.2.3. (Khinchin's inequality) Let $N \in \mathbb{N}$ and fix $a_1, a_2, \dots, a_N \in \mathbb{C}$. Then, for all $p > 0$,

$$\mathbb{E} \left(\left| \sum_{n=1}^N \omega_n a_n \right|^p \right) \sim \left(\sum_{n=1}^N |a_n|^2 \right)^{p/2},$$

where the ω_n are independent random variables, each taking the values ± 1 with equal probability.

In the above, $\mathbb{E}$ denotes expected value. In other words,

$$\mathbb{E} \left(\left| \sum_{n=1}^N \omega_n a_n \right|^p \right) = \frac{1}{\#\Omega} \cdot \sum_{\vec{\omega} \in \Omega} \left| \sum_{n=1}^N \omega_n a_n \right|^p,$$

where $\Omega := \{-1, 1\}^n$. The inequality is trivial (it is an equality in fact) when $p = 2$. We see this by expanding out the square and using that $\mathbb{E}(\omega_n) = 0$ for all $n = 1, \dots, N$:

$$\begin{aligned} \mathbb{E} \left(\left| \sum_{n=1}^N \omega_n a_n \right|^2 \right) &= \mathbb{E} \left(\sum_{i,j} \omega_i \omega_j \cdot a_i \overline{a_j} \right) \\ &= \sum_{i,j} a_i \overline{a_j} \cdot \mathbb{E}(\omega_i \omega_j) \\ &= \sum_i |a_i|^2 \cdot \mathbb{E}(\omega_i^2) + \sum_{i \neq j} a_i \overline{a_j} \cdot \mathbb{E}(\omega_i) \mathbb{E}(\omega_j) \\ &= \sum_{i=1}^N |a_i|^2. \end{aligned}$$

The proof for general $p > 0$ is more complicated². In any case, independently of p and the proof's difficulty, Khinchin's inequality tells us that something **better than the triangle inequality and the Cauchy-Schwarz inequality** holds (and, in fact, something like the **Pythagorean Theorem** holds), as long as we are not too particular about our choice of signs ω_n . Indeed, it implies that there exists a choice of $\vec{\omega} \in \{-1, 1\}^n$ for which

$$\left| \sum_{n=1}^N \omega_n a_n \right| \lesssim \left(\sum_{n=1}^N |a_n|^2 \right)^{1/2}.$$

This is not obvious, for the following reasons:

²The proof for the case that interests us, specifically $p \geq 1$, can be found in Wolff's lecture notes.

- For any given $\vec{\omega} \in \{-1, 1\}^n$,

$$\left| \sum_{n=1}^N \omega_n a_n \right| \leq \sum_{n=1}^N |a_n|$$

(with equality being possible, when for instance the $a_n \in \mathbb{R}$ for all $n = 1, \dots, N$),
and the right-hand side is *at least* $\left(\sum_{n=1}^N |a_n|^2 \right)^{1/2}$.

- For any given $\vec{\omega} \in \{-1, 1\}^n$,

$$\left| \sum_{n=1}^N \omega_n a_n \right| \leq \left(\sum_{n=1}^N |a_n|^2 \right)^{1/2} \left(\sum_{n=1}^N |\omega_n|^2 \right)^{1/2} \stackrel{|\omega_n|=1 \text{ for all } n}{=} N^{1/2} \cdot \left(\sum_{n=1}^N |a_n|^2 \right)^{1/2},$$

where again the right-hand side is much larger than $\left(\sum_{n=1}^N |a_n|^2 \right)^{1/2}$.

Proof of Proposition 1.2.2. We first observe that, since the Fourier transform is 1-1 and onto $\mathcal{S}(\mathbb{R}^n)$, any estimate of the form

$$\|\widehat{f}\|_{p'} \leq C \|f\|_p \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n)$$

is equivalent to the estimate

$$\|f\|_{p'} \leq C \|\widehat{f}\|_p \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n).$$

Therefore, since for $p > 2$ we have $p' < 2$, our problem becomes showing that, **for** $1 \leq p < 2$, the estimate

$$\|f\|_p \leq C \|\widehat{f}\|_{p'} \quad \text{for all } f \in \mathcal{S}(\mathbb{R}^n)$$

is **false** (that is, the Hausdorff-Young inequality cannot be reversed).

Therefore, **fix** $1 \leq p < 2$.

Fix some $\phi_1 \in C_c^\infty$, such that

$$\phi_1 \geq 0, \quad \text{supp } \phi_1 \subset B(0, 1), \quad \phi_1 = 1 \text{ on } B(0, 1/2).$$

Let $N \in \mathbb{N}$, and consider translated copies $\phi_2, \dots, \phi_N$ of ϕ_1 , so that the functions $\phi_1, \dots, \phi_N$ have pairwise disjoint supports. In particular, setting $k_1 := 0$, we achieve this by choosing $k_2, \dots, k_N \in \mathbb{R}^n$ so that the balls $B(k_j, 1)$, $j = 1, \dots, N$, are pairwise disjoint, and we define

$$\phi_j := \phi(\cdot - k_j), \quad \text{for all } j = 1, \dots, N.$$

For an arbitrary choice of $\vec{\omega} \in \{-1, 1\}^N$, we define

$$f := \sum_{j=1}^N \omega_j \phi_j.$$

Note that this f depends on N , as well as on $\vec{\omega}$; and thinking of $\vec{\omega}$ as a random variable as in the statement of Khinchin's inequality, f is a random function.

Observe that $\|f\|_p$ is independent of $\vec{\omega}$. Indeed,

$$f = \begin{cases} \omega_1 \phi_1 & \text{on } B(k_1, 1), \\ \omega_2 \phi_2 & \text{on } B(k_2, 1), \\ \vdots & \\ \omega_N \phi_N & \text{on } B(k_N, 1), \\ 0 & \text{elsewhere} \end{cases}.$$

Therefore,

$$|f| = \begin{cases} \phi_1 & \text{on } B(k_1, 1), \\ \phi_2 & \text{on } B(k_2, 1), \\ \vdots & \\ \phi_N & \text{on } B(k_N, 1), \\ 0 & \text{elsewhere} \end{cases},$$

and the disjointness of the balls $B(k_j, 1)$ implies that

$$\|f\|_p = \left(\sum_{j=1}^N \int_{B(k_j, 1)} |\phi_j|^p \right)^{1/p} = \left(N \int_{B(0, 1)} |\phi_1|^p \right)^{1/p} = N^{1/p} \|\phi_1\|_p.$$

On the other hand, due to Khinchin's inequality, we can show that $\mathbb{E}\|\hat{f}\|_{p'}$ is *very small* (so small that the estimate $\|f\|_p \leq C\|\hat{f}\|_{p'}$ cannot hold over all the possible choices of f). Indeed,

$$f = \sum_{j=1}^N \omega_j \phi_j = \sum_{j=1}^N \phi_1(\cdot - k_j),$$

so

$$\hat{f}(\xi) = \sum_{j=1}^N \omega_j e^{-2\pi i \langle \xi, k_j \rangle} \widehat{\phi_1}(\xi) \quad \text{for all } \xi \in \mathbb{R}^n.$$

Therefore,

$$\begin{aligned} \mathbb{E}\|\hat{f}\|_{p'}^{p'} &= \mathbb{E} \left(\int_{\xi \in \mathbb{R}^n} \left| \sum_{j=1}^N \omega_j e^{-2\pi i \langle \xi, k_j \rangle} \widehat{\phi_1}(\xi) \right|^{p'} d\xi \right) \\ &= \mathbb{E} \left(\int_{\xi \in \mathbb{R}^n} \left| \sum_{j=1}^N \omega_j e^{-2\pi i \langle \xi, k_j \rangle} \right|^{p'} \cdot |\widehat{\phi_1}(\xi)|^{p'} d\xi \right) \\ &= \int_{\xi \in \mathbb{R}^n} |\widehat{\phi_1}(\xi)|^{p'} \cdot \mathbb{E} \left(\left| \sum_{j=1}^N \omega_j e^{-2\pi i \langle \xi, k_j \rangle} \right|^{p'} \right) d\xi \\ &\stackrel{\text{Khinchin}}{\lesssim} \int_{\xi \in \mathbb{R}^n} |\widehat{\phi_1}(\xi)|^{p'} \left(\sum_{j=1}^N |e^{-2\pi i \langle \xi, k_j \rangle}|^2 \right)^{p'/2} d\xi \\ &= N^{p'/2} \|\widehat{\phi_1}\|_{p'}^{p'}, \end{aligned}$$

hence there exists a choice of $\vec{\omega} \in \{-1, 1\}^n$ such that the corresponding f satisfies

$$\|\hat{f}\|_{p'} \lesssim N^{1/2} \|\widehat{\phi_1}\|_{p'}.$$

For every N , fix such an f and denote it by f_N (so that we remember the dependence of f on N). We have proved that, for every $N \in \mathbb{N}$,

$$\|f_N\|_p = N^{1/p} \|\phi_1\|_p \quad \text{and} \quad \|\hat{f}_N\|_{p'} \lesssim N^{1/2} \|\widehat{\phi_1}\|_{p'}.$$

Therefore, if there existed $C > 0$ such that

$$(8) \quad \|f_N\|_p \leq C \|\hat{f}_N\|_{p'} \quad \text{for all } N \in \mathbb{N},$$

it would follow that

$$N^{1/p} \leq C' \cdot N^{1/2} \quad \text{for all } N \in \mathbb{N},$$

or equivalently

$$N^{\frac{1}{2} - \frac{1}{p}} \geq C'' \quad \text{for all } N \in \mathbb{N},$$

for a constant $C'' > 0$ that depends only on ϕ_1 (and not on N). This is true as $N \rightarrow +\infty$ only when $\frac{1}{2} - \frac{1}{p} \geq 0$, or equivalently $p \geq 2$. However, our p is smaller than 2 (see the condition in blue earlier). Therefore, (8) does not hold.

□

CHAPTER 2

An interlude: The inverse Fourier transform

Let $\mathcal{F}$ be the Fourier transform on $L^1 + L^2$. Just as we usually denote $\mathcal{F}$ by $\widehat{\cdot}$, we denote the inverse $\mathcal{F}^{-1}$ of $\mathcal{F}$ by $\check{\cdot}$. In other words,

$$(\widehat{f})^\check{} = f, \quad \text{for all } f \in L^1 + L^2.$$

Recall the Fourier inversion formula

$$f(x) = \widehat{\widehat{f}}(-x) \quad \text{for a.e. } x \in \mathbb{R}^n,$$

which is valid for every $f \in L^2$, and for every $f \in L^1$ such that $\widehat{f} \in L^1$. It follows that, for any $g \in L^2$, as well as for any $g \in L^1$ that is the Fourier transform of an L^1 function,

$$\check{g}(x) = \widehat{g}(-x), \quad \text{for a.e. } x \in \mathbb{R}^n.$$

If in particular $g \in L^1$ is the Fourier transform of an L^1 function, it holds that

$$\check{g}(x) = \int e^{2\pi i \langle x, \xi \rangle} g(\xi) d\xi, \quad \text{for a.e. } x \in \mathbb{R}^n.$$

Note that the above integral also makes sense for every $g \in L^1$, even if g is not the Fourier transform of an L^1 function. Therefore, we may extend the definition of $\check{\cdot}$ to L^1 , as follows:

$$\check{g}(x) := \int e^{2\pi i \langle x, \xi \rangle} g(\xi) d\xi = \widehat{g}(-x), \quad \text{for all } x \in \mathbb{R}^n \text{ and all } g \in L^1.$$

We now recall some properties of the Fourier transform, and translate them into properties of $\check{\cdot}$.

→ *Properties of $\widehat{\cdot}$ and $\check{\cdot}$*

(P1) For every $f, g \in L^2$, the following hold:

$$\int \widehat{f}g = \int f\widehat{g},$$

$$\int \check{f}g = \int f\check{g}.$$

(P2) For every $f, g \in L^2$,

$$\int fg = \int \widehat{f}(\xi) \widehat{g}(-\xi) d\xi = \int \widehat{f} \check{g}.$$

(P3) For every $g \in L^2$,

$$\widehat{\widehat{g}} = \check{g}, \quad \check{\check{g}} = g.$$

(P4) For every $f, g \in L^1 + L^2$, it holds that

$$\widehat{f * g} = \widehat{f} \cdot \widehat{g}, \quad \widetilde{f * g} = \check{f} \cdot \check{g},$$

and, for every $f, g \in L^2$, it holds that

$$\widehat{f \cdot g} = \widehat{f} * \widehat{g}, \quad \widetilde{f \cdot g} = \check{f} * \check{g}.$$

(P4) For every $f \in L^1$ and any $a \in \mathbb{R}^n$, the equalities

$$\begin{aligned}\left(f(\cdot + a)\right)^{\widehat{}}(\xi) &= e^{2\pi i \langle \xi, a \rangle} \widehat{f}(\xi), \\ \left(f(\cdot + a)\right)^{\check{}}(\xi) &= e^{-2\pi i \langle \xi, a \rangle} \check{f}(\xi), \\ \left(e^{2\pi i \langle a, \cdot \rangle} f\right)^{\widehat{}}(\xi) &= \widehat{f}(\xi - a), \\ \left(e^{2\pi i \langle a, \cdot \rangle} f\right)^{\check{}}(\xi) &= \check{f}(\xi + a)\end{aligned}$$

hold for every $\xi \in \mathbb{R}^n$. The same estimates hold a.e. when $f \in L^2$.

CHAPTER 3

Do partial Fourier integrals converge?

3.1. An initial reduction

The Fourier inversion formula

$$f(x) = \int_{\mathbb{R}^n} e^{2\pi i \langle x, \xi \rangle} \hat{f}(\xi) \, d\xi$$

holds a.e. for every $f \in L^1$ such that $\hat{f} \in L^1$. **What if $\hat{f} \notin L^1$?** The above integral is not even well-defined. Can we still make some sense of the inversion formula?

An important observation is that, for all f for which we have defined the Fourier transform, that is, for all $f \in L^p$ with $p \in [1, 2]$, it holds that $\hat{f} \in L^{p'}$. So, even though $\hat{f}$ may not be in L^1 (as $p' > 2$), it is still in L^1_{loc} - and in fact the 'partial' integral

$$\int_U e^{2\pi i \langle x, \xi \rangle} \hat{f}(\xi) \, d\xi$$

is well-defined for every bounded, Lebesgue-measurable $U \subseteq \mathbb{R}^n$, as

$$\int_U |\hat{f}| \leq \|\hat{f}\|_{p'} \cdot |U|^{1/p} < +\infty$$

by Hölder's inequality. Therefore, the following question still makes sense.

Question. Do the integrals of the form

$$\int_U e^{2\pi i \langle x, \xi \rangle} \hat{f}(\xi) \, d\xi$$

converge to $f(x)$ as U gets larger and larger?
And what about other types of convergence?

Note that, as the sets U increase so that their union is $\mathbb{R}^n$, the answer to the above question is **yes when $f \in L^1$ and $\hat{f} \in L^1$** , by the Fourier inversion formula and the Dominated Convergence Theorem.

We will be particularly interested in the integrals

$$S_R f(x) = \int_{[-R, R]^n} e^{2\pi i \langle x, \xi \rangle} \hat{f}(\xi) \, d\xi$$

and

$$S_R f(x) = \int_{B(0, R)} e^{2\pi i \langle x, \xi \rangle} \hat{f}(\xi) \, d\xi,$$

which a priori we probably expect to have similar behaviour. Note that in the case $n = 1$, these partial Fourier integrals are the same:

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

We reiterate that is important here is that the functions $S_R f$ make sense for all $f \in L^p$ with $p \in [1, 2]$ (while $\int e^{2\pi i \langle x, \xi \rangle} \widehat{f}(\xi) d\xi$ may not). So:

We ask the following questions, all of which aim to make sense of the Fourier inversion formula when the integral $\int e^{2\pi i \langle x, \xi \rangle} \widehat{f}(\xi) d\xi$ may not exist.

- For $p \in [1, 2]$, S_R is an operator on the whole of L^p . **Let $f \in L^p$.**

$$\left\{ \begin{array}{l} \text{Do we have that } S_R f \xrightarrow{R \rightarrow +\infty} f \text{ pointwise a.e.?} \\ \text{Do we have that } S_R f \in L^p \text{ for all } R \geq 1, \text{ and that } \|S_R f - f\|_p \xrightarrow{R \rightarrow +\infty} 0? \end{array} \right.$$

- For $p > 2$, S_R is a well-defined operator on the dense subspaces $L^1 \cap L^p$ and $L^2 \cap L^p$ of L^p . Our analysis will work no matter which of these spaces we start with, but, for consistency with Fourier multiplier theory (upcoming), we view S_R as an operator on $L^p \cap L^2$ (which contains $\mathcal{S}(\mathbb{R}^n)$). **Let $f \in L^2 \cap L^p$.**

$$\left\{ \begin{array}{l} \text{Do we have that } S_R f \xrightarrow{R \rightarrow +\infty} f \text{ pointwise a.e.?} \\ \text{Do we have that } S_R f \in L^p \text{ for all } R \geq 1, \text{ and that } \|S_R f - f\|_p \xrightarrow{R \rightarrow +\infty} 0? \end{array} \right.$$

And

$$\left\{ \begin{array}{l} \text{Can we extend } S_R \text{ to the whole of } L^p, \text{ and ask the same questions on } L^p? \end{array} \right.$$

3.2. Pointwise convergence

We will only state the results. For $f \in L^1$ such that $\widehat{f} \in L^1$ (e.g., for $f \in \mathcal{S}(\mathbb{R}^n)$), we of course have that

$$S_R f(x) \xrightarrow{R \rightarrow +\infty} \int e^{-2\pi i \langle x, \xi \rangle} \widehat{f}(\xi) d\xi \quad (= f(x) \text{ for a.e. } x)$$

by the Fourier inversion formula and the Dominated Convergence Theorem. So, for such $f \in L^1$, the answer to our pointwise convergence question is yes.

It turns out that things can go very badly for other $f \in L^1$, even in one dimension:

THEOREM 3.2.1. (Kolmogorov, 1922) *There exists $f \in L^1(\mathbb{R})$ such that*

$$S_R f(x) \not\rightarrow f(x) \text{ as } R \rightarrow +\infty, \quad \text{for a.e. } x \in \mathbb{R}.$$

The following celebrated theorem states that $p = 1$ is exceptional in terms of the above phenomenon.

THEOREM 3.2.2. (Carleson-Hunt, 1966) *Let $p \in (1, 2]$. Then, for every $f \in L^p(\mathbb{R})$,*

$$S_R f \xrightarrow{R \rightarrow +\infty} f \text{ a.e.}$$

The proof follows by proving L^p -boundedness of the Carleson maximal operator $\mathcal{C}$, defined by

$$\mathcal{C}f(x) := \sup_{R \geq 1} \left| \int_{-R}^R e^{2\pi i \langle x, \xi \rangle} \widehat{f}(\xi) d\xi \right|, \quad x \in \mathbb{R}.$$

There is also a way to extend the definition of $\mathcal{C}$ and the proof of these results¹ to all $p > 2$.

3.3. L^p -convergence

Let $1 \leq p < +\infty$. We denote

$$\|S_R\|_{p \rightarrow p} := \sup_{\|f\|_p=1, f \in L^p \cap L^2} \|S_R f\|_p.$$

This quantity is finite if and only if S_R is a bounded operator from $L^p \cap L^2$ to L^p .

The following proposition reduces our earlier questions to a single question about L^p -boundedness of S_R on $L^p \cap L^2$.

PROPOSITION 3.3.1. *Let $1 \leq p < +\infty$. The following are equivalent.*

(i) *It holds that*

$$\sup_{R \geq 1} \|S_R\|_{p \rightarrow p} < +\infty.$$

(ii) *For all $R \geq 1$, S_R extends to a bounded, linear operator $S_R : L^p \rightarrow L^p$, so that, for every $f \in L^p$, it holds that*

$$\|S_R f - f\|_p \xrightarrow{R \rightarrow +\infty} 0.$$

Note that, if (i) holds, the extension in (ii) is unique (by density of $L^p \cap L^2$ in L^p).

Observation: By a change of variables, it is easy to see that, for all $R \geq 1$,

$$\|S_R\|_{p \rightarrow p} = \|S_1\|_{p \rightarrow p}.$$

Therefore,

$$\sup_{R \geq 1} \|S_R\|_{p \rightarrow p} = \|S_1\|_{p \rightarrow p}.$$

PROOF OF PROPOSITION 3.3.1. (i) $\Rightarrow$ (ii): Since

$$\sup_{R \geq 1} \sup_{\|f\|_p=1, f \in L^p \cap L^2} \|S_R f\|_p =: C < +\infty,$$

it follows that, for all $R \geq 1$, $S_R f \in L^p$ for every $f \in L^p \cap L^2$, and that the operator $S_R : L^p \cap L^2 \rightarrow L^p$ is bounded, with

$$\|S_R f\|_p \leq C \cdot \|f\|_p \quad \text{for all } f \in L^p \cap L^2.$$

(It will be useful here that C is independent of R .) Since each S_R is a bounded linear operator on the dense subspace $L^p \cap L^2$ of L^p , it has a unique extension to a bounded linear operator $S_R : L^p \rightarrow L^p$, with the same norm.

¹See L. Grafakos's book *Modern Fourier Analysis*, Chapter 11.

Now, let $f \in L^p$, and let $\epsilon > 0$. There exists $g \in \mathcal{S}(\mathbb{R}^n)$, with $\hat{g} \in C_c^\infty(\mathbb{R}^n)$, such that²

$$\|f - g\|_p < \epsilon.$$

The reason we pick specifically such $g \in \mathcal{S}(\mathbb{R}^n)$ is that the desired estimate

$$\|S_R g - g\|_p \xrightarrow{R \rightarrow +\infty} 0$$

holds for it - in fact, $\|S_R g - g\|_p = 0$ for large R , as $\text{supp } \hat{g} \subseteq B(0, R)$ and thus

$$S_R g(x) = \int_{B(0, R)} e^{2\pi i \langle x, \xi \rangle} \hat{g}(\xi) d\xi = \int e^{2\pi i \langle x, \xi \rangle} \hat{g}(\xi) d\xi \stackrel{g \in \mathcal{S}(\mathbb{R}^n)}{=} g(x) \quad \text{for all } x \in \mathbb{R}^n.$$

Therefore,

$$\|S_R f - f\|_p \leq \|S_R f - S_R g\|_p + \|S_R g - g\|_p + \|g - f\|_p \leq (C + 1)\|f - g\|_p \leq (C + 1)\epsilon,$$

for all large R . It follows that $\|S_R f - f\|_p \xrightarrow{R \rightarrow +\infty} 0$.

(ii) $\Rightarrow$ (i): This follows by the Uniform Boundedness Principle. The operators

$$S_N : (L^p, \|\cdot\|_p) \rightarrow (L^p, \|\cdot\|_p), \quad N \in \mathbb{N},$$

are all bounded operators, and their domain is a Banach space. Moreover, for every $f \in L^p$, the set $\{\|S_N f\|_p\}_{N \in \mathbb{N}}$ is bounded, as $(\|S_N f\|_p)_{N \in \mathbb{N}}$ is a convergent sequence in $\mathbb{R}$ (it converges to $\|f\|_p$). By the Uniform Boundedness Principle, the norms $\|S_N\|_{p \rightarrow p}$, $N \in \mathbb{N}$, are uniformly bounded. However, all these norms are the same as each other, and equal to $\|S_R\|_{p \rightarrow p}$ for any $R \geq 1$. We conclude that the $\|S_R\|_{p \rightarrow p}$ are all the same number. □

By all the above, our questions reduce to the following.³

Question: Let $1 < p < +\infty$. Is $\|S_1\|_{p \rightarrow p} < +\infty$?

The answer is **YES in one dimension**, in which case of course

$$[-R, R]^n = B(0, R) = [-R, R] :$$

THEOREM 3.3.2. (Riesz) Let $1 < p < +\infty$. When $n = 1$, in which case

$$S_R f(x) = \int_{-R}^R e^{2\pi i \langle x, \xi \rangle} \hat{f}(\xi) d\xi, \quad x \in \mathbb{R},$$

it holds that

$$\|S_1\|_{p \rightarrow p} < +\infty.$$

The proof follows from the L^p -boundedness properties of the Hilbert transform, and we will not focus on it here.

So, in one dimension, the situation is fully resolved: For all $1 < p < +\infty$, S_R can be defined on the whole of $L^p(\mathbb{R})$, and

$$\|S_R f - f\|_p \rightarrow 0 \quad \text{for all } f \in L^p(\mathbb{R}).$$

What about higher dimensions?

²This is left as an exercise.

³We cannot hope for L^1 -boundedness of S_1 , as, via the above analysis, that would contradict Kolmogorov's counterexample.

→ **The case of cubic Fourier integrals.** In the case where

$$S_R f(x) := \int_{[-R,R]^n} e^{2\pi i \langle x, \xi \rangle} \widehat{f}(\xi) d\xi,$$

the situation in all dimensions $n \geq 2$ follows directly by induction on the dimension, using the $n = 1$ case and Fubini's theorem. So, in this case, again we have that, for all $1 < p < +\infty$, S_R can be defined on the whole of $L^p(\mathbb{R}^n)$, and

$$\|S_R f - f\|_p \rightarrow 0 \quad \text{for all } f \in L^p(\mathbb{R}^n).$$

→ **The case of spherical Fourier integrals.** In the case where

$$S_R f(x) := \int_{B(0,R)} e^{2\pi i \langle x, \xi \rangle} \widehat{f}(\xi) d\xi,$$

the above tensorisation argument does not work - we cannot 'separate' variables this time. So, we have to follow a different avenue. We view S_1 from a different perspective:

For every $1 < p < +\infty$ and $f \in L^p \cap L^2$,

$$S_1 f(x) = \int_{B(0,1)} e^{2\pi i \langle x, \xi \rangle} \widehat{f}(\xi) d\xi = \int e^{2\pi i \langle x, \xi \rangle} \chi_{B(0,1)}(\xi) \widehat{f}(\xi) d\xi = \left(\chi_{B(0,1)} \widehat{f} \right)^\vee(x).$$

Therefore,

$$\widehat{S_1 f} = \chi_{B(0,1)} \widehat{f}.$$

Every operator T defined on $L^p \cap L^2$ by $\widehat{Tf} = m \widehat{f}$, for some bounded, Lebesgue-measurable function m , is called a *Fourier multiplier operator*, and there is a whole theory developed for these objects. The operator S_1 is clearly a Fourier multiplier operator, for $m = \chi_{B(0,1)}$.

3.4. Fourier multipliers: the basics

Let $n \in \mathbb{N}$ and $1 \leq p \leq +\infty$. Let $m : \mathbb{R}^n \rightarrow \mathbb{C}$ be a bounded, Lebesgue-measurable function, and define an operator T_m on $L^2(\mathbb{R}^n)$ by

$$\widehat{T_m f} = m \widehat{f} \quad \text{for all } f \in L^2.$$

Observe that T_m is a well-defined, linear operator from L^2 to L^2 . Indeed, for all $f \in L^2$, $\widehat{f} \in L^2$ as well, and, since m is bounded and Lebesgue-measurable, we also have that $m \widehat{f} \in L^2$. And, since the Fourier transform is 1-1 and onto L^2 , there exists a unique $g \in L^2$ so that $\widehat{g} = m \widehat{f}$. This $g \in L^2$ is precisely $T_m f$. In other words,

$$T_m f = (m \widehat{f})^\vee.$$

DEFINITION 3.4.1. (L^p -Fourier multiplier) Let $m : \mathbb{R}^n \rightarrow \mathbb{C}$ be a bounded, Lebesgue-measurable function. We say that m is an **L^p -multiplier** (or an **L^p -Fourier multiplier**) if the operator T_m , restricted on $L^p(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, is L^p -bounded. That is, m is an L^p -multiplier if

$$T_m f \in L^p \quad \text{for all } f \in L^p \cap L^2$$

and the operator $T_m : (L^p \cap L^2, \|\cdot\|_p) \rightarrow (L^p, \|\cdot\|_p)$ is bounded, i.e.

$$\|T_m f\|_p \leq C_p \|f\|_p \quad \text{for all } f \in L^p \cap L^2.$$

In this case, we also call T_m an **L^p -Fourier multiplier operator**.

Thanks to its L^p -boundedness property, an L^p -Fourier multiplier operator T_m can be uniquely linearly extended to an L^p -bounded operator on the whole of L^p (via the classical density argument that we have already seen in different contexts). This is true even for $p > 2$, in which case $\widehat{f}$ does not exist in general.

PROPOSITION 3.4.2. *Every Lebesgue-measurable, bounded function $m : \mathbb{R}^n \rightarrow \mathbb{C}$ is an L^2 -multiplier.*

PROOF. We have already explained that T_m is an operator from L^2 to L^2 . It remains to show that it is L^2 -bounded. This holds by Plancherel's theorem: for all $f \in L^2$,

$$\|T_m f\|_2 = \|\widehat{T_m f}\|_2 = \|m \widehat{f}\|_2 \leq \|m\|_\infty \|\widehat{f}\|_2 = \|m\|_\infty \|f\|_2.$$

□

PROPOSITION 3.4.3. *Let $m : \mathbb{R}^n \rightarrow \mathbb{C}$ be a Lebesgue-measurable, bounded function, and let $1 \leq p \leq +\infty$. Then,*

$$m \text{ is an } L^p\text{-multiplier} \iff m \text{ is an } L^{p'}\text{-multiplier}.$$

PROOF. The proof that follows shows that, when $1 \leq p < +\infty$, $T_{\overline{m}} : L^{p'} \rightarrow L^{p'}$ is the adjoint of $T_m : L^p \rightarrow L^p$. The result directly follows for such p , as then

$$\|T_m\|_{p \rightarrow p} = \|T_{\overline{m}}\|_{p' \rightarrow p'} = \|T_m\|_{p' \rightarrow p'}$$

(the last equality is trivial to show). However, this doesn't work for $p = \infty$ (as L^1 is not the dual of L^∞), so we present the proof via the well-known formulae for p -norms of operators and functions (which hold irrespective of p). In particular,

$$(9) \quad \begin{aligned} \|T_m\|_{p \rightarrow p} &= \sup_{\|f\|_p=1, f \in L^p \cap L^2} \|T_m f\|_p \\ &= \sup_{\|f\|_p=1, f \in L^p \cap L^2} \sup_{\|g\|_{p'}=1, g \in L^{p'} \cap L^2} \langle T_m f, g \rangle, \end{aligned}$$

where, for all $f \in L^p \cap L^2$ and all $g \in L^{p'} \cap L^2$,

$$\begin{aligned} \langle T_m f, g \rangle &= \int T_m f \cdot \bar{g} \\ &\stackrel{(P1)}{=} \int \widehat{T_m f} \cdot \check{\bar{g}} = \int m \widehat{f} \cdot \check{\bar{g}} \\ &\stackrel{\check{\bar{g}}=\widehat{\bar{g}}}{=} \int \widehat{f} \cdot m \widehat{\bar{g}} = \int \widehat{f} \cdot \overline{(\widehat{m} \widehat{g})} = \int \widehat{f} \cdot \overline{(T_{\overline{m}} g)}^\wedge \\ &\stackrel{\check{\widehat{h}}=\widehat{\check{h}}}{=} \int \widehat{f} \cdot \left(\overline{T_{\overline{m}} g} \right)^\vee \\ &= \int f \cdot \overline{T_{\overline{m}} g} \\ &= \langle f, T_{\overline{m}} g \rangle. \end{aligned}$$

Plugging this into (9) and swapping the order of the suprema, we obtain

$$\begin{aligned} \|T_m\|_{p \rightarrow p} &= \sup_{\|g\|_{p'}=1, g \in L^{p'} \cap L^2} \sup_{\|f\|_p=1, f \in L^p \cap L^2} \langle f, T_{\overline{m}} g \rangle \\ &= \sup_{\|g\|_{p'}=1, g \in L^{p'} \cap L^2} \|T_{\overline{m}} g\|_{p'} \\ &= \|T_{\overline{m}}\|_{p' \rightarrow p'}. \end{aligned}$$

It is easy to see that $\|T_{\overline{m}}\|_{p' \rightarrow p'} = \|T_m\|_{p' \rightarrow p'}$, and it is left as an exercise.

□

PROPOSITION 3.4.4. Let $m \in \mathcal{S}(\mathbb{R}^n)$. Then, m is an L^p -multiplier, for every $1 \leq p \leq +\infty$.

PROOF. Let $1 < p < +\infty$. Since m is in L^2 , it follows that

$$T_m f = (m \hat{f})^\vee = \check{m} * f, \quad \text{for all } f \in L^2 \cap L^p.$$

Now, $m \in \mathcal{S}(\mathbb{R}^n)$, hence $\check{m} \in \mathcal{S}(\mathbb{R}^n) \subset L^1$, so $\|\check{m}\|_1 < +\infty$. Therefore, for all $f \in L^p \cap L^2$,

$$\|T_m f\|_p = \|\check{m} * f\|_p \leq \|\check{m}\|_1 \cdot \|f\|_p,$$

where $\|\check{m}\|_1 < +\infty$. It follows that T_m is L^p -bounded, hence m is an L^p -multiplier. □

More about Fourier multipliers can be found in Stein's book *Singular Integrals and Differentiability Properties of Functions*, Chapter 3.

3.5. Back to S_1 : Fefferman's ball multiplier theorem

In this section we focus on $\mathbb{R}^n$ when $n \geq 2$.

For all $1 < p < +\infty$, the operator S_1 is well-defined on $L^p \cap L^2$ via the formula

$$\widehat{S_1 f} = \chi_{B(0,1)} \hat{f}, \quad f \in L^p \cap L^2.$$

Since $\chi_{B(0,1)}$ is a bounded, Lebesgue-measurable function, S_1 is L^p -bounded if and only if $\chi_{B(0,1)}$ is an L^p -multiplier.

- Since $\chi_{B(0,1)}$ is a bounded, Lebesgue-measurable function, S_1 is automatically L^2 -bounded (that is, $\chi_{B(0,1)}$ is an L^2 -multiplier).
- Since $\chi_{B(0,1)}$ is not smooth, we cannot use Proposition 3.4.4 to deduce that it is an L^p -multiplier for all other p .

Even so, it was an established belief in the community that $\chi_{B(0,1)}$ must be an L^p -multiplier for a wider range of exponents p . So, it came as a bit of a shock when Fefferman disproved it⁴ in 1971.

Fefferman: $\chi_{B(0,1)}$ is **not** an L^p -multiplier, for any $p \neq 2$!

It follows that on $\mathbb{R}^n$, with $n \geq 2$, the spherical Fourier partial integrals S_R behave badly:

- S_R is **not** L^p -bounded, for any $p \neq 2$.
- For every $p \neq 2$, we **cannot** extend S_R to a bounded linear operator on L^p so that $\|S_R f - f\|_p \rightarrow 0$ as $R \rightarrow +\infty$ for every $f \in L^p$.

Fefferman's proof is based on the Perron tree construction of Kakeya sets of measure 0. Other tools in the version of the proof that we will present here are the uncertainty principle and Khinchin's inequality.

⁴C. Fefferman, 'The multiplier problem for the ball', Ann. of Math. (2), vol. 94., no. 2, pp. 330-336, 1971.

3.6. The uncertainty principle

The uncertainty principle is the following heuristic, which can be made precise in several ways, according to our goals every time.

The uncertainty principle (heuristic): If f is localised in a ball of radius R , then $|\hat{f}|$ is locally constant at scale R^{-1} .

When we say that f is localised in a ball of radius R , we mean that the support of f is contained in such a ball.

The heuristic can be made precise, for instance, via the statement of Heisenberg's uncertainty principle and Bernstein's inequality. We will see a different, simpler perspective here, which **morally** lets us discretise $\hat{f}$ at scale R^{-1} , i.e. to write

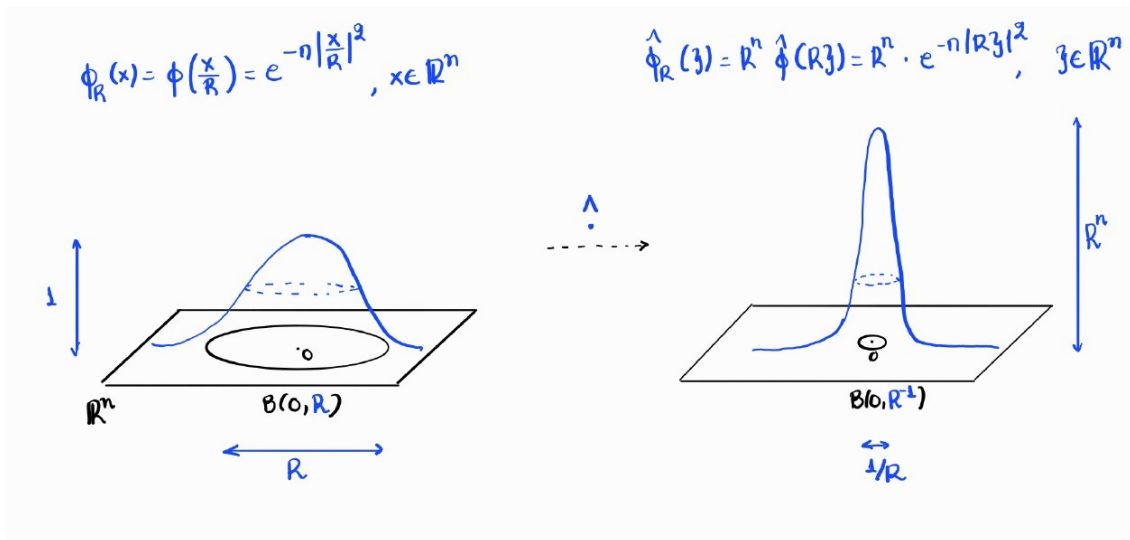
$$(10) \quad |\hat{f}| \text{ " = " } \sum_{B_k} c_k \chi_{B_k},$$

where the B_k are finitely overlapping balls of radius R^{-1} .

A warmup. The uncertainty principle can be easily understood when f is a Gaussian (which is *essentially* supported on $B(0, 1)$). In particular, let

$$\phi(x) = e^{-\pi|x|^2}, \quad x \in \mathbb{R}^n.$$

Recall that $\hat{\phi} = \phi$. So, defining $\phi_R := \phi(\frac{\cdot}{R})$, which is practically supported in $B(0, R)$, we see that $\hat{\phi}_R = R^n \hat{\phi}(R \cdot) = R^n \phi(R \cdot)$, which is practically supported in $B(0, R^{-1})$.



In this example, we see that for every $\epsilon > 0$ and $N \in \mathbb{N}$, due to the fast decay properties of the Schwartz function ϕ (specifically, $|\phi(\xi)| \leq C_N \cdot \frac{1}{(1+|\xi|)^N}$), it holds that

$$|\hat{\phi}_R(\xi)| \leq C_{N,\epsilon} R^{-N} \quad \text{for } |\xi| \geq R^{-1+\epsilon}.$$

So, we can think of $\hat{\phi}_R$ as being roughly constant (and non-zero) on each of the $R^{n\epsilon} = R^{O(\epsilon)}$ balls $B(c, R^{-1})$ with centres c at distance up to $R^{1+\epsilon}$ from the origin, and roughly 0 on the rest of the space.

In this warmup, there is nothing too special about the Gaussian - we could have replaced it with any bump function ϕ whose Fourier transform is non-negative. Note that such functions exist, and they will be useful to us:

OBSERVATION 3.6.1. *There exists a non-zero function $\phi \in C_c^\infty(\mathbb{R}^n)$, such that*

$$\text{supp}\phi \subseteq B(0, 1), \quad \phi \geq 0, \quad \int \phi = 1$$

and

$$\widehat{\phi} \geq 0.$$

PROOF. Let $\psi \in C_c^\infty(\mathbb{R}^n)$, with

$$\text{supp}\psi \subseteq B(0, 1/2), \quad \psi \geq 0, \quad \int \psi = 1.$$

Then, the function $\phi := \psi * \psi(-\cdot)$ (where $\phi(-\cdot)$ is the reflection of ψ around 0) has the desired properties:

- $\text{supp}\phi \subseteq \text{supp}\psi + \text{supp}\psi(-\cdot) = B(0, 1/2) + B(0, 1/2) = B(0, 1)$.
- $\phi \geq 0$, as it is the convolution of two non-negative functions.
- We have that

$$\int \phi = \int \psi \cdot \int \psi(-\cdot) = \int \psi \cdot \int \psi = 1.$$

- The Fourier transform $\widehat{\phi}$ is given by the formula

$$\widehat{\phi}(\xi) = \widehat{\psi}(\xi) \cdot \widehat{\psi(-\cdot)}(\xi) = \widehat{\psi}(\xi) \cdot \overline{\widehat{\psi}(\xi)} = |\widehat{\psi}(\xi)|^2 \geq 0,$$

where in the penultimate equality we used that ψ is real-valued:

$$\begin{aligned} \widehat{\psi(-\cdot)}(\xi) &= \int e^{-2\pi i \langle \xi, x \rangle} \psi(-x) dx \\ &\stackrel{-x=y}{=} \int e^{2\pi i \langle \xi, y \rangle} \psi(y) dy \\ &\stackrel{\psi(x) \in \mathbb{R} \ \forall x}{=} \overline{\int e^{-2\pi i \langle \xi, y \rangle} \psi(y) dy} = \overline{\widehat{\psi}(\xi)}. \end{aligned}$$

□

→ **From balls to rectangles.** In the above analysis there is nothing too special about the ball $B(0, 1)$ - things would have worked similarly if $B(0, 1)$ was replaced by the rectangle $[-1, 1]^n$. The only difference would be that the Gaussian ϕ_R is roughly supported on $[-R, R]^n$ and its Fourier transform is roughly constant on $[-R^{-1}, R^{-1}]^n$, i.e. the rectangle with same main directions but reciprocal dimensions. We can similarly adapt the analysis to any rectangle with dimensions $r_1, \dots, r_n$, via a rescaling and a rotation.

Indeed, we first adapt it to $\prod_{j=1}^n [-r_j, r_j]^n$, by again using the Gaussian

$$\phi(x) = e^{-\pi|x|^2}, \quad x \in \mathbb{R}^n$$

and defining

$$\phi_{r_1, \dots, r_n}(x) := \phi\left(\frac{x_1}{r_1}, \dots, \frac{x_n}{r_n}\right), \quad x \in \mathbb{R}^n,$$

a Gaussian now adapted on $\prod_{j=1}^n [-r_j, r_j]^n$. Now,

$$\widehat{\phi_{r_1, \dots, r_n}}(\xi) = r_1 \cdots r_n \cdot \widehat{\phi}(r_1 \xi_1, \dots, r_n \xi_n), \quad \xi \in \mathbb{R}^n,$$

which is a Gaussian roughly supported over $\prod_{j=1}^n [-r_j^{-1}, r_j^{-1}]^n$, i.e. the rectangle through 0 with same directions but reciprocal dimensions.

Now, if $\mathcal{R}$ is any rectangle centred at 0, it is the rotation $S\mathcal{R}_0$ of some rectangle $\mathcal{R}_0 = \prod_{j=1}^n [-r_j, r_j]^n$. We adapt $\phi_{r_1, \dots, r_n}$ to this rotated rectangle $\mathcal{R}$, defining

$$\phi(x) := \phi_{r_1, \dots, r_n}(S^{-1}x), \quad x \in \mathbb{R}^n.$$

Then, using the formula

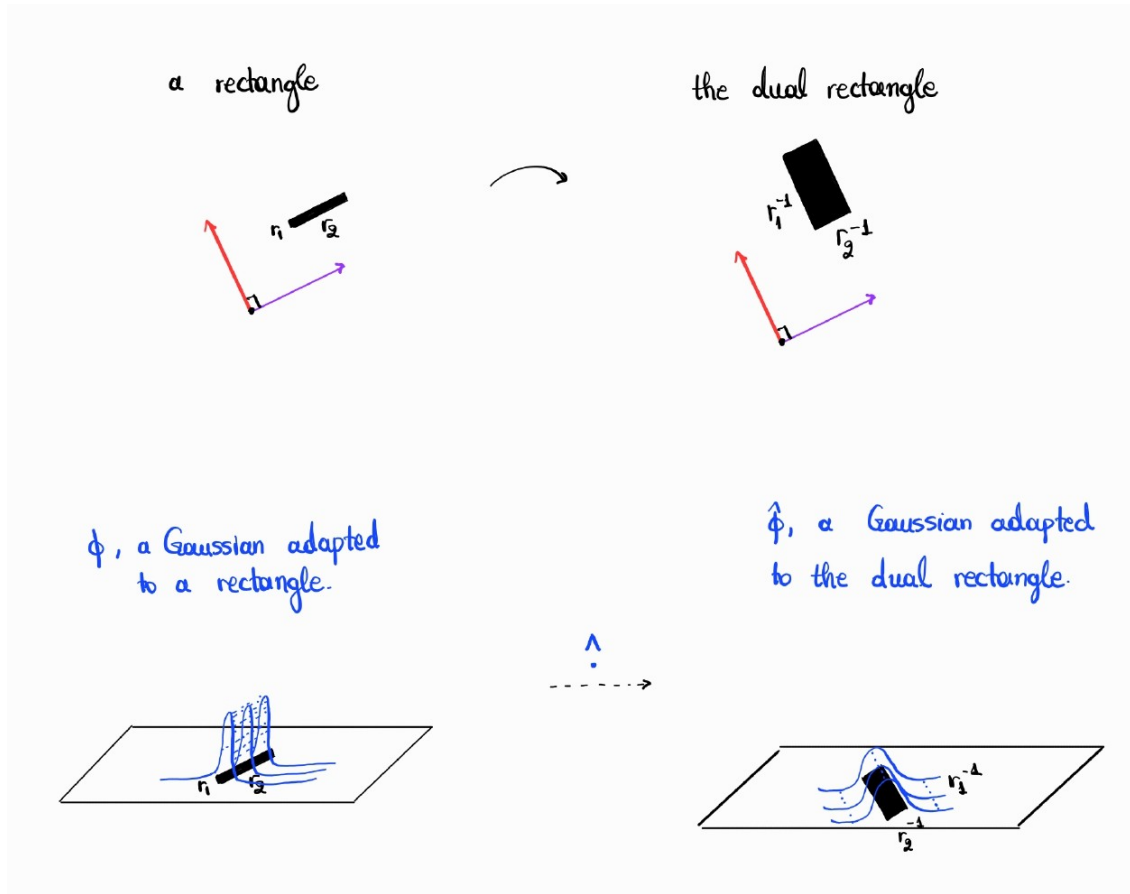
$$\widehat{f \circ M} = |\det M|^{-1} \cdot \widehat{f} \circ (M^{-1})^T$$

for $M = S^{-1}$, we obtain

$$\widehat{\phi}(\xi) = |\det S^{-1}|^{-1} \cdot \widehat{\phi_{r_1, \dots, r_n}} \circ S^T(\xi) \stackrel{|\det S|=1, S^T=S^{-1}}{=} \widehat{\phi_{r_1, \dots, r_n}}(S^{-1}\xi), \quad \xi \in \mathbb{R}^n,$$

so $\widehat{\phi}$ is a Gaussian adapted to the rectangle $S\mathcal{R}_0^*$, where $\mathcal{R}_0^* := \prod_{j=1}^n [-r_j^{-1}, r_j^{-1}]^n$. As before, $\mathcal{R}_0^*$ is the rectangle with centre 0, same directions as $\mathcal{R}_0$, and reciprocal dimensions. We call this *the dual rectangle* to $\mathcal{R}_0$. We may also refer to the dual rectangles to $\mathcal{R}_0$ as all the rectangles parallel to $\mathcal{R}_0^*$.

DEFINITION 3.6.2. Let $\mathcal{R}$ be a rectangle in $\mathbb{R}^n$. We denote by $\mathcal{R}^*$, and call the dual rectangle to $\mathcal{R}$, the rectangle in $\mathbb{R}^n$ with centre 0, same directions as $\mathcal{R}$, but reciprocal dimensions. We may also refer to any translation of $\mathcal{R}^*$ as a dual rectangle to $\mathcal{R}$.



Motivated by this we write the following form of the uncertainty principle heuristic, which still needs to be explained.

The uncertainty principle (heuristic): If f is localised in a rectangle $\mathcal{R}_0$, then $|\hat{f}|$ is roughly constant on all rectangles dual to $\mathcal{R}_0$.

This will allow us to write

$$(11) \quad |\hat{f}| \text{ " = " } \sum_T c_T \chi_T,$$

over a family of disjoint rectangles (or tubes) T that cover $\mathbb{R}^n$ and are all dual to $\mathcal{R}_0$. From now on we will be precise and try to make sense of the above.

Facts for non-negative functions with bounded support.

OBSERVATION 3.6.3. Consider non-zero $\phi \in L^1(\mathbb{R}^n)$, with

$$\text{supp} \phi \subseteq B(0, R), \quad \phi \geq 0.$$

Then, for $c := \hat{\phi}(0) (= \int \phi)$,

$$|\hat{\phi} - c| \leq \frac{c}{2} \quad \text{on } B\left(0, \frac{1}{50R}\right).$$

This means that $\hat{\phi}$ is roughly c on $B\left(0, \frac{1}{50R}\right)$.

PROOF. Observe that $\int \phi = \hat{\phi}(0)$. Therefore, we wish to show that, roughly R^{-1} -close to 0, the values of $\hat{\phi}$ do not change too much.

For every $\xi \in \mathbb{R}^n$,

$$\begin{aligned} |\hat{\phi}(\xi) - \hat{\phi}(0)| &= \left| \int e^{2\pi i \langle \xi, x \rangle} \phi(x) dx - \int e^{2\pi i \langle 0, x \rangle} \phi(x) dx \right| \\ &\leq \int |e^{2\pi i \langle \xi, x \rangle} - e^{2\pi i \langle 0, x \rangle}| \phi(x) dx \\ &\leq 16 \int |\langle \xi, x \rangle - \langle 0, x \rangle| \phi(x) dx \\ &\leq 16 \int |\xi| \cdot |x| \phi(x) dx. \end{aligned}$$

So, if in particular $\xi \in B\left(0, \frac{1}{50R}\right)$, and since $\text{supp} \phi \subseteq B(0, R)$, it follows that

$$|\hat{\phi}(\xi) - \hat{\phi}(0)| \leq 16 \cdot \frac{1}{50R} \cdot R \cdot \int \phi \leq \frac{1}{2} \hat{\phi}(0).$$

□

We can easily derive a similar fact by considering rectangles instead of balls.

OBSERVATION 3.6.4. Let $\mathcal{R}$ be a rectangle in $\mathbb{R}^n$, centred at 0. Consider non-zero $\phi \in L^1(\mathbb{R}^n)$, with

$$\text{supp} \phi \subseteq \mathcal{R}, \quad \phi \geq 0.$$

Then, for $c := \hat{\phi}(0) (= \int \phi)$,

$$|\hat{\phi} - c| \leq \frac{c}{2} \quad \text{on } \frac{1}{50n} \mathcal{R}^*.$$

PROOF. Applying a rotation, we may assume that $\mathcal{R}$ has axis-parallel directions, so

$$\mathcal{R} = \prod_{j=1}^n [-r_j, r_j]$$

for some $r_1, \dots, r_n > 0$. For every $\xi \in \mathbb{R}^n$,

$$\hat{\phi}(\xi) = \int_{\mathcal{R}} e^{2\pi i(\xi_1 x_1 + \dots + \xi_n x_n)} \phi(x) dx.$$

For every $x \in \mathcal{R}$ and every $\xi \in \frac{1}{50n} \mathcal{R}^* = \prod_{j=1}^n \left[-\frac{1}{r_j}, \frac{1}{r_j}\right]$,

$$\begin{aligned} |\xi_1 x_1 + \dots + \xi_n x_n| &\leq |\xi_1| \cdot |x_1| + \dots + |\xi_n| \cdot |x_n| \\ &\leq r_1 \cdot \frac{1}{50n} \cdot \frac{1}{r_1} + \dots + r_n \cdot \frac{1}{50n} \cdot \frac{1}{r_n} \\ &= \frac{1}{50}, \end{aligned}$$

therefore, every $\xi \in \frac{1}{50n} \mathcal{R}^*$,

$$\begin{aligned} |\hat{\phi}(\xi) - \hat{\phi}(0)| &\leq \int |e^{2\pi i \langle \xi, x \rangle} - e^{2\pi i \langle 0, x \rangle}| \phi(x) dx \\ &\leq 16 \int |\langle \xi, x \rangle| \phi(x) dx \\ &\leq \frac{16}{50} \cdot \int \phi \\ &\leq \frac{1}{2} \cdot \hat{\phi}(0). \end{aligned}$$

□

A preliminary form of the uncertainty principle. Combining Observation 3.6.1 with the above two observations, we readily obtain a precise, simple form of the uncertainty principle for all compactly supported functions, whose interpretation agrees with our earlier heuristics.

PROPOSITION 3.6.5. (The uncertainty principle via convolution)

(i) (Case of balls) *There exists a $\psi \in \mathcal{S}(\mathbb{R}^n)$, with*

$$\psi \geq 0, \quad \psi \sim 1 \text{ on } B(0, 1/100),$$

so that for any $R > 0$ and any $f \in C_c^\infty(\mathbb{R}^n)$ with $\text{supp } f \subseteq B(0, R)$, it holds that

$$\hat{f} = \hat{f} * \psi_{1/R},$$

where $\psi_{1/R} := R^n \cdot \psi(R \cdot)$. Note that

$$\psi_{1/R} \sim R^n \quad \text{on } B\left(0, \frac{1}{100R}\right).$$

(ii) (Case of rectangles) *There exists a $\psi \in \mathcal{S}(\mathbb{R}^n)$, with*

$$\psi \geq 0, \quad \psi \sim 1 \text{ on } \left[-\frac{1}{100}, \frac{1}{100}\right]^n,$$

so that for any rectangle $\mathcal{R}$ in $\mathbb{R}^n$ and any $f \in C_c^\infty(\mathbb{R}^n)$ with $\text{supp } f \subseteq \mathcal{R}$, it holds that

$$\hat{f} = \hat{f} * \psi_{\mathcal{R}*},$$

where, if $\mathcal{R} = S \left(\prod_{j=1}^n [-r_j, r_j] \right)$, where S is some rotation, then $\psi_{\mathcal{R}^*}(\xi) = r_1 \cdots r_n \cdot \psi \left(r_1 \cdot (S^{-1}\xi)_1, \dots, r_n \cdot (S^{-1}\xi)_n \right)$. Note that

$$\psi_{\mathcal{R}^*} \sim |\mathcal{R}| = \frac{1}{|\mathcal{R}^*|} \quad \text{on } \frac{1}{50n} \mathcal{R}^*.$$

We emphasise that here, and usually in whatever follows, for any rectangle $\mathcal{R}$ in $\mathbb{R}^n$ and any constant $c > 0$, we will be denoting by $c\mathcal{R}$ the rectangle with same centre as $\mathcal{R}$, same directions, but dimensions those of $\mathcal{R}$ multiplied by c .

Interpretation. In the case of balls, $\psi_{1/R}$ is roughly $R^n \cdot \chi_{B(0, (100R)^{-1})} \sim \frac{\chi_{B(0, (100R)^{-1})}}{|B(0, (100R)^{-1})|}$. So, convolution of $\hat{f}$ with $\psi_{1/R}$ is roughly averaging $\hat{f}$ at scale $1/R$. The above tells us that this averaging doesn't make any difference to $\hat{f}$, therefore $\hat{f}$ must be roughly constant at scale $1/R$.

Similarly, in the case of rectangles, $\psi_{\mathcal{R}^*}$ is roughly $\frac{\chi_{\mathcal{R}^*}}{|\mathcal{R}^*|}$. So, convolution of $\hat{f}$ with $\psi_{\mathcal{R}^*}$ is roughly averaging $\hat{f}$ of every translation of the rectangle $\mathcal{R}^*$. Since this action makes no difference to the values of $\hat{f}$, $\hat{f}$ must be roughly constant on every rectangle dual to $\mathcal{R}$.

PROOF OF PROPOSITION 3.6.5. (i) Fix some $\phi \in C_c^\infty(\mathbb{R}^n)$, with

$$\text{supp } \phi \subseteq B(0, 2), \quad \phi \sim 1 \text{ on } B(0, 1), \quad \phi \geq 0$$

so that

$$\hat{\phi} \geq 0$$

(we know that such ϕ exists, by Observation 3.6.1). Define $\phi_R := \phi(\cdot/R)$; this is ϕ adapted to the ball $B(0, 2R)$, and in particular

$$\phi_R = 1 \quad \text{on } B(0, R).$$

So, for any $R > 0$ and any $f \in C_c^\infty(\mathbb{R}^n)$ with $\text{supp } f \subseteq B(0, R)$, it holds that

$$f = f \cdot \phi_R,$$

hence

$$\hat{f} = \hat{f} * \hat{\phi}_R.$$

It is easy to check that the function $\psi := \hat{\phi}$ has all the desired properties.

Assertion (ii) follows similarly, by Observation 3.6.4. This is left as an exercise. $\square$

Discretisation of $\hat{f}$ for f supported on a ball. We have already seen what happens *close to 0*, when f is *non-negative* and has compact support (Observations 3.6.3 and 3.6.4). We will now use this to see what happens in the general case.

→ **When f is supported in $B(0, 1)$.** Let $f_0 \in L^2$ be an arbitrary function supported in $B(0, 1)$.

There exists $\phi \in C_c^\infty(\mathbb{R}^n)$ with

$$\text{supp } \phi \subseteq B(0, 2), \quad \phi \geq 0, \quad \phi = 1 \text{ on } B(0, 1), \quad \int \phi = 1.$$

We fix this ϕ from now on in this section. Since $\text{supp } f_0 \subseteq B(0, 1)$ and $\phi = 1$ on $B(0, 1)$, it follows that

$$f_0 = f_0 \cdot \phi.$$

By the Carleson-Hunt theorem⁵, the function f_0 , which is also supported on $[-2, 2]^n$, satisfies

$$f_0(x) = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k e^{2\pi i \langle k, x \rangle} \quad \text{for a.e. } x \in [-2, 2]^n,$$

hence

$$f_0(x) = f_0(x) \cdot \phi(x) = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k e^{2\pi i \langle k, x \rangle} \phi(x) \quad \text{for a.e. } x \in \mathbb{R}^n.$$

Here, we may view the summands $c_k e^{2\pi i \langle k, \cdot \rangle} \phi$ as mollified versions of the summands $c_k e^{2\pi i \langle k, \cdot \rangle} \chi_{B(0,1)}$, which we would get from the Fourier series of f_0 if we had not multiplied with the smooth function ϕ . And we expect the mollified summands $c_k e^{2\pi i \langle k, \cdot \rangle} \phi$ to have much better behaved Fourier transforms.

It follows by the above that

$$\widehat{f_0} = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k \widehat{\phi}(\cdot - k).$$

Therefore, denoting

$$\psi := \widehat{\phi}$$

and recalling by Observation 3.6.3 that⁶

$$\widehat{\phi} \sim 1 \quad \text{on } B(0, 1/100)$$

(that is, in practice, $\widehat{\phi}$ is a mollified version of $\chi_{B(0,1)}$, and hence $\widehat{\phi}(\cdot - k)$ is a mollified version of $\chi_{B(k,1)}$), we obtain the following.

PROPOSITION 3.6.6. (An uncertainty principle for functions supported on $B(0, 1)$.)
There exists a function $\psi \in \mathcal{S}(\mathbb{R}^n)$, with

$$\psi \sim 1 \text{ on } B(0, 1/100),$$

such that, for every $f_0 \in C_c^\infty(\mathbb{R}^n)$ with $\text{supp } f_0 \subseteq B(0, 1)$,

$$\widehat{f_0} = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k \psi_k,$$

where

$$\psi_k := \psi(\cdot - k) \quad \text{for all } k \in \frac{1}{4}\mathbb{Z}^n.$$

The above is a precise version of the heuristic (10), in which the rapidly decreasing (i.e., Schwartz) functions ψ_k are replaced by unit balls centred at k .

→ **When f is supported on an arbitrary ball centred at 0.**

Let $f \in L^2$ be an arbitrary function supported in $B(0, R)$, for some $R > 0$. The rescaling $f_0 := f(R \cdot)$ of f is supported on $B(0, 1)$, so, by Proposition 3.6.6, there exists ψ as in the proposition (and independent of f and R) so that

$$\widehat{f(R \cdot)} = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k \psi(\cdot - k).$$

⁵Here we are referring to the original form of the theorem, which was for Fourier series of 1-periodic functions in $L^2(\mathbb{T}^n)$.

⁶The notation $\widehat{\phi} \sim 1$ here means that $\widehat{\phi}(0) > 0$ and $|\widehat{\phi}(\xi) - \widehat{\phi}(0)| \leq \frac{\widehat{\phi}(0)}{2}$ on $B(0, 1/100)$.

Since $\widehat{f(R \cdot)} = \frac{1}{R^n} \cdot \widehat{f}(\cdot)$, it follows that

$$\widehat{f}(\xi) = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} R^n c_k \cdot \psi(R\xi - k) = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c'_k \cdot \psi\left(R \cdot \left(\xi - \frac{k}{R}\right)\right) = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c'_k \psi_{1/R}\left(\xi - \frac{k}{R}\right),$$

where $c'_k := R^n c_k$ and $\psi_{1/R} := \psi(R \cdot)$. Observe that each $\psi_{1/R}(\cdot - k/R)$ is the translation of $\psi_{1/R}$ by $k/R \in \frac{1}{4R}\mathbb{Z}^n$. Therefore, taking the sum over all elements of the form $k/R \in \frac{1}{4R}\mathbb{Z}^n$, and relabeling the coefficients, we obtain

$$\widehat{f} = \sum_{k \in \frac{1}{4R}\mathbb{Z}^n} c_k \psi_{1/R}(\cdot - k)$$

for some coefficients $c_k \in \mathbb{C}$.

Now, since $\psi \sim 1$ on $B(0, 1/100)$, it follows that

$$\psi_{1/R} = \psi(R \cdot) \sim 1 \text{ on } B\left(0, \frac{1}{100R}\right).$$

Moreover,

$$\int \psi_{1/R} = \int \psi(R \cdot) = \frac{1}{R^n} \cdot \int \psi \sim \frac{1}{R^n} \sim \int \chi_{B(0, \frac{1}{R})}.$$

(In the above, the implicit constant technically depends on ψ , however ψ is fixed throughout this analysis.)

Therefore, $\psi_{1/R}$ may be viewed as a mollified version of $\chi_{B(0, \frac{1}{R})}$, and $\psi_{1/R}(\cdot - k)$ as a mollified version of $\chi_{B(k, \frac{1}{R})}$. Therefore, relabeling $\psi_{1/R}$ by ψ , we obtain the following analogue of (10).

PROPOSITION 3.6.7. (An uncertainty principle for functions supported on $B(0, R)$.)
There exists a function $\psi \in \mathcal{S}(\mathbb{R}^n)$, with

$$\psi \sim 1 \text{ on } B\left(0, \frac{1}{100R}\right) \quad \text{and} \quad \int \psi \sim \frac{1}{R^n} \quad \left(= \int \chi_{B(0, R^{-1})}\right),$$

such that, for every $f \in C_c^\infty(\mathbb{R}^n)$ with $\text{supp } f \subseteq B(0, R)$,

$$\widehat{f} = \sum_{k \in \frac{1}{4R}\mathbb{Z}^n} c_k \psi_k,$$

where

$$\psi_k := \psi(\cdot - k) \quad \text{for all } k \in \frac{1}{4R}\mathbb{Z}^n.$$

It is of course useful to remember that, in the above analysis, ψ in the above proposition is actually a relabeling of $\psi(R \cdot)$, for the original ψ from Proposition 3.6.6.

→ **When f is supported on any ball.**

Let $f \in L^2$ be an arbitrary function supported on $B(\omega, R)$, for some $\omega \in \mathbb{R}^n$ and $R > 0$. The translation $f(\cdot + \omega)$ is supported on $B(0, R)$, therefore there exists ψ as in the statement of Proposition 3.6.7 so that

$$\left(\widehat{f(\cdot + \omega)}\right) = \sum_{k \in \frac{1}{4R}\mathbb{Z}^n} c_k \psi(\cdot - k).$$

However, $\left(f(\cdot + \omega)\right)^\wedge(\xi) = e^{2\pi i \langle \omega, \xi \rangle} \hat{f}(\xi)$, for all $\xi \in \mathbb{R}^n$. Therefore,

$$\hat{f} = \sum_{k \in \frac{1}{4R}\mathbb{Z}^n} c_k e^{-2\pi i \langle \omega, \cdot \rangle} \psi(\cdot - k).$$

Note that, just like before, each $\psi(\cdot - k)$ is a mollified version of $\chi_{B(k, R^{-1})}$. The only difference is that the fact that the centre of the ball is not 0 is interpreted as a modulation $e^{-2\pi i \langle \omega, \cdot \rangle}$ on any of these ‘characteristic function approximants’.

Heuristic: When $\text{supp } f \subseteq B(\omega, R)$, roughly speaking, we can write

$$\hat{f} = \sum_k c_k e^{-2\pi i \langle \omega, \cdot \rangle} \chi_B,$$

where the B are essentially disjoint balls of radius R^{-1} , covering $\mathbb{R}^n$. In other words, we may think of $\hat{f}$ as a sum of waves that do not interact much with each other, each wave being supported on a different ball of radius R^{-1} .

The precise statement is the following.

PROPOSITION 3.6.8. (An uncertainty principle for functions supported on $B(\omega, R)$.)
There exists a function $\psi \in \mathcal{S}(\mathbb{R}^n)$, with

$$\psi \sim 1 \text{ on } B\left(0, \frac{1}{100R}\right) \quad \text{and} \quad \int \psi \sim \frac{1}{R^n} \quad \left(= \int \chi_{B(0, R^{-1})}\right),$$

such that, for every $\omega \in \mathbb{R}^n$, for every $f \in C_c^\infty(\mathbb{R}^n)$ with $\text{supp } f \subseteq B(\omega, R)$,

$$\hat{f} = \sum_{k \in \frac{1}{4R}\mathbb{Z}^n} c_k e^{-2\pi i \langle \omega, \cdot \rangle} \psi_k,$$

where

$$\psi_k := \psi(\cdot - k) \quad \text{for all } k \in \frac{1}{4R}\mathbb{Z}^n.$$

We emphasise that the function ψ here can be taken to be the same for all $\omega \in \mathbb{R}^n$: it is an adaptation, to $B(0, R)$, of the function ψ in Proposition 3.6.6.

Discretisation of $\hat{f}$ for f supported on a rectangle. The analysis here is completely analogous to the one for balls. To avoid tedious technicalities, we go through the analysis quickly and simply state the results.

→ **When f is supported on $[-1, 1]^n$.** Let $f_0 \in L^2$ be an arbitrary function supported on $[-1, 1]^n$.

By Observation 3.6.1, there exists a non-zero $\phi \in C_c^\infty(\mathbb{R}^n)$, such that

$$\text{supp } \phi \subseteq [-2, 2]^n, \quad \phi \geq 0, \quad \phi = 1 \text{ on } [-1, 1]^n, \quad \int \phi \sim 1.$$

Fix this ϕ from now on. Since $\text{supp } f_0 \subseteq [-1, 1]^n$ while $\phi = 1$ on $[-1, 1]^n$, it follows that

$$f_0 = f_0 \cdot \phi.$$

By the Carleson-Hunt theorem, the function f_0 , which is also supported on $[-2, 2]^n$, satisfies

$$f_0(x) = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k e^{2\pi i \langle k, x \rangle} \quad \text{for a.e. } x \in [-2, 2]^n.$$

Since both f_0 and ϕ are zero outside $[-2, 2]^n$, it follows that

$$f_0(x) \stackrel{\text{earlier}}{=} f_0(x) \cdot \phi(x) = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k e^{2\pi i \langle k, x \rangle} \phi(x) \quad \text{for a.e. } x \in \mathbb{R}^n.$$

We have therefore obtained something like a Fourier series of f_0 , however with summands $c_k e^{2\pi i \langle k, \cdot \rangle} \phi$, which are modified versions of $c_k e^{2\pi i \langle k, \cdot \rangle} \chi_{[-1, 1]^n}$.

It follows that

$$\hat{f}_0 = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k \hat{\phi}(\cdot - k).$$

Therefore, denoting

$$\psi := \hat{\phi}$$

and recalling by Observation 3.6.4 that

$$\hat{\phi} \sim 1 \quad \text{on} \quad \left[-\frac{1}{100}, \frac{1}{100} \right],$$

we deduce the following.

PROPOSITION 3.6.9. An uncertainty principle for functions supported on $[-1, 1]^n$.
There exists a function $\psi \in \mathcal{S}(\mathbb{R}^n)$, with

$$\psi \sim 1 \text{ on } \left[-\frac{1}{100}, \frac{1}{100} \right]^n,$$

such that, for every $f_0 \in L^2$ with $\text{supp } f_0 \subseteq [-1, 1]^n$,

$$\hat{f}_0 = \sum_{k \in \frac{1}{4}\mathbb{Z}^n} c_k \psi_k,$$

where

$$\psi_k := \psi(\cdot - k) \quad \text{for all } k \in \frac{1}{4}\mathbb{Z}^n.$$

→ **When f is supported on any rectangle centred at 0.** Let $f \in L^2$ be an arbitrary function supported on $\mathcal{R} := \prod_{j=1}^n [-r_j, r_j]$, for some $r_1, \dots, r_n > 0$. The rescaling $f_0(x) := f(r_1 x_1, \dots, r_n x_n)$ is supported on $[-1, 1]^n$. Therefore, applying Proposition 3.6.9 to this f_0 and rescaling back, we conclude the following.

PROPOSITION 3.6.10. (An uncertainty principle for functions supported on $\mathcal{R} := \prod_{j=1}^n [-r_j, r_j]$.)
There exists a function $\psi \in \mathcal{S}(\mathbb{R}^n)$, with

$$\psi \sim 1 \text{ on } \frac{1}{50n} \mathcal{R}^* \quad \text{and} \quad \int \psi \sim |\mathcal{R}^*|,$$

so that

$$\hat{f} = \sum_{k \in \frac{1}{4r_1}\mathbb{Z} \times \dots \times \frac{1}{4r_n}\mathbb{Z}} c_k \psi_k,$$

where

$$\psi_k := \psi(\cdot - k) \quad \text{for all } k \in \frac{1}{4r_1}\mathbb{Z} \times \dots \times \frac{1}{4r_n}\mathbb{Z}.$$

In fact, this ψ is a relabeling of the function $r_1 \cdots r_n \psi(r_1 \xi_1, \dots, r_n \xi_n)$, for the function ψ from the previous Proposition 3.6.9.

→ **When f is supported on any rectangle.** Let $\mathcal{R}$ be a rectangle centred at 0. Then, $\mathcal{R} = S \left(\prod_{j=1}^n [-r_j, r_j] \right)$, for some rotation S and some $r_1, \dots, r_n > 0$.

Let $f \in L^2$ be supported on $\mathcal{R}$. Then, the function

$$f_0 := f(S \cdot)$$

is supported on $\prod_{j=1}^n [-r_j, r_j]$, so its Fourier transform satisfies

$$\hat{f}_0 = \sum_{k \in \frac{1}{4r_1} \mathbb{Z} \times \cdots \times \frac{1}{4r_n} \mathbb{Z}} c_k \psi(\cdot - k)$$

by Proposition 3.6.10. Since

$$\hat{f}_0 = |\det S|^{-1} \cdot \hat{f} \circ (S^{-1})^T = \hat{f}(S \cdot),$$

it follows that

$$\begin{aligned} \hat{f}(\xi) &= \sum_{k \in \frac{1}{4r_1} \mathbb{Z} \times \cdots \times \frac{1}{4r_n} \mathbb{Z}} c_k \psi(S^{-1} \xi - k) \\ &= \sum_{k \in \frac{1}{4r_1} \mathbb{Z} \times \cdots \times \frac{1}{4r_n} \mathbb{Z}} c_k (\psi \circ S^{-1})(\xi - Sk) \\ &= \sum_{k \in S(\mathcal{G})} c'_k (\psi \circ S^{-1})(\xi - k), \end{aligned}$$

for the grid $\mathcal{G} = \frac{1}{4r_1} \mathbb{Z} \times \cdots \times \frac{1}{4r_n} \mathbb{Z}$.

Relabeling $\psi \circ S^{-1} = \psi(S^{-1} \cdot)$ by ψ , we conclude the following.

PROPOSITION 3.6.11. (An uncertainty principle for functions supported on rectangles.

Let $\mathcal{R}$ be a rectangle with centre 0. Then, $\mathcal{R} = S \left(\prod_{j=1}^n [-r_j, r_j] \right)$, for some rotation S and some $r_1, \dots, r_n > 0$. Let

$$\mathcal{G} = \frac{1}{4r_1} \mathbb{Z} \times \cdots \times \frac{1}{4r_n} \mathbb{Z}.$$

There exists a function $\psi \in \mathcal{S}(\mathbb{R}^n)$, with

$$\psi \sim 1 \text{ on } \frac{1}{50n} \mathcal{R}^* \quad \text{and} \quad \int \psi \sim |\mathcal{R}^*|,$$

so that the following hold.

(i) For any $f \in L^2$ supported on $\mathcal{R}$,

$$\hat{f} = \sum_{k \in S(\mathcal{G})} c_k \psi(\cdot - k).$$

(ii) For any $f \in L^2$ supported on $\mathcal{R} + \omega$, for some $\omega \in \mathbb{R}^n$,

$$\hat{f} = \sum_{k \in S(\mathcal{G})} c_k e^{-2\pi i \langle \omega, \cdot \rangle} \psi(\cdot - k).$$

Each $\psi(\cdot - k)$ is a mollified version of $\chi_{\mathcal{R}^* + k}$. Moreover, the rectangles $\mathcal{R}^* - k$, over all $k \in S(\mathcal{G})$, are practically disjoint translations of each other. Therefore, we may view the uncertainty principle in this case as follows.

Heuristic: When f is supported on a rectangle $\mathcal{R}$, then, roughly speaking, we can write

$$\hat{f} = \sum_T c_T e^{-2\pi i \langle \omega, \cdot \rangle} \chi_T,$$

where the T are practically disjoint rectangles, which are all translated copies of the dual rectangle $\mathcal{R}^*$, and together ‘tile’ $\mathbb{R}^n$. In other words, we may think of $\hat{f}$ as a sum of waves that do not interact much with each other, each wave being supported on a different one of these parallel rectangles T .

3.7. The proof of Fefferman’s ball multiplier theorem

Let $1 < p < +\infty$. Recall that the operator S_1 is well-defined on $L^p \cap L^2$, by

$$\widehat{S_1 f} = \chi_{B(0,1)} \hat{f}.$$

The question was: For which $p \neq 2$ is S_1 L^p -bounded? In other words, for which $p \neq 2$ is $\chi_{B(0,1)}$ an L^p -multiplier?

THEOREM 3.7.1. (Fefferman, 1971) *The function $\chi_{B(0,1)}$ is an L^p -multiplier only when $p = 2$.*

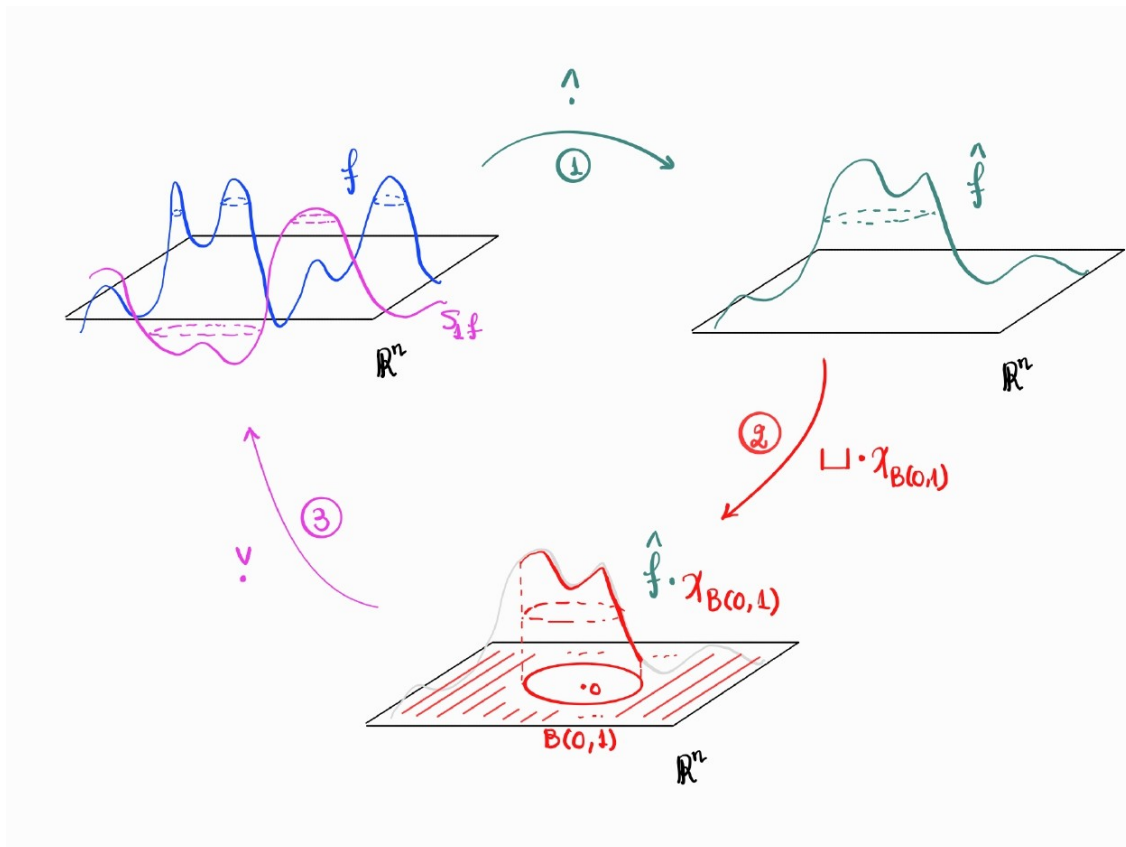
The rest of the section is dedicated to the proof of this theorem. We know that, for any $1 < p < +\infty$, $\chi_{B(0,1)}$ (and in fact any bounded, Lebesgue-measurable function) is an L^p -multiplier if and only if it is an $L^{p'}$ multiplier. Therefore, it suffices to show that:

$\chi_{B(0,1)}$ is **not** an L^p -multiplier, for any $p > 2$.

So, fix $p > 2$. We wish to show that $\|S_1\|_{p \rightarrow p} = +\infty$. Therefore, we have the following goal.

Goal. For every $M > 0$, find $f_M \in L^p \cap L^2$ so that $\|S_1 f_M\|_p > M \cdot \|f_M\|_p$.

→ **What does S_1 do?** It starts with a function $f \in L^p \cap L^2$, then considers $\hat{f}$, then alters $\hat{f}$ by restricting it to $B(0, 1)$, and then returns back by Fourier-inverting $\hat{f} \chi_{B(0,1)}$.



The point is to find functions f for which the final function $S_1 f$ has much larger p norm than f - e.g. has way more 'peaks' than f .

For simplicity we will work in two dimensions ($n = 2$), and then we will show that this case easily implies Fefferman's theorem in all dimensions.

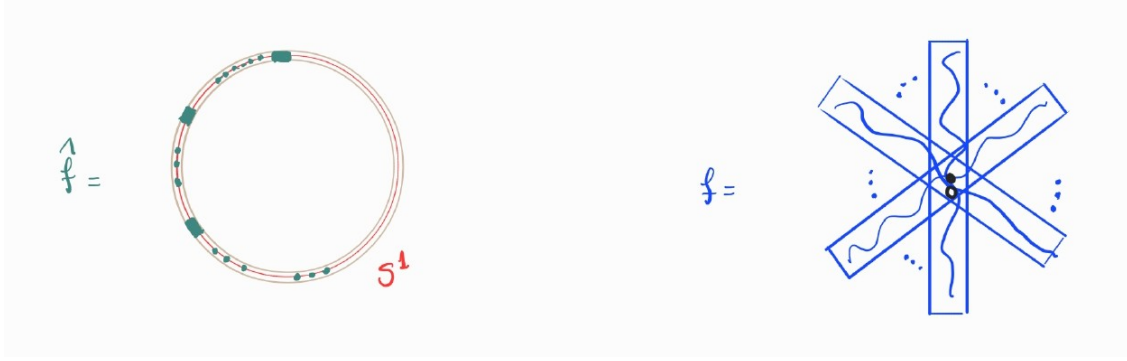
Main idea. Suppose that $\hat{f}$ is supported in an annulus A which is a thin neighbourhood of $\mathbb{S}^{n-1}$; in particular, with centre 0, inner radius $1 - R^{-1}$ and outer radius $1 + R^{-1}$.

Multiplying with $\chi_{B(0,1)}$ restricts $\hat{f}$ to *half the annulus*, in particular to the annulus with centre 0, same inner radius, but outer radius 1. So, $\hat{f}\chi_{B(0,1)}$ has support 'half' of the support of $\hat{f}$. Note that this support reduction happens to all little rectangles θ comprising A (that is, along all directions normal to the sphere $\mathbb{S}^{n-1}$).

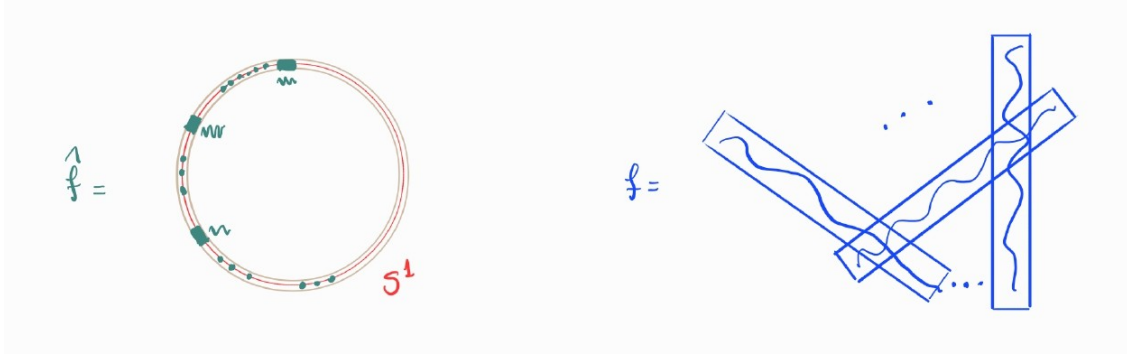
So, taking inverse Fourier transforms of all the rectangular pieces of both $\hat{f}$ and $\hat{f}\chi_{B(0,1)}$ (to compare f with $S_1 f$), the uncertainty principle tells us that $S_1 f$ should be 'doubly spread out' relative to f on all the dual rectangles (which point to all directions normal to $\mathbb{S}^{n-1}$).

So, the **strategy** will be the following:

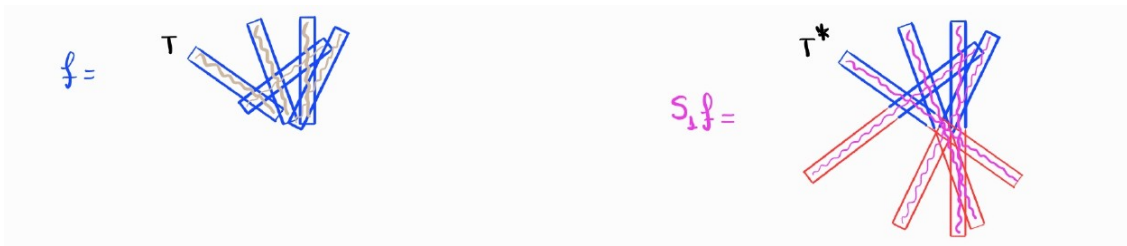
→ Start with $\hat{f}$ that is the sum of 'indicators' of little rectangles, which together form a thin neighbourhood of $\mathbb{S}^1$. Then, f is a sum of waves, each evolving within a tube through 0.



→ Giving appropriate oscillation to $\hat{f}$ within each little rectangle, we spread out the above tubes as we wish: For any family of rectangles T , we can make sure that f is a sum of waves f_T , each evolving in the rectangle T .



→ Multiplying $\chi_{B(0,1)}$ with $\hat{f}$ (to obtain $\widehat{S_1 f}$), each of the original little rectangles is split in half. Therefore, by the uncertainty principle, while each f_T was large in magnitude within the rectangle T , now $S_1 f_T$ should be large in magnitude within T^* , where T^* has the same width as T , but double the length.



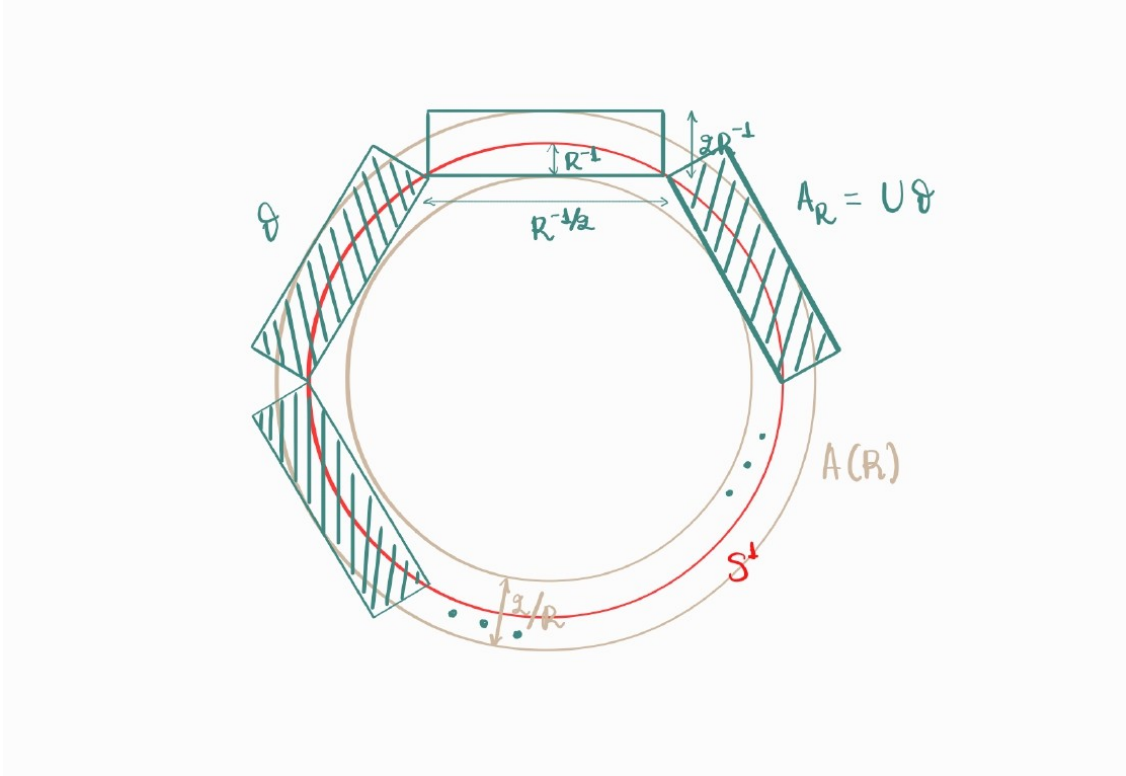
→ So, if we manage to find a configuration of rectangles T such that the T are **disjoint**, while the T^* **overlap a lot**, then the waves f_T don't interact with each other, while the waves Sf_T interact a lot with each other, allowing for the possibility of constructive interference (which would mean that Sf has a lot of peaks, while f does not). In such a case, it is possible to have $\|Sf\|_p \gg \|f\|_p$.

Let us now make formal sense of the above. Fix some large parameter $R \gg 1$.

Step 1: Defining the support A_R of $\hat{f}$. Consider the annulus

$$A(R) := \{x \in \mathbb{R}^n : 1 - R^{-1} \leq |x| \leq 1 + R^{-1}\}.$$

We approximate $A(R)$ by a union of disjoint $R^{-1/2} \times 2R^{-1}$ rectangles (or ‘caps’) θ .



The base of each rectangle is a chord of length $R^{-1/2}$ - then, the rectangle with this base and height R^{-1} is the smallest rectangle that contains the arc of S^1 with the same end-points as the chord. We simply double each rectangle in height, so that essentially the top half of each is outside $B(0, 1)$. (So, when we restrict to $B(0, 1)$ later, each of these rectangles will be essentially halved.)

Let Θ be the family of these rectangles (caps), which are $R^{1/2}$ in total. Define

$$A_R := \cup_{\theta \in \Theta} \theta.$$

This A_R (which is practically the original annulus) will be the support of our $\hat{f}$.

Defining $\hat{f}$. For every $\theta \in \Theta$, denote its centre by ω_θ .

Observation. For every $\theta \in \Theta$,

$$\left(\frac{\chi_\theta}{|\theta|} \right)^\sim \text{ “=” } e^{2\pi i \langle \cdot, \omega_\theta \rangle} \chi_{\theta^*}.$$

Therefore, if

$$\hat{f} \text{ “=” } \sum_{\theta \in \Theta} \frac{\chi_\theta}{|\theta|},$$

then

$$f \text{ “=” } \sum_{\theta \in \Theta} e^{2\pi i \langle \cdot, \omega_\theta \rangle} \chi_{\theta^*}.$$

The θ^* are $R^{-1/2}$ -separated $R^{1/2} \times R$ -rectangles with centre 0. This becomes precise when the χ_θ and the χ_{θ^*} are replaced by Schwartz versions.

To be precise, fix a function $\psi \in C_c^\infty(\mathbb{R}^n)$, with

$$\text{supp} \psi \subseteq [-1, 1]^2, \quad \psi \geq 0, \quad \psi = 1 \text{ on } \left[-\frac{1}{2}, \frac{1}{2}\right]^n.$$

This function does not depend on R , and will be fixed from now on. By our earlier uncertainty principle analysis, we know that⁷

$$\tilde{\psi} \sim 1 \quad \text{on } \left[-\frac{1}{100}, \frac{1}{100}\right]^n.$$

Since additionally $\tilde{\psi}$ is rapidly decreasing (as it is a Schwartz function), and $\|\tilde{\psi}\|_2 = \|\psi\|_2 \sim 1$, we may think of $\tilde{\psi}$ as a mollified version of $\chi_{[-1,1]^2}$.

For each $\theta \in \Theta$, denote by $\theta_0 := \theta - \omega_\theta$ the translation of θ through 0.

We adapt ψ to each θ_0 , over all $\theta \in \Theta$, via a rescaling and a rotation of ψ . So, for each θ_0 , we obtain a $\psi_{\theta_0} \in C_c^\infty(\mathbb{R}^n)$, with

$$\text{supp} \psi_{\theta_0} \subseteq \theta_0, \quad \psi_{\theta_0} \geq 0, \quad \psi_{\theta_0} = 1 \text{ on } \frac{1}{2}\theta_0, \quad \int \psi_{\theta_0} \sim |\theta_0|.$$

In other words, ψ_{θ_0} is a Schwartz version of χ_{θ_0} . By Observation 3.6.4,

$$(\psi_{\theta_0})^\sim \sim (\psi_{\theta_0})^\sim(0) = \int \psi \sim |\theta_0| \quad \text{on } \frac{1}{100}\theta_0^*.$$

Therefore, the function

$$\tilde{\psi}_{\theta_0^*} := \left(\frac{\psi_{\theta_0}}{|\theta_0|} \right)^\sim$$

is a Schwartz version of $\chi_{\theta_0^*}$:

$$|\tilde{\psi}_{\theta_0^*}| \sim 1 \text{ on } \frac{1}{100}\theta_0^*, \quad \int |\tilde{\psi}_{\theta_0^*}| \sim |\theta_0^*|.$$

Note that $\theta_0^* = \theta^*$.

The cap θ is a translation of θ_0 : $\theta = \theta_0 + \omega_\theta$, where ω_θ is the centre of θ . Therefore, we can adapt each ψ_{θ_0} to θ , via a translation:

$$\psi_\theta := \psi_{\theta_0}(\cdot - \omega_\theta).$$

It follows that

$$\text{supp} \frac{\psi_\theta}{|\theta|} \subseteq \theta, \quad \frac{\psi_\theta}{|\theta|} \geq 0, \quad \psi_\theta \sim 1 \text{ on } \frac{1}{2}\theta,$$

so $\frac{\psi_\theta}{|\theta|}$ is a Schwartz version of $\frac{\chi_\theta}{|\theta|}$, while

$$\left(\frac{\psi_\theta}{|\theta|} \right)^\sim = e^{2\pi i \langle \cdot, \omega_\theta \rangle} \left(\frac{\psi_{\theta_0}}{|\theta_0|} \right)^\sim = e^{2\pi i \langle \cdot, \omega_\theta \rangle} \tilde{\psi}_{\theta_0^*} \stackrel{\theta_0^* = \theta^*}{=} e^{2\pi i \langle \cdot, \omega_\theta \rangle} \tilde{\psi}_{\theta^*},$$

which is a Schwartz version of $e^{2\pi i \langle \cdot, \omega_\theta \rangle} \chi_{\theta^*}$.

Defining

$$\hat{f} := \sum_{\theta \in \Theta} \frac{\psi_\theta}{|\theta|},$$

it follows that

$$f = \sum_{\theta \in \Theta} \left(\frac{\psi_\theta}{|\theta|} \right)^\sim = \sum_{\theta \in \Theta} e^{2\pi i \langle \cdot, \omega_\theta \rangle} \tilde{\psi}_{\theta^*},$$

⁷As earlier, $\tilde{\psi}$ may take complex values, but they are all roughly 1 on this rectangle.

which is a sum of waves, each practically supported in one of the rectangles θ^* , which all pass through 0 and their directions are $R^{-1/2}$ -separated (just like the normals to any two caps of Θ).

$$f = \sum_{\theta \in \Theta} e^{2ni\langle \cdot, w_\theta \rangle} \tilde{\psi}_{\theta^*}$$

$$|f| = \sum_{\theta \in \Theta} e^{2ni\langle \cdot, w_\theta \rangle} \chi_{\theta^*}$$

$$\hat{f} = \sum_{\theta \in \Theta} \frac{\psi_\theta}{|\theta|}$$

$$|\hat{f}| = \sum_{\theta \in \Theta} \frac{\chi_\theta}{|\theta|}$$

Now observe that, by replacing each $\frac{\psi_\theta}{|\theta|}$ with an appropriate modulation of itself, we can translate the waves in the above picture anywhere we wish. Moreover, we can randomise f , by multiplying each summand with ± 1 at random. In other words, let $\{p_\theta\}_{\theta \in \Theta}$ be a collection of independent random variables, such that, for every $\theta \in \Theta$, p_θ takes the values ± 1 with equal probability.

$$f = \sum_{\theta \in \Theta} p_\theta \cdot e^{2ni\langle \cdot, -v_\theta, w_\theta \rangle} \chi_{\theta^* + v_\theta}$$

$$|\hat{f}| = \sum_{\theta \in \Theta} p_\theta \cdot e^{-2ni\langle \cdot, v_\theta \rangle} \frac{\chi_\theta}{|\theta|}$$

Heuristic. For any choice of vectors $v_\theta \in \mathbb{R}^n$, over all $\theta \in \Theta$, if

$$\hat{f} \text{ “ = ” } \sum_{\theta \in \Theta} p_\theta e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\chi_\theta}{|\theta|},$$

then

$$f \text{ “ = ” } \sum_{\theta \in \Theta} p_\theta e^{2\pi i \langle \cdot - v_\theta, \omega_\theta \rangle} \chi_{\theta^* + v_\theta}.$$

The $\theta^* + v_\theta$ are $R^{-1/2}$ -separated $R^{1/2} \times R$ -rectangles, each with centre v_θ .

To be precise, for every choice of vectors $v_\theta \in \mathbb{R}^n$, over all $\theta \in \Theta$, it holds that

$$\begin{aligned} \left(e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\psi_\theta}{|\theta|} \right)^\sim &= \left(\frac{\psi_\theta}{|\theta|} \right)^\sim (\cdot - v_\theta) \\ &= e^{2\pi i \langle \cdot - v_\theta, \omega_\theta \rangle} \tilde{\psi}_{\theta^*}(\cdot - v_\theta) \\ &=: e^{2\pi i \langle \cdot - v_\theta, \omega_\theta \rangle} \tilde{\psi}_{\theta^* + v_\theta}, \end{aligned}$$

where $\tilde{\psi}_{\theta^* + v_\theta} := \tilde{\psi}_{\theta^*}(\cdot - v_\theta)$ is a Schwartz version of $\chi_{\theta^* + v_\theta}$, the characteristic function of the rectangle $\theta^* + v_\theta$ (which has centre v_θ). Therefore, the function

$$\hat{f} := \sum_{\theta \in \Theta} p_\theta e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\psi_\theta}{|\theta|}$$

satisfies

$$f = \sum_{\theta \in \Theta} p_\theta e^{2\pi i \langle \cdot - v_\theta, \omega_\theta \rangle} \psi_{\theta^* + v_\theta}.$$

Of course, the rectangles $\theta^* + v_\theta$, over all $\theta \in \Theta$, still form a maximal $R^{-1/2}$ -separated family of $R^{1/2} \times R$ rectangles in $\mathbb{R}^n$.

Now, we switch our perspective to the rectangles instead. For every $R^{1/2} \times R$ rectangle T in $\mathbb{R}^n$, denote by $e(T)$ the direction of T , by $v(T)$ the centre of T , and by $\theta(T)$ the rectangle in Θ whose normal is the direction of T . The earlier analysis can be translated to the following:

Let $\mathbb{T}$ be a maximal family of $R^{-1/2}$ -separated $R^{1/2} \times R$ -rectangles in $\mathbb{R}^n$. Let $\{\omega_T\}_{T \in \mathbb{T}}$ be a collection of independent random variables, each taking the values ± 1 with equal probability. Then, we have shown a formal version of the following.

Heuristic. If

$$\hat{f} \text{ “ = ” } \sum_{T \in \mathbb{T}} \omega_T e^{-2\pi i \langle \cdot, v(T) \rangle} \frac{\chi_{\theta(T)}}{|\theta(T)|},$$

then

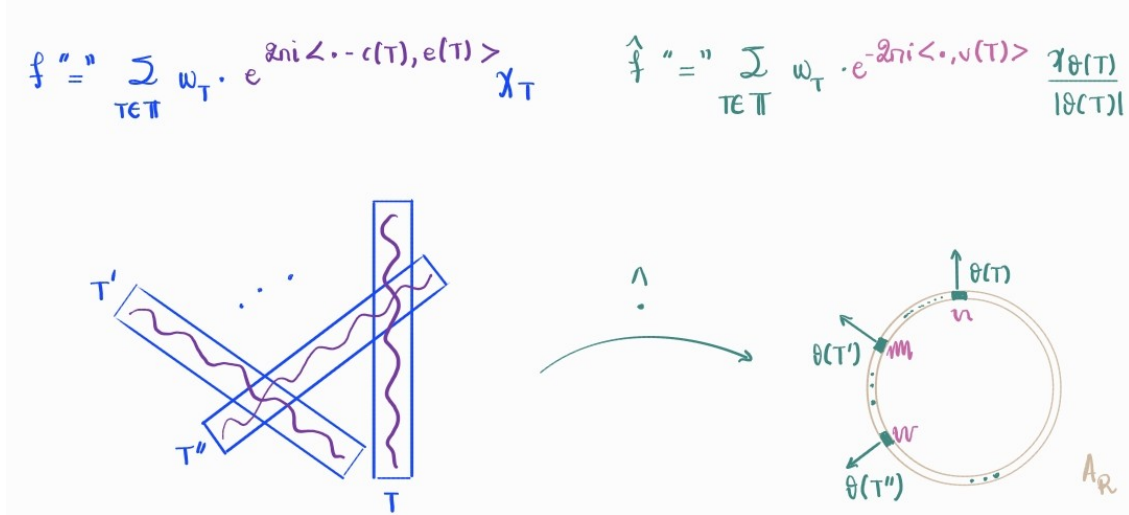
$$f \text{ “ = ” } \sum_{T \in \mathbb{T}} \omega_T e^{2\pi i \langle \cdot - v(T), e(T) \rangle} \chi_T.$$

Formally, we define

$$\hat{f} := \sum_{T \in \mathbb{T}} \omega_T e^{2\pi i \langle \cdot, v(T) \rangle} \frac{\psi_{\theta(T)}}{|\theta(T)|},$$

from which we conclude that

$$f = \sum_{T \in \mathbb{T}} \omega_T e^{-2\pi i \langle \cdot - v(T), e(T) \rangle} \tilde{\psi}_T.$$



We have a lot of freedom here: f depends on our choice $\mathbb{T}$ of maximal collection of $R^{-1/2}$ -separated rectangles, as well as on the signs we choose each time. The goal is to make a clever choice of these parameters, so that

$$\|S_1 f\|_p \gg \|f\|_p.$$

Understanding $S_1 f$.

→ *A heuristic explanation.*

Ignoring the randomisation of f , the above analysis is based on the fact that, for every $\theta \in \Theta$ and $v_\theta \in \mathbb{R}^n$, it holds that

$$\hat{f}|_\theta = e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\psi_\theta}{|\theta|}$$

and

$$\left(\hat{f}|_\theta\right)^\sim = \left(e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\psi_\theta}{|\theta|}\right)^\sim = e^{2\pi i \langle \cdot, -v_\theta, \omega_\theta \rangle} \tilde{\psi}_T,$$

where $T := \theta^* + v_\theta$ and the functions $\psi_\theta, \tilde{\psi}_T$ are Schwartz approximations of χ_θ and χ_T , respectively. Since

$$\hat{f} = \sum_{\theta \in \Theta} \hat{f}|_\theta,$$

it follows that

$$f = (\hat{f})^\sim = \sum_{\theta \in \Theta} \left(\hat{f}|_\theta\right)^\sim,$$

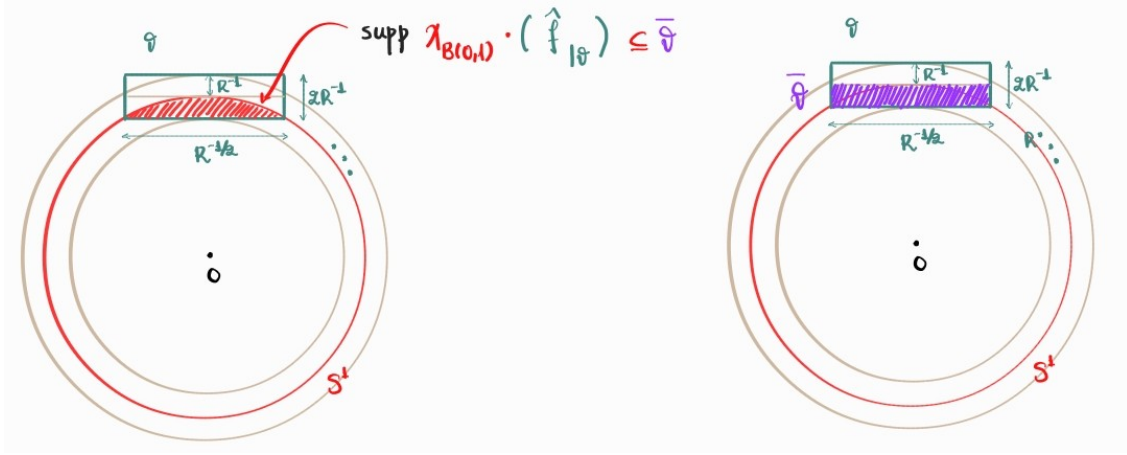
while

$$S_1 f = (\chi_{B(0,1)} \hat{f})^\sim = \sum_{\theta \in \Theta} \left(\chi_{B(0,1)} \hat{f}|_\theta\right)^\sim.$$

So, to compare f with $S_1 f$, we need to compare $\left(\hat{f}|_\theta\right)^\sim$ with $\left(\chi_{B(0,1)} \hat{f}|_\theta\right)^\sim$ for each $\theta \in \Theta$.

Let $\theta \in \Theta$. We begin by comparing $\hat{f}|_\theta$ with $\chi_{B(0,1)} \hat{f}|_\theta$. In other words, we compare $e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\psi_\theta}{|\theta|}$ with $\chi_{B(0,1)} e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\psi_\theta}{|\theta|}$.

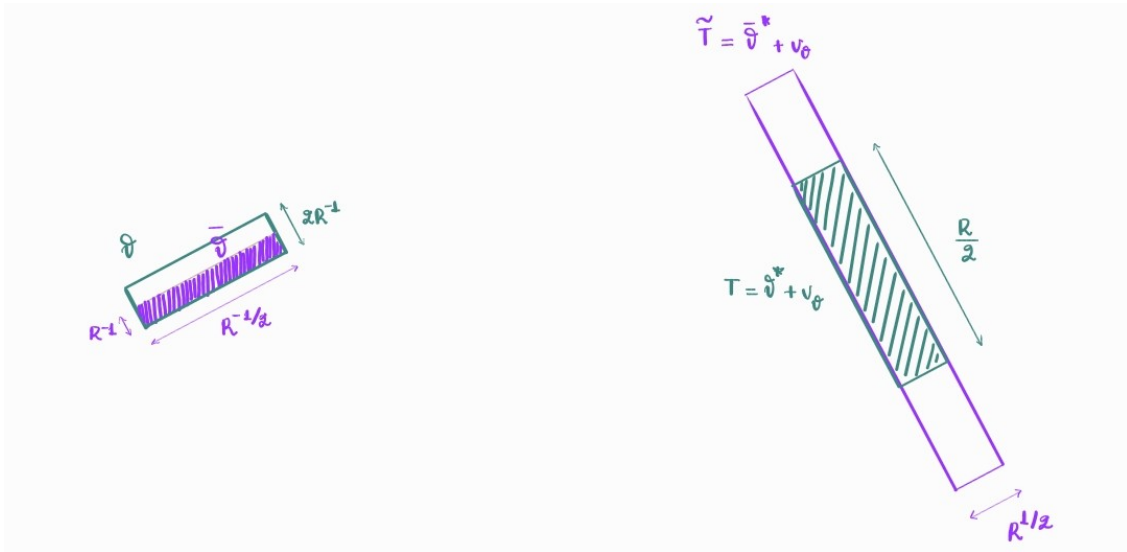
While $e^{-2\pi i \langle \cdot, v_\theta \rangle} \frac{\psi_\theta}{|\theta|}$ is localised on θ , multiplying it with $\chi_{B(0,1)}$ localises it to the bottom half of θ , i.e. to the rectangle $\bar{\theta}$ with the same base as θ , but with half the height.



By the uncertainty principle, we expect the function

$$\left(\chi_{B(0,1)} \cdot \hat{f}|_{\theta} \right)^{\vee} = \left(\chi_{B(0,1)} \cdot e^{-2\pi i \langle \cdot, v_{\theta} \rangle} \frac{\psi_{\theta}}{|\theta|} \right)^{\vee}$$

to be roughly constant in magnitude on all rectangles dual to $\bar{\theta}$, and thus on $\tilde{T} := \bar{\theta}^* + v_{\theta}$, the rectangle with the same centre as T , same width $R^{1/2}$, but *double the height*:



Observe that, if $\chi_{B(0,1)}\psi_{\theta}$ (which is supported on $\bar{\theta}$) was a smooth function $\phi_{\bar{\theta}}$, then the above would be obvious: we would conclude by the above analysis that

$$\left(\chi_{B(0,1)} \cdot \hat{f}|_{\theta} \right)^{\vee} = \left(\chi_{B(0,1)} \cdot e^{-2\pi i \langle \cdot, v_{\theta} \rangle} \frac{\psi_{\theta}}{|\theta|} \right)^{\vee} = \frac{1}{2} \left(e^{-2\pi i \langle \cdot, v_{\theta} \rangle} \frac{\phi_{\bar{\theta}}}{|\bar{\theta}|} \right)^{\vee} = \frac{1}{2} e^{2\pi i \langle \cdot, -v_{\theta}, \omega_{\bar{\theta}} \rangle} \tilde{\phi}_{\tilde{T}},$$

where $\tilde{\phi}_{\tilde{T}}$ is a Schwartz approximation of $\chi_{\tilde{T}}$. So, $\left| \left(\chi_{B(0,1)} \cdot \hat{f}|_{\theta} \right)^{\vee} \right|$ would be roughly constant on $\tilde{T}$.

While $\chi_{B(0,1)}\psi_{\theta}$ is not smooth, we will still be able to deduce that

$$\left| \left(\chi_{B(0,1)} \cdot \hat{f}|_{\theta} \right)^{\vee} \right| \text{ is large on } \tilde{T},$$

by Observation 3.6.4. Hence, for all $T \in \mathbb{T}$, it will hold that

$$\left| \left(\widehat{f}_{|\theta(T)} \right)^\vee \right| \quad \text{is large on } \widetilde{T}.$$

→ *A formal analysis.* For every $T \in \mathbb{T}$, denote

$$f_T := \omega_T e^{2\pi i \langle \cdot - v(T), e(T) \rangle} \widetilde{\psi}_T.$$

We know that

$$f = \sum_{T \in \mathbb{T}} f_T$$

and

$$\widehat{f} = \sum_{T \in \mathbb{T}} \widehat{f}_T,$$

where

$$\widehat{f}_T = \omega_T e^{-2\pi i \langle \cdot, v(T) \rangle} \frac{\psi_{\theta(T)}}{|\theta(T)|}.$$

Of course,

$$S_1 f = \sum_{T \in \mathbb{T}} S_1 f_T = \sum_{T \in \mathbb{T}} \left(\chi_{B(0,1)} \cdot \widehat{f}_T \right)^\vee,$$

so we will try to understand each of these pieces separately.

CLAIM 3.7.2. *For every $T \in \mathbb{T}$, it holds that*

$$|S_1 f_T| \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \widetilde{T},$$

where $\widetilde{T}$ is the rectangle with the same centre and axis as T , same width as T (so, with width $R^{1/2}$), but double the height (so, with height $2R$).

PROOF. The proof follows by Observation 3.6.4. For simplicity, we denote $\theta := \theta(T)$, $v := v(T)$, $e := e(T)$. Note that $T = \theta^* + v$. We can also ignore the sign ω_T , so that

$$f_T = e^{2\pi i \langle \cdot - v, e \rangle} \widetilde{\psi}_T$$

and

$$\widehat{f}_T = e^{-2\pi i \langle \cdot, v \rangle} \frac{\psi_\theta}{|\theta|},$$

where $\psi_\theta = \psi_{\theta_0}(\cdot - \omega_\theta)$ (recall, θ_0 is the translation of θ so that it has centre 0). We are interested in understanding

$$S_1 f_T = \left(e^{-2\pi i \langle \cdot, v \rangle} \chi_{B(0,1)} \frac{\psi_\theta}{|\theta|} \right)^\vee.$$

Observe that $\chi_{B(0,1)} \frac{\psi_\theta}{|\theta|}$ is a translation of $\phi := \chi_{B(0,1)-\omega_\theta} \frac{\psi_{\theta_0}}{|\theta_0|}$. Specifically,

$$\chi_{B(0,1)} \frac{\psi_\theta}{|\theta|} = \phi(\cdot - \omega_\theta).$$

The function $\phi = \chi_{B(0,1)-\omega_\theta} \frac{\psi_{\theta_0}}{|\theta_0|}$ is practically $\frac{\psi_{\theta_0}}{|\theta_0|}$, but with its support ‘cut in half’. In particular,

$$\text{supp } \phi \subseteq \overline{\theta_0}, \quad \phi \geq 0, \quad \int \phi \sim 1.$$

By Observation 3.6.4, if $\overline{\theta_0}$ was centred at 0, we would obtain

$$\check{\phi} \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \overline{\theta_0}^*, \quad \text{where } \overline{\theta_0}^* = \widetilde{T} - v.$$

However, the rectangle $\overline{\theta_0}$ is almost, but not quite, centred at 0: in truth, $\overline{\theta_0}$, which is very thin, touches 0 on its boundary. So, we first consider the translated function $\phi(\cdot - \omega_{\overline{\theta_0}})$, which has all the above nice properties of ϕ , but is also supported on the rectangle $\overline{\theta_0} - \omega_{\overline{\theta_0}}$ which is centred at 0. By Observation 3.6.4,

$$\left(\phi(\cdot - \omega_{\overline{\theta_0}}) \right)^\sim \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \overline{\theta_0}^*.$$

Since ϕ is a translation of $\phi(\cdot - \omega_{\overline{\theta_0}})$, it follows that

$$|\check{\phi}| \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \overline{\theta_0}^*.$$

And $\chi_{B(0,1)} \frac{\psi_\theta}{|\theta|}$ is in turn a translation of ϕ , so

$$\left| \left(\chi_{B(0,1)} \frac{\psi_\theta}{|\theta|} \right)^\sim \right| \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \overline{\theta_0}^*.$$

Finally, $e^{-2\pi i \langle \cdot, v \rangle} \chi_{B(0,1)} \frac{\psi_\theta}{|\theta|}$ is a modulation of $\chi_{B(0,1)} \frac{\psi_\theta}{|\theta|}$ by $e^{-2\pi i \langle \cdot, v \rangle}$, so

$$\left| \left(e^{-2\pi i \langle \cdot, v \rangle} \chi_{B(0,1)} \frac{\psi_\theta}{|\theta|} \right)^\sim \right| \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \overline{\theta_0}^* + v = \frac{1}{100} \tilde{T}.$$

We have thus concluded that

$$|S_1 f_T| \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \tilde{T}.$$

A small adjustment. There is a small issue here: For each $T \in \mathbb{T}$, we wanted $|S_1 f_T|$ to be large on a tube twice as long as T (or at least elongated relative to T), so that, while the f_T may not really interact along the original tubes T , the $S_1 f_T$ interact a lot along their elongations.

However, the statement of Claim 3.7.2 does not guarantee this: T is an $R^{1/2} \times R$ rectangle, while $\frac{1}{100} \tilde{T}$ is a rectangle with the same centre as T , and dimensions $\frac{1}{100} R^{1/2} \times \frac{1}{50} R$. So, $\tilde{T}$ is in fact contained in T .

We easily solve this issue by adjusting a little our original definition of $\hat{f}$ (and thus of f). In particular, instead of the original $R^{-1/2} \times R^{-1}$ rectangles θ in the definition of $\hat{f}$, we use instead the $R^{-1/2} \times 30R^{-1}$ rectangles with the same bases as the originals, but 30 times the height.

The analysis in this case is the same as before, with the following differences:

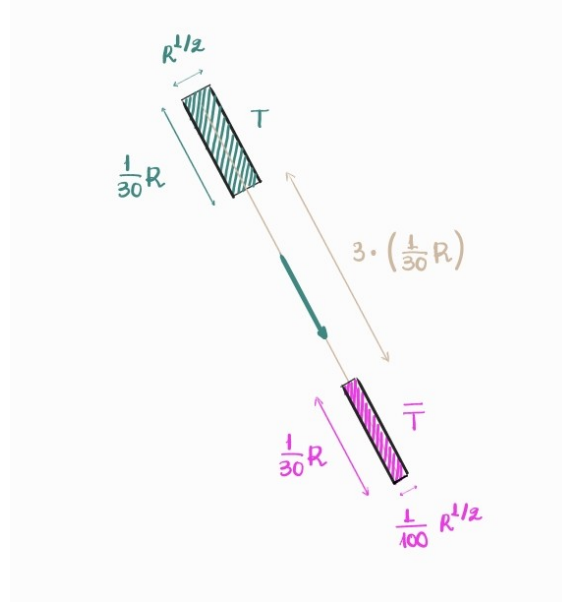
- The tubes $T \in \mathbb{T}$ have dimensions $R^{1/2} \times \frac{1}{30} R$.
- Restricting each θ to $B(0, 1)$ now reduces the height of θ by a factor of $\frac{1}{30}$, so, by the same uncertainty principle considerations as earlier,

$$|S_1 f_T| \gtrsim 1 \quad \text{on} \quad \frac{1}{100} \tilde{T},$$

where $\tilde{T}$ is the rectangle with the same centre and directions as T , but with dimensions $R^{1/2} \times 30R$. Therefore, $\frac{1}{100} \tilde{T}$ has the same centre as T , but dimensions $\frac{1}{100} R^{1/2} \times \frac{30}{100} R$.

So, even though $\frac{1}{100} \tilde{T}$ is a little thinner than T , it is considerably longer: it has length 9 times the length of T .

We denote by $\bar{T}$ the tube we obtain as follows: We translate T downwards along its axis, so that its bottom side touches the bottom side of $\frac{1}{100}\tilde{T}$. We then multiply the width of this tube by $\frac{1}{100}$, while preserving its central axis (so that it fits within $\frac{1}{100}\tilde{T}$).



We therefore have that $\bar{T}$ is a $\frac{1}{100}R^{1/2} \times \frac{1}{30}R$ tube (it is a translated copy of T , only a little thinner), and

$$|S_1 f_T| \gtrsim 1 \quad \text{on} \quad \bar{T}.$$

The tube $\bar{T}$ is often referred to as ‘the reach’ of T . It is a piece of an elongated version of T where something good happens. We record this information:

CLAIM 3.7.3. *For every $T \in \mathbb{T}$, it holds that*

$$|S_1 f_T| \gtrsim 1 \quad \text{on} \quad \bar{T}.$$

Assuming disjointness of the tubes T and calculating $\|f\|_p$. We assume from now on that the tubes T in $\mathbb{T}$ are disjoint. This helps us calculate $\|f\|_p$. Indeed, as we mentioned earlier, our f depends on the family $\mathbb{T}$ of tubes T we have chosen:

$$f = \sum_{T \in \mathbb{T}} \omega_T f_T,$$

where

$$f_T = e^{2\pi i \langle \cdot, -v(T), e(T) \rangle} \tilde{\psi}_T.$$

Due to the fact that $\tilde{\psi}_T$ is a Schwartz function adapted to T (it is a mollified version of χ_T), it decays relatively fast outside T , so in practice it is as if the f_T have disjoint supports. Therefore,

$$(12) \quad \|f\|_p^p \sim \sum_{T \in \mathbb{T}} \int |f_T|^p = \sum_{T \in \mathbb{T}} \int |\psi_T|^p \sim \sum_{T \in \mathbb{T}} \int |\chi_T|^p = \sum_{T \in \mathbb{T}} |T|.$$

Let us show the details. For every $T \in \mathbb{T}$ and every $x \in T$, there are $O_\psi(1)$ $T' \in \mathbb{T}$ for which $f_{T'}(x)$ is not tiny (those that intersect the $C_\psi R^{1/2}$ -neighbourhood of T , for a constant C_ψ that depends only on ψ - which is fixed throughout this analysis). Indeed, let us call this family of tubes $\mathbb{T}_T$. For the rest of the tubes $T' \in \mathbb{T} \setminus \mathbb{T}_T$, we can arrange them in increasing order according to their distance from T . So, due to the

disjointness of the tubes and the fact that we work in two dimensions, we can write $\mathbb{T} \setminus \mathbb{T}_T := \{T_1, T_2, \dots, T_m\}$ so that

$$\text{dist}(x, T_j) \geq (C_\psi + j)R^{1/2}, \quad \text{for all } j = 1, \dots, m.$$

Therefore, for some constant $c_\psi > 0$ (independent of c_ψ),

$$\begin{aligned} \left| \sum_{j=1}^m \omega_{T_j} f_{T_j}(x) \right| &\leq \sum_{j=1}^m |\omega_{T_j} f_{T_j}(x)| = \sum_{j=1}^m |\tilde{\psi}_T(x)| \\ &\leq \sum_{j=1}^m \frac{c_\psi}{(R^{-1/2} |\text{dist}(x, T_j)|)^2} \\ &\leq \sum_{j=1}^m \frac{c_\psi}{(C_\psi + j)^2} \leq \frac{c_\psi}{C_\psi} \sim 1. \end{aligned}$$

Therefore, for each $T \in \mathbb{T}$,

$$\int_T \left| \sum_{T' \in \mathbb{T} \setminus \mathbb{T}_T} \omega_{T'} f_{T'} \right|^p \lesssim |T| \sim \int_T |f_T|^p \lesssim \int |f_T|^p.$$

On the other hand, $\#(\mathbb{T} \setminus \mathbb{T}_T) \lesssim 1$, so, by Hölder's inequality,

$$\int_T \left| \sum_{T' \in \mathbb{T} \setminus \mathbb{T}_T} \omega_{T'} f_{T'} \right|^p \lesssim \sum_{T' \in \mathbb{T} \setminus \mathbb{T}_T} \int |f_{T'}|^p.$$

Therefore, for each $T \in \mathbb{T}$,

$$\int_T |f|^p \lesssim \sum_{T' \in \mathbb{T} \setminus \mathbb{T}_T} \int |f_{T'}|^p.$$

Since the tubes $T \in \mathbb{T}$ are disjoint, it follows that

$$\begin{aligned} \int |f|^p &= \sum_{T \in \mathbb{T}} \int |f|^p \lesssim \sum_{T \in \mathbb{T}} \sum_{T' \in \mathbb{T} \setminus \mathbb{T}_T} \int |f_{T'}|^p \\ &= \sum_{T' \in \mathbb{T}} \#\{T \in \mathbb{T} : T' \in \mathbb{T}_T\} \int |f_{T'}|^p \\ &\sim \sum_{T' \in \mathbb{T}} \int |f_{T'}|^p. \end{aligned}$$

Estimating $\|S_1 f\|_p$. We do this using Claim 3.7.3 and Khinchin's inequality. Since

$$f = \sum_{T \in \mathbb{T}} \omega_T f_T,$$

it follows that

$$\begin{aligned} \mathbb{E} \|S_1 f\|_p^p &= \mathbb{E} \int_{x \in \mathbb{R}^n} \left| \sum_{T \in \mathbb{T}} \omega_T \cdot S_1 f_T(x) \right|^p dx \\ &= \int_{x \in \mathbb{R}^n} \mathbb{E} \left| \sum_{T \in \mathbb{T}} \omega_T \cdot S_1 f_T(x) \right|^p dx \\ &\stackrel{\text{Khinchin}}{=} \int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} |S_1 f_T(x)|^2 \right)^{p/2} dx \\ &= \left\| \sum_{T \in \mathbb{T}} |S_1 f_T|^2 \right\|_{p/2}^{p/2} \end{aligned}$$

For every $T \in \mathbb{T}$, $|S_1 f_T| \geq 1$ on $\overline{T}$. Therefore,

$$\sum_{T \in \mathbb{T}} |S_1 f_T|^2 \gtrsim \sum_{T \in \mathbb{T}} \chi_{\overline{T}}$$

pointwise, where both functions above are non-negative. Hence,

$$\left\| \sum_{T \in \mathbb{T}} |S_1 f_T|^2 \right\|_{p/2}^{p/2} \geq \left\| \sum_{T \in \mathbb{T}} \chi_{\overline{T}} \right\|_{p/2}^{p/2}.$$

We conclude that

$$\mathbb{E} \|S_1 f\|_p^p \gtrsim \left\| \sum_{T \in \mathbb{T}} \chi_{\overline{T}} \right\|_{p/2}^{p/2},$$

hence there exists a choice of signs $\omega_T \in \{-1, 1\}$, over all $T \in \mathbb{T}$, so that

$$\|S_1 f\|_p^p \gtrsim \left\| \sum_{T \in \mathbb{T}} \chi_{\overline{T}} \right\|_{p/2}^{p/2},$$

irrespective of our choice of family $\mathbb{T}$ for which

$$f = \sum_{T \in \mathbb{T}} \omega_T f_T.$$

□

Determining $\mathbb{T}$ so that $\|S_1 f\|_p \gg \|f\|_p$. We have shown that, irrespective of our choice of our family $\mathbb{T}$ of rectangles, there exists a choice of signs ω_T , over all $T \in \mathbb{T}$, so that the function

$$f = \sum_{T \in \mathbb{T}} \omega_T f_T$$

satisfies

$$\|f\|_p^p \sim \sum_{T \in \mathbb{T}} |T|,$$

while

$$\|S_1 f\|_p^p \gtrsim \left\| \sum_{T \in \mathbb{T}} \chi_{\overline{T}} \right\|_{p/2}^{p/2} = \int \left(\sum_{T \in \mathbb{T}} \chi_{\overline{T}} \right)^{p/2}.$$

Therefore, in order to have

$$\|S_1 f\|_p \gg \|f\|_p,$$

it suffices to find a family $\mathbb{T}$ with the so far properties we have imposed (a maximal family of **disjoint** $R^{-1/2}$ -separated $R^{1/2} \times R$ -rectangles in $\mathbb{R}^2$), so that

$$(13) \quad \sum_{T \in \mathbb{T}} |T| \ll \int \left(\sum_{T \in \mathbb{T}} \chi_{\overline{T}} \right)^{p/2}.$$

Note that this is a purely *combinatorial* problem: it has nothing to do with waves.

Let us try to understand our desired inequality (13) a little better. Its LHS equals

$$\sum_{T \in \mathbb{T}} |T| = \sum_{T \in \mathbb{T}} \int \chi_T = \int \left(\sum_{T \in \mathbb{T}} \chi_T \right).$$

Observe that $\sum_{T \in \mathbb{T}} \chi_T(x)$ simply counts the number of tubes T that pass through each $x \in \mathbb{R}^n$, which in this specific case (where the tubes T are disjoint) equals either 0 or 1. Therefore, specifically for the families $\mathbb{T}$ that we are studying, the LHS of (13) equals

$$\sum_{T \in \mathbb{T}} |T| = \int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{p/2}$$

(as $0^{p/2} = 0^1$ and $1^{p/2} = 1^1$).

This expression is very similar to the one on the RHS; the only difference is that the RHS involves the tubes $\bar{T}$, the ‘reaches’ of the tubes T . That is, the RHS also counts the overlap of tubes (this time the tubes $\bar{T}$) with multiplicity, however this multiplicity will be much larger in general than 0 or 1, **if the tubes $\bar{T}$ overlap a lot** (unlike the tubes T): in that case, not only will the quantity $\sum_{T \in \mathbb{T}} \chi_T(x)$ be considerably larger than 1 for many x , but the quantity $(\sum_{T \in \mathbb{T}} \chi_T(x))^{p/2}$ will be even larger for those x , as it equals a natural number raised to the power $p/2 > 1$. Therefore, in order to create an example

Idea. In order to force (13), we will look for a maximal family $\mathbb{T}$ of **disjoint** $R^{-1/2}$ -separated $R^{1/2} \times R$ -rectangles T , so that **their ‘reaches’ $\bar{T}$ overlap a lot.**

The existence of such a family $\mathbb{T}$, for which (13) holds, is guaranteed by the Perron tree construction of Kakeya sets of arbitrarily small measure.

Finishing the argument. The key fact that we need is the following. It says that it is possible for the reaches $\bar{T}$ of the tubes T to cover much less space than the tubes T , which are disjoint. This must mean that they overlap a lot.

PROPOSITION 3.7.4. *For every $M > 0$, there exists a maximal family $\mathbb{T}$ of disjoint, $R^{-1/2}$ -separated, $R^{1/2} \times \frac{1}{30}R$ -tubes in $\mathbb{R}^2$, so that*

$$(14) \quad \left| \bigcup_{T \in \mathbb{T}} \bar{T} \right| \leq \frac{1}{M} \cdot \left| \bigcup_{T \in \mathbb{T}} T \right|.$$

We postpone the discussion of the proof of this proposition, to show that (14) implies the following quantified version of (13) (as $\frac{p}{2} > 1$):

$$(15) \quad \sum_{T \in \mathbb{T}} |T| \lesssim \left(\frac{1}{M} \right)^{\frac{p}{2}-1} \cdot \int \left(\sum_{T \in \mathbb{T}} \chi_{\bar{T}} \right)^{p/2}.$$

By our earlier analysis, this will mean that, for every $M > 0$, there exists (non-zero) f so that

$$\|S_1 f\|_p^p \gtrsim \left(\frac{1}{M} \right)^{\frac{p}{2}-1} \cdot \|f\|_p^p,$$

whence $\|S_1\|_{p \rightarrow p} < +\infty$. **And we will be done!**

→ (14)⇒(15): It holds that

$$\begin{aligned}
\sum_{T \in \mathbb{T}} |T| &\stackrel{|T| \sim |\bar{T}|}{\forall T \in \mathbb{T}} \sum_{T \in \mathbb{T}} |\bar{T}| = \sum_{T \in \mathbb{T}} \int \chi_{\bar{T}} \\
&= \int \left(\sum_{T \in \mathbb{T}} \chi_{\bar{T}} \right) \\
&\stackrel{\frac{p}{2} > 1, \text{ Hölder}}{\leq} \left(\int \left(\sum_{T \in \mathbb{T}} \chi_{\bar{T}} \right)^{\frac{p}{2}} \right)^{\frac{2}{p}} \cdot \left| \bigcup_{T \in \mathbb{T}} \bar{T} \right|^{1 - \frac{2}{p}} \\
&\stackrel{(14)}{\leq} \left(\int \left(\sum_{T \in \mathbb{T}} \chi_{\bar{T}} \right)^{\frac{p}{2}} \right)^{\frac{2}{p}} \left(\frac{1}{M} \right)^{1 - \frac{2}{p}} \left| \bigcup_{T \in \mathbb{T}} T \right|^{1 - \frac{2}{p}}.
\end{aligned}$$

Using that $\left| \bigcup_{T \in \mathbb{T}} T \right| = \sum_{T \in \mathbb{T}} |T|$ and rearranging, we deduce that

$$\left(\sum_{T \in \mathbb{T}} |T| \right)^{\frac{2}{p}} \lesssim \left(\int \left(\sum_{T \in \mathbb{T}} \chi_{\bar{T}} \right)^{\frac{p}{2}} \right)^{\frac{2}{p}} \left(\frac{1}{M} \right)^{1 - \frac{2}{p}},$$

hence

$$\sum_{T \in \mathbb{T}} |T| \lesssim \left(\frac{1}{M} \right)^{\frac{p}{2} - 1} \int \left(\sum_{T \in \mathbb{T}} \chi_{\bar{T}} \right)^{\frac{p}{2}},$$

as desired.

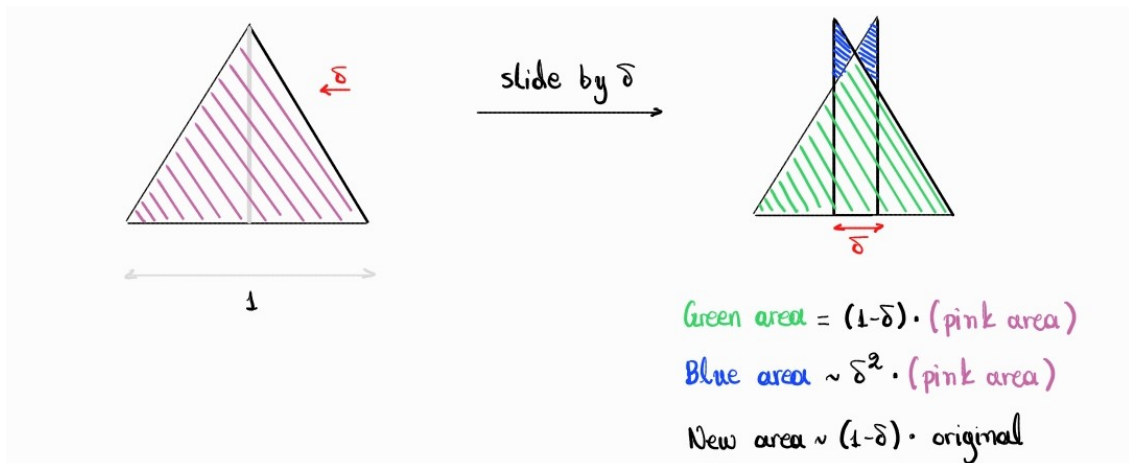
So, what remains is to prove Proposition 3.7.4. This is done easily once one knows about the Perron tree construction of *Kakeya sets of arbitrarily small measure*.

Kakeya sets of tiny measure: the Perron tree construction. A *Kakeya set* in $\mathbb{R}^n$ is a compact set in $\mathbb{R}^n$ that contains a unit line segment in every direction. In 1917, the Japanese mathematician Soichi Kakeya asked how large Kakeya sets can be.

It turns out that Kakeya sets can have (Lebesgue) measure 0. The first construction of Kakeya sets of arbitrarily small measure is due to Besicovitch (1919). The construction is iterative and is known as a Perron tree. We explain the construction in two dimensions - it can then easily be generalised to all dimensions.

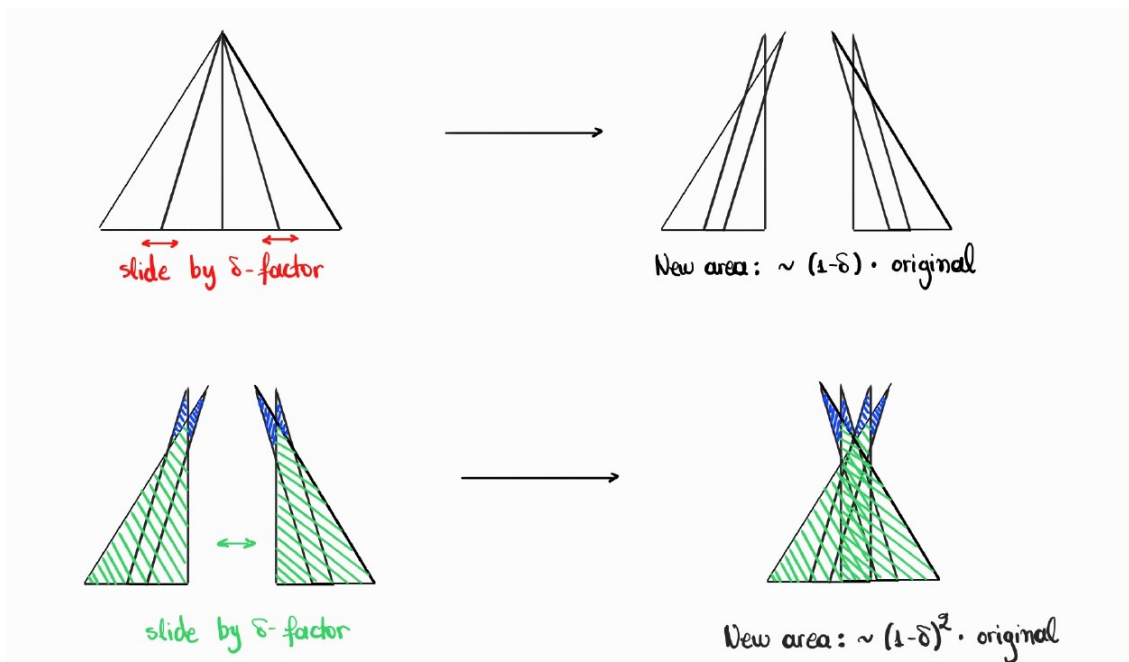
We begin with an equilateral triangle. All the line segments emanating from its top vertex and arriving at its base cover a 60° -worth of directions. We need to rearrange these line segments, so that their union has very small area. Then, rotating the final shape twice (so that we cover the remaining 120° -worth of directions), we will have created a Kakeya set of very small area.

→ We cut the triangle in half, and slide the two triangles a little bit along their bases. Then, the original line segments will simply be rearranged, but their union (the union of the two translated triangles) has smaller area than the original triangle. This is all quantifiable, according to the amount by which we choose to slide, and in fact holds for any triangle we begin with (independently of whether it is equilateral or not). In particular, if we slide the two triangles by a factor of δ (meaning that we slide until their overlap has length δ times the original triangle's base), then the new area will be smaller than the original by a factor of roughly $(1 - \delta)$.



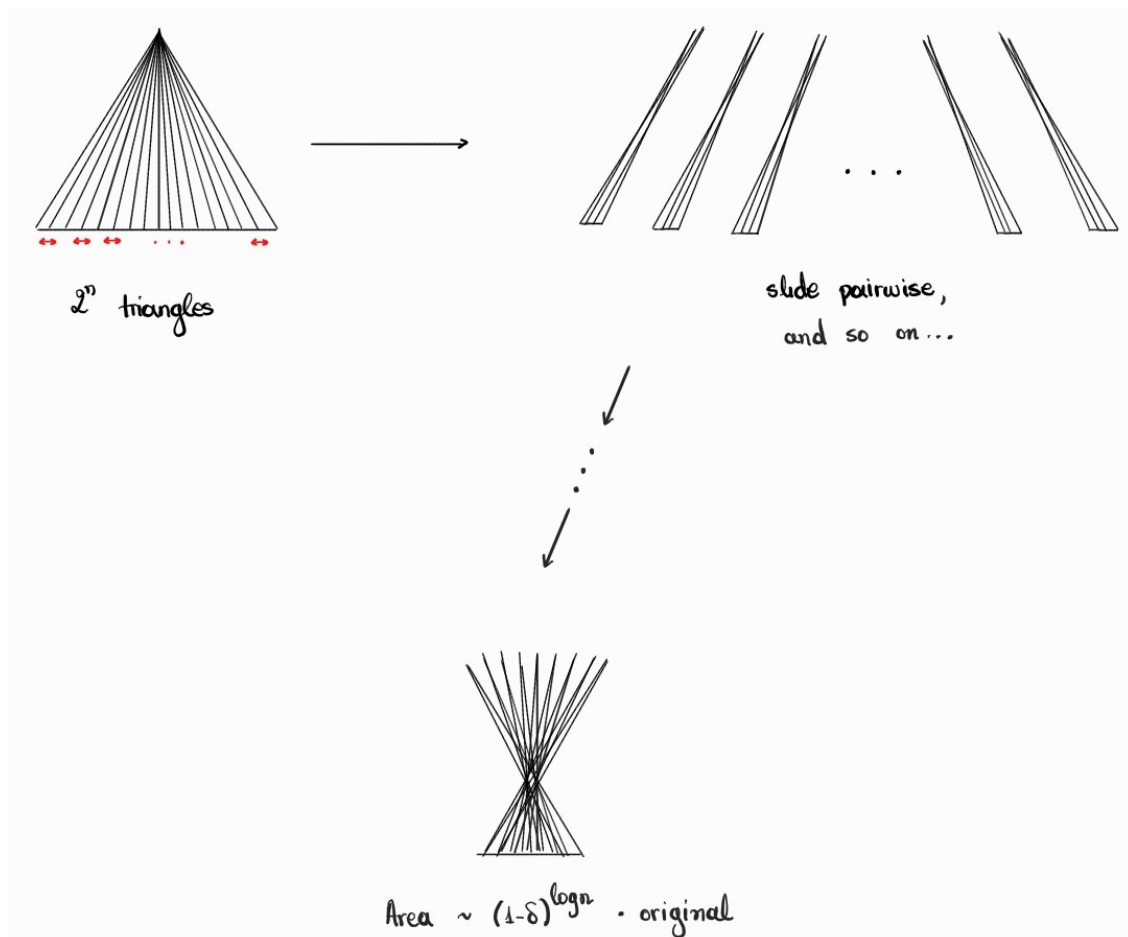
At this first stage, we have managed to create a Kakeya set (technically, involving a 60° -worth of directions), of smaller area than the original; in particular, area $(1 - \delta)$ times the original.

→ We now perform the above idea on each of the two pieces of the triangle independently, and then slide the two shapes together. More precisely, we cut the original triangle in four pieces: two subtriangles in its left half triangle T_1 , two subtriangles in its right half triangle T_2 . In each half, we slide the two subtriangles along their bases, by the same factor δ as before. This way, we create two shapes, each with smaller area than the original halves T_1 and T_2 (by a factor of roughly $(1 - \delta)$). Moreover, the left shape contains a triangle T'_1 similar to T_1 , and the right shape contains a triangle T'_2 , similar to T_2 . We now slide the two shapes along their bases, so that the bases of T'_1 and T'_2 overlap by a factor of δ . One can then see that the final shape has area roughly a $(1 - \delta)^2$ -portion of the original.



We have thus created a Kakeya set (OK, involving a 60° -worth of directions), with area much smaller than before, specifically roughly $(1 - \delta)^2$ times the original.

→ We apply the above idea, but this time splitting the original triangle into 2^n sub-triangles, rather than just 4, and iteratively splitting in consecutive pairs that we slide together. Letting n go to infinity, we create a Kakeya set of arbitrarily small area. There are $\log n$ steps until we end up with a single base and we stop sliding, and the final area will be roughly $(1 - \delta)^{\log n}$ times the original.



Details can be found in K. Falconer's book 'The Geometry of Fractal Sets', Section 7.2, or in T. Tao's Math 245B Fourier restriction course, Lecture Notes 3.

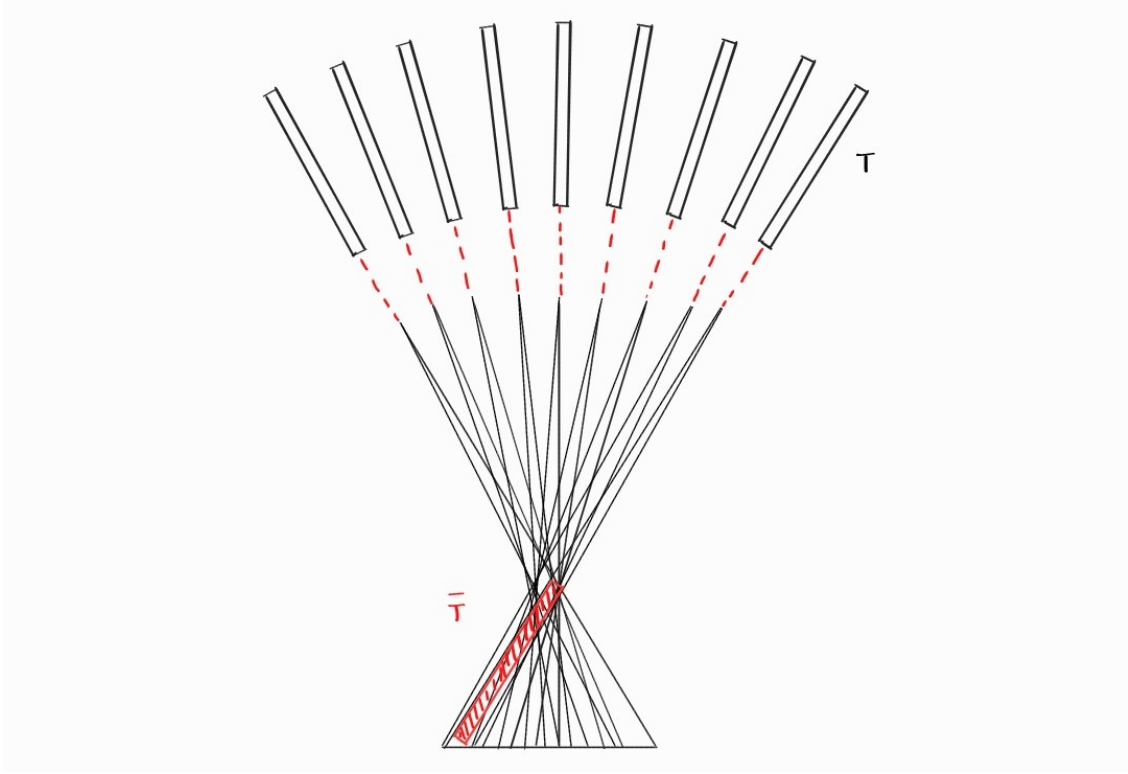
Proof of Proposition 3.7.4 Let $M > 0$. There exists a Kakeya set K as in the above construction, with

$$|K| < \frac{1}{M}.$$

This Kakeya set K comprises of 2^n triangles (for some appropriate $n \in \mathbb{N}$), emanating from a common base. All these triangles have base length 2^{-n} , and height 1. Moreover, the directions of any of the central axes of these triangles are 2^{-n} -separated. So, if for each of these triangles Δ_i we consider the tube $\overline{T}_i$, with the same central axis as Δ_i , but with half the base and half the height, then we get a collection of 2^n tubes, which are 2^{-n} -separated and have dimensions $\frac{1}{2}2^{-n} \times \frac{1}{2}$, and are all contained in K . Therefore,

$$\left| \bigcup_{i=1}^{2^n} \overline{T}_i \right| \leq \frac{1}{M}.$$

We have not chosen the notation $\overline{T}_i$ by chance: each $\overline{T}_i$ is the ‘reach’ of a tube T_i . We ‘thin’ each $\overline{T}_i$, by multiplying all their widths by the same appropriate constant, so that T_i , over all $i = 1, \dots, 2^n$, are pairwise disjoint. This is easy to achieve: if the bases of two tubes $\overline{T}_i, \overline{T}_j$ do not overlap, then it is trivial that the T_i, T_j do not meet either, as their directions point outwards; if the bases of the tubes $\overline{T}_i, \overline{T}_j$ overlap, then we use their angular separation to see that further away they will be disjoint.



Therefore, the family $\mathbb{T} := \{T_i : i = 1, \dots, 2^n\}$ is a disjoint family of 2^n tubes, which are 2^{-n} -separated and have dimensions $c2^{-n} \times \frac{1}{2}$. Therefore,

$$|\bigcup_{T \in \mathbb{T}} T| = \sum_{T \in \mathbb{T}} |T| \sim 1.$$

It follows that

$$|\bigcup_{T \in \mathbb{T}} \overline{T}| \lesssim \frac{1}{M} \cdot |\bigcup_{T \in \mathbb{T}} T|.$$

There exists R so that $R^{-1/2} = 2^{-n}$. Dilating the space $\mathbb{R}^2$ by R , we obtain a family $\mathbb{T}$ of $R^{1/2}$ tubes, which are $R^{-1/2}$ -separated and have dimensions $cR^{-1/2} \times \frac{R}{2}$, so that their reaches $\overline{T}$ satisfy

$$|\bigcup_{T \in \mathbb{T}} \overline{T}| \lesssim \frac{1}{M} \cdot |\bigcup_{T \in \mathbb{T}} T|.$$

We slightly adjust the sizes of the original tubes to get our desired statement. This is left as an exercise.

CHAPTER 4

The Kakeya conjectures

We have so far seen that, if we start with a function f supported in a small neighbourhood of a sphere, we can chop it up in its pieces f_θ , each supported on a different spherical cap θ . By the uncertainty principle, $\hat{f}_\theta$ is a sum of waves, each evolving inside a different tube, dual to θ . As θ varies, the corresponding dual tubes change direction. So, in total, we obtain a big family $\mathbb{T}$ of tubes, so that $\hat{f}$ is a sum of waves, each evolving inside a different tube of $\mathbb{T}$. Whenever any tubes of $\mathbb{T}$ meet (which can only happen if they have different directions), the waves that they carry interfere with each other - and understanding this interference leads us to understanding $\hat{f}$.

It appears therefore that, in harmonic analytic problems, tubes that point in different directions appear naturally - and that understanding their overlap helps with the original harmonic analytic problems as well.

Understanding the overlap of tubes that point in different directions is the aim of the **Kakeya conjectures**. These conjectures aim to control the quantities

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^p$$

for different p (for any family $\mathbb{T}$ of δ -separated $\delta \times \cdots \times \delta \times 1$ -tubes in $\mathbb{R}^n$), and aim to show that, *in some appropriate sense, such tubes behave as if they are disjoint*.

This of course needs to be explained (if only to understand why it doesn't contradict the existence of Kakeya sets of measure 0, which played an important role in the proof of Fefferman's ball multiplier theorem).

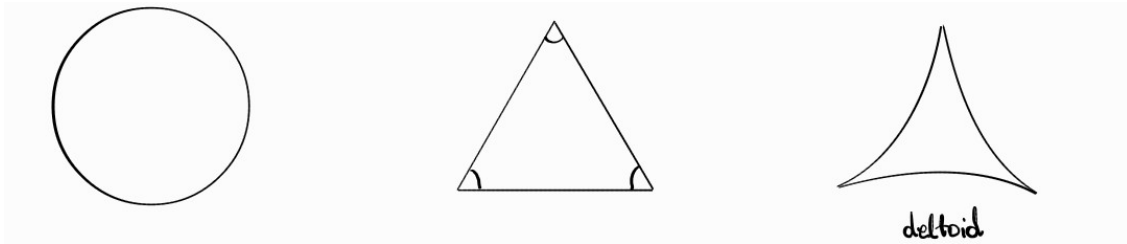
In this chapter, we will focus on the *Kakeya set conjecture* (which, if true, implies that Kakeya sets have full Hausdorff dimension) and on the stronger *Kakeya maximal operator conjecture*.

4.1. The Kakeya set conjecture

It all started in 1917, when Soichi Kakeya asked the following question:

Kakeya, 1917: How much area do we need, in order to continuously rotate a needle in the plane in all 360° ?

Shapes within which such continuous rotation is possible are a disc of diameter 1, and an equilateral triangle of side-length 1. The triangle has smaller area than the disc. Kakeya had come up with a third space of even smaller area with this property, the deltoid.

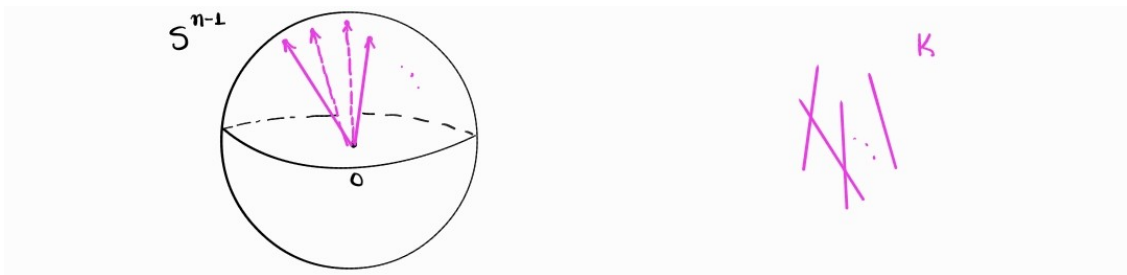


He believed that in fact the deltoid is optimal, in the sense that it achieves the smallest area in which rotation of the needle in all directions is possible. He was quite wrong:

Besicovitch, 1919: We can continuously rotate a needle in all 360° in as little area as we like.

In fact, if we don't insist for the rotation of the needle to be performed in a *continuous* manner, the above can even happen in zero area.

DEFINITION 4.1.1. (Kakeya set) A Kakeya set in $\mathbb{R}^n$ is a compact subset of $\mathbb{R}^n$, which contains a unit line segment in every direction.



THEOREM 4.1.2. There exist Kakeya sets in $\mathbb{R}^n$ of Lebesgue measure 0.

Sketch of proof. There are several constructions of such sets, such as the following by Kahane: in $\mathbb{R}^2$, take $C \times \{0\}$ and $C \times \{1\}$, where C is the middle thirds Cantor set. Connect with line segments every point of $C \times \{0\}$ with every point of $C \times \{1\}$. The union K of these line segments (united with appropriate $O(1)$ rotations of itself) is a Kakeya set in $\mathbb{R}^2$ with measure 0. Moreover, for any $n \geq 3$, the set $K \times \mathbb{R}^{n-2}$ is a Kakeya set in $\mathbb{R}^n$ of measure 0.

□

Even so, it feels that Kakeya sets should still be large: for every direction in $\mathbb{S}^{n-1}$ (that is, for an $(n - 1)$ -dimensional worth of directions), there exists a full unit line segment (1-dimensional) within K in that direction. In other words, it feels that the dimension of any Kakeya set in $\mathbb{R}^n$ should be $(n - 1) + 1 = n$, at least for an appropriate notion of dimension.

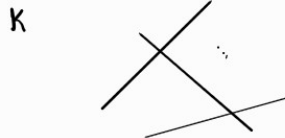
CONJECTURE 4.1.3. (The Kakeya set conjecture) Every Kakeya set in $\mathbb{R}^n$ has Hausdorff (and thus Minkowski) dimension equal to n .

Note that this conjecture is interesting exactly because there exist Kakeya sets in any dimension with Lebesgue measure 0. Indeed, for any Kakeya set K with $\lambda(K) > 0$, we automatically know that $\dim_H K = n$ (and thus $\dim_M K = n$).

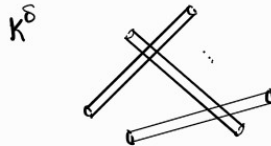
→ **Focusing on the Minkowski dimension.** Let K be a Kakeya set in $\mathbb{R}^n$. In order to show that K has Minkowski dimension n , we wish to show that, even though K may have volume 0, any δ -neighbourhood of K has relatively large volume. In particular:

Kakeya set conjecture: $|K^\delta| \gtrsim 1$ for every $\delta > 0$.

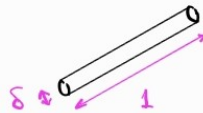
To achieve this, we want to use the fact that K contains a unit line segment in every direction.



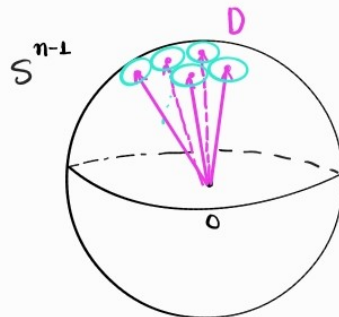
This implies that K^δ contains a δ -tube in every direction,



where a δ -tube is any tube with cross-section of diameter δ and with length 1:



• **A sufficient condition.** By the above, considering D to be any maximal set of δ -separated directions on $\mathbb{S}^{n-1}$ (by covering $\mathbb{S}^{n-1}$ by $O(1)$ -overlapping δ -caps and considering their centres as the elements of D), we see that in particular the δ -tubes with directions in D live inside K^δ .



Therefore, K^δ contains a maximal family $\mathbb{T}$ of δ -separated δ -tubes. Note that

$$\#\mathbb{T} = \#D = \frac{\text{surface area of } \mathbb{S}^{n-1}}{\text{surface area of a } \delta\text{-cap}} \sim \frac{1}{\delta^{n-1}}.$$

Therefore, in order to show that $\dim_M K = n$, it suffices to show that, for any $\delta > 0$ and any maximal family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$, it holds that

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \gtrsim 1.$$

• **A necessary condition.** We will now see that the above condition is also necessary. Indeed, consider any maximal family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$. Then, every $T \in \mathbb{T}$ contains a line segment of length ≥ 1 in every direction $e \in \mathbb{S}^{n-1}$ with $|e - e(T)| \lesssim \delta$.

Therefore, $\bigcup_{T \in \mathbb{T}} T$ contains a Kakeya set. So, for the Kakeya set conjecture to hold, it is necessary for any δ -neighbourhood of $\bigcup_{T \in \mathbb{T}} T$ to have volume $\gtrsim 1$. However, this δ -neighbourhood is in turn contained in $2 \bigcup_{T \in \mathbb{T}} T$ (for small δ , which are anyway the ones that matter).

We conclude that, if the Kakeya set conjecture holds, then necessarily $\left| 2 \bigcup_{T \in \mathbb{T}} T \right| \gtrsim 1$, and hence

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \gtrsim 1,$$

for any $\delta > 0$ and any maximal family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$.

We have thus found an equivalent form of the (Minkowski version of the) Kakeya set conjecture.

CONJECTURE 4.1.4. (The Kakeya set conjecture, an equivalent formulation) *For any $\delta > 0$, any family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$ satisfies the estimate*

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \gtrsim 1.$$

→ Let us now understand the conjecture a little better. The quantity 1 on the right-hand side equals

$$1 = \frac{1}{\delta^{n-1}} \cdot \delta^{n-1} \sim \#\mathbb{T} \cdot (\text{volume of each tube in } \mathbb{T}) = \sum_{T \in \mathbb{T}} |T|,$$

the sum of the volumes of the tubes, which is equal to the volume their union would have **if the tubes were pairwise disjoint**.

So, the Kakeya set conjecture claims that, for any maximal family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$,

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \gtrsim \sum_{T \in \mathbb{T}} |T|.$$

This seems counter-intuitive, as in fact the *reverse* inequality

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \leq \sum_{T \in \mathbb{T}} |T|$$

trivially holds: the volume of the union is at most what it would be if the tubes were disjoint. What the Kakeya set conjecture claims is that $\left| \bigcup_{T \in \mathbb{T}} T \right|$ **is still comparable to the bigger quantity** $\sum_{T \in \mathbb{T}} |T|$, if we are not bothered to ‘slightly’ decrease $\sum_{T \in \mathbb{T}} |T|$ by multiplying it with a power of $(\log \frac{1}{\delta})^{-1}$. In particular, it claims that there exist constants $c_n, C_n > 0$ so that

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \geq c_n \left(\frac{1}{\log \frac{1}{\delta}} \right)^{C_n} \cdot \sum_{T \in \mathbb{T}} |T|.$$

The same is conjectured to be true for any family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$, whether it is maximal or not. We thus write a final form of the Kakeya set conjecture.

The Kakeya set conjecture. For any $\delta > 0$, any family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$ satisfies

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \approx \sum_{T \in \mathbb{T}} |T|.$$

Moral of the Kakeya set conjecture: *Tubes that point in (truly) different directions behave as if they are disjoint, in that their union is not too much smaller (in the $\approx$ sense) than what it would be if they were disjoint.*

The Kakeya set conjecture is only known in dimensions $n = 2$ (Davies, 1971) and $n = 3$ (Wang and Zahl, 2025).

4.2. The Kakeya maximal operator conjecture

We will now see a stronger conjecture than the Kakeya set conjecture, known as the *Kakeya maximal operator conjecture*. This conjecture claims again that tubes that point in different directions should behave as if they are disjoint, but this time in terms of their overlap.

In particular, let $\delta > 0$ and let $\mathbb{T}$ be a family of δ -separated δ -tubes in $\mathbb{R}^n$. Let us slowly formulate the conjecture.

- Roughly speaking, the conjecture will claim that, in principle, $\lesssim 1$ tube of $\mathbb{T}$ will pass through each point of $\mathbb{R}^n$.
- For every $x \in \mathbb{R}^n$,

$$\sum_{T \in \mathbb{T}} \chi_T(x) = \#\{\text{tubes in } \mathbb{T} \text{ through } x\}.$$

Thus, a way to understand whether few or many tubes pass through x is to compare the quantities

$$\left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^p \quad \text{and} \quad \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^1$$

for some $p > 1$. These quantities are equal exactly when either 0 or 1 tubes pass through x , however they start to differ significantly the more tubes pass through x (with the p -th power becoming much larger than the 1-power). Similarly, the quantities are comparable in the $\approx$ -sense, i.e.

$$\left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^p \approx \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^1,$$

exactly when $\lesssim 1$ tubes of $\mathbb{T}$ pass through x .

- The Kakeya maximal operator conjecture states that, for $p = \frac{n}{n-1}$ (which is > 1), it indeed holds that

$$\left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^{\frac{n}{n-1}} \approx \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^1 \quad \text{for 'most' } x \in \mathbb{R}^n,$$

and that therefore indeed $\lesssim 1$ tubes of $\mathbb{T}$ should pass through 'most' $x \in \mathbb{R}^n$.

Precisely, it claims that

$$\int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^{\frac{n}{n-1}} dx \approx \int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^1 dx.$$

- Observe that the $\geq$ direction is trivial: it always holds that

$$\left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^{\frac{n}{n-1}} \geq \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^1 \quad \text{for all } x \in \mathbb{R}^n,$$

hence

$$\int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^{\frac{n}{n-1}} dx \geq \int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^1 dx.$$

Therefore, the interesting direction is the opposite one: to show that, even though the LHS is larger than the RHS, we can still reduce it enough by multiplying it with an appropriate power of $(\log \frac{1}{\delta})^{-1}$ so that it falls below the RHS. Therefore, the essence of the Kakeya maximal operator conjecture is the claim that

$$\int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^{\frac{n}{n-1}} dx \lesssim \int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^1 dx.$$

- Finally, we observe that the RHS equals

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right) = \sum_{T \in \mathbb{T}} \int \chi_T = \sum_{T \in \mathbb{T}} |T| = \#\mathbb{T} \cdot \delta^{n-1}.$$

We are now ready to formulate the conjecture.

The Kakeya maximal operator conjecture. For any $\delta > 0$, any family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$ satisfies

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{\frac{n}{n-1}} \lesssim \#\mathbb{T} \cdot \delta^{n-1}.$$

As we have already explained, the $\geq$ direction trivially holds.

Moral of the Kakeya maximal operator conjecture: Tubes that point in (truly) different directions behave as if they are disjoint, in the sense that, in principle, $\lesssim 1$ tube of $\mathbb{T}$ passes through each point of $\mathbb{R}^n$.

The **Keakeya** maximal operator conjecture is only known to be true in $\mathbb{R}^2$ (Cordoba, 1977).

→ **Remark: The Keakeya maximal operator.** It is probably unclear to the reader where the expression ‘maximal operator’ is coming from here, as there is no maximal function featuring in the above analysis. It turns out that there exists an *equivalent* formulation of the conjecture, that involves the *Keakeya maximal operator* (or *function*).

In particular, for any $\delta > 0$, the Keakeya maximal operator K^δ is an operator that maps any measurable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ to the function $K^\delta f : \mathbb{S}^{n-1} \rightarrow [0, +\infty]$, with

$$K^\delta f(e) := \sup_{\delta\text{-tubes } T \parallel e} \frac{\int_T |f|}{|T|}, \quad \text{for all } e \in \mathbb{S}^{n-1}.$$

In other words, $K^\delta f$, evaluated at the direction e , gives the largest average of $|f|$ on any δ -tube in $\mathbb{R}^n$ that points in direction e . The original formulation of the Keakeya maximal operator conjecture states, for every $n \geq 2$, K^δ is a bounded operator from $L^n(\mathbb{R}^n)$ to $L^n(\mathbb{S}^{n-1})$, and in particular that

$$(16) \quad \|K^\delta f\|_n \lesssim \|f\|_n \quad \text{for all } f \in L^n(\mathbb{R}^n).$$

Now, L^n -boundedness of the operator K^δ is equivalent to $L^{\frac{n}{n-1}}$ -boundedness of the dual operator (as $\frac{n}{n-1} = n'$). When one finds the dual operator (which is easy), it is not too hard to show that its $L^{\frac{n}{n-1}}$ -boundedness is equivalent to our formulation of the Keakeya maximal operator conjecture. Therefore, our formulation (which involves only counting how often tubes in $\mathbb{R}^n$ that point in different directions meet) is equivalent to (16), for any dimension $n \geq 2$. □

4.3. Keakeya max $\implies$ Keakeya

The Keakeya maximal operator conjecture implies the Keakeya set conjecture. We will here prove this implication for the Minkowski dimension version of the Keakeya set conjecture. There exists a proof for the Hausdorff dimension version as well, however it is more complicated and uses the Keakeya maximal function formulation that we just discussed (so we choose not to present it here).¹

PROPOSITION 4.3.1. *The Keakeya maximal operator conjecture in $\mathbb{R}^n$ implies the Keakeya set conjecture in $\mathbb{R}^n$. In particular, it implies that Keakeya sets in $\mathbb{R}^n$ have Hausdorff (and Minkowski) dimension n .*

Proof for Minkowski dimension. Let $n \geq 2$. Let $\delta > 0$ and let $\mathbb{T}$ be a family of δ -separated δ -tubes in $\mathbb{R}^n$, for which

$$(17) \quad \int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{\frac{n}{n-1}} \lesssim \#\mathbb{T} \cdot \delta^{n-1} \quad \left(= \sum_{T \in \mathbb{T}} |T| \right).$$

We will show that

$$(18) \quad \left| \bigcup_{T \in \mathbb{T}} T \right| \gtrsim \sum_{T \in \mathbb{T}} |T|.$$

¹Details can be found in Wolff’s lecture notes, p.74.

Indeed, we begin with our desired RHS and aim to bound it by the LHS above:

$$\begin{aligned} \sum_{T \in \mathbb{T}} |T| &= \sum_{T \in \mathbb{T}} \int \chi_T = \int \left(\sum_{T \in \mathbb{T}} \chi_T \right) \\ &\stackrel{\text{Hölder}}{\lesssim} \left[\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{\frac{n}{n-1}} \right]^{\frac{n-1}{n}} \cdot \left| \bigcup_{T \in \mathbb{T}} T \right|^{\frac{1}{n}} \\ &\stackrel{(17)}{\approx} \left(\sum_{T \in \mathbb{T}} |T| \right)^{\frac{n-1}{n}} \cdot \left| \bigcup_{T \in \mathbb{T}} T \right|^{\frac{1}{n}}, \end{aligned}$$

and, by rearranging,

$$\left(\sum_{T \in \mathbb{T}} |T| \right)^{\frac{1}{n}} \lesssim \left| \bigcup_{T \in \mathbb{T}} T \right|^{\frac{1}{n}},$$

which is the desired estimate (18). □

4.4. The Kakeya maximal operator conjecture holds in $\mathbb{R}^2$.

When $n = 2$, $\frac{n}{n-1} = 2$ and $\delta^{n-1} = \delta$. So, the Kakeya maximal operator conjecture states that

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^2 \lesssim \#\mathbb{T} \cdot \delta,$$

for all $\delta > 0$ and every family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^2$.

It turns out that this can be easily proved, exactly because the power $\frac{n}{n-1}$ is an integer, and thus has a combinatorial interpretation (in this case, what we essentially need is that, if we have n objects, n^2 is the number of pairs of these objects). The argument also involves a dyadic decomposition, something very common in harmonic analysis.

THEOREM 4.4.1. (Córdoba, 1977) *The Kakeya maximal operator conjecture holds in $\mathbb{R}^2$. It follows that Kakeya sets in $\mathbb{R}^2$ have Hausdorff (and Minkowski) dimension equal to 2.*

PROOF. Let $\delta > 0$ and let $\mathbb{T}$ be a family of δ -separated δ -tubes in $\mathbb{R}^2$. We wish to show that

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^2 \lesssim \#\mathbb{T} \cdot \delta \quad \left(= \sum_{T \in \mathbb{T}} |T| \right).$$

By expanding out the square in the LHS, we obtain

$$\begin{aligned} \int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^2 &= \int \left(\sum_{T \in \mathbb{T}} \chi_T \right) \cdot \left(\sum_{T' \in \mathbb{T}} \chi_{T'} \right) \\ &= \sum_{(T, T') \in \mathbb{T}^2: T \neq T'} \int \chi_T \chi_{T'} + \sum_{(T, T') \in \mathbb{T}^2: T = T'} \int \chi_T \chi_{T'} \\ &= \sum_{(T, T') \in \mathbb{T}^2: T \neq T'} |T \cap T'| + \sum_{T \in \mathbb{T}} |T|. \end{aligned}$$

We have thus split into diagonal terms ($T = T'$) and off-diagonal terms ($T \neq T'$). The contribution of the diagonal terms is

$$\sum_{T \in \mathbb{T}} |T| \quad (= \#\mathbb{T} \cdot \delta),$$

which is our desired RHS. So, the diagonal terms are harmless, and it remains to show that the off-diagonal contribution is just as small (up to logarithms); that is, to show that

$$\sum_{(T, T') \in \mathbb{T}^2: T \neq T'} |T \cap T'| \lesssim \#\mathbb{T} \cdot \delta$$

as well. It is easy to see that, for any two δ -separated δ -tubes T, T' that intersect,

$$|T \cap T'| \lesssim \frac{\delta^2}{\alpha},$$

where α is the acute angle between the directions of the two tubes (with equality in the $\sim$ -sense if the tubes fully intersect). In other words,

$$(19) \quad \sum_{(T, T') \in \mathbb{T}^2: T \neq T'} |T \cap T'| \lesssim \sum_{T \in \mathbb{T}} \sum_{T' \in \mathbb{T}: T' \neq T} \frac{\delta^2}{\text{Angle}(T, T')}.$$

For any given tube T , the acute angle α at which any other tube in $\mathbb{T}$ can meet T is in the interval $[\delta, \pi/2]$. We split this interval into M dyadic groups $[\delta, 2\delta), [2\delta, 2^2\delta), [2^2\delta, 2^3\delta), \dots, [2^{M-1}\delta, 2^M\delta)$, where $2^M\delta$ is the smallest dyadic multiple of δ larger than $\pi/2$. This means that the number M of groups satisfies

$$2^M\delta \sim 1,$$

and thus²

$$M \sim \log \frac{1}{\delta} \approx 1.$$

Therefore, for every fixed $T \in \mathbb{T}$,

$$(20) \quad \sum_{T' \in \mathbb{T}: T' \neq T} \frac{\delta^2}{\text{Angle}(T, T')} \sim \sum_{k=1}^M \sum_{T' \in \mathbb{T}: \text{Angle}(T, T') \in [2^{k-1}\delta, 2^k\delta]} \frac{\delta^2}{2^k\delta} \\ \sim \sum_{k=1}^M \#\{T' \in \mathbb{T} : \text{Angle}(T, T') \sim 2^k\delta\} \cdot \frac{\delta}{2^k}.$$

It remains to see how many $T' \in \mathbb{T}$ can contribute to the inner sum above, that is, to control

$$\#\{T' \in \mathbb{T} : \text{Angle}(T, T') \sim 2^k\delta\}.$$

These tubes are as many as their directions (as their directions are distinct). Now, these directions all create angle $\sim 2^k\delta$ with the direction $e(T)$ of T , therefore they lie in an arc of $\mathbb{S}^1$ with length $\sim 2^k\delta$ (and centre $e(T)$). At the same time, they are δ -separated along this arc; so they are in total

$$\leq \frac{\text{length of the } \sim 2^k\delta\text{-arc}}{\delta} \sim \frac{2^k\delta}{\delta} = 2^k.$$

Hence,

$$\#\{T' \in \mathbb{T} : \text{Angle}(T, T') \sim 2^k\delta\} \lesssim 2^k, \quad k = 1, \dots, M.$$

Substituting this into (20), and recalling that $M \sim \log \frac{1}{\delta}$, we deduce that, for every fixed $T \in \mathbb{T}$,

$$\sum_{T' \in \mathbb{T}: T' \neq T} \frac{\delta^2}{\text{Angle}(T, T')} \lesssim \sum_{k=1}^M 2^k \cdot \frac{\delta}{2^k} = M\delta \sim \log \frac{1}{\delta} \cdot \delta \approx \delta.$$

²This practically means that, even though there are many different angles at which tubes $T' \in \mathbb{T}$ can intersect T , if we pretend that powers of $\log \frac{1}{\delta}$ are unimportant, it is as if all T' intersect T at practically *one* angle, so the calculations should be easier.

Adding over all $T \in \mathbb{T}$ and recalling (19), it follows that the off-diagonal contribution is also small:

$$\sum_{(T,T') \in \mathbb{T}^2: T \neq T'} |T \cap T'| \lesssim \sum_{T \in \mathbb{T}} \delta = \#\mathbb{T} \cdot \delta,$$

as desired. □

In any dimension $n \geq 3$, the exponent $\frac{n}{n-1}$ in the LHS of the Kakeya maximal operator conjecture is **not an integer**, so it does not automatically grant the conjecture a combinatorial flavour.

Of course, the real reason why the conjecture is harder to solve in higher dimensions is that, the higher the dimension n , the less we expect that tubes pointing in different directions should meet. This is reflected in the fact that $\frac{n}{n-1}$ decreases as n increases.

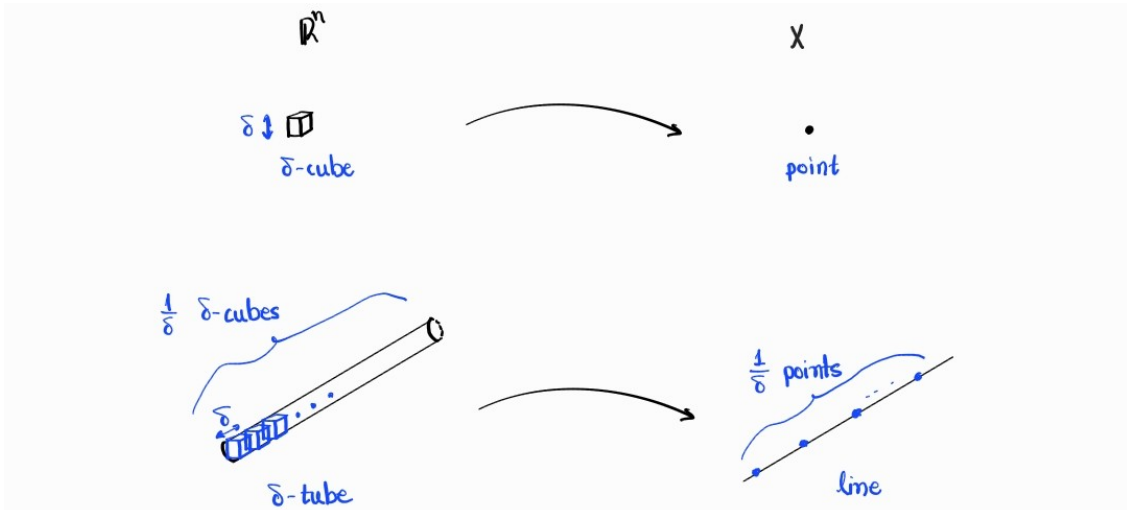
4.5. The Kakeya problem in finite fields

As we have seen, the Kakeya set conjecture claims that, for any $\delta > 0$ and any maximal family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$, the tubes in $\mathbb{T}$ behave as if they are disjoint, in the sense that

$$\left| \bigcup_{T \in \mathbb{T}} T \right| \approx \sum_{T \in \mathbb{T}} |T|.$$

The fact that the tubes have *width* really complicates matters, because the overlap of any two tubes that meet depends on the *angle* between them. This is not true for lines: any pair of intersecting lines meet *at only 1 point*.

We would therefore like to create an analogue of the problem in another appropriate universe, where every δ -tube has been replaced by a line, and every δ -cube by a point. However, every δ -tube in $\mathbb{R}^n$ can be covered by $\sim \frac{1}{\delta}$ essentially disjoint δ -cubes. So, every line in X must consist of $\sim \frac{1}{\delta}$ (and hence finitely many) points. We are thus looking for a space X so that:



It follows that an appropriate universe X for the above to work is $\mathbb{F}^n$, where $\mathbb{F}$ is a *finite field* with

$$\#\mathbb{F} \sim \frac{1}{\delta}.$$

Indeed:

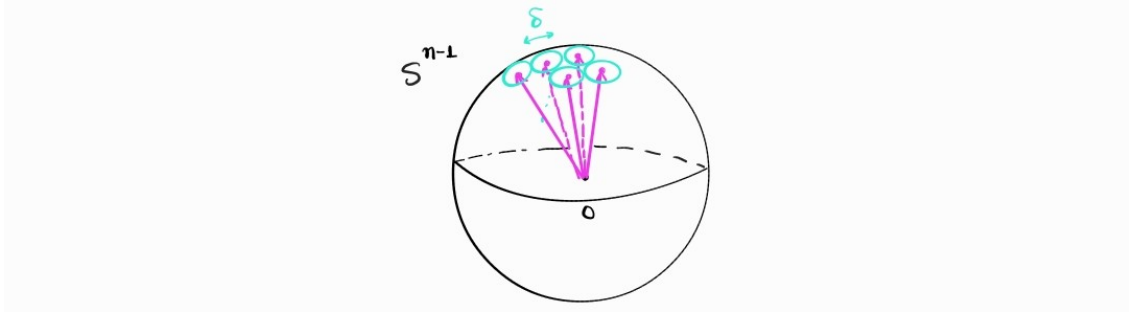
- In this universe, every line is of the form

$$\ell_{\vec{a}, \vec{b}} := \{\vec{a} + t\vec{b} : t \in \mathbb{F}\}$$

for some $\vec{a} \in \mathbb{F}^n$ and some $\vec{b} \in \mathbb{F}^n \setminus \{0\}$, where $\vec{a}$ is a *base point* of the line (that is, a point though which the line passes), and $\vec{b}$ is the *direction* of the line. Therefore, each line in $\mathbb{F}^n$ consists of $\#\mathbb{F} \sim \frac{1}{\delta}$ points.

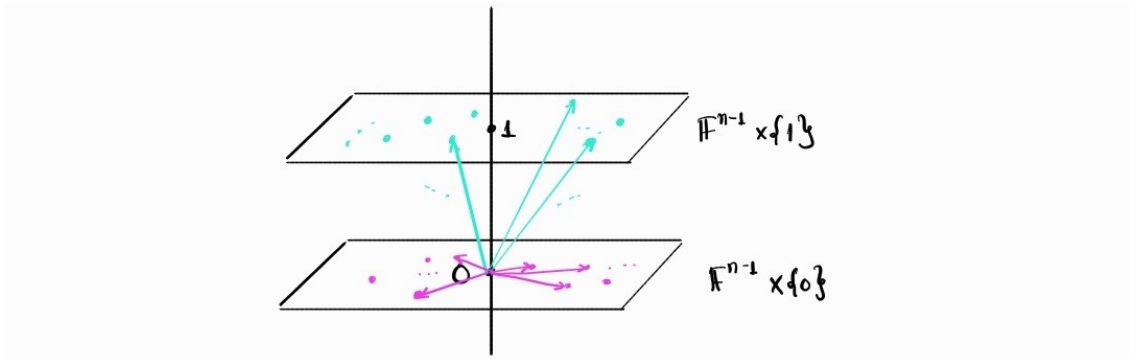
- The desired analogies with the Kakeya set conjecture in $\mathbb{R}^n$ continue to hold in terms of the number of directions involved:

→ In $\mathbb{R}^n$, we consider an arbitrary maximal family of δ -separated δ -tubes, hence the directions involved are $\sim \frac{1}{\delta^{n-1}}$ (as many as the δ -caps needed to cover $\mathbb{S}^{n-1}$).



→ In $\mathbb{F}^n$, there exist $\sim \#\mathbb{F}^{n-1}$ ($\sim \frac{1}{\delta^{n-1}}$) distinct directions:

- the directions $\vec{b} = (b_1, \dots, b_n) \in \mathbb{F}^n$ with $b_n \neq 0$, which via rescaling (multiplying each such $\vec{b}$ with the scalar $b_n^{-1} \in \mathbb{F}$) become vectors on the hyperplane $\mathbb{F}^{n-1} \times \{1\}$, and
- the directions $\vec{b} = (b_1, \dots, b_n) \in \mathbb{F}^n$ with $b_n = 0$, which live on $\mathbb{F}^{n-1} \times \{0\}$.



By the above, the analogue of the Kakeya set conjecture in finite fields should be as follows (replacing δ -tubes in all δ -discretised directions in $\mathbb{R}^n$ with lines in all possible directions in $\mathbb{F}^n$):

Let $\mathbb{F}$ be any finite field, and let $\mathcal{L}$ be any family of lines in $\mathbb{F}^n$, which contains one line in each direction. Then,

$$(21) \quad \#\left(\bigcup_{\ell \in \mathcal{L}} \ell\right) \sim \sum_{\ell \in \mathcal{L}} \#\ell.$$

Note that the reason we propose $\sim$ in this discrete setting, rather than $\approx$, is that, in the Kakeya set conjecture we expect the logarithmic losses of $\frac{1}{\delta}$ to come only from the fact that the tubes have thickness (which forces their overlap to depend on their angle)³. In the case of lines this is no longer the case, so we expect no logarithmic losses in terms of $\#\mathbb{F}$.

Now, the RHS in (21) equals the number of lines in $\mathcal{L}$ (i.e., the number of distinct directions in $\mathbb{F}^n$) times the number of points each line in $\mathbb{F}^n$ can contain (i.e. $\#\mathbb{F}$). In other words,

$$\sum_{\ell \in \mathcal{L}} \#\ell = \#\mathcal{L} \cdot \#\mathbb{F} \sim \#\mathbb{F}^{n-1} \cdot \#\mathbb{F} = \#\mathbb{F}^n.$$

Therefore, the Kakeya set conjecture in finite fields should claim that

$$\#\left(\bigcup_{\ell \in \mathcal{L}} \ell\right) \sim \#\mathbb{F}^n,$$

that is, that **a set containing a line in each direction should cover a definite proportion of the whole space $\mathbb{F}^n$.**

We are now ready to formulate the analogue of the Kakeya set conjecture in finite fields.

DEFINITION 4.5.1. (Kakeya sets in finite fields) Let $\mathbb{F}$ be a finite field and $n \geq 2$. A Kakeya set in $\mathbb{F}^n$ is any subset of $\mathbb{F}^n$ which contains a line in each direction.

CONJECTURE 4.5.2. (The Kakeya conjecture in finite fields) Let $n \in \mathbb{N}$. Then, there exists a constant $c_n > 0$, which depends only on n , so that

$$\#K \geq c_n \cdot \#\mathbb{F}^n,$$

for every finite field $\mathbb{F}$ and any Kakeya set K in $\mathbb{F}^n$.

The finite field Kakeya conjecture was proposed by Wolff in the 1990s, and was solved by Dvir in 2008. Dvir's solution induced a revolution in the areas of discrete geometry and harmonic analysis, because the method that he used was **algebraic**, and therefore revealed that the Kakeya problem (and thus related questions, such as the Fourier restriction conjecture) have an underlying *algebraic nature*. This was a surprise, as the statement of the Kakeya conjecture doesn't seem to have anything to do with algebra.

Dvir's solution was based on *the polynomial method*, which was first developed in the early 1900s in the context of number theory and had been later used in the 1980s in computer science. Dvir, a computer scientist himself, applied it in discrete geometry. Since then, it has been further developed to have numerous applications in incidence geometry, harmonic analysis and geometric measure theory.

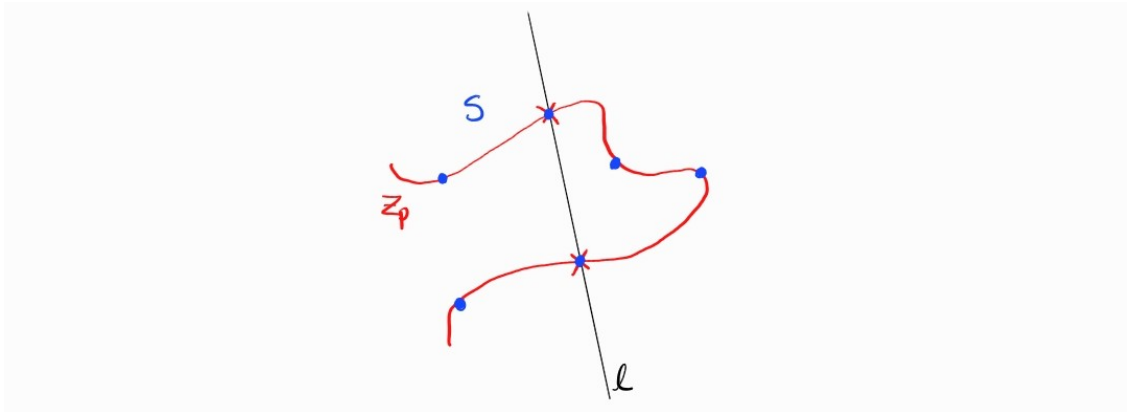
³Recall from the proof of the Kakeya set conjecture in two dimensions that indeed the $\log(\frac{1}{\delta})$ factor came from the fact that we split dyadically the range of angles at which any two δ -tubes can meet, and the number of these dyadic groups was $\sim \log(\frac{1}{\delta})$.

4.6. The polynomial method: the basic idea

The basic idea behind the polynomial method is the following.

Suppose that we wish to count points in some finite set $S \subseteq \mathbb{F}^n$ (where $\mathbb{F}$ is *any* field).

To that end, it might help to find a non-zero **polynomial** $p \in \mathbb{F}[x_1, \dots, x_n]$, **of low degree**, which **vanishes on S** .



(that is, so that $S \subseteq Z_p := \{\vec{x} \in \mathbb{F}^n : p(\vec{x}) = 0\}$).

This surface Z_p can perhaps reveal a lot of hidden structure that we couldn't see before, and this may help us count our points. Indeed, in the case where S lives on some lines provided by our problem (e.g. the case of Kakeya sets), then:

For every line ℓ , $p|_\ell$ is a polynomial of *one variable*; so, it is either the zero polynomial (in which case $\ell \subseteq Z_p$), or it has $\leq \deg p|_\ell \leq \deg p$ roots (in which case $\#(S \cap \ell) \leq \deg p$).

In other words, either ℓ contains a small number of points (quantitative information about S), or ℓ fully lives inside the zero set Z_p of p (structural information).

Therefore, the original problem of counting the points of S reduces to counting the points of S that live in lines fully inside Z_p . This is much stronger information than the one we started with (which was that simply S lives inside Z_p).

We can then study each point x in $S \cap Z_p$ using the structure of the algebraic surface Z_p (for instance, according to whether the tangent space to Z_p at x exists or not). So:

General polynomial method strategy: To count points in some set S that lives on lines in a family $\mathcal{L}$ within some $\mathbb{F}^n$:

1. Find a **low-degree** non-zero $p \in \mathbb{F}[x_1, \dots, x_n]$ that **vanishes on S** .
2. Check how the **vanishing** property **propagates across the lines in $\mathcal{L}$** .
3. Exploit the structure Z_p reveals to **count the points in S** .

4.7. The proof of the finite field Kakeya conjecture

Dvir's solution was based of the following.

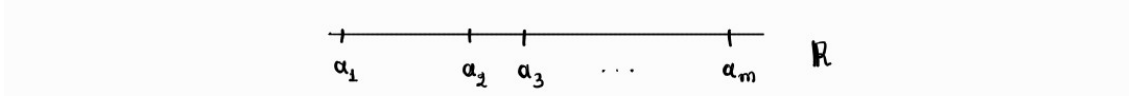
LEMMA 4.7.1. (**Dvir's vanishing lemma**) Let $\mathbb{F}$ be an arbitrary field, $n \in \mathbb{N}$, and S be a finite subset of $\mathbb{F}^n$. Then, there exists a non-zero polynomial $p \in \mathbb{F}[x_1, \dots, x_n]$, with $\deg p \lesssim \#S^{1/n}$, which vanishes on S .

We emphasise that the implicit multiplicative constant here depends only on the dimension n , and not on $\mathbb{F}$ or S .

Observations.

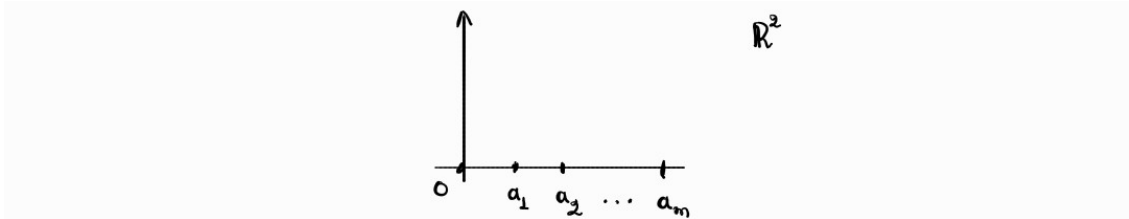
- The upper bound $\sim \#S^{1/n}$ on the degree of the vanishing polynomial depends only on the number of points of S (and the ambient dimension n), but not on how S is distributed in space.
- The upper bound $\sim \#S^{1/n}$ on the degree of the vanishing polynomial decreases as the dimension n increases.

For instance, let $a_1 < a_2 < \dots < a_m$ be m points in $\mathbb{R}$. Let $S := \{a_1, \dots, a_m\}$, a subset of $\mathbb{R}$:



We know that every $p \in \mathbb{R}[x]$ that vanishes at $a_1, \dots, a_m$ has to be divisible by $x - a_1, \dots, x - a_m$. Therefore, the non-zero polynomial $p \in \mathbb{R}[x]$ of minimal degree which vanishes on S is (any scalar multiple of) $(x - a_1) \cdots (x - a_m)$, which has degree $m = \#S$. So, for $n = 1$, the upper bound $\sim \#S$ on $\deg p$ cannot be decreased.

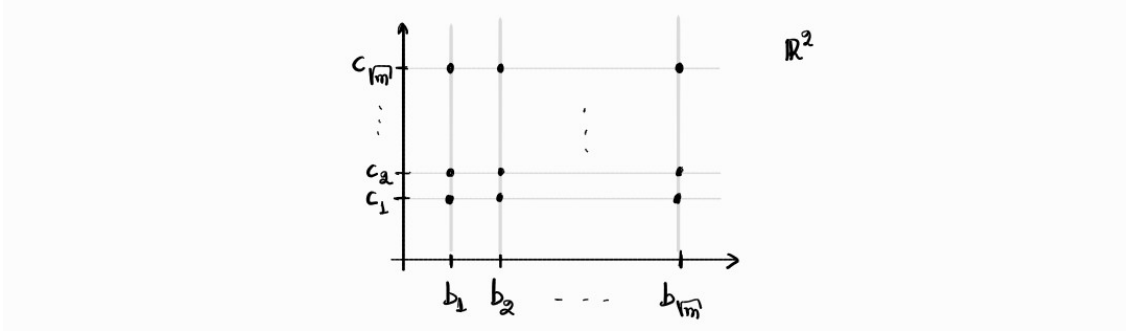
However, we may view S as a subset of $\mathbb{R}^2$ as well (that is, we can consider $S = \{(a_1, 0), \dots, (a_m, 0)\}$):



Then, we see that the non-zero polynomial $p \in \mathbb{R}[x, y]$, with $p(x, y) = y$, vanishes on S , and has degree 1 (which is much smaller than $\#S$).

- When S is a grid of points in $\mathbb{R}^n$, it is easy to see that there indeed exists a non-zero $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \lesssim \#S^{1/n}$, vanishing on S .

For example, in $\mathbb{R}^2$ we consider $S = B \times C$, where $B = \{b_1 < \dots < b_{\sqrt{m}}\}$ and $C = \{c_1 < \dots < c_{\sqrt{m}}\}$, for some square $m \in \mathbb{N}$:



Then, $\#S = m$, and a non-zero $p \in \mathbb{R}[x, y]$ that vanishes on S is

$$p(x, y) = (x - b_1) \cdots (x - b_{\sqrt{m}}),$$

which has degree $\sqrt{m} = \#S^{1/2}$. Indeed, it is easy to see that

$$Z_p = \bigcup_{i=1}^{\sqrt{m}} \{(x, y) : x = b_i\},$$

the union of the vertical lines passing through $(b_1, 0), \dots, (b_{\sqrt{m}}, 0)$.

→ Before we proceed to the proof of the vanishing lemma, we clarify what we mean by *zero polynomial*.

DEFINITION 4.7.2. We say that a polynomial $p \in \mathbb{F}[x_1, \dots, x_n]$ is the *zero polynomial* if all of its coefficients are 0 (that is, if it is the zero element of the ring $\mathbb{F}[x_1, \dots, x_n]$).

For $\mathbb{F} = \mathbb{R}$, this is the same as saying that p vanishes on the whole of $\mathbb{R}^n$. However, in fields with positive characteristic, it is possible for a polynomial to vanish on the whole of $\mathbb{F}^n$ *without being the zero polynomial*, i.e. with having at least one non-zero coefficient. An example is $x^p - x \in \mathbb{Z}_p[x]$, which, by Fermat's little theorem, vanishes at every $x \in \mathbb{F}$, while it is clearly not the zero element of $\mathbb{F}[x]$.

With this definition, we are now ready to prove one of the simplest statements at the core of the polynomial method.

LEMMA 4.7.3. If a polynomial $p \in \mathbb{F}[x_1, \dots, x_n]$ vanishes along a line $\ell = \ell_{\vec{a}, \vec{b}}$ (strictly) more than $\deg p$ times, then the one-variable polynomial

$$p|_{\ell}(t) := p(\vec{a} + t\vec{b}) \in \mathbb{F}[t]$$

is the zero polynomial (i.e., the zero element of $\mathbb{F}[t]$).

Note that in this case p vanishes at every point of ℓ . However, what the lemma asserts is the *stronger* statement that all of the coefficients of $p|_{\ell}$ are 0.

PROOF. It is easy to see that $\deg p|_{\ell} \leq \deg p$. Therefore, if the one-variable polynomial $p|_{\ell}$ has more than $\deg p \geq \deg p|_{\ell}$ roots, then it is the zero polynomial (by the fundamental theorem of algebra). $\square$

PROOF OF DVIR'S VANISHING LEMMA. Let $S := \{y_1, \dots, y_m\}$ be an arbitrary finite set in $\mathbb{F}^n$. We wish to show that there exists a non-zero $p \in \mathbb{F}[x_1, \dots, x_n]$, with $\deg p \lesssim m^{1/n}$, that vanishes on S .

A general polynomial $p \in \mathbb{F}[x_1, \dots, x_n]$ of degree at most d is a sum of monomials of degree at most d , i.e. has the form

$$p(x_1, \dots, x_n) = \sum_{a_1 + \dots + a_n \leq d} c_{a_1, \dots, a_n} x_1^{a_1} \cdots x_n^{a_n}$$

for some coefficients $c_{a_1, \dots, a_n} \in \mathbb{F}$. We want to find d as small as possible and coefficients $c_{a_1, \dots, a_n} \in \mathbb{F}$, not *all* zero, so that

$$(22) \quad \begin{cases} p(y_1) = 0 \\ p(y_2) = 0 \\ \vdots \\ p(y_m) = 0. \end{cases}$$

Then, the corresponding polynomial p will be non-zero, and will vanish on S .

For each $j = 1, \dots, m$, the equation $p(y_j) = 0$ is a linear equation with unknowns the coefficients $c_{a_1, \dots, a_n}$ (as the variables $x_1, \dots, x_n$ have been replaced by the precise values of the coordinates of the point $y_j \in \mathbb{F}^n$). Therefore, (22) is a system of m linear equations, with unknowns the coefficients, which are as many as the elements of the set

$$\{(a_1, \dots, a_n) \in \mathbb{Z}_{\geq 0}^n : a_1 + \dots + a_n \leq d\},$$

and thus $\binom{n+d}{n}$ in total.

A system of linear equations has for sure a non-trivial (i.e. non-zero) solution if

$$\# \text{ unknowns} > \# \text{ equations}.$$

Therefore, our system of m linear equations with $\binom{d+n}{n}$ unknowns has a non-trivial solution if

$$(23) \quad \binom{d+n}{n} > m.$$

That is, for any d satisfying (23), there exist coefficients $c_{a_1, \dots, a_n} \in \mathbb{F}$, that are not *all* zero, so that the corresponding polynomial p vanishes on $y_1, \dots, y_m$. This p is non-zero (as it has at least one non-zero coefficient), and has degree at most d .

In other words, for any d within the range of (23), there exists a non-zero $p \in \mathbb{F}[x_1, \dots, x_n]$, with $\deg p \leq d$, that vanishes on S . To complete the proof, we apply this for d as low as possible in the range (23). In particular,

$$\binom{d+n}{n} \sim d^n,$$

therefore the smallest d satisfying (23) also satisfies that

$$d^n \sim m \implies d \sim m^{1/n}.$$

Therefore, there exists a non-zero $p \in \mathbb{F}[x_1, \dots, x_n]$, with $\deg p \lesssim m^{1/n} = \#S^{1/n}$, vanishing on S .

□

We now present Dvir's proof to the finite field Kakeya conjecture.

THEOREM 4.7.4. (Dvir, 2008) *The Kakeya conjecture holds in finite fields. That is, for every $n \in \mathbb{N}$, there exists a constant $c_n > 0$, which only depends on n , so that*

$$\#K \geq c_n \cdot \#\mathbb{F}^n$$

for every finite field $\mathbb{F}$ and any Kakeya set K in $\mathbb{F}^n$.

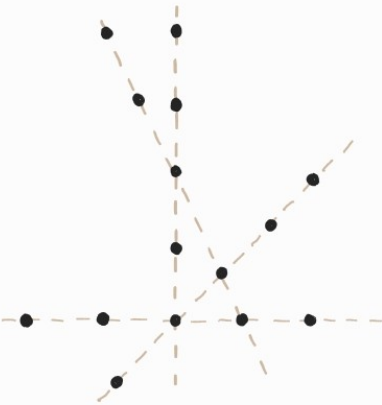
PROOF. Let $n \geq 2$, $\mathbb{F}$ a finite field, and K a Kakeya set in $\mathbb{F}^n$. Assume for contradiction that

$$(24) \quad \#K < c_n \# \mathbb{F}^n,$$

for a constant $c_n > 0$ that will be determined soon.

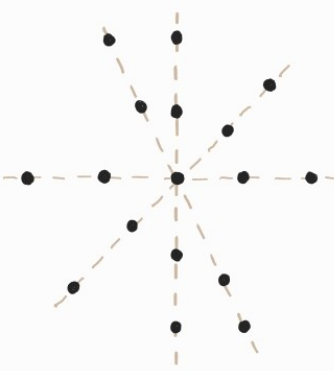
Proof strategy: We find a polynomial p of very low degree, which vanishes on the lines in K . The degree can be very low due to Dvir's vanishing lemma, as we are assuming here that K itself is very small. Then, we will show that another polynomial H_p , with the same degree as p , vanishes on *all the translations of these lines through* $0 \in \mathbb{F}^n$, and therefore vanishes on the whole of $\mathbb{F}^n$. It turns out though that the degree of H_p is too small for this to be possible (just like a low-degree polynomial in one variable cannot have too many roots).

Create low-degree p
vanishing on all lines in K



Create low-degree H_p
vanishing on all lines through 0

$\Rightarrow H_p$ vanishes everywhere. But...





Too low degree for this to be true!

- **1st step:** Find a polynomial of low degree vanishing on the points we want to count. By Dvir's vanishing lemma, there exists a non-zero $p \in \mathbb{F}[x_1, \dots, x_n]$, with

$$\deg p \leq C_n \#K^{1/n} \leq C_n c_n^{1/n} \cdot \# \mathbb{F},$$

vanishing on K (where C_n is a constant depending only on n). We now fix c_n to be so small that $C_n c_n^{1/n} < 1$. Then,

$$(25) \quad \deg p < \# \mathbb{F}.$$

- **2nd step:** See how the vanishing property propagates to the lines carrying the set. Let ℓ be a line with $\ell \subseteq K$. Since p vanishes on K , it also vanishes on ℓ , which consists of $\# \mathbb{F}$ points. So, p vanishes on ℓ more times than its degree, and thus

$$p|_{\ell} = 0 \in \mathbb{F}[t].$$

- We study more carefully what this implies. Let H_p be the homogeneous part of p of largest degree (that is, $\deg p$). In other words, we know that

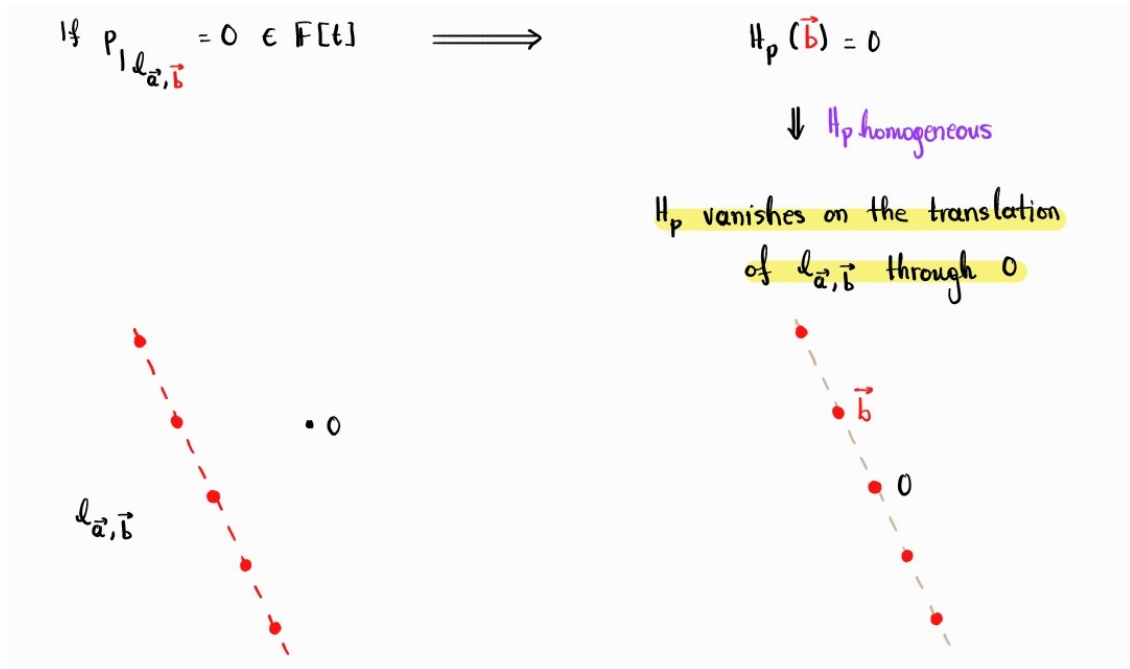
$$p(x_1, \dots, x_n) = \sum_{d_1 + \dots + d_n = \deg p} c_{d_1, \dots, d_n} x_1^{d_1} \cdots x_n^{d_n} + \sum_{d_1 + \dots + d_n < \deg p} c_{d_1, \dots, d_n} x_1^{d_1} \cdots x_n^{d_n},$$

for some coefficients $c_{d_1, \dots, d_n} \in \mathbb{F}$. We define

$$H_p(x_1, \dots, x_n) := \sum_{d_1 + \dots + d_n = \deg p} c_{d_1, \dots, d_n} x_1^{d_1} \cdots x_n^{d_n}.$$

This is a homogeneous polynomial in $\mathbb{F}[x_1, \dots, x_n]$, with $\deg H_p = \deg p$. We will now show the following:

If $p|_\ell$ is the zero polynomial, then H_p vanishes on the direction of the line ℓ .



Indeed, consider the line

$$\ell_{\vec{a}, \vec{b}} = \{\vec{a} + t\vec{b} : t \in \mathbb{F}\},$$

which has direction $\vec{b} \in \mathbb{F}^n \setminus \{0\}$ and passes through $\vec{a} \in \mathbb{F}^n$. We will now show that

$$(26) \quad p(\vec{a} + t\vec{b}) = 0 \in \mathbb{F}[t] \implies H_p(\vec{b}) = 0.$$

The reason is that $H_p(\vec{b})$ is the coefficient of $t^{\deg p}$ in the expansion of $p(\vec{a} + t\vec{b})$ as a polynomial in t (and must therefore vanish, as the latter is the zero polynomial). Indeed,

$$\begin{aligned} p(\vec{a} + t\vec{b}) &= p(a_1 + tb_1, \dots, a_n + tb_n) \\ &= \sum_{d_1 + \dots + d_n \leq \deg p} c_{d_1, \dots, d_n} (a_1 + tb_1)^{d_1} \cdots (a_n + tb_n)^{d_n}. \end{aligned}$$

Therefore, the coefficient of $t^{\deg p}$ on the RHS equals

$$\sum_{d_1 + \dots + d_n = \deg p} c_{d_1, \dots, d_n} b_1^{d_1} \cdots b_n^{d_n} = H_p(\vec{b}).$$

And since $p(\vec{a} + t\vec{b})$ is the zero polynomial, it follows that

$$H_p(\vec{b}) = 0.$$

- We combine the last two bullet points: For every line $\ell_{\vec{a}, \vec{b}}$ that is contained in K , it holds that $H_p(\vec{b}) = 0$. And since H_p is homogeneous, it vanishes on the whole line through 0 in direction $\vec{b}$. In other words:

If a line is contained in K , then H_p vanishes on the translation of the line through 0.

- Since K is a Kakeya set, it contains a line in any direction. Therefore:

H_p vanishes on all the lines of $\mathbb{F}^n$ through 0, and thus on the whole of $\mathbb{F}^n$.

- However, H_p has **too small degree** for this to be possible (and this leads us to a contradiction). Indeed, the Schwartz-Zippel lemma guarantees that a polynomial in $\mathbb{F}[x_1, \dots, x_n]$ cannot vanish at too many points if it has small degree:

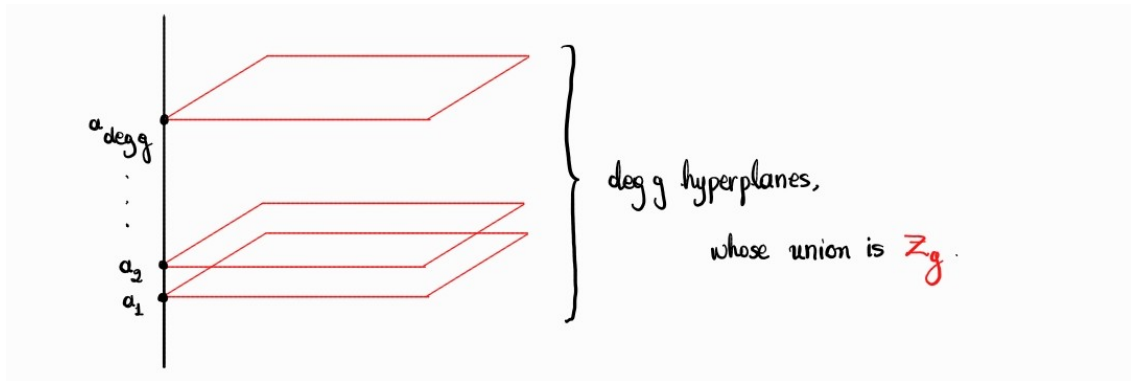
LEMMA 4.7.5. (**Schwartz-Zippel**) Let $g \in \mathbb{F}[x_1, \dots, x_n]$ be a non-zero polynomial. Then,

$$\#Z_g \leq \deg g \cdot \#\mathbb{F}^{n-1}.$$

We will not prove this, but let us discuss it:

→ For $n = 1$, this is known: it says that a polynomial of one variable can have at most as many roots as its degree.

→ For $n \geq 2$, observe that $\#\mathbb{F}^{n-1}$ is the cardinality of any hyperplane in $\mathbb{F}^n$. Therefore, $\deg g \cdot \#\mathbb{F}^{n-1}$ is the cardinality of $\deg g$ parallel hyperplanes in $\mathbb{F}^n$. So, the Schwartz-Zippel lemma states that g can have at most as many roots as it would have if it was a product of linear terms. (Note that the zero set of the polynomial $(x_n - a_1) \cdots (x_n - a_m)$ is the union of m horizontal hyperplanes, $\mathbb{F}^{n-1} \times \{a_1\}, \dots, \mathbb{F}^{n-1} \times \{a_m\}$.)



Returning to H_p : By the Schwartz-Zippel lemma, H_p can have at most

$$\deg H_p \cdot \#\mathbb{F}^{n-1} \leq \deg p \cdot \#\mathbb{F}^{n-1} \stackrel{(25)}{<} \#\mathbb{F} \cdot \#\mathbb{F}^{n-1} = \#\mathbb{F}^n$$

roots inside $\mathbb{F}^n$, so it cannot vanish on the whole of $\mathbb{F}^n$.

This is a contradiction to the previous bullet point. The contradiction is due to the fact that $\deg p < \#\mathbb{F}$, which followed from our initial assumption (24) that

$$\#K < c_n \cdot \#\mathbb{F}^n$$

for some constant c_n depending only on n . It follows that, for the constant c_n in this proof,

$$\#K \geq c_n \cdot \#\mathbb{F}^n,$$

for every finite field $\mathbb{F}$, and every Kakeya set K in $\mathbb{F}^n$.

□

CHAPTER 5

The Fourier restriction conjecture

5.1. The Fourier restriction problem: the basics

We know that, for all $p \in [1, 2]$, the Fourier transform of any L^p -function is well-defined. However, there is a crucial difference between the two extremes, $p = 1$ and $p = 2$:

For any $f \in L^1(\mathbb{R}^n)$, $\hat{f}$ is well-defined pointwise, via the formula

$$\hat{f}(\xi) = \int e^{-2\pi i \langle \xi, x \rangle} f(x) dx, \quad \xi \in \mathbb{R}^n,$$

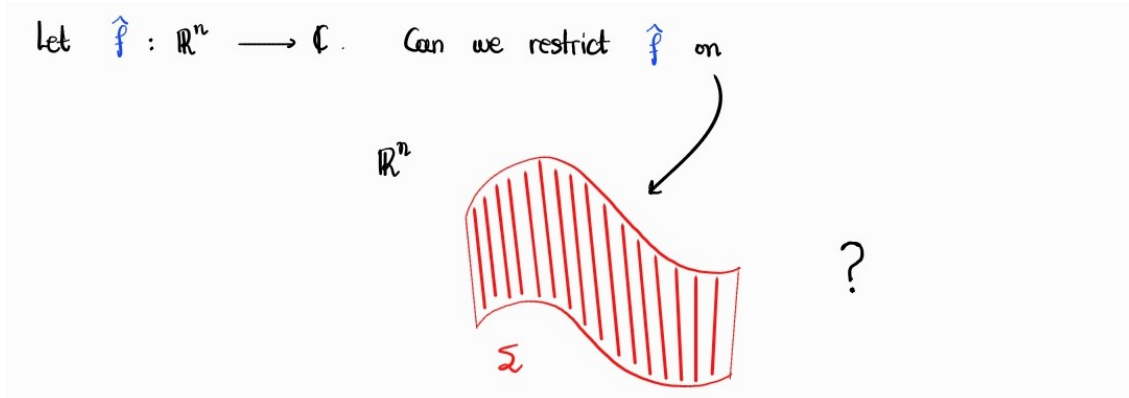
and is in fact a *continuous* function. Therefore, **$\hat{f}$ makes sense on any subset of $\mathbb{R}^n$ of measure 0.**

On the other hand, **for any $f \in L^2(\mathbb{R}^n)$, $\hat{f}$ is defined as an L^2 -limit via density of Schwartz functions in $L^2(\mathbb{R}^n)$, and is thus defined *almost everywhere* (in other words, it is a whole equivalence class of functions that are equal a.e.).** Moreover, $\hat{\cdot} : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ is onto, therefore $\hat{f}$ can be any L^2 -function in this case (so, we cannot expect it to satisfy additional agreeable properties, such as continuity). So, **in general, $\hat{f}$ cannot be meaningfully restricted on any subset of $\mathbb{R}^n$ of measure 0.**

For every intermediate $p \in (1, 2)$, the Fourier transform is again defined via density of Schwartz functions in $L^p(\mathbb{R}^n)$ (or equivalently by writing every L^p -function as a sum of an L^1 and an L^2 -function, and adding up the corresponding Fourier transforms). Therefore, for every $f \in L^p(\mathbb{R}^n)$, $\hat{f}$ is again defined almost everywhere. Moreover, by the Hausdorff-Young inequality, we know that $\hat{\cdot} : L^p(\mathbb{R}^n) \rightarrow L^{p'}(\mathbb{R}^n)$. However, this time the operator $\hat{\cdot}$ is not necessarily onto $L^{p'}(\mathbb{R}^n)$, which means that **we may expect $\hat{f}$ to have some better properties**, which would allow us to restrict it meaningfully on at least certain subsets of $\mathbb{R}^n$ of measure 0. Subsets of interest here are (compact patches of) hypersurfaces Σ of $\mathbb{R}^n$.

The Fourier restriction problem. For which $p \in [1, 2]$ and for which hypersurfaces Σ in $\mathbb{R}^n$ can we meaningfully restrict $\hat{f}$ on Σ , for all $f \in L^p(\mathbb{R}^n)$?

Since $\hat{f}|_{\Sigma}$ makes sense for all $f \in L^1$, but in general not for $f \in L^2$, we expect that, **some-where on the road between 1 and 2, $\hat{f}|_{\Sigma}$ should stop making sense.** And, theoretically, this threshold could change according to Σ .



To even begin thinking about this question, we need to understand what $\hat{f}|_{\Sigma}$ could even mean when $\hat{f}$ is not defined pointwise. We need to come up with a more ‘canonical’ definition, than simply choosing any values we wish for $\hat{f}$ on the 0-measure set Σ : this would not be a meaningful restriction of $\hat{f}$ on Σ - on the contrary, it would be entirely random. In particular, we need a definition of $\hat{f}|_{\Sigma}$ that respects what happens when f is well-behaved, e.g. an L^1 -function. For such functions f , $\hat{f}|_{\Sigma}$ is clearly well-defined (as $\hat{f}$ is defined pointwise). And, when f is close to some $g \in L^p$, we would like $\hat{g}|_{\Sigma}$ to be close to $\hat{f}|_{\Sigma}$. That is, for general $g \in L^p$ we could attempt to define $\hat{g}|_{\Sigma}$ via density.

→ How would we define $\hat{f}|_{\Sigma}$ meaningfully, when $\hat{f}$ is defined almost everywhere?

Let $p \in [1, 2]$ and Σ be (a compact patch of) a hypersurface in $\mathbb{R}^n$. Denote by $d\sigma$ the surface measure on Σ - that is, the Hausdorff measure H^{n-1} restricted on the Borel subsets of Σ (multiplied with an appropriate constant). From now on, we consider Σ equipped with the measure $d\sigma$.

Since all Schwartz functions are in L^1 , $\hat{f}|_{\Sigma}$ is definitely well-defined (pointwise in fact) for all $f \in S(\mathbb{R}^n)$.

Idea: Suppose that there exists some $q \in [1, +\infty]$ so that

$$(27) \quad \|\hat{f}|_{\Sigma}\|_{L^q(\Sigma)} \leq C_{\Sigma,p,q} \|f\|_{L^p(\mathbb{R}^n)}, \quad \text{for all } f \in S(\mathbb{R}^n).$$

Then, we can meaningfully define $\hat{f}|_{\Sigma}$ for all $f \in L^p(\mathbb{R}^n)$.

PROOF. Let $f \in L^p(\mathbb{R}^n)$. There exists a sequence $(g_k)_{k \in \mathbb{N}}$ of Schwartz functions, with

$$\|g_k - f\|_{L^p(\mathbb{R}^n)} \xrightarrow{k \rightarrow +\infty} 0.$$

Since $(g_k)_{k \in \mathbb{N}}$ converges in $(L^p(\mathbb{R}^n), \|\cdot\|_p)$, it is a Cauchy sequence in $(L^p(\mathbb{R}^n), \|\cdot\|_p)$. Due to (27), $(\hat{g}_k|_{\Sigma})_{k \in \mathbb{N}}$ is also a Cauchy sequence in $(L^q(\Sigma), \|\cdot\|_q)$. And since $(L^q(\Sigma), \|\cdot\|_q)$ is a Banach space, it follows that there exists $\tilde{f} \in L^q(\Sigma)$ so that

$$\|\hat{g}_k|_{\Sigma} - \tilde{f}\|_{L^q(\Sigma)} \xrightarrow{k \rightarrow +\infty} 0.$$

Given this, it makes sense to define

$$\hat{f}|_{\Sigma} := \tilde{f}.$$

This is what one naturally perceives as the restriction of $\hat{f}$ on Σ , and it is an L^q -limit with respect to the surface measure $d\sigma$ of Σ (so, it is defined almost everywhere on Σ with respect to this measure).¹

□

Observe that, to be able to meaningfully restrict $\hat{f}$ on Σ for all $f \in L^p(\mathbb{R}^n)$ (at least via the above argument), we need (27) to hold for *some* q . **The precise value of q is not important.** In other words, we need there to exist *some* q so that the operator $\hat{\cdot}$ is bounded from $L^p(\mathbb{R}^n)$ to $L^q(\Sigma)$. The important exponent is the p (which appears in the domain), not the q (which appears in the target space).

We are now ready to reformulate the restriction problem.

The Fourier restriction problem. Let Σ be a (compact patch of a) hypersurface in $\mathbb{R}^n$. For which $p \in [1, 2]$ does there exist $q \in [1, +\infty]$ so that

$$(27) \quad \|\hat{f}|_{\Sigma}\|_{L^q(\Sigma)} \leq C_{\Sigma,p,q} \|f\|_{L^p(\mathbb{R}^n)}, \quad \text{for all } f \in S(\mathbb{R}^n)?$$

(These are the p for which the Fourier restriction $\hat{f}|_{\Sigma}$ makes sense, for all $f \in L^p(\mathbb{R}^n)$.)

The Fourier restriction conjecture (which we will not state yet) specifies the pairs (p, q) for which (27) should hold, according to how curved Σ is (that is, according to the number of non-zero principal curvatures of Σ).

- For $p = 1$: We already know that, independently of Σ , $\hat{f}$ is defined pointwise (and is continuous) for every $f \in L^1(\mathbb{R}^n)$, so $\hat{f}|_{\Sigma}$ is definitely well-defined. And indeed, (27) holds for $q = +\infty$: by the triangle inequality, for every $\xi \in \mathbb{R}^n$,

$$|\hat{f}(\xi)| \leq \|f\|_1,$$

so

$$\|\hat{f}\|_{\infty} \leq \|f\|_1.$$

This bound holds on the whole of $\mathbb{R}^n$ (the domain of $\hat{f}$), so it holds in particular on Σ :

$$\|\hat{f}|_{\Sigma}\|_{\infty} \leq \|f\|_1.$$

- For $p = 2$: We know that, independently of Σ , there should not exist any meaningful restriction of $\hat{f}$ on Σ for general $f \in L^2(\mathbb{R}^n)$. This means that there should *not* exist any $q \in [1, +\infty]$ so that

$$\|\hat{f}|_{\Sigma}\|_q \leq C \|f\|_2 \quad \text{for all } f \in S(\mathbb{R}^n).$$

We will soon see a proof of this, via a dual formulation of the problem (which has to do with the *Fourier extension operator*, rather than the Fourier restriction operator $\hat{\cdot}|_{\Sigma}$).

- By the Riesz-Thorin interpolation theorem, if (27) holds for some p_0 between 1 and 2, then it should hold for all $p \in [1, p_0]$. Therefore, we are looking for that critical exponent p_0 , on the left of which (27) holds, and on the right of which it doesn't.

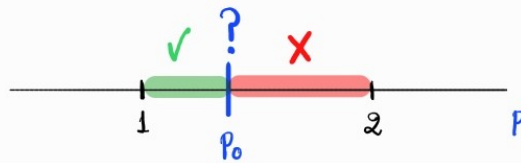
To sum up:

¹It was via a very similar argument that we extended the definition of the Fourier transform from Schwartz functions to the whole of L^p , using the Hausdorff-Young inequality instead of (27).

For which $p \in [1, 2]$ does there exist $q \in [1, \infty]$ so that

$$\|\hat{f}\|_q \lesssim \|f\|_p, \quad \forall f \in \mathcal{S}(\mathbb{R}^n) \quad ?$$

Aim: Push p as high as possible !

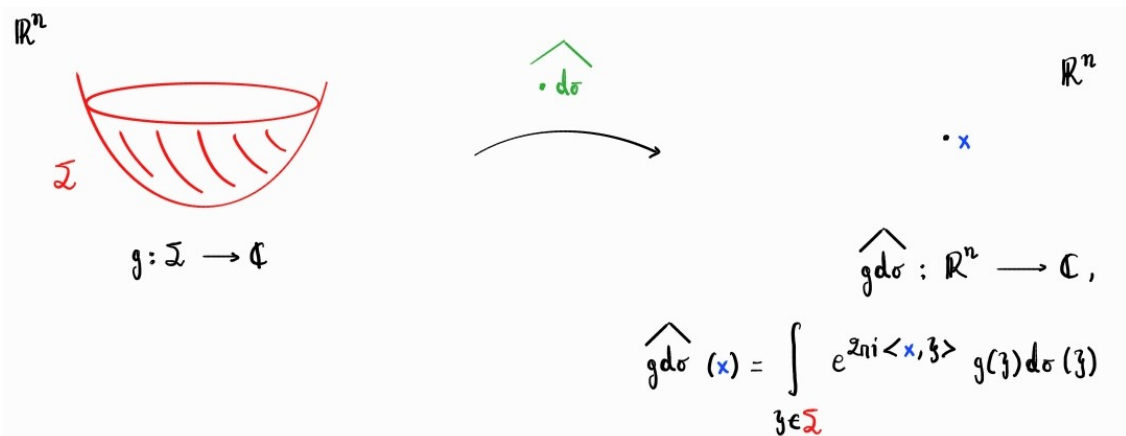


5.2. A dual formulation via the Fourier extension operator

Instead of thinking about boundedness of the Fourier restriction operator $\hat{\cdot}|_{\Sigma}$, we can instead think boundedness of its dual operator. It turns out that this dual operator, known as the *Fourier extension operator*, starts instead with functions defined on Σ , and converts them into functions defined on the whole of $\mathbb{R}^n$.

DEFINITION 5.2.1. (The Fourier extension operator $\widehat{g d\sigma}$) Let Σ be a (compact patch of a) hypersurface in $\mathbb{R}^n$, and let $d\sigma$ denote the Lebesgue surface measure on Σ . The Fourier extension operator associated to Σ is the operator $\widehat{g d\sigma}$, which maps each $g \in L^1(\Sigma)$ to the function $\widehat{g d\sigma} : \mathbb{R}^n \rightarrow \mathbb{C}$, defined by

$$\widehat{g d\sigma}(x) := \int_{\Sigma} e^{2\pi i \langle x, \xi \rangle} g(\xi) d\sigma(\xi), \quad \text{for all } x \in \mathbb{R}^n.$$



Observe the following:

- $\widehat{g d\sigma}$ is simply the inverse Fourier transform of g , with respect to the measure of the surface on which g lives.
- While each g in the definition is defined on Σ , which is $(n-1)$ -dimensional, the inverse Fourier transform $\widehat{g d\sigma}$ is defined on $\mathbb{R}^n$, which is a higher-dimensional space (it

has dimension n). This motivates the term ‘extension’ operator for the inverse Fourier transform $\widehat{\cdot d\sigma}$.

- Since Σ is a compact set, it has finite measure $\sigma(\Sigma)$. Therefore, $L^1(\Sigma)$ contains all the other spaces $L^p(\Sigma)$, for all $p > 1$. Therefore, $\widehat{\cdot d\sigma}$, as defined above, acts on any $L^p(\Sigma)$ -function, for all $p \geq 1$.

PROPOSITION 5.2.2. *Let $p, q \geq 1$. The estimate*

$$\|\widehat{f}|_{\Sigma}\|_{L^q(\Sigma)} \leq A \cdot \|f\|_{L^p(\mathbb{R}^n)} \quad \text{for all } f \in L^p(\mathbb{R}^n)$$

holds if and only if

$$\|\widehat{gd\sigma}\|_{L^{p'}(\mathbb{R}^n)} \leq A \cdot \|g\|_{L^{q'}(\Sigma)} \quad \text{for all } g \in L^{q'}(\Sigma).$$

PROOF. For every $f \in L^p(\mathbb{R}^n)$ and every $g \in L^{q'}(\Sigma)$,

$$\begin{aligned} \int_{\xi \in \Sigma} \widehat{f}|_{\Sigma}(\xi) \overline{g(\xi)} d\sigma(\xi) &= \int_{\xi \in \Sigma} \left(\int_{x \in \mathbb{R}^n} e^{-2\pi i \langle \xi, x \rangle} f(x) dx \right) \overline{g(\xi)} d\sigma(\xi) \\ &\stackrel{\text{Fubini}}{=} \int_{x \in \mathbb{R}^n} \left(\int_{\xi \in \Sigma} e^{-2\pi i \langle \xi, x \rangle} \overline{g(\xi)} d\sigma(\xi) \right) f(x) dx \\ &= \int_{x \in \mathbb{R}^n} f(x) \overline{\left(\int_{\xi \in \Sigma} e^{2\pi i \langle \xi, x \rangle} g(\xi) d\sigma(\xi) \right)} dx \\ &= \int_{x \in \mathbb{R}^n} f(x) \overline{\widehat{gd\sigma}(x)} dx. \end{aligned}$$

Therefore, the Fourier restriction operator

$$\widehat{\cdot}|_{\Sigma} : L^p(\mathbb{R}^n) \rightarrow L^q(\Sigma)$$

is (well-defined and) bounded if and only if the Fourier extension operator

$$\widehat{\cdot d\sigma} : L^{q'}(\Sigma) \rightarrow L^{p'}(\mathbb{R}^n)$$

is (well-defined and) bounded, and their norms are the same.

□

Note that $p \in [1, 2]$ if and only if $p' \in [2, +\infty]$. And **increasing p as much as possible is equivalent to lowering p' as much as possible**. Therefore, an equivalent form of the Fourier restriction problem is the following.

The Fourier restriction problem. Let Σ be a (compact patch of a) hypersurface in $\mathbb{R}^n$. For which $p \in [2, +\infty]$ does there exist $q \in [1, +\infty]$ so that

$$(28) \quad \|\widehat{fd\sigma}\|_{L^p(\mathbb{R}^n)} \leq C_{p,q} \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma)?$$

(These are the p for which the restriction $\widehat{f}|_{\Sigma}$ makes sense, for all $f \in L^{p'}(\mathbb{R}^n)$.)

- In this dual formulation of the problem, the exponent that matters is the p on the LHS, and the aim is to lower it as much as possible. The q on the RHS doesn't matter: as soon as (28) holds for a given p and for some q on the RHS, then $\widehat{f}|_{\Sigma}$ makes sense for all $f \in L^{p'}(\Sigma)$.

- We know that, independently of Σ , (28) holds at $p = \infty$ (for some q), as the Fourier transform of all $L^1(\mathbb{R}^n)$ -functions can be restricted on Σ . A reality check:

$$|\widehat{fd\sigma}(x)| \leq \int |f(\xi)| d\sigma(\xi) = \|f\|_{L^1(\Sigma)} \quad \text{for all } x \in \mathbb{R}^n,$$

and thus

$$\|\widehat{fd\sigma}\|_{L^\infty(\mathbb{R}^n)} \leq \|f\|_{L^1(\Sigma)}, \quad \text{for all } f \in L^1(\Sigma).$$

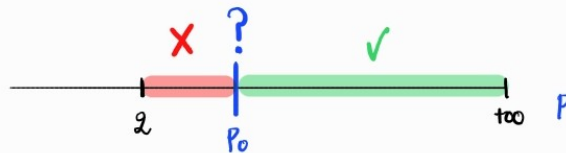
- For $p = 2$, we know that, independently of Σ , there does not exist any q for which (28) holds, as otherwise we would be able to restrict the Fourier transform of all $L^2(\mathbb{R}^n)$ -functions on Σ . We will prove this in a moment in another way (see Proposition 5.2.3 below).
- By Riesz-Thorin interpolation, if (28) holds for some p_0 , then it holds in the whole interval $[p_0, +\infty]$. So, we are looking for the critical exponent $p_0 \in (1, +\infty]$, so that (28) holds on the right of it, and fails on the left.

Summing up, the question is:

For which $p \in [2, +\infty]$ does there exist $q \in [1, +\infty]$ so that

$$\|\widehat{fd\sigma}\|_p \lesssim \|f\|_q, \quad \forall f \in L^q(\Sigma)?$$

Aim: Push p as low as possible !

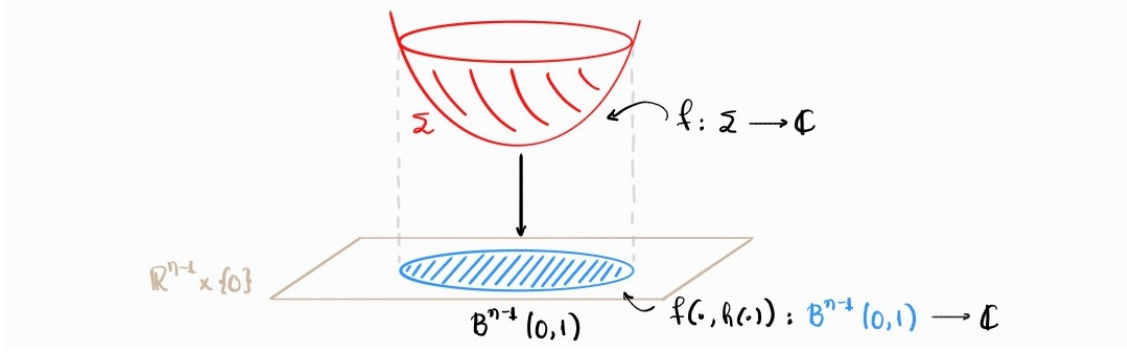


→ **Understanding the extension operator.** A way to perhaps simplify calculations and our view of the extension operator is to parametrise Σ . Without loss of generality, since Σ is compact, we may assume that Σ is the graph of some differentiable function h , over the unit ball $B^{n-1}(0, 1)$ with centre 0 in $\mathbb{R}^{n-1}$:

$$\Sigma := \{(\omega, h(\omega)) : \omega \in B^{n-1}(0, 1)\}.$$

We may also assume that $|\nabla h| \lesssim_\Sigma 1$ on $B^{n-1}(0, 1)$ (where the implicit multiplicative constant is allowed to depend on Σ), as otherwise we may split the compact set Σ in smaller, $O_\Sigma(1)$ patches, so that an appropriate rotation of each patch is indeed a graph of a function with controlled gradient.

Then, we may think of every function defined on Σ as a function defined on the set $B^{n-1}(0, 1)$ 'below' Σ , instead. And, on $B(0, 1)$, integration is easier in terms of calculations, as it is done with respect to the usual Lebesgue measure in $\mathbb{R}^{n-1}$.



Indeed, for any $f \in L^1(\Sigma)$, a change of variables gives

$$\widehat{f d\sigma}(x) = \int_{\xi \in \Sigma} e^{2\pi i \langle x, \xi \rangle} f(\xi) d\sigma(\xi)$$

$$\stackrel{\xi = (\omega, h(\omega))}{=} \int_{\omega \in B^{n-1}(0,1)} e^{2\pi i \langle x, (\omega, h(\omega)) \rangle} f(\omega, h(\omega)) \sqrt{1 + |\nabla h(\omega)|^2} d\omega.$$

In the integrand, we have the modulation $e^{2\pi i \langle x, (\omega, h(\omega)) \rangle}$, multiplied with a function that now only depends on ω (which lives in $\mathbb{R}^{n-1}$). We call this function $g : \mathbb{R}^{n-1} \rightarrow \mathbb{C}$:

$$g(\omega) := f(\omega, h(\omega)) \sqrt{1 + |\nabla h(\omega)|^2} \chi_{B^{n-1}(0,1)}(\omega), \quad \omega \in \mathbb{R}^{n-1}.$$

We have thus show that, for this g (which, given Σ , only depends on f),

$$\widehat{f d\sigma}(x) = \int_{\omega \in \mathbb{R}^{n-1}} e^{2\pi i \langle x, (\omega, h(\omega)) \rangle} g(\omega) d\omega, \quad \text{for all } x \in \mathbb{R}^n.$$

This is reminiscent of the standard inverse Fourier transform of g (in $n-1$ variables), but there is an important difference: in the phase function (the exponent of the exponential) we have the inner product of the frequency x with the arbitrary element of Σ (rather than the arbitrary element of $\mathbb{R}^{n-1}$). And, of course, the frequencies x in question live in $\mathbb{R}^n$, rather than $\mathbb{R}^{n-1}$.

We now go one step further: we write each frequency $x \in \mathbb{R}^n$ as (x', t) , where $x' \in \mathbb{R}^{n-1}$ and $t \in \mathbb{R}$. We may think in fact of t as a time variable. And something interesting happens when we fix a 'time' t and think of the values $\widehat{f d\sigma}$ takes at that 'time' (that is, when we consider the values of $\widehat{f d\sigma}$ on the horizontal hyperplane $\{x_n = t\}$ inside $\mathbb{R}^n$): For each $t \in \mathbb{R}$,

$$\widehat{f d\sigma}(x', t) = \int_{\omega \in \mathbb{R}^{n-1}} e^{2\pi i \langle x', \omega \rangle} e^{2\pi i t h(\omega)} g(\omega) d\omega = \check{g}_t(x'), \quad x' \in \mathbb{R}^{n-1},$$

where $g_t := e^{2\pi i t h(\cdot)} g : \mathbb{R}^{n-1} \rightarrow \mathbb{C}$, and $\check{\cdot}$ is the standard inverse Fourier transform on functions defined on $\mathbb{R}^{n-1}$.

In other words, the extension operator $\widehat{f d\sigma}(\cdot, t)$ of f at any time t is the usual inverse Fourier transform (in $n-1$ variables) of a function g_t that depends on f and Σ in a very precise way. Specifically at time $t = 0$, $\widehat{f d\sigma}(\cdot, 0) = \check{g} = \widehat{g}(-\cdot)$.

This means that, at time 0 (that is, on the horizontal hyperplane $\{x_n = 0\}$), $\widehat{f d\sigma}$ is simply the inverse Fourier transform of g (and g is a slightly altered version of f). So, **the values of $\widehat{f d\sigma}$ at time 0 fully determine f** (via Fourier inversion). Therefore, the values of $\widehat{f d\sigma}$ at all other times t are also entirely determined by what happens at time 0. Even so, the norm of $\widehat{f d\sigma}$ is dramatically hard to control.

We now see how this viewpoint can come in handy in calculations.

PROPOSITION 5.2.3. *For every $q \in [1, +\infty]$ and every non-zero $f \in L^q(\Sigma)$,*

$$\|\widehat{f d\sigma}\|_{L^2(\mathbb{R}^n)} = +\infty.$$

Therefore, there exists no $q \in [1, +\infty]$ for which

$$\|\widehat{f d\sigma}\|_{L^2(\mathbb{R}^n)} \lesssim_q \|f\|_{L^q(\Sigma)}.$$

PROOF. Retaining the notation of the preceding discussion, we recall that, for every $t \in \mathbb{R}$,

$$\widehat{f d\sigma}(x', t) = \int_{\omega \in \mathbb{R}^{n-1}} e^{2\pi i \langle x', \omega \rangle} e^{2\pi i t h(\omega)} g(\omega) d\omega = \check{g}_t(x') \quad \text{for all } x' \in \mathbb{R}^{n-1},$$

where

$$g(\omega) := f(\omega, h(\omega)) \sqrt{1 + |\nabla h(\omega)|^2} \chi_{B^{n-1}(0,1)}(\omega), \quad \omega \in \mathbb{R}^{n-1}$$

and

$$g_t := e^{2\pi i t h(\cdot)} g : \mathbb{R}^{n-1} \rightarrow \mathbb{C}.$$

Therefore,

$$\int_{\mathbb{R}^{n-1}} |\widehat{f d\sigma}(\cdot, t)|^2 = \int |\check{g}_t|^2 \stackrel{\text{Plancherel}}{=} \int |g_t|^2 \stackrel{|g_t|=|g|}{=} \int |g|^2,$$

a non-zero quantity (as f is non-zero), independent of t . Therefore, further integrating over all $t \in \mathbb{R}$, we obtain

$$\int |\widehat{f d\sigma}|^2 = +\infty.$$

□

→ **A reformulation of the restriction problem, using a parametrisation of Σ .**

As we have seen, we may assume that

$$\Sigma = \{(\omega, h(\omega)) : \omega \in B^{n-1}(0, 1)\}$$

for some differentiable h , with $|\nabla h| \lesssim_\Sigma 1$. And that, parametrising Σ like this,

$$\widehat{f d\sigma}(x) = \int_{\omega \in B^{n-1}(0,1)} e^{2\pi i \langle x, (\omega, h(\omega)) \rangle} f(\omega, h(\omega)) \sqrt{1 + |\nabla h(\omega)|^2} d\omega.$$

Therefore the operator $\widehat{\cdot d\sigma}$ maps

$$f \in L^1(\Sigma) \mapsto \int_{\omega \in B^{n-1}(0,1)} e^{2\pi i \langle x, (\omega, h(\omega)) \rangle} f(\omega, h(\omega)) \sqrt{1 + |\nabla h(\omega)|^2} d\omega : \mathbb{R}^n \rightarrow \mathbb{C}.$$

The Fourier restriction problem asks for which p there exists q so that this operator is bounded from $L^q(\Sigma)$ to $L^p(\mathbb{R}^n)$. What is quite annoying about the integrand of the RHS is the quantity $\sqrt{1 + |\nabla h|^2}$. However, it turns out that, since $\sqrt{1 + |\nabla h|^2} \sim_\Sigma 1$ pointwise, boundedness of this operator is independent of whether we keep this quantity inside the integrand. In other words, we can delete it altogether.

DEFINITION 5.2.4. **(The Fourier extension operator E)** *Let Σ be a hypersurface in $\mathbb{R}^n$, parametrised by*

$$\Sigma = \{(\omega, h(\omega)) : \omega \in B^{n-1}(0, 1)\}$$

for some differentiable h , with $|\nabla h| \lesssim_\Sigma 1$. We define an operator E , which maps every $f \in L^1(B^{n-1}(0, 1))$ to the function $Ef : \mathbb{R}^n \rightarrow \mathbb{C}$, with

$$Ef(x) := \int_{\omega \in B^{n-1}(0, 1)} e^{2\pi i \langle x, (\omega, h(\omega)) \rangle} f(\omega) d\omega, \quad \text{for all } x \in \mathbb{R}^n.$$

This operator is usually also referred to as the Fourier extension operator associated to Σ .

By the above discussion, the precise relationship between $\widehat{\cdot d\sigma}$ and E is the following: For every $f \in L^1(\Sigma)$,

$$(29) \quad \widehat{fd\sigma} = E\tilde{f},$$

where

$$\tilde{f}(\omega) := f(\omega, h(\omega)) \sqrt{1 + |\nabla h(\omega)|^2}, \quad \text{for all } \omega \in B^{n-1}(0, 1).$$

We will now see that the Fourier restriction problem can be reinterpreted in terms of the (perhaps simpler) operator E .

PROPOSITION 5.2.5. *Let $p, q \in [1, +\infty]$. The operator $\widehat{\cdot d\sigma}$ is bounded from $L^q(\Sigma)$ to $L^p(\mathbb{R}^n)$ if and only if the operator E is bounded from $L^q(B^{n-1}(0, 1))$ to $L^p(\mathbb{R}^n)$. In other words,*

$$(30) \quad \|\widehat{fd\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma, p, q} \|f\|_{L^q(\Sigma)} \quad \text{for all } f \in L^q(\Sigma)$$

if and only if

$$(31) \quad \|Ef\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma, p, q} \|f\|_{L^q(B^{n-1}(0, 1))} \quad \text{for all } f \in L^q(B^{n-1}(0, 1)).$$

PROOF. Suppose that (30) holds. Given (29), this means that

$$\|E\tilde{f}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma, p, q} \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma),$$

where

$$\tilde{f}(\omega) := f(\omega, h(\omega)) \sqrt{1 + |\nabla h(\omega)|^2}, \quad \text{for all } \omega \in B^{n-1}(0, 1).$$

Now, for every $g : B^{n-1}(0, 1) \rightarrow \mathbb{C}$, define its ‘lift’ g_Σ on Σ , as follows: For every $\xi \in \Sigma$, there exists a unique $\omega \in B^{n-1}(0, 1)$ so that $\xi = (\omega, h(\omega))$. Define $g_\Sigma : \Sigma \rightarrow \mathbb{C}$ by

$$g_\Sigma(\xi) \stackrel{\xi=(\omega, h(\omega))}{=} \frac{g(\omega)}{\sqrt{1 + |\nabla h(\omega)|^2}}, \quad \text{for all } \xi = (\omega, h(\omega)) \in \Sigma.$$

Observe that $g = \widetilde{g_\Sigma}$, therefore

$$(32) \quad \|Eg\|_{L^p(\mathbb{R}^n)} = \|E\widetilde{g_\Sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma, p, q} \|g_\Sigma\|_{L^q(\Sigma)}, \quad \text{for all } g \in L^q(B^{n-1}(0, 1)).$$

To show (31), it remains to compare $\|g_\Sigma\|_{L^q(\Sigma)}$ with $\|g\|_{L^q(B^{n-1}(0, 1))}$:

$$\begin{aligned} \int_{\xi \in \Sigma} |g_\Sigma(\xi)|^q d\sigma(\xi) &\stackrel{\xi=(\omega, h(\omega))}{=} \int_{\omega \in B^{n-1}(0, 1)} \frac{|g(\omega)|^q}{\sqrt{1 + |\nabla h(\omega)|^2}^q} \sqrt{1 + |\nabla h(\omega)|^2} d\omega \\ &\sim_\Sigma \int_{B^{n-1}(0, 1)} |g|^q, \end{aligned}$$

where the last estimate is due to the fact that $\sqrt{1 + |\nabla h|^2} \sim_\Sigma 1$ pointwise. Therefore,

$$\|g_\Sigma\|_{L^q(\Sigma)} \sim_\Sigma \|g\|_{L^q(B^{n-1}(0, 1))}.$$

Plugging this into (32), we derive (31).

The implication (31) $\Rightarrow$ (30) is similar and is left as an exercise.

□

The following reformulation of the Fourier restriction problem is therefore equivalent to the earlier versions.

The Fourier restriction problem in terms of E . Let Σ be a (compact patch of a) hypersurface in $\mathbb{R}^n$, parametrised by

$$\Sigma = \{(\omega, h(\omega)) : \omega \in B^{n-1}(0, 1)\},$$

for some differentiable $h : B^{n-1}(0, 1) \rightarrow \mathbb{R}$ with $|\nabla h| \lesssim_\Sigma 1$. For which $p \in [2, +\infty]$ does there exist $q \in [1, +\infty]$ so that

$$(33) \quad \|Ef\|_{L^p(\mathbb{R}^n)} \leq C_{\Sigma, p, q} \|f\|_{L^q(B^{n-1}(0, 1))}, \quad \text{for all } f \in L^q(B^{n-1}(0, 1))?$$

- Just as in the equivalent version of the problem expressed in terms of $\widehat{\cdot d\sigma}$, there exists such a q for $p = +\infty$, but there exists no such q for $p = 2$. Therefore, by Riesz-Thorin interpolation, there exists some p_0 between 2 and $+\infty$, so that the answer is yes on the right of p_0 , and no on the left of p_0 . The problem is to find this critical exponent p_0 . In other words, the problem is to *lower the p on the LHS as much as possible*.

- Both equivalent versions of the Fourier restriction problem (expressed in terms of either $\widehat{\cdot d\sigma}$ or E) are used extensively in the literature - each presents its own advantages, and people choose which one to use according to the context.

5.3. The Fourier extension operator in PDE

The operators E and $\widehat{\cdot d\sigma}$, associated to specific hypersurfaces, have a central role in PDE. In particular:

→ When Σ is the **paraboloid**

$$\Sigma = \{(\omega, |\omega|^2) : \omega \in B^{n-1}(0, 1)\}$$

(in which case $h(\omega) = |\omega|^2$), the corresponding extension operator E satisfies

$$Ef(x', t) = \int_{\omega \in B^{n-1}(0, 1)} e^{2\pi i(\langle x', \omega \rangle + t|\omega|^2)} f(\omega) d\omega, \quad \text{for all } x' \in \mathbb{R}^{n-1}, t \in \mathbb{R}.$$

Given any $f \in L^1(B^{n-1}(0, 1))$, the function $Ef(\cdot, t)$ happens to be a solution $u(\cdot, t)$, in time t , of the linear Schrödinger equation

$$\Delta_x u = 2\pi i \partial_t u,$$

with initial data $u(\cdot, 0) = \check{f}$ (observe that $u = Ef$ satisfies both the PDE and the initial condition).

Now, we know that the solution u to the linear Schrödinger equation has a very nice physics interpretation: If $|u(\cdot, 0)|^2$ is the probability distribution that describes the position of an electron at time 0 (meaning that the probability that the electron lives in some set $A \subseteq \mathbb{R}^{n-1}$ at time 0 is equal to $\int_A |u(\cdot, 0)|^2$), then $|u(\cdot, t)|^2$ is the probability distribution that describes the position of an electron at time t .

So, the extension operator associated to the paraboloid satisfies that, if $|\check{f}|^2 = |Ef(\cdot, 0)|^2$ is the probability distribution that describes the position of an electron at time 0, then $|Ef(\cdot, t)|^2$ is the probability distribution that describes the position of an electron at time t . For a reality check, recall that, for every $t \in \mathbb{R}$,

$$Ef(x', t) = \int_{\omega \in B^{n-1}(0, 1)} e^{2\pi i \langle x', \omega \rangle} e^{2\pi i t |\omega|^2} f(\omega) d\omega = \check{f}_t(x'), \quad \text{for all } x' \in \mathbb{R}^{n-1},$$

where $f_t(\omega) := e^{2\pi i t |\omega|^2}$. Therefore,

$$\int |Ef(\cdot, t)|^2 = \int |\check{f}_t|^2 = \int |f_t|^2 = \int |f|^2$$

independently of t (a property known as ‘preservation of energy’), a quantity which is equal to 1 if $|\check{f}|^2$ is a probability distribution. So, $|Ef(\cdot, t)|^2$ can also play the role of a probability distribution, for every t .

(Note that energy preservation for E has nothing to do with the hypersurface being specifically the paraboloid, and would have worked similarly for any h . It is the reason that L^2 -restriction is not possible - a calculation we saw earlier in terms of $\widehat{\cdot d\sigma}$ instead of E .)

→ When Σ is the **cone**

$$\Sigma = \{(\omega, |\omega|) : \omega \in B := B^{n-1}(0, 1) \setminus B^{n-1}(0, 1/2)\}$$

(in which case $h(\omega) = |\omega|$), the corresponding extension operator E satisfies

$$Ef(x', t) = \int_{\omega \in B} e^{2\pi i (\langle x', \omega \rangle + t|\omega|)} f(\omega) d\omega, \quad \text{for all } x' \in \mathbb{R}^{n-1}, t \in \mathbb{R}.$$

Given any $f \in L^1(B)$, the function $Ef(\cdot, t)$ happens to be a solution $u(\cdot, t)$, at time t , of the half-wave equation

$$\sqrt{-\Delta_x} u = 2\pi i \partial_t u.$$

→ When Σ is the **sphere** $\mathbb{S}^{n-1}$, the functions $\widehat{f d\sigma}$, over all $f \in L^1(\mathbb{S}^{n-1})$, are the solutions $u : \mathbb{R}^n \rightarrow \mathbb{C}$ to the Helmholtz equation

$$\Delta u = -4\pi^2 u$$

(as, taking Fourier transforms, we see that these solutions are exactly those functions u whose Fourier transforms $\hat{u}$ are supported on $\mathbb{S}^{n-1}$, so u is given by Fourier inverting $\hat{u}$ with respect to $d\sigma$).

Via the above examples, we see that showing $L^p - L^q$ -boundedness of the extension operator associated to specific hypersurfaces Σ is equivalent to finding what are known as *Strichartz estimates* for solutions to PDE. These are estimates that control the size (norm) of the solution by the size of the initial data. And, at least from the PDE perspective, the paraboloid, the sphere and the cone should be hypersurfaces of particular interest when it comes to the study of the extension operator.

5.4. The Fourier restriction conjecture

The Fourier restriction conjecture predicts, according to Σ , the exact pairs (p, q) for which the associated Fourier extension operator $\widehat{\cdot d\sigma}$ (or equivalently E) should be $L^p - L^q$ bounded.

It turns out that the curvature of Σ really matters. When Σ is a hyperplane, restriction of $\hat{f}$ on Σ is possible only for $L^1(\mathbb{R}^n)$ -functions f , and not for any other $L^p(\mathbb{R}^n)$ -functions.

PROPOSITION 5.4.1. *There are no non-trivial Fourier restriction estimates when Σ is a hyperplane. In other words, while it is true in this case that*

$$\|Ef\|_{L^\infty(\mathbb{R}^n)} \leq \|f\|_{L^1(B^{n-1}(0,1))} \quad \text{for all } f \in L^1(B^{n-1}(0,1))$$

(as this holds for all Σ), for every $p < \infty$ and every $q \in [1, +\infty]$ the estimate

$$\|Ef\|_{L^p(\mathbb{R}^n)} \leq C_{\Sigma,p,q} \|f\|_{L^q(B^{n-1}(0,1))} \quad \text{for all } f \in L^q(B^{n-1}(0,1))$$

fails. In fact,

$$\|Ef\|_{L^p(\mathbb{R}^n)} = +\infty$$

for all non-zero $f \in L^1(B^{n-1}(0,1))$.

PROOF. Via a translation and a rotation, we may assume that Σ is the horizontal hyperplane through 0 in $\mathbb{R}^n$:

$$\Sigma = \{(\omega, 0) : \omega \in B^{n-1}(0,1)\}.$$

Let $p \in [1, +\infty)$. For every $f \in L^1(B^{n-1}(0,1))$ and every $t \in \mathbb{R}$,

$$Ef(x', t) = \int_{\omega \in B^{n-1}(0,1)} e^{2\pi i \langle x', t \rangle, (\omega, 0)} f(\omega) d\omega = \int_{\omega \in B^{n-1}(0,1)} e^{2\pi i \langle x', \omega \rangle} f(\omega) d\omega = \check{f}(x'),$$

for all $x' \in \mathbb{R}^{n-1}$. In other words, at every 'time' t , $Ef(\cdot, t)$ is the same function:

$$Ef(\cdot, t) = \check{f}, \quad \text{for all } t \in \mathbb{R}.$$

This means that, as long as $\check{f}$ is not the zero function (that is, f is non-zero), the p -norm of Ef (which is given by an integral over x' and t , as opposed to the ∞ -norm of Ef) is infinite:

$$\begin{aligned} \int_{x \in \mathbb{R}^n} |Ef(x)|^p dx &\stackrel{\text{Fubini}}{=} \int_{t \in \mathbb{R}} \left(\int_{x' \in \mathbb{R}^{n-1}} |Ef(x', t)|^p dx' \right) dt \\ &= \int_{t \in \mathbb{R}} \left(\int_{x' \in \mathbb{R}^{n-1}} |\check{f}(x')|^p \right) dt \\ &= +\infty, \end{aligned}$$

as we are integrating, over $\mathbb{R}$, the non-zero constant $\int_{x' \in \mathbb{R}^{n-1}} |\check{f}(x')|^p$. Therefore,

$$\|Ef\|_{L^p(\mathbb{R}^n)} = +\infty \quad \text{for all non-zero } f \in L^1(B^{n-1}(0,1)).$$

□

It was Stein who first noticed that, when Σ has non-vanishing Gaussian curvature (in which case all principal curvatures of Σ are non-zero), things are very different. For instance, the celebrated Stein-Tomas theorem states that the extension operator E associated to the sphere $\mathbb{S}^{n-1}$ satisfies

$$\|Ef\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_2, \quad f \in L^2(B^{n-1}(0,1)),$$

for all $p \geq \frac{2(n+1)}{n-1}$. In other words, in this case we can push the exponent p on the LHS of (33) much lower than $+\infty$. We will see the proof of this in due time.

For now, let us acquire some intuition as to why the curvature of Σ should even play a role.

→ The flatter Σ is, the fewer restriction estimates we get.

More precisely, the flatter Σ is, the higher p is forced to be on the LHS of (33). Flatness of the hypersurface

$$\Sigma = \{(\omega, h(\omega)) : \omega \in B^{n-1}(0,1)\},$$

when h is smooth, can be understood in terms of the order of vanishing of the function h (whose graph is Σ) at 0.

DEFINITION 5.4.2. Let $h : B^{n-1}(0, 1) \rightarrow \mathbb{R}$ be smooth. We say that h vanishes at order k at $0 \in \mathbb{R}^{n-1}$ when all partial derivatives of h of order $0, 1, \dots, k-1$ vanish at 0 .

The higher the order of vanishing at 0 , the flatter Σ is. Equivalently, the lower the order of vanishing at 0 , the more curved Σ is. When Σ is the horizontal hyperplane through $0 \in \mathbb{R}^n$, then $h = 0$, so h vanishes at all orders at 0 (that is, the order of vanishing is $+\infty$). When Σ is the paraboloid of the sphere, h vanishes at 0 at order 2. We can imagine that as we bend the paraboloid more and more, so that it becomes the horizontal hyperplane, we pass through all the intermediate states: h starts to vanish at higher and higher order at 0 , until Σ becomes flat.

Two comments:

- *Why don't we talk about order of vanishing 0 or 1?* In terms of the Fourier restriction problem, we may always assume that our Σ is such that

$$h(0) = 0 \quad \text{and} \quad \nabla h(0) = 0,$$

i.e. that h vanishes at 0 at order ≥ 2 . (This is why, for our purposes, surfaces such as the paraboloid and the sphere are optimally curved.) Indeed, $h(0) = 0$ is simply achieved via a translation of Σ . Moreover, $(\nabla h(0), -1)$ is simply a normal vector to Σ at the point $(0, h(0)) = 0 \in \mathbb{R}^n$. Requiring for $\nabla h(0)$ to be 0 is therefore the same as requiring for the normal to Σ at $0 \in \mathbb{R}^n$ to point in the vertical direction. This can further be achieved via a rotation of Σ .

- *Why do we only consider derivatives at 0 to understand flatness of Σ ?* Derivatives are only controlled by local properties, so Σ could be curved close to 0 , but flat further out, which could mess things up. This is indeed correct - however, to prove a restriction estimate (e.g. of the form (33)) for the compact set Σ , we can cut up Σ in smaller, finitely many pieces (such as caps), on each of which the order of vanishing of the defining function is fixed.

PROPOSITION 5.4.3. Let

$$\Sigma := \{(\omega, h(\omega)) : \omega \in B^{n-1}(0, 1)\}$$

be a smooth hypersurface in $\mathbb{R}^n$, where h vanishes at $0 \in \mathbb{R}^{n-1}$ at order k (with $k \geq 2$). Then, any restriction estimate of the form

$$(34) \quad \|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma, p, q} \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma),$$

may only be true for

$$p \geq \frac{n+k-1}{n-1} q'.$$

The quantity $\frac{n+k-1}{n-1} q'$ increases with k . So, the flatter Σ is (that is, the larger k is), the stronger conditions are imposed on p , and thus the fewer restriction estimates we get. When $k = 2$ (e.g. when Σ is the paraboloid or the sphere, the hypersurfaces of most interest when it comes to Fourier restriction), we get the most lenient conditions on p from the above proposition. Even then, the restriction estimate (34) may only be true when

$$p \geq \frac{n+1}{n-1} q'.$$

PROOF. The proposition is proved via what are known as 'Knapp examples'. A Knapp example is a function supported on a little cap of Σ . In this case, for every small

$\delta > 0$, we consider the δ -cap θ on Σ with centre 0, and a smooth function f , with

$$f = 1 \text{ on } \frac{1}{2}\theta, \quad 0 \leq f \leq 1, \quad \text{supp } f \subseteq \theta.$$

So, $\|f\|_{L^\infty(\Sigma)} = 1$, and, for every $q \in [1, +\infty)$,

$$\|f\|_{L^q(\Sigma)} \sim |\theta|^{1/q} \sim \delta^{\frac{n-1}{q}}.$$

We now wish to understand $\widehat{f d\sigma}$. We know that when a function has support on a little rectangle, then its Fourier transform is large in magnitude on the dual rectangle through 0. We expect something similar to hold here for $\widehat{f d\sigma}$, which is again an (inverse) Fourier transform of sorts, just with respect to the surface measure of Σ . Indeed, we verify this easily, by slightly fattening f , and then comparing $\widehat{f d\sigma}$ with the Fourier transform of the fattened version:

First, we denote by $\bar{\theta}$ the smallest rectangle in $\mathbb{R}^n$ containing θ . This rectangle is horizontal, and has dimensions $\delta \times \cdots \times \delta \times \lambda$, where, crucially, the height λ depends on the curvature of Σ . In particular, since h vanishes at order k at 0, it follows that

$$|h(\omega)| \lesssim |\omega|^k \quad \text{for all } \omega \in B^{n-1}(0, 1).$$

Therefore,

$$\lambda \lesssim \delta^k,$$

and equality is possible (e.g. if $h(\omega) = |\omega|^k$). Therefore, the best we can say is that $\bar{\theta}$ is contained in a horizontal $\delta \times \cdots \times \delta \times \delta^k$ rectangle in $\mathbb{R}^n$. For simplicity, we denote by $\bar{\theta}$ the slightly larger $\delta \times \cdots \times \delta \times 2\delta^k$ -rectangle, with the same base and double the height.

We see that the δ -cap θ is contained in the rectangle $\bar{\theta}$. Moreover, due to its slightly enlarged height, $\bar{\theta}$ also contains all the vertical translations of θ of the form

$$\theta_t := (0, \dots, 0, t) + \theta, \quad 0 \leq t \leq \delta^k.$$

The union θ_{fat} of these translations is a fattened version of θ (fattened by a factor of δ^k). And it lives inside the rectangle $\bar{\theta}$.

We now consider a copy of f on each of these translated copies of θ . More precisely, for every $0 \leq t \leq \delta^k$, we define $f_t : \theta_t \rightarrow \mathbb{C}$, by

$$f_t := f(\cdot - (0, \dots, 0, t)).$$

We have thus created a fattened version of f , $f_{\text{fat}} : \theta_{\text{fat}} \rightarrow \mathbb{C}$, such that

$$f = f_t \text{ on } \theta_t, \quad \text{for all } 0 \leq t \leq \delta^k.$$

By Fubini, $\widetilde{f_{\text{fat}}}(x)$ and $\widehat{f d\sigma}$ are related. Indeed, for every $x \in \mathbb{R}^n$,

$$\begin{aligned} \widetilde{f_{\text{fat}}}(x) &= \int_{y \in \theta_{\text{fat}}} e^{2\pi i \langle x, y \rangle} f_{\text{fat}}(y) dy \\ &\stackrel{y = \xi + (0, \dots, 0, t) : \xi \in \theta, t \in [0, \delta^k]}{=} \int_{t \in [0, \delta^k]} \left(\int_{\xi \in \theta} e^{2\pi i \langle x, \xi \rangle} e^{2\pi i x_n t} f(\xi) J(\xi) d\sigma(\xi) \right) dt \\ &= \int_{t \in [0, \delta^k]} e^{2\pi i x_n t} \left(\int_{\xi \in \theta} e^{2\pi i \langle x, \xi \rangle} f(\xi) J(\xi) d\sigma(\xi) \right) dt \end{aligned}$$

where $J(\xi) = J(\xi, t)$ is the Jacobian factor due to the change of variables, and which we can see is independent of t and satisfies $J \sim_\Sigma 1$ (the implicit constant depends on the

curvature of Σ). First of all, this means that, if we start with f/J , rather than f , $\widehat{fd\sigma}(x)$ will appear in the integrand above:

$$\begin{aligned} \left((f/J)_{\text{fat}} \right)^{\sim}(x) &= \int_{t \in [0, \delta^k]} e^{2\pi i x_n t} \left(\int_{\xi \in \theta} e^{2\pi i \langle x, \xi \rangle} f(\xi) d\sigma(\xi) \right) dt \\ &= \int_{t \in [0, \delta^k]} e^{2\pi i x_n t} \widehat{fd\sigma}(x) dt \\ &= \widehat{fd\sigma}(x) \cdot \left(\int_{t \in [0, \delta^k]} e^{2\pi i x_n t} dt \right). \end{aligned}$$

Considering $x \in \mathbb{R}^n$ with $|x_n \delta^k| \leq 1$, i.e. $|x_n| \leq \frac{1}{\delta^k}$, we see that

$$\left| \int_{t \in [0, \delta^k]} e^{2\pi i x_n t} dt \right| \sim \frac{1}{|x_n|} \cdot |e^{2\pi i x_n \delta^k} - e^{2\pi i x_n 0}| \sim \frac{1}{|x_n|} \cdot |x_n \delta^k - x_n 0| \sim \delta^k.$$

Therefore,

$$\left| \left((f/J)_{\text{fat}} \right)^{\sim}(x) \right| \sim \delta^k \cdot |\widehat{fd\sigma}(x)|, \quad \text{for all } x \in \mathbb{R}^n \text{ with } |x_n| \leq \frac{1}{\delta^k}.$$

Moreover, f/J is not too different from f , as $J \sim_{\Sigma} 1$. In particular,

$$\text{supp}(f/J)_{\text{fat}} \subseteq \bar{\theta}, \quad \int_{\bar{\theta}} (f/J)_{\text{fat}} \sim |\bar{\theta}|.$$

Thanks to our analysis on the uncertainty principle, we know that

$$\left| \left((f/J)_{\text{fat}} \right)^{\sim}(x) \right| \gtrsim |\bar{\theta}| \quad \text{on } \frac{1}{100} \cdot \bar{\theta}^*,$$

where $\bar{\theta}^*$ is the dual rectangle of $\bar{\theta}$ through 0. Therefore,

$$|\widehat{fd\sigma}| \gtrsim \delta^{-k} \cdot |\bar{\theta}| \quad \text{on } \frac{1}{100} \cdot \bar{\theta}^*.$$

It follows that

$$\begin{aligned} \|\widehat{fd\sigma}\|_{L^p(\mathbb{R}^n)} &\geq \left(\int_{\frac{1}{100} \cdot \bar{\theta}^*} |\widehat{fd\sigma}|^p \right)^{1/p} \\ &\gtrsim (\delta^{-k} |\bar{\theta}|) \cdot |\bar{\theta}^*|^{1/p} \\ &\sim \delta^{-k} \cdot \left(\frac{1}{\delta^{n-1}} \cdot \frac{1}{\delta^k} \right)^{1/p-1} \\ &= \left(\frac{1}{\delta} \right)^{k - \frac{n+k-1}{p'}}. \end{aligned}$$

Therefore, if (34) holds for some $p \in [1, +\infty)$ and some $q \in [1, +\infty]$, then there exists a constant $C = C(\Sigma, p, q)$ so that

$$\left(\frac{1}{\delta} \right)^{k - \frac{n+k-1}{p'}} \leq C \cdot \delta^{\frac{n-1}{q}}, \quad \text{for all small } \delta > 0.$$

Equivalently, we should have

$$\left(\frac{1}{\delta} \right)^{k - \frac{n+k-1}{p'} + \frac{n-1}{q}} \leq C, \quad \text{for all small } \delta > 0.$$

Letting $\delta \rightarrow 0$, we see that this is only possible when the exponent on the LHS is ≤ 0 . So, (34) fails whenever

$$k - \frac{n+k-1}{p'} + \frac{n-1}{q} > 0,$$

or equivalently

$$p < \frac{n+k-1}{n-1} q',$$

as the proposition asserts. □

For simplicity, we denote from now on by $E_\Sigma(q \rightarrow p)$ a statement of the form

$$\|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma, p, q} \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma).$$

Moreover, we focus henceforth on hypersurfaces Σ such as the [sphere](#) or the [paraboloid](#), which vanish only at order 2 at 0, and for which we know from the above that $E_\Sigma(q \rightarrow p)$ may hold for

$$(35) \quad p \geq \frac{n+1}{n-1} q'$$

(in other words, these are the hypersurfaces for which the above proposition imposes the fewest conditions on p , and thus we hope to get the most Fourier restriction estimates).

—→ **An additional condition.**

At the same time, there is another condition on p that shows up, when we consider $f = 1$ identically on Σ . This condition shows up for all smooth Σ of non-vanishing Gaussian curvature, but traditionally it arose for $\Sigma = \mathbb{S}^{n-1}$.

PROPOSITION 5.4.4. *Let $d\sigma$ be the Lebesgue surface measure on the sphere $\mathbb{S}^{n-1}$. Then,*

$$|\widehat{d\sigma}(x)| \sim \frac{1}{|x|^{\frac{n-1}{2}}} \quad \text{for all } |x| \gtrsim 1.$$

The proof of this is quite technical, and is based on the method of *stationary phase*, which we aim to see later on in the course.² For now we accept this as a fact, and we notice that this means that, if $E_{\mathbb{S}^{n-1}}(q \rightarrow p)$ holds for some $p \in [1, +\infty)$ (and some $q \in [1, +\infty]$), then equivalently we have

$$\|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma, p, q} \|f\|_{L^q(\mathbb{S}^{n-1})}, \quad \text{for all } f \in L^q(\mathbb{S}^{n-1}),$$

so the LHS should be finite for every $f \in L^q(\mathbb{S}^{n-1})$, and in particular for $f = 1$. In other words, we should have that

$$\|\widehat{d\sigma}\|_{L^p(\mathbb{R}^n)} < +\infty,$$

and thus, by the above proposition, that

$$\int_{|x| \gtrsim 1} \frac{1}{|x|^{\frac{(n-1)p}{2}}} dx < +\infty.$$

This works if and only if $\frac{(n-1)p}{2} > n$, or equivalently

$$(36) \quad p > \frac{2n}{n-1}.$$

Notice that this lower bound is *strictly larger* than 2, which was our previous absolute lower bound. So, this is a *stronger condition* than our original $p > 2$. This therefore

²The proof can be found in T. Tao's Lecture Notes 1 (Proposition 5.3).

becomes, after (35), the second necessary condition we have found for $E_{\mathbb{S}^{n-1}}(q \rightarrow p)$ to hold for some p .

It turns out that (36) is necessary not only when Σ is the sphere, but also any smooth hypersurface of non-vanishing Gaussian curvature (such as the paraboloid). Indeed, the method of non-stationary phase again gives that

$$|\widehat{d\sigma}(x)| \sim \frac{1}{|x|^{\frac{n-1}{2}}}$$

for enough x far from 0 so that the above argument still applies.

→ The Fourier restriction conjecture.

It is conjectured that conditions (35) and (36) are *the only ones* required for $E_{\Sigma}(q \rightarrow p)$ to hold, when Σ is a smooth hypersurface of non-vanishing Gaussian curvature.

The Fourier restriction conjecture (Stein, 1960s). Let Σ be a smooth hypersurface in $\mathbb{R}^n$, of non-vanishing Gaussian curvature. Then, the estimate

$$(37) \quad \|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma,p,q} \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma)$$

holds for some $p, q \in [1, +\infty]$ if and only if

$$p > \frac{2n}{n-1} \quad \text{and} \quad p \geq \frac{n+1}{n-1} q'.$$

Some comments:

- Recall that the goal was to push p on the LHS of the restriction estimate (37) *as low as possible*. The q was not so important in terms of the Fourier restriction problem (even though, of course, the lower the q , the stronger (37) becomes).
- We already know that we *cannot* push p on the LHS of (37) below $\frac{2n}{n-1}$ (in fact, p cannot even touch $\frac{2n}{n-1}$).
- The additional condition $p \geq \frac{n+1}{n-1} q'$ may force p to be even larger (which is something we don't want). So, the most useful q for Fourier restriction are those for which

$$\frac{2n}{n-1} \geq \frac{n+1}{n-1} q',$$

i.e. the q for which

$$q \geq \frac{2n}{n-1}.$$

Focusing on these q , the Fourier restriction conjecture simply states that

$$\|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma,p,q} \|f\|_{L^q(\Sigma)} \quad \text{for all } f \in L^q(\Sigma),$$

for all (p, q) with $p > \frac{2n}{n-1}$ and $q \geq \frac{2n}{n-1}$. Therefore, the strongest form of the conjecture states that

$$\|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma,p,q} \|f\|_{L^{\frac{2n}{n-1}}(\Sigma)} \quad \text{for all } f \in L^{\frac{2n}{n-1}}(\Sigma)$$

whenever $p > \frac{2n}{n-1}$.

However, since the focus of the Fourier restriction problem is on the p on the LHS, we can relax (that is, increase) the q on the RHS. This is why it is common to see the conjecture in the literature as saying that

$$\|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma,p,q} \|f\|_{L^p(\Sigma)} \quad \text{for all } f \in L^p(\Sigma)$$

whenever $p > \frac{2n}{n-1}$, or even in its weakest form, which asserts that

$$\|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_{\Sigma,p,q} \|f\|_{L^\infty(\Sigma)} \quad \text{for all } f \in L^\infty(\Sigma)$$

whenever $p > \frac{2n}{n-1}$. At least in terms of the Fourier restriction problem (which focuses on the p alone), all these forms are of similar interest.

→ ***The local version.***

Instead of aiming to control $\widehat{f d\sigma}$ on the whole of $\mathbb{R}^n$, one can aim to control it on every big ball $B_R := B(0, R)$ in $\mathbb{R}^n$. In fact, we can ask for which p and q the estimate

$$(38) \quad \|\widehat{f d\sigma}\|_{L^p(B_R)} \lesssim_{\Sigma,p,q,\epsilon} R^\epsilon \|f\|_{L^q(\Sigma)} \quad \text{for all } f \in L^q(\Sigma)$$

holds, for all $R \geq 1$ and all $\epsilon > 0$. With the same argument as earlier, we again see that

$$p \geq \frac{n+1}{n-1} q'$$

is necessary, whenever Σ has non-vanishing Gaussian curvature. However, the earlier condition $p > \frac{2n}{n-1}$ can now be relaxed to

$$p \geq \frac{2n}{n-1}$$

(that is, in this case perhaps p can even touch the critical exponent $\frac{2n}{n-1}$). Indeed, the condition $p > \frac{2n}{n-1}$ arose earlier from the fact that

$$|\widehat{d\sigma}(x)| \sim \frac{1}{|x|^{\frac{n-1}{2}}} \quad \text{for all } |x| \gtrsim 1$$

when $\Sigma = \mathbb{S}^{n-1}$. For $\widehat{d\sigma}$ to live in $L^p(\mathbb{R}^n)$, in other words for $\frac{1}{|x|^{\frac{n-1}{2}}}$ to be L^p -integrable on the whole of $\mathbb{R}^n$, we needed $\frac{p(n-1)}{2}$ to be *strictly larger* than n . However, for the local version (38) we only need to integrate on B_R . And requiring (38) to hold for $f = 1$ (and all $R \geq 1$ and $\epsilon > 0$) yields

$$\int_{1 \leq |x| \leq R} \frac{1}{|x|^{\frac{p(n-1)}{2}}} dx \lesssim_{\Sigma,q,\epsilon} R^\epsilon.$$

The LHS above equals

$$\int_{\xi \in \mathbb{S}^{n-1}} \int_{r=1}^R \frac{1}{r^{\frac{p(n-1)}{2}}} r^{n-1} dr d\sigma(\xi) = \int_1^R \frac{1}{r^{\frac{p(n-1)}{2} - (n-1)}} dr,$$

and is $\lesssim_{\Sigma,p,q} R^\epsilon$ for all $\epsilon > 0$ and all $R \geq 1$ if and only if

$$\frac{p(n-1)}{2} - (n-1) \geq 1$$

(as, even when it equals 1, the integral is $\log R$), or equivalently

$$p \geq \frac{2n}{n-1}.$$

The Fourier restriction conjecture (local version). Let Σ be a smooth hypersurface in $\mathbb{R}^n$, of non-vanishing Gaussian curvature. Then, the estimate

$$\|\widehat{f d\sigma}\|_{L^p(B_R)} \lesssim_{\Sigma,p,q,\epsilon} R^\epsilon \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma)$$

holds for all $R \geq 1$ and $\epsilon > 0$, for all

$$p \geq \frac{2n}{n-1} \quad \text{and} \quad p \geq \frac{n+1}{n-1} q'.$$

The strongest form of the conjecture in this local form is acquired for the lowest possible p and q , i.e. for $p = q = \frac{2n}{n-1}$.

It turns out that **the local and global version of the Fourier restriction conjecture are equivalent**. This is due an ϵ -removal lemma by Tao, who showed that if the local estimate

$$\|\widehat{f d\sigma}\|_{L^{p_0}(B_R)} \lesssim_{\Sigma, p, q, \epsilon} R^\epsilon \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma), \ R \geq 1 \text{ and } \epsilon > 0$$

holds for *some* p_0 and q , then the global estimate

$$\|\widehat{f d\sigma}\|_{L^p(B_R)} \lesssim_{\Sigma, p, q} \|f\|_{L^q(\Sigma)}, \quad \text{for all } f \in L^q(\Sigma)$$

holds for all $p > p_0$ (and the same q).

In other words, the estimate

$$\|\widehat{f d\sigma}\|_{L^{\frac{2n}{n-1}}(B_R)} \lesssim_{\Sigma, p, q, \epsilon} R^\epsilon \|f\|_{L^{\frac{2n}{n-1}}(\Sigma)}, \quad \text{for all } f \in L^{\frac{2n}{n-1}}(\Sigma), \ R \geq 1 \text{ and } \epsilon > 0$$

is equivalent to the Fourier restriction conjecture.

Connections with other problems. The Fourier restriction conjecture has been proved for $n = 2$, but is open in all higher dimensions.

The Fourier restriction conjecture in $\mathbb{R}^n$ implies the Kakeya maximal operator conjecture in $\mathbb{R}^n$, and thus that Kakeya sets have full Hausdorff dimension (for all $n \geq 2$). It also has other implications in **geometric measure theory** (such as in Falconer's distance set conjecture).

It has been shown by Tao and Carbery that, to an extent, the Fourier restriction conjecture is equivalent to the **Bochner-Riesz conjecture**.

Apart from its connection to **PDE**, the study of the Fourier extension operator E associated to the paraboloid recently led to the resolution of the main conjecture in Vinogradov's Mean Value Theorem from **number theory** (a revolutionary proof by Bourgain, Demeter and Guth from 2015).

5.5. Restriction implies Kakeya

We prove here that the Fourier restriction conjecture implies the maximal Kakeya operator conjecture (and thus that Kakeya sets in $\mathbb{R}^n$ have Hausdorff dimension equal to n).

PROPOSITION 5.5.1. *Let $n \geq 2$, and consider the sphere $\mathbb{S}^{n-1}$ with associated surface measure $d\sigma$. Suppose that*

$$\|\widehat{f d\sigma}\|_{L^{\frac{2n}{n-1}}(B_R)} \lesssim_\epsilon R^\epsilon \|f\|_{L^{\frac{2n}{n-1}}(\Sigma)}, \quad \text{for all } f \in L^{\frac{2n}{n-1}}(\Sigma), \ R \geq 1 \text{ and } \epsilon > 0.$$

Then, for $\delta > 0$ and every family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$, it holds that

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{\frac{n}{n-1}} \lesssim_\epsilon \delta^{-\epsilon} \delta^{n-1} \#\mathbb{T}, \quad \text{for all } \epsilon > 0.$$

PROOF. Recall that the RHS $\delta^{n-1} \#\mathbb{T}$ in the Kakeya maximal operator estimate we are aiming for equals $\sum_{T \in \mathbb{T}} |T|$. Therefore, this Kakeya estimate is dilation-invariant. Moreover, we may assume that the tubes it involves (which are all of width δ and length

1) all live in B_1 . So, it suffices instead to show that, for every $R \geq 1$ and every family $\mathbb{T}$ of $R^{-1/2}$ -separated $R^{1/2} \times \dots \times R^{1/2} \times R$ tubes in B_R , it holds that

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{\frac{n}{n-1}} \lesssim_\epsilon R^\epsilon \sum_{T \in \mathbb{T}} |T|, \quad \text{for all } \epsilon > 0.$$

The proof strategy is the same as the one in Fefferman's theorem. Let $\mathbb{T}$ be such a family of tubes. We will create a function f on $\mathbb{S}^{n-1}$, so that $\widehat{f d\sigma}$ is a sum of waves, each evolving inside a different tube of $\mathbb{T}$.

Indeed, consider a collection Θ of $R^{-1/2}$ -caps on $\mathbb{S}^{n-1}$, with normals the directions of the tubes in $\mathbb{T}$. Since the tubes are $R^{-1/2}$ -separated, the normals to the caps in Θ are also $R^{-1/2}$ -separated. Due to the geometry of the sphere, this means that the caps in Θ are essentially disjoint.

Denote by T_θ the tube in $\mathbb{T}$ in direction normal to θ . For every $\theta \in \Theta$, let ψ_θ a smooth function, which is a mollified version of χ_θ ; in particular,

$$\text{supp } \psi_\theta \subseteq \theta, \quad \psi_\theta = 1 \text{ on } \frac{1}{2}\theta, \quad 0 \leq \psi_\theta \leq 1.$$

Denote by $\bar{\theta}$ the smallest rectangle containing the $R^{-1/2}$ -cap θ (which we know has dimensions roughly $R^{-1/2} \times \dots \times R^{-1/2} \times R^{-1}$).

Then, just like in the case where θ is an actual rectangle (rather than a cap), we have that

$$\left| \left(\frac{\psi_\theta}{|\theta|} d\sigma \right)^\wedge \right| \gtrsim 1 \quad \text{on } \frac{1}{100}\bar{\theta}^*$$

and that, modulating the function $\frac{\psi_\theta}{|\theta|}$ appropriately, its extension operator is large in magnitude in whichever translation of the tube $\frac{1}{100n}\bar{\theta}^*$ we wish. The tube $\frac{1}{100n}T_\theta$ is such a translation.

Indeed, let ω_θ be the centre of θ and define $\theta_0 := \theta - \omega_\theta$ and $\psi_{\theta_0} := \psi(\cdot + \omega_\theta)$ to be the translations of θ and ψ_θ at 0, respectively.

• To begin with,

$$\begin{aligned} \left| \left(\frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \right)^\wedge (x) - \left(\frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \right)^\wedge (0) \right| &= \left| \int e^{2\pi i \langle x, \xi \rangle} \frac{\psi_{\theta_0}(\xi)}{|\theta_0|} d\sigma(\xi) - \int e^{2\pi i \langle 0, \xi \rangle} \frac{\psi_{\theta_0}(\xi)}{|\theta_0|} d\sigma(\xi) \right| \\ &\leq \frac{1}{|\theta_0|} \int |e^{2\pi i \langle x, \xi \rangle} - e^{2\pi i \langle 0, \xi \rangle}| \psi_{\theta_0}(\xi) d\sigma(\xi) \\ &\leq 16 \cdot \frac{1}{|\theta_0|} \int_{\theta_0} |\langle x, \xi \rangle| \psi_{\theta_0} d\sigma(\xi) \\ &\leq \frac{1}{2} \int \frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \\ &= \frac{1}{2} \left(\frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \right)^\wedge (0) \quad \text{for all } x \in \frac{1}{100n}\bar{\theta}^*, \end{aligned}$$

where the penultimate estimate is due to the fact that $|\langle x, \xi \rangle| \leq \frac{1}{50}$ for all $\frac{1}{100n}\bar{\theta}^*$ and $\xi \in \theta_0$. And since

$$\left(\frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \right)^\wedge(0) = \int \frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \sim 1,$$

it follows that

$$\left| \left(\frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \right)^\wedge(x) \right| \gtrsim 1 \quad \text{on } \frac{1}{100n}\bar{\theta}^*.$$

- Since $\frac{\psi_\theta}{|\theta|}$ is a translation of $\frac{\psi_{\theta_0}}{|\theta_0|}$, its extension operator should be a modulation of $\frac{\psi_{\theta_0}}{|\theta_0|}$ (and thus should also be large in magnitude on $\frac{1}{100n}\bar{\theta}^*$). Indeed, for all $x \in \mathbb{R}^n$,

$$\begin{aligned} \left(\frac{\psi_\theta}{|\theta|} d\sigma \right)^\wedge(x) &= \int_{\xi \in \theta} e^{2\pi i \langle x, \xi \rangle} \frac{\psi_\theta(\xi)}{|\theta|} d\sigma(\xi) \\ &\stackrel{\xi=y+\omega_\theta}{=} \int_{y \in \theta_0} e^{2\pi i \langle x, y+\omega_\theta \rangle} \frac{\psi_{\theta_0}(y)}{|\theta_0|} d\sigma(y) \\ &= e^{2\pi i \langle x, \omega_\theta \rangle} \int_{y \in \theta_0} e^{2\pi i \langle x, y \rangle} \frac{\psi_{\theta_0}(y)}{|\theta_0|} d\sigma(y) \\ &= e^{2\pi i \langle x, \omega_\theta \rangle} \left(\frac{\psi_{\theta_0}}{|\theta_0|} d\sigma \right)^\wedge(x). \end{aligned}$$

Therefore,

$$\left| \left(\frac{\psi_\theta}{|\theta|} d\sigma \right)^\wedge \right| \gtrsim 1 \quad \text{on } \frac{1}{100n}\bar{\theta}^*.$$

- Finally, let $v(\theta)$ be the centre of the tube T_θ . We consider the modulation

$$g_\theta(\xi) := e^{-2\pi i \langle v(\theta), \xi \rangle} \frac{\psi_\theta(\xi)}{|\theta|}, \quad \xi \in \mathbb{S}^{n-1}.$$

Then, for all $x \in \mathbb{R}^n$,

$$\begin{aligned} \widehat{g_\theta d\sigma}(x) &= \int e^{2\pi i \langle x, \xi \rangle} e^{-2\pi i \langle v(\theta), \xi \rangle} \frac{\psi_\theta(\xi)}{|\theta|} d\sigma(\xi) \\ &= \int e^{2\pi i \langle x-v(\theta), \xi \rangle} \frac{\psi_\theta(\xi)}{|\theta|} d\sigma(\xi) \\ &= \widehat{\psi_\theta d\sigma}(x - v(\theta)). \end{aligned}$$

Therefore,

$$|\widehat{g_\theta d\sigma}| \gtrsim 1 \quad \text{on } v(\theta) + \frac{1}{100n}\bar{\theta}^* = \frac{1}{100n}T_\theta.$$

- We wanted $|\widehat{g_\theta d\sigma}|$ to be large on $\mathbb{T}_\theta$, rather than on its small dilation $\frac{1}{100n}T_\theta$. This is simply achieved by applying the above for caps θ of dimensions $cR^{-1/2} \times \dots \times cR^{-1/2} \times c^2R^{-1}$, for an appropriate constant c (which depends only on n). This way, for each such θ , we create a function g_θ supported on θ , with

$$|\widehat{g_\theta d\sigma}| \gtrsim 1 \quad \text{on } T_\theta.$$

Now, we switch our perspective to the tubes in $\mathbb{T}$. For each $T \in \mathbb{T}$, we rename $g_T := g_\theta$, where θ is the unique cap $\theta \in \Theta$ with $T = T_\theta$. Denote this cap by $\theta(T)$. Then, for $(\omega_T)_{T \in \mathbb{T}}$ a random choice of signs ± 1 (where the ω_T are independent random variables, each taking values ± 1 with probability $1/2$), we consider the randomised function

$$g := \sum_{T \in \mathbb{T}} \omega_T g_T \quad \text{on } \mathbb{S}^{n-1},$$

which we have shown satisfies

$$\text{supp } g_T \subseteq \theta(T), \quad \int_{\theta(T)} |g_T| \sim 1, \quad \text{and} \quad |\widehat{g_T d\sigma}| \gtrsim 1 \text{ on } T, \quad \text{for each } T \in \mathbb{T}.$$

Suppose that the Fourier restriction conjecture holds for $\mathbb{S}^{n-1}$. Then, we must have that, for all $\epsilon > 0$, there exists constant $C_\epsilon > 0$ so that

$$(39) \quad \int_{B_R} |\widehat{g d\sigma}|^{\frac{2n}{n-1}} \leq C_\epsilon R^\epsilon \int_{\mathbb{S}^{n-1}} |g|^{\frac{2n}{n-1}},$$

for all choices $(\omega_T)_{T \in \mathbb{T}}$ of signs.

On the one hand, $g = \sum_{T \in \mathbb{T}} \omega_T g_T$, where all the functions g_T have disjoint supports $\theta(T)$. Moreover, for each $T \in \mathbb{T}$, $|g_T| = \frac{\psi_{\theta(T)}}{|\theta(T)|}$. Therefore,

$$\begin{aligned} \int_{\mathbb{S}^{n-1}} |g|^{\frac{2n}{n-1}} d\sigma &= \sum_{T \in \mathbb{T}} \int_{\theta(T)} |g_T|^{\frac{2n}{n-1}} d\sigma \\ &= \sum_{T \in \mathbb{T}} \int_{\theta(T)} \left(\frac{\psi_{\theta(T)}}{|\theta(T)|} \right)^{\frac{2n}{n-1}} d\sigma \\ &\sim \sum_{T \in \mathbb{T}} |\theta(T)|^{1 - \frac{2n}{n-1}} = \sum_{T \in \mathbb{T}} |\theta(T)|^{-\frac{n+1}{n-1}} \\ &= \sum_{T \in \mathbb{T}} (R^{-1/2})^{-(n-1) \cdot \frac{n+1}{n-1}} = \sum_{T \in \mathbb{T}} R^{\frac{n+1}{2}} \\ &= \sum_{T \in \mathbb{T}} |T|, \end{aligned}$$

which is our desired RHS.

On the other hand, we can control the average of the integrals $\int_{B_R} |\widehat{g d\sigma}|^{\frac{2n}{n-1}}$, over all choices $(\omega_T)_{T \in \mathbb{T}}$ of signs, using Khinchin's inequality:

$$\begin{aligned} \mathbb{E} \left(\int_{x \in B_R} |\widehat{g d\sigma}(x)|^{\frac{2n}{n-1}} dx \right) &\stackrel{\text{Fubini}}{=} \int_{x \in \mathbb{B}_R} \mathbb{E} \left(|\widehat{g d\sigma}(x)|^{\frac{2n}{n-1}} \right) dx \\ &= \int_{x \in B_R} \mathbb{E} \left(\left| \sum_{T \in \mathbb{T}} \widehat{g_T d\sigma}(x) \right|^{\frac{2n}{n-1}} \right) dx \\ &\stackrel{\text{Khinchin}}{\sim} \int_{x \in B_R} \left(\sum_{T \in \mathbb{T}} |\widehat{g_T d\sigma}(x)|^2 \right)^{\frac{n}{n-1}} dx \\ &\stackrel{|\widehat{g_T d\sigma}| \gtrsim \chi_T, \forall T}{\gtrsim} \int_{x \in B_R} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^{\frac{n}{n-1}} dx \\ &\stackrel{T \subseteq B_R, \forall T}{=} \int_{x \in \mathbb{R}^n} \left(\sum_{T \in \mathbb{T}} \chi_T(x) \right)^{\frac{n}{n-1}} dx. \end{aligned}$$

So, there exists a choice of $(\omega_T)_{T \in \mathbb{T}}$, so that the corresponding $g = \sum_{T \in \mathbb{T}} \omega_T g_T$ satisfies

$$\int_{x \in B_R} |\widehat{g d\sigma}(x)|^{\frac{2n}{n-1}} dx \gtrsim \int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{\frac{n}{n-1}},$$

which is our desired LHS. Plugging this g into (39), we obtain

$$\int \left(\sum_{T \in \mathbb{T}} \chi_T \right)^{\frac{n}{n-1}} \leq C_\epsilon R^\epsilon \sum_{T \in \mathbb{T}} |T|,$$

as we wished. □

5.6. The Stein-Tomas restriction theorem

In this section we will prove, away from the endpoint, a celebrated Fourier restriction theorem by Stein and Tomas.

THEOREM 5.6.1. (Stein-Tomas, 1975) *Let $n \geq 2$, and $d\sigma$ be the Lebesgue surface measure on $\mathbb{S}^{n-1}$. Then,*

$$\|\widehat{f d\sigma}\|_{L^{\frac{2(n+1)}{n-1}}(\mathbb{R}^n)} \lesssim \|f\|_{L^2(\mathbb{S}^{n-1})}, \quad \text{for all } f \in L^2(\mathbb{S}^{n-1}).$$

In other words, the Fourier restriction estimate $E_{\mathbb{S}^{n-1}} \left(2 \rightarrow \frac{2(n+1)}{n-1} \right)$ holds.

Some comments:

- *How close is Stein-Tomas to the full Fourier restriction conjecture?* Recall that the Fourier restriction problem asked for which p there exists q so that

$$(40) \quad \|\widehat{f d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^q(\mathbb{S}^{n-1})}, \quad \text{for all } f \in L^q(\mathbb{S}^{n-1}).$$

In fact, the goal was to lower p as much as possible. The Fourier restriction conjecture claims that we can make p as low as $\frac{2n}{n-1}$ ($+\epsilon$, for all $\epsilon > 0$). The Stein-Tomas theorem asserts that we can lower p as much as $\frac{2(n+1)}{n-1}$. The gap between $\frac{2(n+1)}{n-1}$ and $\frac{2n}{n-1}$ gets smaller and smaller as the dimension n increases, however it is very difficult to bridge.

- *How sharp is Stein-Tomas?* We know that the Fourier restriction conjecture (40) is expected to hold if and only if

$$p > \frac{2n}{n-1} \quad \text{and} \quad p \geq \frac{n+1}{n-1} q'.$$

When $p = \frac{2(n+1)}{n-1}$, the admissible q satisfy

$$q' \leq \frac{n-1}{n+1} p = \frac{n-1}{n+1} \cdot \frac{2(n+1)}{n-1} = 2,$$

hence $q \geq 2$. Therefore, given that the Stein-Tomas theorem focuses on $p = \frac{2(n+1)}{n-1}$, it is sharp: the q on its RHS cannot be lowered. Equivalently, if one insists on studying only functions $f \in L^2(\mathbb{S}^{n-1})$ in terms of Fourier restriction (and fixing $q = 2$ on the RHS of (40)), then the lowest possible p they can hope for in (40) is the Stein-Tomas exponent $\frac{2(n+1)}{n-1}$.

In other words:

The Stein-Tomas restriction estimate is the best possible estimate of the form $E_{\mathbb{S}^{n-1}}(2 \rightarrow p)$ that one can hope for.

The proof of the Stein-Tomas theorem away from the endpoint is done using what is known as ‘the TT^* method’, which is widely used in analysis. The idea is that an operator $T : L^2(A) \rightarrow L^p(B)$ (for some measure spaces A and B) is bounded if and only if the operator $TT^* : L^{p'}(B) \rightarrow L^2(A)$ is bounded. For us, T is the Fourier extension operator E associated to the sphere $\mathbb{S}^{n-1}$ (in which case of course the dual E^* is the Fourier restriction operator $\hat{\cdot}|_{\mathbb{S}^{n-1}}$), and $p = \frac{2(n+1)}{n-1}$.

—→ **The TT^* method.**

Let A and B be measure spaces, and T an operator that acts on functions $f : A \rightarrow \mathbb{C}$ and returns functions $Tf : B \rightarrow \mathbb{C}$. We know that

$$T : L^2(A) \rightarrow L^p(B) \text{ is well-defined and bounded}$$

if and only if

$$T^* : L^{p'}(B) \rightarrow L^2(A) \text{ is well-defined and bounded,}$$

as

$$\|T\|_{L^2(A) \rightarrow L^p(B)} = \|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}.$$

The proof of this is very simple: it is based on the fact that

$$\langle Tf, g \rangle = \langle f, T^*g \rangle$$

for all $f \in L^2(A)$ and $g \in L^{p'}(B)$.

(Here, we slightly abuse notation, to denote by $\langle \phi, \psi \rangle$ the integral $\int \phi \cdot \bar{\psi}$ with respect to the measure of the space on which ϕ and ψ are defined, as long as $\phi \cdot \bar{\psi}$ is integrable.)

We now use this to associate the norm of T with the an appropriate norm of TT^* .

THEOREM 5.6.2. (The TT^* method) *It holds that*

$$T : L^2(A) \rightarrow L^p(B) \text{ is well-defined and bounded}$$

if and only if

$$TT^* : L^{p'}(B) \rightarrow L^2(A) \text{ is well-defined and bounded.}$$

In fact,

$$\|T\|_{L^2(A) \rightarrow L^p(B)} = \sqrt{\|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}}.$$

PROOF. It suffices to focus on the claimed relationship between the norms of the two operators T and TT^* . And since $\|T\|_{L^2(A) \rightarrow L^p(B)} = \|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}$, it suffices to show that

$$\|T^*\|_{L^{p'}(B) \rightarrow L^2(A)} = \sqrt{\|TT^*\|_{L^{p'}(B) \rightarrow L^2(A)}}.$$

- $\|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}^2 \leq \|TT^*\|_{L^{p'}(B) \rightarrow L^2(A)}:$

$$\begin{aligned}
\|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}^2 &= \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \|T^*f\|_{L^2(A)}^2 \\
&= \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} |\langle T^*f, T^*f \rangle| \\
&= \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} |\langle f, TT^*f \rangle| \\
&\stackrel{\text{Hölder}}{\leq} \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \|f\|_{L^{p'}(B)} \|TT^*f\|_{L^p(B)} \\
&\leq \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \|f\|_{L^{p'}(B)} \cdot \|TT^*\|_{L^{p'}(B) \rightarrow L^p(B)} \|f\|_{L^{p'}(B)} \\
&= \|TT^*\|_{L^{p'}(B) \rightarrow L^p(B)}.
\end{aligned}$$

Therefore, if $TT^* : L^{p'}(B) \rightarrow L^p(B)$ is well-defined and bounded, then $T^* : L^{p'}(B) \rightarrow L^2(A)$ is also well-defined and bounded, and $\|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}^2 \leq \|TT^*\|_{L^{p'}(B) \rightarrow L^p(B)}$.

$$\bullet \|TT^*\|_{L^{p'}(B) \rightarrow L^p(B)} \leq \|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}^2:$$

$$\begin{aligned}
\|TT^*\|_{L^{p'}(B) \rightarrow L^p(B)} &= \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \|TT^*f\|_{L^p(B)} \\
&= \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \sup_{g \in L^{p'}(B), \|g\|_{p'}=1} \langle TT^*f, g \rangle \\
&= \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \sup_{g \in L^{p'}(B), \|g\|_{p'}=1} \langle T^*f, T^*g \rangle \\
&\stackrel{\text{Hölder}}{\leq} \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \sup_{g \in L^{p'}(B), \|g\|_{p'}=1} \|T^*f\|_{L^2(A)} \cdot \|T^*g\|_{L^2(A)} \\
&\leq \sup_{f \in L^{p'}(B), \|f\|_{p'}=1} \sup_{g \in L^{p'}(B), \|g\|_{p'}=1} \|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}^2 \|f\|_{L^2(A)} \|g\|_{L^2(A)} \\
&= \|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}^2.
\end{aligned}$$

Therefore, if $T^* : L^{p'}(B) \rightarrow L^2(A)$ is well-defined and bounded, then $TT^* : L^{p'}(B) \rightarrow L^p(B)$ is also well-defined and bounded, and $\|TT^*\|_{L^{p'}(B) \rightarrow L^p(B)} \leq \|T^*\|_{L^{p'}(B) \rightarrow L^2(A)}^2$. $\square$

→ The proof of the Stein-Tomas theorem.

We will here prove the Stein-Tomas theorem just off the endpoint. That is, we will show that

$$E_{\mathbb{S}^{n-1}}(2 \rightarrow p) \text{ holds, for all } p > \frac{2(n+1)}{n-1}.$$

The calculation to achieve the endpoint estimate is a bit more complicated (it uses complex interpolation, rather than Riesz-Thorin interpolation).

Let $n \geq 2$ and $p > \frac{2(n+1)}{n-1}$. From now on, let $T := \widehat{\cdot d\sigma}$, the extension operator associated to the sphere (in which case T^* is the Fourier restriction operator $\widehat{\cdot}|_{\mathbb{S}^{n-1}}$). We wish to show that

$$\|T\|_{L^2(\mathbb{S}^{n-1}) \rightarrow L^p(\mathbb{R}^n)} < +\infty,$$

or equivalently that

$$\|TT^*\|_{L^{p'}(\mathbb{R}^n) \rightarrow L^p(\mathbb{R}^n)} < +\infty$$

(by the TT^* method).

To prove this, it suffices to work with Schwartz functions, rather than the whole of $L^{p'}(\mathbb{R}^n)$. So, our first job is to find a formula for the image TT^*f , for every $f \in S(\mathbb{R}^n)$. Fix such f . Then,

$$T^*f = \widehat{f}|_{\mathbb{S}^{n-1}}$$

(a well-defined function), so T^*f is the usual Fourier transform of f , just restricted on $\mathbb{S}^{n-1}$:

$$T^*f(\xi) = \int_{y \in \mathbb{R}^n} e^{-2\pi i \langle \xi, y \rangle} f(y) dy, \quad \xi \in \mathbb{S}^{n-1}.$$

So,

$$\begin{aligned} TT^*f(x) &= \int_{\xi \in \mathbb{S}^{n-1}} e^{2\pi i \langle x, \xi \rangle} T^*f(\xi) d\sigma(\xi) \\ &= \int_{\xi \in \mathbb{S}^{n-1}} e^{2\pi i \langle x, \xi \rangle} \left(\int_{y \in \mathbb{R}^n} e^{-2\pi i \langle \xi, y \rangle} f(y) dy \right) d\sigma(\xi) \\ &\stackrel{\text{Fubini}}{=} \int_{y \in \mathbb{R}^n} \left(\int_{\xi \in \mathbb{S}^{n-1}} e^{2\pi i \langle x-y, \xi \rangle} d\sigma(\xi) \right) f(y) dy \\ &= \int_{y \in \mathbb{R}^n} \widehat{d\sigma}(x-y) f(y) dy \\ &= f * \widehat{d\sigma}(x). \end{aligned}$$

In other words, TT^* is simply convolution with the function $\widehat{d\sigma}$ (the Fourier extension operator of the function 1 on $\mathbb{S}^{n-1}$). And our goal is to show that this operator is bounded from $(S(\mathbb{R}^n), \|\cdot\|_{p'})$ to $(L^p(\mathbb{R}^n), \|\cdot\|_p)$, i.e. that

$$(41) \quad \|f * \widehat{d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^{p'}(\mathbb{R}^n)}, \quad \text{for all } f \in S(\mathbb{R}^n).$$

The function $\widehat{d\sigma}$ is smooth, radial and bounded, and moreover

$$|\widehat{d\sigma}(x)| \sim \frac{1}{|x|^{\frac{n-1}{2}}} \quad \text{for all } |x| \geq 1.$$

We chop up $\widehat{d\sigma}$ in pieces K_j , one supported on $B(0, 1)$, the rest supported in larger and larger dyadic annuli (where $|x| \sim 2^j$, and thus $|\widehat{d\sigma}(x)|$ is controlled). It is helpful though for these pieces to be smoothed out. So, we start with a smooth function ψ_0 , with

$$\text{supp} \psi_0 \subseteq B(0, 1), \quad \psi_0 = 1 \text{ on } B(0, 1/2), \quad 0 \leq \psi_0 \leq 1,$$

and for each $j \in \mathbb{N}$ we define

$$\psi_j(x) := \psi_0\left(\frac{x}{2^j}\right) - \psi_0\left(\frac{x}{2^{j-1}}\right), \quad x \in \mathbb{R}^n,$$

which is a smooth function with

$$\text{supp} \psi_j \subseteq \{x \in \mathbb{R}^n : 2^{j-2} \leq |x| \leq 2^j\}.$$

Moreover, all the ψ_j are dilations of the same smooth function $\psi(x) := \psi_0(x) - \psi_0(2x)$ (whose support is in the annulus $\{1/2 \leq |x| \leq 1\}$):

$$\psi_j(x) = \psi\left(\frac{x}{2^j}\right), \quad x \in \mathbb{R}^n,$$

for all $j \in \mathbb{N}$. It is easy to see that

$$1 = \psi_0 + \sum_{j=1}^{+\infty} \psi_j,$$

so

$$\widehat{d\sigma} = \psi_0 \widehat{d\sigma} + \sum_{j=1}^{+\infty} \psi_j \widehat{d\sigma}.$$

Let $f \in S(\mathbb{R}^n)$; recall that we are aiming for (41). By the above,

$$f * \widehat{d\sigma} = f * (\psi_0 \widehat{d\sigma}) + \sum_{j=1}^{+\infty} f * (\psi_j \widehat{d\sigma}).$$

Since each $\psi_j \widehat{d\sigma}$ is now supported on an annulus on which $|\widehat{d\sigma}|$ is controlled, we hope to be able to control well each of the above convolutions independently, and then add up. Indeed, by the triangle inequality,

$$(42) \quad \|f * \widehat{d\sigma}\|_{L^{p'}(\mathbb{R}^n)} \leq \|f * (\psi_0 \widehat{d\sigma})\|_{L^{p'}(\mathbb{R}^n)} + \sum_{j=1}^{+\infty} \|f * (\psi_j \widehat{d\sigma})\|_{L^{p'}(\mathbb{R}^n)}.$$

Controlling $f * (\psi_0 \widehat{d\sigma})$. Young's convolution inequality states that, for all $p, q, r \in [1, +\infty]$ with

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

the estimate

$$\|f * g\|_{L^r(\mathbb{R}^n)} \leq \|f\|_{L^p(\mathbb{R}^n)} \cdot \|g\|_{L^q(\mathbb{R}^n)}$$

holds, for all $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$.

Applying this for our $f \in S(\mathbb{R}^n)$, for $g = \psi_0 \widehat{d\sigma}$ (which is a smooth, compactly supported function), $r = p'$ and $q = \frac{p'}{2}$, we obtain

$$\|f * (\psi_0 \widehat{d\sigma})\|_{L^{p'}(\mathbb{R}^n)} \leq \|f\|_{L^p(\mathbb{R}^n)} \cdot \|\psi_0 \widehat{d\sigma}\|_{L^q(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}.$$

Controlling each $f * (\psi_j \widehat{d\sigma})$. The above argument still works for each $\psi_j \widehat{d\sigma}$ in place of $\psi_0 \widehat{d\sigma}$. However, if we then add up over all $j \in \mathbb{N}$, we will get an infinite quantity. Instead, we wish to find an estimate of the form

$$(43) \quad \|f * (\psi_j \widehat{d\sigma})\|_{L^{p'}(\mathbb{R}^n)} \lesssim 2^{-j\epsilon} \|f\|_{L^p(\mathbb{R}^n)}$$

for *some* constant $\epsilon > 0$ (independent of f and j), so that applying (42) and adding up these upper bounds we will obtain

$$\|f * \widehat{d\sigma}\|_{L^{p'}(\mathbb{R}^n)} \lesssim \left(1 + \sum_{j=1}^{+\infty} 2^{-j\epsilon}\right) \|f\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)},$$

as desired. We will achieve (43) by showing that the convolution operator with each $\psi_j \widehat{d\sigma}$ is bounded from L^1 to L^∞ , and from L^2 to L^2 , and then applying Riesz-Thorin interpolation.

So, let $j \in \mathbb{N}$.

• *Boundedness of the convolution operator $\cdot * (\psi_j \widehat{d\sigma})$ from L^1 to L^∞ .* Again by Young's convolution inequality, for our arbitrary $f \in S(\mathbb{R}^n)$ it holds that

$$\|f * (\psi_j \widehat{d\sigma})\|_\infty \leq \|f\|_{L^1(\mathbb{R}^n)} \|\psi_j \widehat{d\sigma}\|_{L^\infty(\mathbb{R}^n)}.$$

Now, $\text{supp} \psi_j \subseteq \{x \in \mathbb{R}^n : 2^{j-2} \leq |x| \leq 2^j\}$. Moreover, $0 \leq \psi_j \leq 1$, while $|\widehat{d\sigma}| \sim 2^{-\frac{j(n-1)}{2}}$ on this annulus. Therefore,

$$\|\psi_j \widehat{d\sigma}\|_{L^\infty(\mathbb{R}^n)} \lesssim 2^{-\frac{j(n-1)}{2}}.$$

It follows that

$$\|f * (\psi_j \widehat{d\sigma})\|_{L^\infty(\mathbb{R}^n)} \lesssim 2^{-\frac{j(n-1)}{2}} \|f\|_{L^1(\mathbb{R}^n)}$$

(where the implicit multiplicative constant is independent of f).

• *Boundedness of the convolution operator $\cdot * (\psi_j \widehat{d\sigma})$ from L^2 to L^2 .* By Plancherel's theorem,

$$\begin{aligned} (44) \quad \|f * (\psi_j \widehat{d\sigma})\|_{L^2(\mathbb{R}^n)} &= \|(f * (\psi_j \widehat{d\sigma}))^\wedge\|_{L^2(\mathbb{R}^n)} \\ &= \|\widehat{f} \cdot \widehat{\psi_j \widehat{d\sigma}}\|_{L^2(\mathbb{R}^n)} \\ &\leq \|\widehat{f}\|_{L^2(\mathbb{R}^n)} \cdot \|\widehat{\psi_j \widehat{d\sigma}}\|_{L^\infty(\mathbb{R}^n)} \\ &= \|f\|_{L^2(\mathbb{R}^n)} \cdot \|\widehat{\psi_j} * d\sigma\|_{L^\infty(\mathbb{R}^n)}, \end{aligned}$$

where by $g * d\sigma$ we denote the function defined by

$$g * d\sigma(x) := \int_{\xi \in \mathbb{S}^{n-1}} g(x - \xi) d\sigma(\xi), \quad x \in \mathbb{R}^n,$$

for any continuous $g : \mathbb{R}^n \rightarrow \mathbb{C}$. Indeed, let us verify the formula

$$\widehat{g \widehat{d\sigma}} = \widehat{g} * d\sigma,$$

for any such g . For all $x \in \mathbb{R}^n$,

$$\begin{aligned} \widehat{g \widehat{d\sigma}}(x) &= \int_{y \in \mathbb{R}^n} e^{-2\pi i \langle x, y \rangle} g(y) \widehat{d\sigma}(y) dy \\ &= \int_{y \in \mathbb{R}^n} e^{-2\pi i \langle x, y \rangle} g(y) \left(\int_{\xi \in \mathbb{S}^{n-1}} e^{2\pi i \langle y, \xi \rangle} d\sigma(\xi) \right) dy \\ &\stackrel{\text{Fubini}}{=} \int_{\xi \in \mathbb{S}^{n-1}} \left(\int_{y \in \mathbb{R}^n} e^{-2\pi i \langle x - \xi, y \rangle} g(y) dy \right) d\sigma(\xi) \\ &= \int_{\xi \in \mathbb{S}^{n-1}} \widehat{g}(x - \xi) d\sigma(\xi) \\ &= \widehat{g} * d\sigma(x), \end{aligned}$$

as desired.

We thus need to control $\|\widehat{\psi_j} * d\sigma\|_{L^\infty(\mathbb{R}^n)}$ in an efficient way. Recall that all the functions ψ_j satisfy $\psi_j = \psi(\frac{\cdot}{2^j})$, for the same Schwartz function ψ . Therefore, $\widehat{\psi_j} = (2^j)^n \widehat{\psi}(2^j \cdot)$ for all $j \in \mathbb{N}$. Now, since $\widehat{\psi}$ is also Schwartz, it satisfies bounds of the form

$$|\widehat{\psi}(x)| \leq \frac{C_N}{1 + |x|^N}, \quad x \in \mathbb{R}^n,$$

for all $N \in \mathbb{N}$. It turns out that our proof will work for $N = n$. In other words, we will use that

$$|\widehat{\psi}(x)| \leq w(x) := \frac{C}{1 + |x|^n}, \quad x \in \mathbb{R}^n$$

for some constant C (which depends only on the fixed function ψ), and thus

$$|\widehat{\psi_j}| = 2^{jn} |\widehat{\psi}(2^j \cdot)| \leq 2^{jn} w(2^j \cdot).$$

It follows that, for all $j \in \mathbb{N}$ and $x \in \mathbb{R}^n$,

$$|\widehat{\psi_j} * d\sigma(x)| \leq |\widehat{\psi_j}| * d\sigma(x) \leq (2^{jn} w(2^j \cdot)) * d\sigma(x).$$

The idea now is that $2^{jn} w(2^j \cdot)$ behaves like an approximate identity (as $j \rightarrow +\infty$), so convolving it with the sphere essentially blurs out the sphere at scale 2^{-j} , yielding its 2^{-j} -neighbourhood $N_{2^{-j}} \mathbb{S}^{n-1}$. Since however the integral of the convolution $(2^{jn} w(2^j \cdot)) * d\sigma$ is ~ 1 (as it is independent of j), $(2^{jn} w(2^j \cdot)) * d\sigma$ should be on average equal to $|N_{2^{-j}} \mathbb{S}^{n-1}|^{-1} \sim 2^j$ pointwise on its support $N_{2^{-j}} \mathbb{S}^{n-1}$. In other words, at best we expect that

$$\|(2^{jn} w(2^j \cdot)) * d\sigma\|_{L^\infty(\mathbb{R}^n)} \lesssim 2^j.$$

Let us prove this formally. For all $x \in \mathbb{R}^n$,

$$\begin{aligned} (2^{jn} w(2^j \cdot)) * d\sigma(x) &= \int_{\xi \in \mathbb{S}^{n-1}} 2^{jn} w(2^j(x - \xi)) d\sigma(\xi) \\ &= C 2^{jn} \int_{\xi \in \mathbb{S}^{n-1}} \frac{1}{1 + |2^j(x - \xi)|^n} d\sigma(\xi) \\ &\leq C 2^{jn} \int_{\xi \in \mathbb{S}^{n-1}} \frac{1}{1 + 2^{jn}|x - \xi|^n} d\sigma(\xi). \end{aligned}$$

We will show that

$$(45) \quad 2^{jn} \int_{\xi \in \mathbb{S}^{n-1}} \frac{1}{1 + 2^{jn}|x - \xi|^n} d\sigma(\xi) \lesssim 2^j$$

(for an implicit multiplicative constant that depends only on n). We calculate the integral by chopping up the sphere $\mathbb{S}^{n-1}$ into disjoint pieces, according to their dyadic distance from x (so that $|x - \xi|$ is controlled for all ξ in each piece). In particular, let

$$\begin{aligned} S_{\text{large}} &:= \{\xi \in \mathbb{S}^{n-1} : |x - \xi| > 1\}, \\ S_{\text{small}} &:= \{\xi \in \mathbb{S}^{n-1} : |x - \xi| \leq 2^{-j}\}, \end{aligned}$$

and

$$S_m := \{\xi \in \mathbb{S}^{n-1} : 2^{-m} < |x - \xi| \leq 2^{-(m-1)}\}, \quad \text{for all } m \in \{1, \dots, j\}.$$

Since $\mathbb{S}^{n-1} \setminus \{x\} = S_{\text{large}} \sqcup S_{\text{small}} \sqcup \bigsqcup_{m=1}^j S_m$ (as every point $\xi \in \mathbb{S}^{n-1}$ different from x has some positive distance from x), it follows that

$$2^{jn} \int_{\xi \in \mathbb{S}^{n-1}} \frac{1}{1 + 2^{jn}|x - \xi|^n} d\sigma(\xi) = I_{\text{large}} + I_{\text{small}} + \sum_{m=1}^j I_m,$$

where I_{large} involves the integral on S_{large} only, I_{small} involves it on S_{small} , and I_m involves it on S_m , for all $m = 1, \dots, j$. It suffices to show that each of the three terms is $\lesssim 2^j$. And indeed: S_{large} is inside $\mathbb{S}^{n-1}$, so $|S_{\text{large}}| \leq |\mathbb{S}^{n-1}| \lesssim 1$. Therefore,

$$I_{\text{large}} = 2^{jn} \int_{\xi \in S_{\text{large}}} \frac{1}{1 + 2^{jn}|x - \xi|^n} d\sigma(\xi) \leq 2^{jn} |S_{\text{large}}| \frac{1}{2^{jn}} = |S_{\text{large}}| \lesssim 1,$$

so trivially $I_{\text{large}} \lesssim 2^j$. Now, S_{small} is the intersection of the sphere with a ball with radius 2^{-j} (centred at x). Therefore, S_{small} lives inside a cap of $\mathbb{S}^{n-1}$ with radius 2^{-j} , hence $|S_{\text{small}}| \lesssim 2^{-j(n-1)}$. Therefore,

$$I_{\text{small}} = 2^{jn} \int_{\xi \in S_{\text{small}}} \frac{1}{1 + 2^{jn}|x - \xi|^n} d\sigma(\xi) \leq 2^{jn} |S_{\text{small}}| \lesssim 2^{-jn} 2^{-j(n-1)} = 2^j,$$

as desired. Finally, for every $m = 1, \dots, j$, S_m is the intersection of $\mathbb{S}^{n-1}$ with a ball of radius $2^{-(m-1)}$ (centred at x). So, S_m lives inside a cap of $\mathbb{S}^{n-1}$ with radius $2^{-(m-1)}$, hence

$$|S_m| \lesssim 2^{-(m-1)(n-1)} \sim 2^{-m(n-1)}, \quad \text{for all } m = 1, \dots, j.$$

Therefore,

$$\begin{aligned}
\sum_{m=1}^j I_m &= 2^{jn} \sum_{m=1}^j \int_{\xi \in S_m} \frac{1}{1 + 2^{jn} |x - \xi|^n} d\sigma(\xi) \\
&\leq 2^{jn} \sum_{m=1}^j |S_m| \cdot \frac{1}{2^{jn} 2^{-mn}} \\
&= \sum_{m=1}^j |S_m| \cdot 2^{mn} \\
&\lesssim \sum_{m=1}^j 2^{-m(n-1)} \cdot 2^{mn} \\
&= \sum_{m=1}^j 2^m \\
&\sim 2^j,
\end{aligned}$$

again as we wished.

We have thus verified (45). It follows that

$$\|\widehat{\psi_j} * d\sigma\|_{L^\infty(\mathbb{R}^n)} \lesssim 2^j \quad \text{for all } j \in \mathbb{N},$$

and plugging this into (44) we obtain

$$\|f * (\psi_j \widehat{d\sigma})\|_{L^2(\mathbb{R}^n)} \lesssim 2^j \|f\|_{L^2(\mathbb{R}^n)},$$

for our arbitrarily chosen Schwartz function f .

• *Interpolation: Boundedness of the convolution operator $\cdot * (\psi_j \widehat{d\sigma})$ from $L^{p'}$ to L^p .* We have shown that

$$\|f * (\psi_j \widehat{d\sigma})\|_{L^\infty(\mathbb{R}^n)} \leq A \|f\|_{L^1(\mathbb{R}^n)}$$

and

$$\|f * (\psi_j \widehat{d\sigma})\|_{L^2(\mathbb{R}^n)} \leq B \|f\|_{L^2(\mathbb{R}^n)}$$

for all $f \in S(\mathbb{R}^n)$ (and thus for all f in $L^1(\mathbb{R}^n)$ and in $L^2(\mathbb{R}^n)$, respectively, via density), with

$$A \sim 2^{-\frac{j(n-1)}{2}} \quad \text{and} \quad B \sim 2^j.$$

By Riesz-Thorin interpolation, it follows that

$$\|f * (\psi_j \widehat{d\sigma})\|_{L^p(\mathbb{R}^n)} \leq A^\theta B^{1-\theta} \|f\|_{L^{p'}(\mathbb{R}^n)}, \quad \text{for all } f \in L^{p'}(\mathbb{R}^n),$$

for all $p \in (2, +\infty)$, for the $\theta \in (0, 1)$ which satisfies

$$\frac{1}{p} = \frac{\theta}{\infty} + \frac{1-\theta}{2},$$

i.e. for $\theta = 1 - \frac{2}{p}$ (in which case $1 - \theta = \frac{2}{p}$). We conclude that for all $p \in (2, +\infty)$ and all $j \in \mathbb{N}$,

$$\begin{aligned}
\|f * (\psi_j \widehat{d\sigma})\|_{L^p(\mathbb{R}^n)} &\lesssim_p 2^{-\frac{j(n-1)}{2} \cdot (1 - \frac{2}{p})} 2^{\frac{2j}{p}} \|f\|_{L^{p'}(\mathbb{R}^n)} \\
&= 2^{\frac{-j((n-1)(p-2)-4)}{2p}} \|f\|_{L^{p'}(\mathbb{R}^n)}, \quad \text{for all } f \in L^{p'}(\mathbb{R}^n).
\end{aligned}$$

Recalling the triangle inequality (42), we will indeed have that

$$\|f * \widehat{d\sigma}\|_{L^p(\mathbb{R}^n)} \lesssim_p \|f\|_{L^{p'}(\mathbb{R}^n)} \quad \text{for all } f \in L^{p'}(\mathbb{R}^n),$$

as desired, as long as the quantity $\frac{(n-1)(p-2)-4}{2p}$ in the exponent above is strictly *positive*, or equivalently whenever

$$p > \frac{2(n+1)}{n-1}.$$

The proof is complete.

5.7. The Fourier restriction conjecture holds in $\mathbb{R}^2$.

We will prove now that the Fourier restriction conjecture is true in two dimensions. Since restriction implies maximal Kekeya, this will be a new proof that the Kekeya maximal operator conjecture holds in $\mathbb{R}^2$, and that Kekeya sets in $\mathbb{R}^2$ have Hausdorff dimension equal to 2.

In two dimensions, a hypersurface with non-vanishing Gaussian curvature is simply a curve Γ , which is either strictly convex or strictly concave, and which we may assume is defined over $[-1, 1]$ and has a parametrisation

$$\Gamma := \{\Gamma(\omega) := (\omega, h(\omega)) : \omega \in [-1, 1]\},$$

where $|h'| \lesssim 1$. We assume from now on that our curve Γ has this property.

THEOREM 5.7.1. *The Fourier restriction conjecture holds in $\mathbb{R}^2$. In other words, for any curve Γ in $\mathbb{R}^2$ with non-vanishing Gaussian curvature, the associated Fourier extension operator E satisfies*

$$\|Eg\|_{L^4(B_R)} \lesssim_{\epsilon, \Gamma} R^\epsilon \|g\|_{L^4([-1, 1])},$$

for all $R \geq 1$, $\epsilon > 0$ and $g \in L^4([-1, 1])$.

An important ingredient for the proof (which in fact is valid for all hypersurfaces of non-vanishing Gaussian curvature in all dimensions) is that Γ has the following property:

For every $\xi_1 = \Gamma(\omega_1)$, $\xi_2 = \Gamma(\omega_2) \in \Gamma$, the normals $N_{\xi_1}\Gamma = (-1, h'(\omega_1))$ and $N_{\xi_2}\Gamma = (-1, h'(\omega_2))$ to Γ at the points ξ_1, ξ_2 , respectively, have angle between them $\sim_\Sigma |\omega_1 - \omega_2|$.

Now, specifically in two dimensions, the Fourier restriction conjecture concerns the integral $\int_{B_R} |Eg|^4$, for all $g \in L^4(\Gamma)$. Since 4 is a multiple of 2, the idea is to reinterpret the question in terms of norms of $|Eg|^2$, and to expand the square. Then, all pairs $(\omega_1, \omega_2) \in [-1, 1]^2$ will appear, and we will arrange them *according to their distance*.

PROOF OF THEOREM 5.7.1. We prove the restriction theorem in steps.

→ *A bilinear reduction.* Observe that

$$\|Eg\|_{L^4(B_R)} = \|(Eg)^2\|_{L^2(B_R)}^{1/2},$$

so our goal becomes to show that

$$(46) \quad \|(Eg)^2\|_{L^2(B_R)} \lesssim_{\epsilon, \Gamma} R^\epsilon \|g\|_{L^4([-1, 1])}^2,$$

for all $R \geq 1$, $\epsilon > 0$ and $g \in L^4([-1, 1])$. In fact, we may go for the stronger *bilinear* estimate

$$(47) \quad \|Eg \cdot Eh\|_{L^2(B_R)} \lesssim_{\epsilon, \Gamma} R^\epsilon \|g\|_{L^4([-1, 1])} \|h\|_{L^4([-1, 1])},$$

for all $R \geq 1$, $\epsilon > 0$ and $g, h \in L^4([-1, 1])$.

From now on, our goal is the bilinear estimate (47).

→ *A Whitney decomposition*. Fix $R \geq 1$, $\epsilon > 0$ and $g, h \in L^2([-1, 1])$. We wish to study $Eg \cdot Eh$, and the idea is analogous to expanding out the square in $(\sum_{T \in \mathbb{T}} \chi_T)^2$ when we were interested in proving $2d$ maximal Kakeya. In particular, we will chop up g and h in pieces, according to the distance between their inputs, and then multiply out the extensions E of the corresponding pieces to understand $Eg \cdot Eh$. More precisely, for all $x \in \mathbb{R}^n$,

$$\begin{aligned}
(48) \quad Eg(x) \cdot Eh(x) &= \left(\int_{\omega \in [-1, 1]} e^{2\pi i \langle x, \Gamma(\omega) \rangle} g(\omega) d\omega \right) \cdot \left(\int_{u \in [-1, 1]} e^{2\pi i \langle x, \Gamma(u) \rangle} h(u) du \right) \\
&= \int_{(\omega, u) \in [-1, 1]^2} e^{2\pi i \langle x, \Gamma(\omega) + \Gamma(u) \rangle} g(\omega) h(u) d\omega du \\
&= \int_{(\omega, u) \in [-1, 1]^2 \setminus D} e^{2\pi i \langle x, \Gamma(\omega) + \Gamma(u) \rangle} g(\omega) h(u) d\omega du,
\end{aligned}$$

where $D := \{(\omega, u) \in [-1, 1]^2 : \omega = u\}$ is the diagonal within the square. Observe that this is now an integral with respect to the 2-dimensional Lebesgue measure on the square $[-1, 1]^2$. So, the reason the last equality holds is that the 2-dimensional Lebesgue measure along the diagonal is 0.

We are now ready to perform what is known as a *Whitney decomposition* on the set $[-1, 1]^2 \setminus D$. This decomposition will split the pairs (ω, u) in this set into groups, according to the distance between ω and u (which is now non-zero):

Note that $[-1, 1]$ has length 2. So, for every $k \in \mathbb{N} \cup \{0\}$, we can split $[-1, 1]$ into 2^{k+1} consecutive intervals, each of length 2^{-k} . We denote by A_k the set of these intervals.

So, at Step 0 we split $[-1, 1]$ into 2 equal pieces. At Step 1, each of the previous 2 pieces is in turn split into 2 equal pieces, so $[-1, 1]$ is split into 4 equal pieces. And so on: at each new step, each of the intervals from the previous step (a *parent*) is split into two intervals of equal length (its two *children*).

Now, let I, J be two intervals in A_k . We say that I and J are *k-close*, and we denote by $I \sim_k J$ (or equivalently $J \sim_k I$) if I, J are *not* adjacent, but their parents are adjacent. Note that, for every $I \in A_k$, there exist at most 3 intervals $J \in A_k$ that are *k-close* to I , and, for each of these, $\text{dist}(I, J) \sim 2^{-k}$.

The crucial observation now is that, for every $(\omega, u) \in [-1, 1]^2 \setminus D$, there exists a *unique* $k \in \mathbb{N}$ so that the intervals I and J in A_k , which contain ω and u , respectively, are *k-close*. Indeed, at Step 0, there are only two adjacent intervals, so these cannot be 0-close (hence $k = 0$ is irrelevant). Now, since $|\omega - u| \neq 0$, there exists a unique $k \in \mathbb{N}$ so that

$$2^{-(k-1)} \leq |\omega - u| < 2^{-(k-2)}.$$

Since the intervals in A_k are all of length 2^{-k} , the lower bound $2^{-(k-1)} = 2 \cdot 2^{-k}$ above is larger than the distance between any two points on consecutive intervals of A_k . So, since $|\omega - u| > 2^{-(k-1)}$, ω and u cannot live on consecutive intervals of A_k . In other words, they live in intervals I, J (respectively) in A_k , which are not adjacent. These intervals of course have different parents I', J' , respectively, in A_{k-1} (otherwise they would be adjacent). The best of our knowledge for these parents I', J' , whose length is $2^{-(k-1)}$, is that $\text{dist}(I', J') < 2^{-(k-2)} = 2 \cdot 2^{-(k-1)} < 2^{-(k-2)}$, i.e. smaller than the length of any interval from A_{k-2} . Therefore, either I', J' are adjacent, in which case $I \sim_k J$, or I', J' are not adjacent but their parents are adjacent, in which case $I' \sim_{k-1} J'$.

We have shown that, for every $(\omega, u) \in [-1, 1]^2 \setminus D$, there exists a unique $k \in \mathbb{N}$ and unique $I, J \in A_k$ so that $\omega \in I, u \in J$ and $I \sim_k J$. Therefore,

$$[-1, 1]^2 \setminus D = \bigsqcup_{k=1}^{+\infty} \bigsqcup_{(I, J): I \sim_k J} I \times J.$$

This is what is known as the Whitney decomposition of $[-1, 1]^2$ off the diagonal.

—→ Returning to $\|Eg \cdot Eh\|_{L^2(B_R)}$. Plugging this into (48), we deduce that

$$\begin{aligned} (49) \quad Eg(x) \cdot Eh(x) &= \sum_{k=1}^{+\infty} \sum_{I \sim_k J} \int_{(\omega, u) \in I \times J} e^{2\pi i \langle x, \Gamma(\omega) + \Gamma(u) \rangle} g(\omega) h(u) d\omega du \\ &= \sum_{k=1}^{+\infty} \sum_{I \sim_k J} \left(\int_{\omega \in I} e^{2\pi i \langle x, \Gamma(\omega) \rangle} g(\omega) d\omega \right) \cdot \left(\int_{\omega \in I} e^{2\pi i \langle x, \Gamma(u) \rangle} h(u) du \right) \\ &= \sum_{k=1}^{+\infty} \sum_{I \sim_k J} E(g\chi_I)(x) \cdot E(h\chi_J)(x) \end{aligned}$$

for all $x \in \mathbb{R}^2$.

The idea now is to bound each $\|E(g\chi_I) \cdot E(h\chi_J)\|_{L^2(B_R)}$ independently. In particular, we will show a *better* estimate than (47) for each $I \sim_k J$, of the form

$$(50) \quad \|E(g\chi_I) \cdot E(h\chi_J)\|_{L^2(B_R)} \lesssim Q(R, k) \|g\|_{L^2(I)} \|h\|_{L^2(J)},$$

for an efficient quantity $Q(R, k)$ that depends only on R and k .

We recall that the extension operator E of any function ϕ is essentially the inverse Fourier transform of a fat version of ϕ . More precisely, we thicken Γ upwards by R^{-1} , creating a set $\Gamma_{R^{-1}}$, and for every $\phi : \Gamma \rightarrow \mathbb{C}$ we define $\phi_{\text{fat}} : \Gamma_{R^{-1}} \rightarrow \mathbb{C}$ by

$$\phi_{\text{fat}}(\xi + (0, t)) = \phi(\xi), \quad \text{for all } \xi \in \Gamma.$$

We have proved that then, for all $x \in \mathbb{R}^2$ with $|x_2| \leq R$, and thus for all $x \in B_R$,

$$|Ef(x)| \sim R |(f/H)_{\text{fat}}^\vee(x)|,$$

where H is a function that depends only on Γ . Therefore, denoting by simplicity

$$(g\chi_I/H)_{\text{fat}} =: g_{I, \text{fat}}, \quad (h\chi_J/H)_{\text{fat}} =: h_{J, \text{fat}},$$

we see that

$$\begin{aligned} \|E(g\chi_I) \cdot E(h\chi_J)\|_{L^2(B_R)} &\sim R^2 \|(g_{I, \text{fat}})^\vee \cdot (h_{J, \text{fat}})^\vee\|_{L^2(B_R)} \\ &= R^2 \|(g_{I, \text{fat}} * h_{J, \text{fat}})^\vee\|_{L^2(B_R)} \\ &\stackrel{\text{Plancherel}}{=} R^2 \|g_{I, \text{fat}} * h_{J, \text{fat}}\|_{L^2(B_R)}, \end{aligned}$$

while also

$$\|g\|_{L^2(I)} = \|g\chi_I\|_{L^2([-1, 1])} \sim R^{1/2} \|g_{I, \text{fat}}\|_2, \quad \|h\|_{L^2(J)} = \|h\chi_J\|_{L^2([-1, 1])} \sim R^{1/2} \|h_{J, \text{fat}}\|_2$$

(where the RHS 2-norms are taken with respect to the usual Lebesgue measure in two dimensions). Therefore, any estimate of the form

$$(51) \quad \|g_{I, \text{fat}} * h_{J, \text{fat}}\|_{L^2(B_R)} \lesssim \tilde{Q}(R, k) \|g_{I, \text{fat}}\|_2 \|h_{J, \text{fat}}\|_2$$

is equivalent to the estimate

$$(52) \quad \|E(g\chi_I) \cdot E(h\chi_J)\|_{L^2(B_R)} \lesssim R\tilde{Q}(R, k) \|g\|_{L^2(I)} \|h\|_{L^2(J)},$$

i.e. to (50) for $Q(R, k) = R\tilde{Q}(R, k)$. It thus suffices to prove (51), for as small a quantity $\tilde{Q}(R, k)$ as we can. Observe that this is just a bilinear estimate on $\mathbb{R}^2$, for functions that are supported in the R^{-1} -neighbourhoods of the arcs I and J of Γ , respectively. We will thus aim to show, more generally, that, for a small $\tilde{Q}(R, k)$,

$$(53) \quad \|g_I * h_J\|_{L^2(B_R)} \lesssim Q(R, k) \|g_I\|_2 \|h_J\|_2,$$

for all functions g_I, h_J that are supported in the R^{-1} -neighbourhoods of I and J , respectively (and then the formula will automatically apply for $g_{I,\text{fat}}, h_{J,\text{fat}}$, giving (51), and thus (50)).

→ *Bilinear interpolation.* Fix $k \in \mathbb{N}$ and intervals I, J with $I \sim_k J$. An estimate of the form (53) will follow by bilinear interpolation, which is based on the following theorem (the analogue of Riesz-Thorin interpolation in the bilinear case).

THEOREM 5.7.2. (Bilinear interpolation) *Let T be a bilinear operator, and let $p_1, p_1, q_1, q_2, r_1, r_2 \in [1, +\infty]$. Suppose that $T : L^{p_1}(\mathbb{R}^n) \times L^{q_1}(\mathbb{R}^n) \rightarrow L^{r_1}(\mathbb{R}^n)$ is well-defined and bounded, with*

$$\|T(f, g)\|_{r_1} \leq A_1 \|f\|_{p_1} \|g\|_{q_1} \quad \text{for all } f, g \in L^{p_1}(\mathbb{R}^n) \times L^{q_1}(\mathbb{R}^n),$$

and that $T : L^{p_2}(\mathbb{R}^n) \times L^{q_2}(\mathbb{R}^n) \rightarrow L^{r_2}(\mathbb{R}^n)$ is well-defined and bounded, with

$$\|T(f, g)\|_{r_2} \leq A_2 \|f\|_{p_2} \|g\|_{q_2} \quad \text{for all } f, g \in L^{p_2}(\mathbb{R}^n) \times L^{q_2}(\mathbb{R}^n).$$

Then, for any $\theta \in (0, 1)$, and for the exponents p, q, r given by

$$\frac{1}{p} = \frac{\theta}{p_1} + \frac{1-\theta}{p_2}, \quad \frac{1}{q} = \frac{\theta}{q_1} + \frac{1-\theta}{q_2}, \quad \frac{1}{r} = \frac{\theta}{r_1} + \frac{1-\theta}{r_2},$$

the operator $T : L^p(\mathbb{R}^n) \times L^q(\mathbb{R}^n) \rightarrow L^r(\mathbb{R}^n)$ is well-defined and bounded, and

$$\|T(f, g)\|_r \leq A_1^\theta A_2^{1-\theta} \|f\|_p \|g\|_q, \quad \text{for all } f, g \in L^p(\mathbb{R}^n) \times L^q(\mathbb{R}^n).$$

Therefore, to show $L^2 \times L^2 \rightarrow L^2$ -boundedness of a bilinear operator T , it suffices to show $L^1 \times L^1 \rightarrow L^1$ -boundedness and $L^\infty \times L^\infty \rightarrow L^\infty$ -boundedness. In our case, our bilinear operator T acts on pairs (g_I, h_J) of functions, where g_I is supported on the R^{-1} -neighbourhood I_{fat} of I , and h_J is supported on the R^{-1} -neighbourhood J_{fat} of J , by convolving the functions:

$$T(g_I, h_J) := g_I * h_J.$$

Indeed, if we show that

$$(54) \quad \|g_I * h_J\|_1 \leq A_1 \|g_I\|_1 \|h_J\|_1$$

and

$$(55) \quad \|g_I * h_J\|_\infty \leq A_2 \|g_I\|_\infty \|h_J\|_\infty$$

for all g_I, h_J as above, it will follow that

$$\|g_I * h_J\|_2 \leq A_1^\theta A_2^{1-\theta} \|g_I\|_2 \|h_J\|_2,$$

for the parameter θ for which

$$\frac{1}{2} = \frac{\theta}{1} + \frac{1-\theta}{\infty},$$

i.e. for $\theta = \frac{1}{2}$. Our objective thus becomes to show the bilinear estimates (54) and (55), for as small A_1, A_2 as possible.

→ $L^1 \times L^1 \rightarrow L^1$ -boundedness. Here, we are aiming for the bilinear estimate (54). This holds for $A_1 = 1$, by Young's convolution inequality:

$$\|g_I * h_J\|_1 \leq \|g_I\|_1 \|h_J\|_1.$$

$\longrightarrow L^\infty \times L^\infty \rightarrow L^\infty$ -boundedness. Here, we are aiming for the bilinear estimate (55). This is based on the geometry of Γ . In particular, for all g_I, h_J bounded functions defined on $I_{\text{fat}}, J_{\text{fat}}$, respectively, and all $x \in \mathbb{R}$,

$$\begin{aligned} |g_I * h_J(x)| &= \left| \int g_I(x-y) h_J(y) dy \right| \\ &\leq \int |g_I(x-y)| \cdot |h_J(y)| dy \\ &\stackrel{\text{supp } g_I \subseteq I_{\text{fat}}, \text{supp } h_J \subseteq J_{\text{fat}}}{\leq} \|g_I\|_\infty \|h_J\|_\infty \int \chi_{I_{\text{fat}}}(x-y) \cdot \chi_{J_{\text{fat}}}(y) dy \\ &\leq \|\chi_{I_{\text{fat}}} * \chi_{J_{\text{fat}}}\|_\infty \|g_I\|_\infty \|h_J\|_\infty. \end{aligned}$$

In other words,

$$\|g_I * h_J\|_\infty \leq \|\chi_{I_{\text{fat}}} * \chi_{J_{\text{fat}}}\|_\infty \|g_I\|_\infty \|h_J\|_\infty,$$

so the $L^\infty \times L^\infty \rightarrow L^\infty$ norm of our operator is at most $\|\chi_{I_{\text{fat}}} * \chi_{J_{\text{fat}}}\|_\infty$. We now control this norm. For every $x \in \mathbb{R}^2$,

$$\begin{aligned} \chi_{I_{\text{fat}}} * \chi_{J_{\text{fat}}}(x) &= \int \chi_{I_{\text{fat}}}(x-y) \cdot \chi_{J_{\text{fat}}}(y) dy \\ &= \int_{y \in J_{\text{fat}}} \chi_{I_{\text{fat}}}(x-y) dy \\ &= |J_{\text{fat}} \cap (x - I_{\text{fat}})|. \end{aligned}$$

So, we are interested in how large this intersection can be, over all possible translations of $-I_{\text{fat}}$ in $\mathbb{R}^2$.

- The set J_{fat} is the R^{-1} -neighbourhood of the arc J of Γ , which has length 2^{-k} and some normal $N_J(\xi)$ at each of its points ξ . So, we can think of J_{fat} as a curved tube, of dimensions $2^{-k} \times R^{-1}$. It is curved along its 2^{-k} -long direction, and its normals along its R^{-1} -long direction are the normals to J .
- The reflection $-I_{\text{fat}}$ around 0 looks a lot like I_{fat} ; it is the R^{-1} -neighbourhood of the arc $-I$, which also has length 2^{-k} , and normal $N_I(\xi)$ at each of its points ξ . The set $x - I_{\text{fat}}$ is simply the arbitrary translation of $-I_{\text{fat}}$. So, we can think of $x - I_{\text{fat}}$ as a curved tube, of dimensions $2^{-k} \times R^{-1}$. It is curved along its 2^{-k} -long direction, and its normals along its R^{-1} -long direction are the normals to I .

Therefore, we need to understand how large the intersection between two curved tubes as above can be. There is of course a difference in the analysis, between which side is longest, 2^{-k} or R^{-1} .

When $R^{-1} \geq 2^{-k}$, in which case $k \gtrsim \log R$: This case it treated trivially. Indeed, $|J_{\text{fat}} \cap (x - I_{\text{fat}})|$ is at most the size of any of the two curved tubes involved, and thus $\lesssim_\Gamma 2^{-k} \times R^{-1}$. In other words, for all such k , and all $I \sim_k J$, we have concluded that

$$\|g_I * h_J\|_\infty \lesssim_\Gamma 2^{-k} R^{-1} \|g_I\|_\infty \|h_J\|_\infty.$$

By bilinear interpolation between this and the earlier trivial $L^1 \times L^1 \rightarrow L^1$ -estimate, we obtain

$$\|g_I * h_J\|_2 \lesssim_\Gamma (2^{-k} R^{-1})^{1/2} \|g_I\|_2 \|h_J\|_2,$$

for all k in this range, and all $I \sim_k J$.

When $R^{-1} < 2^{-k}$, in which case $k \lesssim \log R$: This situation resembles the one where we would be calculating the largest possible intersection between two *straight* tubes, of thickness R^{-1} , any length, and angle $\sim 2^{-k}$ between them.

The reason is that, since $I \sim_k J$, we have that $\text{dist}(I, J) \sim 2^{-k}$, hence any two normals N_I and N_J to the above curved tubes form angle $\sim_\Gamma 2^{-k}$ between them³. Therefore, just as for straight R^{-1} -tubes with angle $\sim 2^{-k}$ between them, we again have that, for all $x \in \mathbb{R}^2$,

$$|J_{\text{fat}} \cap (x - I_{\text{fat}})| \lesssim_\Gamma \frac{R^{-2}}{2^{-k}}.$$

We thus conclude that

$$\|g_I * h_J\|_\infty \lesssim_\Gamma \frac{R^{-2}}{2^{-k}} \|g_I\|_\infty \|h_J\|_\infty,$$

for all k in this range, and all $I \sim_k J$. By bilinear interpolation between this and the earlier trivial $L^1 \times L^1 \rightarrow L^1$ -estimate, we deduce that

$$\|g_I * h_J\|_2 \lesssim_\Gamma \left(\frac{R^{-2}}{2^{-k}} \right)^{1/2} \|g_I\|_2 \|h_J\|_2,$$

for all k in this range, and all $I \sim_k J$.

—→ *Bringing everything together.* To sum up, we have derived an estimate of the form (52). In particular, for all $g, h \in L^2([-1, 1])$, it holds that

$$\|E(g\chi_I) \cdot E(h\chi_J)\|_{L^2(B_R)} \lesssim_\Gamma R\tilde{Q}(R, k) \|g\|_{L^2(I)} \|h\|_{L^2(J)}, \text{ for all } k \in \mathbb{N} \text{ and } I \sim_k J,$$

where

$$R\tilde{Q}(R, k) = \begin{cases} R(2^{-k}R^{-1})^{1/2} = R^{1/2}2^{-k/2}, & \text{when } 2^k \geq R, \\ R\left(\frac{R^{-2}}{2^{-k}}\right)^{1/2} = 2^{k/2}, & \text{when } 2^k < R. \end{cases}$$

We now recall (49). We split k in the above two ranges, so, by the triangle inequality,

$$\begin{aligned} \|Eg \cdot Eh\|_{L^2(B_R)} &\lesssim \left\| \sum_{k \in \mathbb{N}: 2^k \geq R} \sum_{I \sim_k J} E(g\chi_I) \cdot E(h\chi_J) \right\|_{L^2(B_R)} \\ &\quad + \left\| \sum_{k \in \mathbb{N}: 2^k < R} \sum_{I \sim_k J} E(g\chi_I) \cdot E(h\chi_J) \right\|_{L^2(B_R)} \\ &=: I_1 + I_2. \end{aligned}$$

It remains to control each term.

Controlling I_1 , the term with the large k . I_1 is easy to control, again by the triangle inequality. More precisely,

$$\begin{aligned} I_1 &\leq \sum_{k \in \mathbb{N}: 2^k \geq R} \sum_{I \sim_k J} \|E(g\chi_I) \cdot E(h\chi_J)\|_{L^2(B_R)} \\ &\lesssim_\Gamma \sum_{k \in \mathbb{N}: 2^k \geq R} R^{1/2}2^{-k/2} \sum_{I \sim_k J} \|g\|_{L^2(I)} \|h\|_{L^2(J)}. \end{aligned}$$

We now observe that, for all $k \in \mathbb{N}$,

$$\sum_{I \sim_k J} \|g\|_{L^2(I)} \|h\|_{L^2(J)} \lesssim \|g\|_{L^2([-1, 1])} \|h\|_{L^2([-1, 1])}.$$

³This is exactly where we are using that Γ has non-vanishing curvature.

Indeed, for each $I \in A_k$, the J for which $I \sim_k J$ are all inside $10I$. Therefore,

$$\begin{aligned}
\sum_{I \sim_k J} \|g\|_{L^2(I)} \|h\|_{L^2(J)} &\leq \sum_{I \in A_k} \sum_{J \in A_k: J \subseteq 10I} \|g\|_{L^2(I)} \|h\|_{L^2(J)} \\
&\leq \sum_{I \in A_k} \|g\|_{L^2(I)} \|h\|_{L^2(10I)} \\
&\leq \sum_{I \in A_k} \|g\|_{L^2(10I)} \|h\|_{L^2(10I)} \\
&\stackrel{\text{Cauchy-Schwarz}}{\leq} \left(\sum_{I \in A_k} \|g\|_{L^2(10I)}^2 \right)^{1/2} \left(\sum_{I \in A_k} \|h\|_{L^2(10I)}^2 \right)^{1/2} \\
&\sim \|g\|_{L^2([-1,1])} \|h\|_{L^2([-1,1])}.
\end{aligned}$$

Therefore,

$$I_1 \lesssim_\Gamma \left(\sum_{k \in \mathbb{N}: 2^k \geq R} R^{1/2} 2^{-k/2} \right) \|g\|_{L^2([-1,1])} \|h\|_{L^2([-1,1])}.$$

Finally, let M be the smallest k so that $2^k \geq R$. Then, $2^M \sim R$ (so $M \sim \log R$), and

$$\sum_{k \in \mathbb{N}: 2^k \geq R} R^{1/2} 2^{-k/2} = R^{1/2} \sum_{k=M}^{+\infty} \left(\frac{1}{\sqrt{2}} \right)^k \sim R^{1/2} \left(\frac{1}{\sqrt{2}} \right)^M = \frac{R^{1/2}}{\sqrt{2^M}} \sim \frac{R^{1/2}}{R^{1/2}} \sim 1.$$

It follows that

$$I_1 \lesssim_\Gamma \|g\|_{L^2([-1,1])} \|h\|_{L^2([-1,1])}.$$

By Hölder's inequality, this is even better than the desired estimate

$$I_1 \lesssim_\Gamma \|g\|_{L^4([-1,1])} \|h\|_{L^4([-1,1])}$$

Controlling I_2 , the sum with the small k . Applying to I_2 the triangle inequality as before will not work (as this time adding up the corresponding $R\tilde{Q}(R, k)$ will give something huge). We need something more clever, and that comes from Fefferman, who noticed that, for any given k , the functions $E(g\chi_I) \cdot E(h\chi_J)$, over all different I, J with $I \sim_k J$, are essentially orthogonal in L^2 . Indeed, first we indeed apply the triangle inequality, but only in terms of k :

$$I_2 \leq \sum_{k \in \mathbb{N}: 2^k < R} \left\| \sum_{I \sim_k J} E(g\chi_I) \cdot E(h\chi_J) \right\|_{L^2(B_R)}.$$

The crucial observation is that, morally, for every k ,

$$\left\| \sum_{I \sim_k J} E(g\chi_I) \cdot E(h\chi_J) \right\|_{L^2(B_R)}^2 \lesssim \sum_{I \sim_k J} \|E(g\chi_I) \cdot E(h\chi_J)\|_{L^2(B_R)}^2.$$

This inequality is not in fact true, but it is morally true, and it is better than the triangle inequality. What precisely holds is that

$$\begin{aligned}
\int_{B_R} \left| \sum_{I \sim_k J} E(g\chi_I) \cdot E(h\chi_J) \right|^2 &\sim R^4 \int_{B_R} \left| \sum_{I \sim_k J} (g_{I,\text{fat}})^\sim \cdot (h_{J,\text{fat}})^\sim \right|^2 \\
&\leq R^4 \int \left| \sum_{I \sim_k J} (g_{I,\text{fat}})^\sim \cdot (h_{J,\text{fat}})^\sim \right|^2 \\
&= R^4 \int \left| \sum_{I \sim_k J} (g_{I,\text{fat}} * h_{J,\text{fat}})^\sim \right|^2 \\
&\stackrel{\text{Plancherel}}{=} R^4 \int \left| \sum_{I \sim_k J} g_{I,\text{fat}} * h_{J,\text{fat}} \right|^2.
\end{aligned}$$

The important observation here is that each $g_{I,\text{fat}} * h_{J,\text{fat}}$ is supported in the sumset $I_{\text{fat}} + J_{\text{fat}}$, and these sets are essentially disjoint (in fact, finitely overlapping) over all the pairs (I, J) with $I \sim_k J$. Therefore,

$$\begin{aligned}
\left\| \sum_{I \sim_k J} E(g\chi_I) \cdot E(h\chi_J) \right\|_{L^2(B_R)}^2 &\lesssim R^4 \sum_{I \sim_k J} \|g_{I,\text{fat}} * h_{J,\text{fat}}\|_2^2 \\
&\lesssim_\Gamma R^4 \frac{R^{-2}}{2^{-k}} \sum_{I \sim_k J} \|g_{I,\text{fat}}\|_2^2 \|h_{J,\text{fat}}\|_2^2 \\
&\sim R^4 \frac{R^{-2}}{2^{-k}} R^{-2} \sum_{I \sim_k J} \|g\|_{L^2(I)}^2 \|h\|_{L^2(J)}^2 \\
&= 2^k \sum_{I \sim_k J} \|g\|_{L^2(I)}^2 \|h\|_{L^2(J)}^2.
\end{aligned}$$

Note that this is a better estimate than if we had simply applied the triangle inequality (thanks to the existence of the square powers). We can now move towards 4-norms on the RHS (which was our initial goal anyway). In particular, for all k ,

$$\begin{aligned}
\left\| \sum_{I \sim_k J} E(g\chi_I) \cdot E(h\chi_J) \right\|_{L^2(B_R)} &\lesssim_\Gamma 2^{k/2} \left(\sum_{I \sim_k J} \|g\|_{L^2(I)}^2 \|h\|_{L^2(J)}^2 \right)^{1/2} \\
&\leq 2^{k/2} \left(\sum_{I \in A_k} \|g\|_{L^2(10I)}^2 \right)^{1/2} \max_{I \in A_k} \|h\|_{L^2(10I)} \\
&\stackrel{\text{Hölder}}{\lesssim_\Gamma} 2^{k/2} \|g\|_{L^2([-1,1])} \cdot \max_{I \in A_k} |I|^{1/2} \|h\|_{L^4([-1,1])} \\
&\lesssim 2^{k/2} \|g\|_{L^4([-1,1])} 2^{-k/2} \|h\|_{L^4([-1,1])} \\
&= \|g\|_{L^4([-1,1])} \|h\|_{L^4([-1,1])}.
\end{aligned}$$

Thankfully, we are now adding over a small number of k , and in particular those with $2^k < R$. Earlier we denoted by M the smallest k with $2^k \geq R$, so we are now adding up

over $k \in \{1, \dots, M-1\}$, where $M \sim \log R$. Therefore,

$$\begin{aligned} I_2 &\leq \sum_{k=1}^{M-1} \|g\|_{L^4([-1,1])} \|h\|_{L^4([-1,1])} \\ &\lesssim \log R \|g\|_{L^4([-1,1])} \|h\|_{L^4([-1,1])} \\ &\leq C_\epsilon R^\epsilon \|g\|_{L^4([-1,1])} \|h\|_{L^4([-1,1])}. \end{aligned}$$

□

5.8. The wave packet decomposition

In the proof of the 2D Fourier restriction conjecture, the norms of $(Eg)^2$ were controlled by chopping up the domain of $g(\omega)g(u)$ in pieces, according to the distance between the inputs ω and u .

The modern approaches to the Fourier restriction conjecture in *any* dimension involves what is known as a *wave packet decomposition* of g . This simply decomposes g (rather than a tensor product of g with itself). It is a more complicated decomposition, because it doesn't simply chop up the domain of g in disjoint pieces. However, it has a huge advantage: the pieces are essentially pairwise orthogonal, and the extension operator of each piece is practically a wave evolving inside a different tube.

In what follows, let $n \geq 2$, and let

$$\Sigma := \{(\omega, h(\omega)) : \omega \in B^{n-1}(0, 1)\}$$

be a hypersurface in $\mathbb{R}^n$, of non-vanishing Gaussian curvature, with $|\nabla h| \leq 1$.

Let $R \geq 1$. The idea is to decompose any $g \in L^1(\Sigma)$ in a way that will help us understand $\widehat{gd\sigma}$ inside B_R . To do this, we want to split Σ into small caps θ , and to then use some uncertainty principle on each piece $g\chi_\theta$, as it has small support. Of course, we know the uncertainty principle for the standard Fourier transform of functions supported on small rectangles (and not for the extension operator). So, we will first reduce the restriction problem to a problem that studies functions with support in the R^{-1} -neighbourhood of Σ . And then we will decompose those functions.

→ **Reducing the problem to functions supported in an R^{-1} -neighbourhood of Σ .** Recall that, for every integrable $f : \Sigma \rightarrow \mathbb{C}$, we may define the function f_{fat} on

$$\Sigma_{\text{fat}} := \bigcup_{t \in [0, R^{-1}]} \Sigma + (0, \dots, 0, t),$$

with

$$f_{\text{fat}}(\xi + (0, \dots, 0, t)) := f(\xi) \quad \text{for all } \xi \in \Sigma, t \in [0, R^{-1}].$$

Then,

$$\left((f/J)_{\text{fat}} \right)^\sim(x) = \widehat{f d\sigma}(x) \cdot \left(\int_{t \in [0, R^{-1}]} e^{2\pi i x_n t} dt \right) \quad \text{for all } x \in \mathbb{R}^n,$$

where $J \sim 1$ (and depends only on the curvature of Σ), and

$$\left| \int_{t \in [0, R^{-2}]} e^{2\pi i x_n t} dt \right| \sim R^{-1}$$

on the strip

$$S_R := \{x \in \mathbb{R}^n : |x_n| \leq R\}.$$

Therefore, for all $f \in L^1(\Sigma)$,

$$\left| \left((f/J)_{\text{fat}} \right)^\vee \right| \sim R^{-1} |\widehat{f d\sigma}| \quad \text{on } S_R.$$

Thus, for all $p, q \in [1, +\infty)$,

$$\left\| \left((f/J)_{\text{fat}} \right)^\vee \right\|_{L^p(B_R)} \sim_p R^{-1} \|\widehat{f d\sigma}\|_{L^p(B_R)}$$

and

$$\|(f/J)_{\text{fat}}\|_{L^q(\mathbb{R}^n)} \sim_q R^{-\frac{1}{q}} \|f\|_{L^q(\Sigma)}.$$

So, proving any restriction estimate of the form

$$\|\widehat{f d\sigma}\|_{L^p(B_R)} \lesssim_{p,q,\epsilon} R^\epsilon \|f\|_{L^q(\Sigma)} \quad \text{for all } f \in L^q(\Sigma)$$

is equivalent to showing that

$$(56) \quad \left\| \left((f/J)_{\text{fat}} \right)^\vee \right\|_{L^p(B_R)} \lesssim_{p,q,\epsilon} R^{-1+\frac{1}{q}+\epsilon} \|(f/J)_{\text{fat}}\|_{L^q(\mathbb{R}^n)}, \quad \text{for all } f \in L^q(\Sigma).$$

In other words, studying any local form of the Fourier restriction conjecture in B_R is equivalent to studying an inequality of the above form, for the inverse Fourier transform of functions defined on $N_{R^{-1}}\Sigma$, the R^{-1} -neighbourhood of Σ .

→ *Decomposing any $g \in L^1(N_{R^{-1}}\Sigma)$ into pieces supported in different caps θ .*

We split $N_{R^{-1}}\Sigma$ into disjoint sets θ , each of them roughly the R^{-1} -neighbourhood of an $R^{-1/2}$ -cap (so, each of them roughly an $R^{-1/2} \times \dots \times R^{-1/2} \times R^{-1}$ -rectangle). Let Θ be the family of these sets θ . Then, for every $g \in L^1(N_{R^{-1}}\Sigma)$,

$$g = \sum_{\theta \in \Theta} g_\theta,$$

where $g_\theta := g \chi_\theta$. The functions g_θ have disjoint supports, so they are orthogonal in $L^2(\mathbb{R}^n)$. Our goal will be to further decompose each g_θ into functions that are also roughly $L^2(\mathbb{R}^n)$ -orthogonal. From now on, we denote by θ the smallest rectangle containing the original set θ .

→ *Further decomposing each g_θ .* Fix $\theta \in \Theta$. Denote by N_θ the direction normal to θ .

The dual rectangles to θ have dimensions roughly $R^{1/2} \times \dots \times R^{1/2} \times R$, and direction N_θ . So, they are practically $R^{1/2} \times \dots \times R^{1/2} \times R$ -tubes in this direction. Let $\mathbb{T}(\theta)$ be a $O(1)$ -overlapping family of such parallel tubes, which together tile $\mathbb{R}^n$.

Now, we know that since g_θ is supported in the rectangle θ , $\check{g}_\theta$ is a sum of waves, each practically evolving inside a different tube in $\mathbb{T}(\theta)$. We briefly recall the process that gave this result, from our uncertainty principle chapter: We write g_θ as a Fourier series on θ , and then multiply it with a Schwartz function ψ_θ satisfying

$$\text{supp } \psi_\theta \subseteq 2\theta, \quad \psi_\theta = 1 \text{ on } \theta, \quad 0 \leq \psi_\theta \leq 1.$$

Then,

$$g_\theta = g_\theta \cdot \psi_\theta = \sum_k c_k \exp(k) \psi_\theta,$$

where the sum is taken over all the centres of the tubes in $\mathbb{T}(\theta)$, the functions $\exp(k)$ are exponential functions that depend on k , and the c_k are the Fourier coefficients. We have

seen that this implies that

$$\check{g}_\theta = \sum_k \underbrace{c_k \check{\psi}_\theta(\cdot - k)}_{(c_k \exp(k) \psi_\theta)^\sim},$$

where $\check{\psi}_\theta$ is a Schwartz function adapted to the dual rectangle θ^* through 0 (in the sense that it decays rapidly outside it). This rectangle is practically the tube in $\mathbb{T}(\theta)$ through 0. As k ranges over all the centres of the tubes in $\mathbb{T}(\theta)$, $\check{\psi}_\theta(\cdot - k)$ is simply adapted to the tube of $\mathbb{T}(\theta)$ with centre k .

Therefore, denoting each

$$c_k \exp(k) \psi_{\theta_{\text{fat}}} =: g_T,$$

for T the tube in $\mathbb{T}(\theta)$ with centre k , it follows that

$$g_\theta = \sum_{T \in \mathbb{T}(\theta)} g_T,$$

where each g_T is supported on θ , and

$$\check{g}_\theta = \sum_{T \in \mathbb{T}(\theta)} \check{g}_T,$$

where each $\check{g}_T$ is a Schwartz function essentially supported in the tube T , in the sense that it decays fast outside T , and is thus negligible outside of, say, $R^\epsilon T$:

$$|\check{g}_T(x)| \lesssim \left(R^{-1/2} \text{dist}(x', T) + R^{-1} \text{dist}(x_n, T) \right)^{-10^{100}n} \|g\|_2 \quad \text{for all } x = (x', x_n) \notin T.$$

The supports of $\check{g}_T$ are practically the tubes T , which are finitely overlapping. It is thus not very hard to show that the functions $\check{g}_T$, over all $T \in \mathbb{T}(\theta)$, are practically orthogonal, in the sense that

$$\left\| \sum_{T \in \mathbb{W}} \check{g}_T \right\|_2^2 \lesssim \sum_{T \in \mathbb{W}} \|\check{g}_T\|_2^2, \quad \text{for all } \mathbb{W} \subseteq \mathbb{T}(\theta).$$

By Plancherel's theorem, the original functions g_T , over all $T \in \mathbb{T}(\theta)$, are also orthogonal, even though they have the same support θ :

$$\left\| \sum_{T \in \mathbb{W}} g_T \right\|_2^2 \lesssim \sum_{T \in \mathbb{W}} \|g_T\|_2^2, \quad \text{for all } \mathbb{W} \subseteq \mathbb{T}(\theta).$$

→ **Conclusion:** the wave packet decomposition of $g \in L^1(N_{R^{-1}}\Sigma)$. Denoting

$$\mathbb{T} := \bigsqcup_{\theta \in \Theta} \mathbb{T}(\theta),$$

we have shown that every $g \in L^1(\Sigma_{\text{fat}})$ can be written as

$$g = \sum_{\theta \in \Theta} g_\theta = \sum_{\theta \in \Theta} \sum_{T \in \mathbb{T}(\theta)} g_T = \sum_{T \in \mathbb{T}} g_T.$$

This is the *wave packet decomposition of g at scale R* . The tubes in each family $\mathbb{T}(\theta)$ are parallel to each other (pointing in the direction normal to θ). Tubes corresponding to

different families $\mathbb{T}(\theta_1), \mathbb{T}(\theta_2)$ are $\sim R^{-1/2}$ -separated, due to the curvature of Σ . The functions g_T are essentially orthogonal in $L^2(\mathbb{R}^n)$, in the sense that, for every $\mathbb{W} \subseteq \mathbb{T}$,

$$\begin{aligned} \left\| \sum_{T \in \mathbb{W}} g_T \right\|_2^2 &= \left\| \sum_{\theta \in \Theta} \sum_{T \in \mathbb{W} \cap \mathbb{T}(\theta)} g_T \right\|_2^2 \\ &\stackrel{\text{supp } g_T \subseteq \theta, \forall T \in \mathbb{T}(\theta)}{=} \sum_{\theta \in \Theta} \left\| \sum_{T \in \mathbb{W} \cap \mathbb{T}(\theta)} g_T \right\|_2^2 \\ &\lesssim \sum_{\theta \in \Theta} \sum_{T \in \mathbb{W} \cap \mathbb{T}(\theta)} \|g_T\|_2^2 \\ &= \sum_{T \in \mathbb{W}} \|g_T\|_2^2. \end{aligned}$$

Moreover, for each $T \in \mathbb{T}$, $\widetilde{g}_T$ is a Schwartz function, which is negligible outside $R^\epsilon T$. In particular, using the decay that we know each $\widetilde{g}_T$ has, it is not too hard to show that, for every $x \in \mathbb{R}^n$,

$$\check{g}(x) = \sum_{T \in \mathbb{T}: x \in R^\epsilon T} \widetilde{g}_T(x) + c_\epsilon R^{-10^{100}n} \|g\|_2.$$

We record this information in the following proposition.

PROPOSITION 5.8.1. (The wave packet decomposition of any $g \in L^2(N_{R^{-1}}\Sigma)$ at scale R) Let Σ be a hypersurface in $\mathbb{R}^n$ of non-vanishing Gaussian curvature in $\mathbb{R}^n$. Fix $R \geq 1$, and let $\mathbb{T}$ be the family of $R^{1/2} \times \dots \times R^{1/2} \times R$ -tubes in $\mathbb{R}^n$ constructed earlier. Let $\epsilon > 0$. Then, for all $g \in L^1(N_{R^{-1}}\Sigma)$, we can write

$$g = \sum_{T \in \mathbb{T}} g_T,$$

so that the following hold.

- (i) For each $T \in \mathbb{T}$, g_T is supported in the $R^{-1/2}$ -rectangle in $N_{R^{-1}}\Sigma$ whose normal is parallel to T .
- (ii) For each $T \in \mathbb{T}$, $\widetilde{g}_T$ is essentially supported inside T , in the sense that

$$|\widetilde{g}_T(x)| \lesssim \left(R^{-1/2} \text{dist}(x', T) + R^{-1} \text{dist}(x_n, T) \right)^{-10^{100}n} \|g\|_2 \quad \text{for all } x = (x', x_n) \notin T.$$

- (iii) For each $x \in \mathbb{R}^n$, the tubes of $\mathbb{T}$ that contribute to $\check{g}(x)$ are essentially those that pass through x , in the sense that

$$\check{g}(x) = \sum_{T \in \mathbb{T}: x \in R^\epsilon T} \widetilde{g}_T(x) + c_\epsilon R^{-10^{100}n} \|g\|_2.$$

Therefore, for any $p \in [1, +\infty)$,

$$\|\check{g}\|_{L^p(B_R)} = \left\| \sum_{T \in \mathbb{T}: R^\epsilon T \cap B_R \neq \emptyset} \widetilde{g}_T \right\|_{L^p(B_R)} + c_\epsilon R^{-10^{100}n} \|g\|_2.$$

- (iv) The functions g_T are essentially orthogonal, in the sense that, for all $\mathbb{W} \subseteq \mathbb{T}$,

$$\left\| \sum_{T \in \mathbb{W}} g_T \right\|_2^2 \lesssim \sum_{T \in \mathbb{W}} \|g_T\|_2^2.$$

5.9. The polynomial method in Fourier restriction theory: a motivation

Recall (56): to prove that the extension operator $\widehat{\cdot d\sigma}$ is $L^q(\Sigma) \rightarrow L^p(B_R)$ -bounded, it suffices to show that

$$\|\widetilde{g}\|_{L^p(B_R)} \lesssim_{p,q,\epsilon} R^{-1+\frac{1}{q}+\epsilon} \|g\|_q$$

for functions g supported on $N_{R^{-1}}\Sigma$. Considering the wave packet decomposition $g = \sum_{T \in \mathbb{T}} g_T$ of each such g at scale R , this becomes equivalent to showing that

$$\left\| \sum_{T \in \mathbb{T}: R^\epsilon T \cap B_R \neq \emptyset} \widetilde{g_T} \right\|_{L^p(B_R)} \lesssim_{p,q,\epsilon} R^{-1+\frac{1}{q}+\epsilon} \|g\|_q$$

for all such g . In other words:

We wish to control the total mass of a sum of waves, each essentially evolving inside a different tube that intersects B_R .

This demonstrates again how difficult Fourier restriction is: we have tubes, each carrying a different wave. And whenever any of the tubes meet, the waves that they carry interfere with each other. The goal is to show that in principle this interference is *destructive*, so that the total wave contribution in B_R is quite small.

Even though this is very complicated, it is reminiscent of the Kakeya maximal operator conjecture. There, we have tubes pointing in different directions, and the goal is to show that they have small overlap. In this sense, the Kakeya maximal operator conjecture is a simpler analogue of the Fourier restriction conjecture, where oscillation has been removed.

Of course, the Kakeya maximal operator conjecture itself is very hard. So, people have been trying to remove not only oscillation from the problem, but also volume: they shrink the tubes down to lines, and hope to show that, under certain conditions, the lines do not meet very often. This seems like a gross simplification, however we already know that it was in this setting that the big revolution happened in 2008, when Dvir solved the Kakeya problem in finite fields using the polynomial method, revealing that the Kakeya conjectures have an underlying algebraic geometric nature that was not clear before.

The reason that the polynomial method works so well with lines is that *lines interact very well with zero sets of polynomials*: a polynomial p can either vanish along a line at most as many times as its degree $\deg p$, or the line fully lies inside the zero set Z_p of p . And more generally, if p has a certain algebraic property on many points of a line, it tends to have the same property throughout the whole line.

However, in general, *tubes do not interact well with zero sets of polynomials*. For instance, the zero set of a large degree polynomial can be fully compressed inside a small part of a tube. So, we cannot claim that if a polynomial behaves in a certain way at many points in the tube, then it will have behave similarly throughout the whole tube.

So, we need a different polynomial method, one where at least the computational bounds that work for lines also work for tubes. This polynomial method is *polynomial partitioning*, and was developed in 2010 by Guth and Katz, for the solution of the Erdős distinct distances problem in $\mathbb{R}^2$. Since 2014, polynomial partitioning has been used to make tremendous progress in the Fourier restriction problem, as well as related problems (e.g.

Falconer's distance set conjecture, Hörmander's problem on oscillatory integrals). It has also led to major progress or resolution of numerous problems in incidence geometry, as the degree of the polynomial it yields is in general better than the one coming from Dvir's vanishing lemma.

CHAPTER 6

Polynomial partitioning

6.1. The advantage

Polynomial partitioning is a polynomial method developed by Guth and Katz in 2010, for the solution of the distinct distances problem by Erdős. The method works only in Euclidean space, and can be applied either to a finite set of points (like Dvir's vanishing lemma), or to the mass of an L^1 -function (like $|\widehat{gd\sigma}|^p$ on B_R).

Let us compare Dvir's polynomial method with polynomial partitioning.

→ Dvir's vanishing lemma claims that, for any finite set S in $\mathbb{F}^n$, there exists a non-zero polynomial $p \in \mathbb{F}[x_1, \dots, x_n]$, with $\deg p \lesssim (\#S)^{1/n}$, so that S lives on its zero set Z_p .

→ Polynomial partitioning aims to find a *much lower degree* polynomial, achieving roughly the same thing. It works only in Euclidean space.

At first glance, it is perhaps not clear how this works. Polynomial partitioning ensures that, no matter which finite set S of points we start with in $\mathbb{R}^n$, and *no matter which degree d we choose* (ideally, $d \ll (\#S)^{1/n}$), we can find a polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, whose zero set Z_p splits S in smaller parts, in a nice, controlled way. More precisely, Z_p splits $\mathbb{R}^n$ in a controlled number of pieces (the *cells*, i.e. the connected components of $\mathbb{R}^n \setminus Z_p$), each containing $\lesssim \frac{\#S}{\#\text{cells}}$ points of S (so, each cell contains at most its fair share of points). Then, arguments tend to go this way:

→ **Cellular case.** More than half of the points of S lie inside the cells. Then, we focus our attention on a single cell of average behaviour. This is like a microscopic picture of what we started from: we started with a picture of many points and perhaps many lines through them, and we now focus on considerably fewer points and lines through them. If we are lucky, this case is settled (we control $\#S$) via some inductive argument (e.g. on the number of lines).

→ **Algebraic case.** More than half of the points of S lie on Z_p . This is *better* than what we got from Dvir's vanishing lemma, as now the degree d of the vanishing polynomial is much smaller than $(\#S)^{1/n}$!

6.2. Statement and proof

We first write down the polynomial partitioning theorem in either case (points or masses). The proof is particularly elegant, and relies on the topology of Euclidean space (specifically, on the Borsuk-Ulam theorem).

THEOREM 6.2.1. (Polynomial partitioning for points, Guth-Katz, 2010) *Let $n \in \mathbb{N}$, $d \geq 1$ and S a finite set of points in $\mathbb{R}^n$. Then, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, so that Z_p splits $\mathbb{R}^n$ into $\sim d^n$ cells (these are the connected components of $\mathbb{R}^n \setminus Z_p$, and some may be empty), each containing $\lesssim \frac{\#S}{d^n}$ points of S .*

THEOREM 6.2.2. (Polynomial partitioning for a mass) Let $n \in \mathbb{N}$, $d \geq 1$ and $f \in L^1(\mathbb{R}^n)$. Then, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, so that Z_p splits $\mathbb{R}^n$ into $\sim d^n$ cells C (these are the connected components of $\mathbb{R}^n \setminus Z_p$, and some may be empty), each satisfying

$$\int_C |f| = \frac{\int_{\mathbb{R}^n} |f|}{\#\text{cells}} \sim \frac{\int_{\mathbb{R}^n} |f|}{d^n}.$$

The proof for masses is in fact a little simpler than the one for points. They are both based on the following continuous analogue of Dvir's vanishing lemma.

THEOREM 6.2.3. (Polynomial ham-sandwich theorem, Stone-Tukey, 1942) Let $M \geq 2$, and let $S_1, \dots, S_M$ be subsets of $\mathbb{R}^n$, of finite, positive volume. Then, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \lesssim M^{1/n}$, whose zero set Z_p simultaneously bisects all the $S_1, \dots, S_M$ volume-wise, meaning that

$$|S_i \cap \{p > 0\}| = |S_i \cap \{p < 0\}|, \quad \text{for all } i = 1, \dots, M.$$

Notice that, since we are already talking about the signs of the values of p , we need to be at least in a field with an order. The proof is in turn based on the Borsuk-Ulam theorem, which is true thanks to the topology of Euclidean space.

THEOREM 6.2.4. (Borsuk-Ulam) Let $f : \mathbb{S}^{n-1} \rightarrow \mathbb{R}^{n-1}$ be continuous and odd. Then, there exists $x \in \mathbb{S}^{n-1}$ so that $f(x) = 0$.

Observe that an important condition for this to work is that the map f collapses the sphere of $\mathbb{R}^n$ to the horizontal hyperplane through 0.

PROOF OF THE POLYNOMIAL HAM-SANDWICH THEOREM. We know that each polynomial of degree $\leq d$ has the form

$$p(x_1, \dots, x_n) = \sum_{a_1 + \dots + a_n \leq d} c_{a_1, \dots, a_n} x_1^{a_1} \dots x_n^{a_n},$$

so it can be identified with its sequence of $\binom{d+n}{n}$ coefficients, i.e. with a vector in $\mathbb{R}^{\binom{d+n}{n}}$.

We equip this space with the Euclidean norm, and we consider its sphere $\mathbb{S}^{\binom{d+n}{n}-1}$. This is the set of polynomials of degree $\leq d$, whose vector of coefficients has Euclidean length 1. So, the polynomials in $\mathbb{S}^{\binom{d+n}{n}-1}$ are all *non-zero*.

We now define a map $f : \mathbb{S}^{\binom{d+n}{n}-1} \rightarrow \mathbb{R}^M$, with

$$p \rightarrow \left(|S_1 \cap \{p > 0\}| - |S_1 \cap \{p < 0\}|, \dots, |S_M \cap \{p > 0\}| - |S_M \cap \{p < 0\}| \right).$$

We see that f is continuous, because if the coefficients of p change by very little, then the sets where p is positive and negative, respectively, also change little. Also, f is odd, as $\{-p > 0\} = \{p < 0\}$ and $\{-p < 0\} = \{p > 0\}$. So, as long as

$$(57) \quad \binom{d+n}{n} - 1 \geq M,$$

the Borsuk-Ulam asserts that there exists a $p \in \mathbb{S}^{\binom{d+n}{n}-1}$ so that $f(p) = 0$. This polynomial p is non-zero, has degree $\leq d$, and bisects each S_i volume-wise.

In other words, for every d satisfying (57), there exists a non-zero $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, bisecting each S_i . This still works for the smallest d in the range (57). Since $\binom{d+n}{n} \sim d^n$, this smallest d satisfies

$$d^n \sim M,$$

and thus $d \sim M^{1/n}$. The proof is complete. $\square$

THEOREM 6.2.5. (Discrete polynomial ham-sandwich theorem, Guth-Katz, 2010) *Let $S_1, \dots, S_M$ be finite sets of points in $\mathbb{R}^n$. Then, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \lesssim M^{1/n}$, whose zero set Z_p bisects all the $S_1, \dots, S_M$ simultaneously, in the sense that*

$$\#(S_i \cap \{p > 0\}) \leq \frac{\#S_i}{2} \quad \text{and} \quad \#(S_i \cap \{p < 0\}) \leq \frac{\#S_i}{2}, \quad \text{for all } i = 1, \dots, M.$$

PROOF. For each $\delta > 0$, we consider the δ -neighbourhoods $S_1^\delta, \dots, S_M^\delta$ of the sets $S_1, \dots, S_M$. As long as δ is small enough, each S_i^δ is the union of $\#S_i$ disjoint δ -balls. By the polynomial ham-sandwich theorem, for $d \sim M^{1/n}$, there exists a polynomial $p_\delta \in \mathbb{S}^{\binom{d+n}{n}-1}$, bisecting each S_i volume-wise. We choose a sequence $\delta_n \rightarrow 0$ (e.g. $\delta = \frac{1}{n}$, for $n \in \mathbb{N}$). Then, the sequence $(p_{\delta_n})_{n \in \mathbb{N}}$ lives inside the compact set $\mathbb{S}^{\binom{d+n}{n}-1}$, therefore it has a subsequence that converges to some $p \in \mathbb{S}^{\binom{d+n}{n}-1}$ (with respect to the Euclidean norm). We will show that this p , which has $\deg p \leq d \sim M^{1/n}$, has the desired bisecting property.

For simplicity, we denote the above subsequence by $(p_n)_{n \in \mathbb{N}}$. The fact that $p_n \rightarrow p$ means that the vectors of coefficients of the polynomials p_n converge to the vector of coefficients of p (coordinate-wise), so $p_n \rightarrow p$ uniformly on compact sets.

Suppose now for contradiction that, for some $i \in \{1, \dots, M\}$,

$$\#(S_i \cap \{p > 0\}) > \frac{\#S_i}{2}.$$

In other words, p is positive on more than half of the points of S_i . Let us denote these points by $x_1, \dots, x_k$ (where $k > \frac{\#S_i}{2}$). Since p is continuous on $\mathbb{R}^n$, and these points are finitely many, there exists $\epsilon > 0$ so that p is positive on the balls $B(x_1, \epsilon), \dots, B(x_k, \epsilon)$. Since $p_n \rightarrow p$ uniformly on $B(x_1, \epsilon) \cup \dots \cup B(x_k, \epsilon)$, for large n each p_n is also positive on $B(x_1, \epsilon) \cup \dots \cup B(x_k, \epsilon)$, and thus on $B(x_1, \delta_n), \dots, B(x_k, \delta_n)$ (as, for large n , $\delta_n < \epsilon$). Moreover, these balls are disjoint for large n (as $\delta_n \rightarrow 0$). This is a contradiction, as it would imply that p_n is positive on more than half the volume of $S_i^{\delta_n}$. $\square$

We are now ready to prove the discrete polynomial partitioning theorem (whose statement we recall):

THEOREM 6.2.6. (Polynomial partitioning for points, Guth-Katz, 2010) *Let $n \in \mathbb{N}$, $d \geq 1$ and S a finite set of points in $\mathbb{R}^n$. Then, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, so that Z_p splits $\mathbb{R}^n$ into $\sim d^n$ cells (these are the connected components of $\mathbb{R}^n \setminus Z_p$, and some may be empty), each containing $\lesssim \frac{\#S}{d^n}$ points of S .*

PROOF. We apply the discrete polynomial ham-sandwich theorem iteratively: First we bisect S , splitting it into two parts, then we bisect each of these two parts, then we bisect each of the four resulting parts, and so on. Each bisection happens with a polynomial of low degree, and doubles the number of parts. We stop the process as soon as the degree we have accumulated is roughly d .

Now for the details:

By Theorem 6.2.5 applied to S , there exists a non-zero polynomial $p_1 \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p_1 \lesssim 1^{1/n}$, whose zero set Z_{p_1} bisects S (in the sense of the theorem)¹. I.e., $\mathbb{R}^n \setminus Z_{p_1}$ consists of 2^1 disjoint cells, $\{p_1 > 0\}$ and $\{p_1 < 0\}$, each of which contains $\leq \frac{\#S}{2}$ points of S .

By Theorem 6.2.5 applied to the two sets $S \cap \{p > 0\}$ and $S \cap \{p < 0\}$, there exists a non-zero $p_2 \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p_2 \lesssim 2^{1/n}$, whose zero set Z_{p_2} bisects both $S \cap \{p > 0\}$ and $S \cap \{p < 0\}$. I.e., $\mathbb{R}^n \setminus (Z_{p_1} \cup Z_{p_2})$ consists of 2^2 disjoint cells, $\{p_1 > 0\} \cap \{p_2 > 0\}$, $\{p_1 > 0\} \cap \{p_2 < 0\}$, $\{p_1 < 0\} \cap \{p_2 > 0\}$ and $\{p_1 < 0\} \cap \{p_2 < 0\}$, each of which contains $\leq \frac{\#S}{2^2}$ points of S .

We continue in a similar way. By the end of the j -th step, we have produced non-zero polynomials $p_1, \dots, p_j \in \mathbb{R}[x_1, \dots, x_n]$, of degrees $\deg p_1 \lesssim 2^{(1-1)/n}$, $\deg p_2 \lesssim 2^{(2-1)/n}, \dots, \deg p_j \lesssim 2^{(j-1)/n}$, so that $\mathbb{R}^n \setminus (Z_{p_1} \cup \dots \cup Z_{p_j})$ consists of 2^j disjoint cells, each determined by a choice of signs for the $p_1, \dots, p_j$, and each containing $\leq \frac{\#S}{2^j}$ points of S .

An important observation is that $Z_{p_1} \cup \dots \cup Z_{p_j} = Z_{p_1 \cdots p_j}$, so the above partitioning of $\mathbb{R}^n$ is achieved by the zero set of the single non-zero polynomial $p_1 \cdots p_j$, whose degree is

$$\deg(p_1 \cdots p_j) = \deg p_1 + \dots + \deg p_j \leq C_n(2^{(1-1)/n} + \dots + 2^{(j-1)/n}),$$

for a constant C_n depending only on the dimension n .

We stop the partitioning process at the J -th step, where J is the largest integer so that

$$C_n(2^{(1-1)/n} + \dots + 2^{(J-1)/n}) \leq d.$$

For this J ,

$$2^{(1-1)/n} + \dots + 2^{(J-1)/n} \sim d.$$

Since

$$2^{(1-1)/n} + \dots + 2^{(J-1)/n} = \sum_{k=0}^{J-1} (2^{1/n})^k \sim (2^{1/n})^{J-1} \sim 2^{J/n},$$

the J we are interested in satisfies $2^{J/n} \sim d$, hence $2^J \sim d^n$. This is in fact the number of cells at the end of the J -th step.

Therefore, at the end of this process, we have acquired a non-zero polynomial $p := p_1 \cdots p_J \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, whose zero set Z_p splits $\mathbb{R}^n$ into $2^J \sim d^n$ cells, each containing $\leq \frac{\#S}{2^J} \sim \frac{\#S}{d^n}$ points of S .

□

Sketch of proof of Theorem 6.2.2. The proof of polynomial partitioning for a non-negative mass $f \in L^1(\mathbb{R}^n)$ is a small adaptation of the above argument. We simply need to adjust the polynomial ham-sandwich theorem to the measure $f(x)dx$, rather than the Lebesgue measure dx . With a similar argument, we prove that, for any $S_1, \dots, S_M$ in $\mathbb{R}^n$, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \lesssim M^{1/n}$, which bisects each S_i with respect to the measure $f(x)dx$, meaning that

$$\int_{S_i \cap \{p > 0\}} f(x) dx = \int_{S_i \cap \{p < 0\}} f(x) dx, \quad \text{for all } i = 1, \dots, M$$

(with the implicit constant in the bound for $\deg p$ depending only on the dimension n).

¹In fact, it is easy to see that p_1 may be taken to be a hyperplane.

Then, polynomial partitioning for any non-negative $f \in L^1(\mathbb{R}^n)$ is proved similarly to the discrete case: we first bisect the mass of f (on the whole of $\mathbb{R}^n$), then we bisect each of the two resulting masses, then each of the four resulting masses, and so on, each time using a low-degree polynomial. We stop when we accumulate degree roughly d .

□

6.3. The numerology of the cellular case

Suppose that we wish to control the size of some finite set S in $\mathbb{R}^n$, relative to $\#\mathcal{L}$, for some family $\mathcal{L}$ of lines containing S . Polynomial partitioning may help us with this, especially if the points of S lie on certain lines given by our problem. It asserts the following:

There exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, so that Z_p splits $\mathbb{R}^n$ into $\sim d^n$ cells, each containing $\lesssim \frac{\#S}{d^n}$ points of S .

Then, independently of what d is, there are only two possible cases:

→ **The cellular case.** $\gtrsim \#S$ points of S (say, $> 1\%$ of the points in S) lie outside of Z_p (i.e., inside the union of the cells).

→ **The algebraic case.** $\gtrsim \#S$ points of S (say, $> 90\%$ of the points in S) lie on Z_p .

If for some $d \ll (\#S)^{1/n}$ we can somehow deal with the cellular case, either by contradicting it or by solving our problem in that case, then we can assume that we are in the algebraic case. That is, the problem is reduced to a case where most of our points cluster on the zero set of a polynomial of degree considerably smaller than $(\#S)^{1/n}$. This is *much stronger* than what we would get from Dvir's vanishing lemma.

Note that, a priori, there is no reason at all why one should be able to 'deal' with the cellular case for small d . Indeed, if there was a standard way to reduce to the algebraic case no matter what problem we started with, it would mean that all point-line incidence problems have an underlying algebraic nature, which does not have to be true.

There exists, however, a standard way to at least attempt to deal with the cellular case: working with a cell with average behaviour, i.e. a cell that contains its fair share of points, and is crossed by its fair share of lines in $\mathcal{L}$.

The existence of such a cell is ensured by the following lemma.

LEMMA 6.3.1. (Numerology of average cell) *Suppose that we have partitioned $S \subset \mathbb{R}^n$ using a polynomial of degree at most d , and that we are in the cellular case. Then, for any finite family $\mathcal{L}$ of L lines in $\mathbb{R}^n$, there exists a cell C of average behaviour, i.e. such that*

$$C \text{ contains } \sim \frac{\#S}{d^n} \text{ points of } S$$

and

$$C \text{ is crossed by } \lesssim \frac{L}{d^{n-1}} \text{ lines in } \mathcal{L}.$$

→ *Why might this be useful?* In a cell C of average behaviour, we have a microscopic picture of what we started with. In particular, we have fewer points (but not too few) lying in fewer lines than those we started from. Therefore, induction on the number of lines crossing C , or applying a known estimate inside C , may resolve the cellular case.

The rest of this section is dedicated to the proof of Lemma 6.3.1. The claim directly follows from the following two:

- (i) A definite proportion (say, 1%) of the cells are such that each contains $\sim \frac{\#S}{d^n}$ points of S , and
- (ii) A large proportion of the cells are such that each is crossed by $\lesssim \frac{L}{d^{n-1}}$ lines in $\mathcal{L}$. [By large proportion here we mean large enough so that, when added to the proportion above, we get more than 100% (say, 99.5%).]

If (i) and (ii) hold, then there must exist a cell that satisfies both properties.

We now prove precise versions of (i) and (ii). By our assumption that we are in the cellular case of polynomial partitioning, we know that there exist constants $c_1, c_2 > 0$, depending only on the dimension n , so that

$$\text{each cell contains} \leq c_1 \frac{\#S}{\#\text{cells}} \text{ points of } S$$

and

$$\geq c_2 \#S \text{ points of } S \text{ are in the cells.}$$

Proving (i). For the above constants c_1 and c_2 , the following holds.

CLAIM 6.3.2. *In the cellular case, $> \frac{c_2}{c_1}\%$ of the cells have the property that each contains $\sim \frac{\#S}{d^n}$ points of S .*

PROOF. Let us prove this claim. It suffices to show that $> \frac{c_2}{c_1}\%$ of the cells have the property that each contains $> 0.0001c_2 \frac{\#S}{\#\text{cells}}$ points of S (as $\#\text{cells} \sim d^n$).

The proof is a simple counting argument. We will assume that this is false, and reach a contradiction. Let

$$\mathcal{C}_{\text{good}} := \left\{ \text{cells each containing} > 0.0001c_2 \frac{\#S}{\#\text{cells}} \text{ points of } S \right\}$$

and

$$\mathcal{C}_{\text{bad}} := \left\{ \text{cells each containing} \leq 0.0001c_2 \frac{\#S}{\#\text{cells}} \text{ points of } S \right\}.$$

Our goal is to show that $\#\mathcal{C}_{\text{good}} > \frac{c_2}{c_1}\% \#\text{cells}$.

A large proportion of S lies inside the cells:

$$(58) \quad \sum_{\text{cells in } \mathcal{C}_{\text{good}}} \#\{\text{points of } S \text{ in cell}\} + \sum_{\text{cells in } \mathcal{C}_{\text{bad}}} \#\{\text{points of } S \text{ in cell}\} \geq c_2 \#S.$$

On the other hand, the bad cells contribute very few points in total, as each contains very few points of S :

$$\begin{aligned} \sum_{\text{cells in } \mathcal{C}_{\text{bad}}} \#\{\text{points of } S \text{ in cell}\} &\leq \#\mathcal{C}_{\text{bad}} \cdot 0.0001c_2 \frac{\#S}{\#\text{cells}} \\ &\leq \#\text{cells} \cdot 0.0001c_2 \frac{\#S}{\#\text{cells}} \\ &= 0.0001c_2 \cdot \#S. \end{aligned}$$

If we now assume for contradiction that the good cells are very few, and in particular that $\#\mathcal{C}_{\text{good}} \leq \frac{c_2}{c_1} \% \# \text{cells}$, it will follow that also the good cells contribute very few points:

$$\begin{aligned} \sum_{\text{cells in } \mathcal{C}_{\text{good}}} \#\{\text{points of } S \text{ in cell}\} &\leq \#\mathcal{C}_{\text{good}} \cdot c_1 \frac{\#S}{\# \text{cells}} \\ &\leq \frac{c_2}{c_1} \cdot 0.01 \# \text{cells} \cdot c_1 \frac{\#S}{\# \text{cells}} \\ &= 0.01 c_2 \cdot \#S. \end{aligned}$$

This is a contradiction, as the good and bad cell contributions are not enough to yield (58). So, our assumption that the good cells were very few must be false. In other words, $\#\mathcal{C}_{\text{good}} > \frac{c_2}{c_1} \% \# \text{cells}$, as desired. $\square$

Proving (ii). We know that there exists a constant $c_0 > 0$, depending only on the dimension, so that

$$\# \text{cells} \geq c_0 d^n.$$

For this constant c_0 , the following holds (and in fact independently of whether we are in the cellular case or not).

CLAIM 6.3.3. *For any family $\mathcal{L}$ of L lines in $\mathbb{R}^n$, and for any $\alpha \in (0, 100)$, $> \alpha \%$ of the cells have the property that each is crossed by $\leq A \frac{L}{d^{n-1}}$ lines of $\mathcal{L}$, where $A := \frac{10^{10}}{(100-\alpha)\% c_0}$.*

The precise value of A is unimportant here; all that matters is that it depends only on α (which we are free to choose as we please) and on the ambient dimension n . Observe that we are free to make the percentage of these nice cells as large as we wish (as long as we appropriately increase the factor A).

The proof of Claim 6.3.3 follows by a very simple double counting argument, which in turned is based on the following lemma, which lies at the heart of the polynomial method. **This is why polynomial partitioning (that is, partitioning with the zero set of a polynomial, rather than any other type of function), is so useful:**

LEMMA 6.3.4. *Any line in $\mathbb{R}^n$ enters at most $d + 1$ cells.*

PROOF. Let ℓ be a line in $\mathbb{R}^n$.

As ℓ passes from one cell to the next, it hits the zero set Z_p of p (where p is the partitioning polynomial). Indeed, recall that p is the product of polynomials $p_1, \dots, p_J$ which appeared in the polynomial partitioning iteration, and that each cell is determined by a different choice of signs for the $p_1, \dots, p_J$. Therefore, for any two cells C_1, C_2 , one of the polynomials p_i is, say, positive on C_1 and negative on C_2 . Now, suppose that ℓ enters both C_1 and C_2 . Then, the one-variable polynomial $(p_i)|_{\ell}$ is positive on $\ell \cap C_1$ and negative on $\ell \cap C_2$, so, by the intermediate value theorem (it is continuous) it must vanish in the transition between the cells. So, $p|_{\ell}$ must also vanish in the transition between the cells.

Therefore, if ℓ entered $\geq d + 2$ cells, it would hit $Z_p \geq d + 1 \geq \deg p + 1$ times, so ℓ would lie fully inside Z_p . This is impossible, as in this case ℓ would enter 0 cells (rather than $d + 2$ of them).

It follows that ℓ enters $\leq d + 1$ cells. $\square$

We are now ready to prove our earlier claim.

PROOF OF CLAIM 6.3.3. Fix $\alpha \in (0, 100)$, and let $\mathcal{L}$ be a family of L lines in $\mathbb{R}^n$. The proof is a double counting argument: since each line in $\mathcal{L}$ enters a small number of cells, the average cell must be entered by a small number of lines.

Let us first make a ‘careless’ attempt at this. By the above lemma,

$$\sum_{\ell \in \mathcal{L}} \#\{\text{cells crossed by } \ell\} \leq L \cdot (d+1) \lesssim Ld.$$

On the other hand, this sum also equals $\sum_{\text{cells}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\}$, so

$$\sum_{\text{cells}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} \lesssim Ld.$$

Since however there are $\sim d^n$ cells in total, some cell must contribute less than the average $\sim \frac{Ld}{d^n} \sim \frac{L}{d^{n-1}}$ to the sum above. In other words, there must exist at least one cell that is crossed by $\lesssim \frac{L}{d^{n-1}}$ lines.

This is not enough for us though; we wanted to show that a large proportion of the cells (rather than just one cell) will have this property. So, we now run the argument with a bit more care.

For A as in the statement, define

$$\mathcal{C}_{\text{good}} := \left\{ \text{cells each crossed by } \leq A \cdot \frac{L}{d^{n-1}} \text{ lines of } \mathcal{L} \right\}$$

and

$$\mathcal{C}_{\text{bad}} := \left\{ \text{cells each crossed by } > A \cdot \frac{L}{d^{n-1}} \text{ lines of } \mathcal{L} \right\}.$$

Our goal is to show that $\#\mathcal{C}_{\text{good}} > \alpha\% \#\text{cells}$.

By the above,

$$\begin{aligned} \sum_{\text{cells in } \mathcal{C}_{\text{good}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} + \sum_{\text{cells in } \mathcal{C}_{\text{bad}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} \\ (59) \quad &= \sum_{\text{cells}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} \\ &= \sum_{\ell \in \mathcal{L}} \#\{\text{cells crossed by } \ell\} \\ &\leq L(d+1) \leq 2Ld. \end{aligned}$$

In other words, both good and bad cells must contribute little to the sum. Let us now assume for contradiction that $\#\mathcal{C}_{\text{good}} \leq \alpha\% \#\text{cells}$. This must force the bad cells to be many:

$$\#\mathcal{C}_{\text{bad}} \geq (100 - \alpha)\% \#\text{cells}.$$

But this forces the bad cells’ contribution to (59) to be too large. Indeed,

$$\begin{aligned} \sum_{\text{cells in } \mathcal{C}_{\text{bad}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} &\geq \#\mathcal{C}_{\text{bad}} \cdot A \cdot \frac{L}{d^{n-1}} \\ &\geq (100 - \alpha)\% \#\text{cells} \cdot \frac{10^{10}}{(100 - \alpha)\% c_0} \cdot \frac{L}{d^{n-1}} \\ &\stackrel{\#\text{cells} \geq c_0 d^n}{\geq} (100 - \alpha)\% c_0 d^n \cdot \frac{10^{10}}{(100 - \alpha)\% c_0} \cdot \frac{L}{d^{n-1}} \\ &= 10^{10} Ld, \end{aligned}$$

which contradicts (59). So, the assumption that $\#\mathcal{C}_{\text{good}} \leq \alpha\% \#\text{cells}$ is false.

□

Consider $\alpha \in (0, 100)$ so that $\alpha\% + \frac{c_2}{c_1}\% > 100\%$. By the pigeonhole principle, there must exist a cell satisfying the bounds in both Claims 6.3.2 and 6.3.3. This completes the proof of Lemma 6.3.1.

6.4. Dealing with the cellular case via induction

In this section we see how one may deal with the cellular case **via induction**, to then reduce the problem of controlling $\#S$ to a problem where most of S lies on the zero set Z_p of the partitioning polynomial p .

REMARK 6.4.1. We emphasise here that induction is not the only way to deal with the cellular case, and in fact it may be a crude way to do so. It may be more useful to apply known, different estimates inside the average cell, rather than some inductive hypothesis. We will soon see both approaches via a specific example (the joints problem, a discrete analogue of the Kakeya maximal operator conjecture). However, since the inductive proofs are by now streamlined, and are also used a lot in harmonic analysis, we choose to dedicate this section to them alone.

□

Suppose that we have the following kind of problem.

Problem. Let (P) be some property, and let $q > 0$. For every family $\mathcal{L}$ of L lines in $\mathbb{R}^n$, denote

$$S(\mathcal{L}) := \{x \in \mathbb{R}^n : \text{the lines of } \mathcal{L} \text{ meeting at } x \text{ satisfy (P)}\}.$$

Show that

$$\#S(\mathcal{L}) \lesssim L^q,$$

where the implicit constant depends only on the dimension n and on q .

→ **Examples.**

- When

$$S(\mathcal{L}) = \{x \in \mathbb{R}^n : \text{at least 2 lines of } \mathcal{L} \text{ pass through } x\}.$$

In this case, (P) is the property of having cardinality ≥ 2 . And we know that

$$\#S(\mathcal{L}) \leq \binom{L}{2} \sim L^2,$$

so indeed a formula of the desired form works, for $q = 2$.

- When

$$S(\mathcal{L}) = \{x \in \mathbb{R}^n : \text{the directions of the lines of } \mathcal{L} \text{ through } x \text{ span } \mathbb{R}^n\}.$$

These points are called *joints* formed by $\mathcal{L}$, and we will soon see that

$$\#S(\mathcal{L}) \lesssim L^{\frac{n}{n-1}},$$

so again a formula of the desired form holds, for the magic exponent $q = \frac{n}{n-1}$ of the Kakeya maximal operator conjecture (whose double also features in the Fourier restriction conjecture).

When trying to control $S(\mathcal{L})$, we can of course apply polynomial partitioning to $S(\mathcal{L})$, using a partitioning polynomial p of any degree d we wish. We will see that, as long as $q > \frac{n}{n-1}$ and d is simply a large constant, we can deal with the cellular case by induction,

and reduce to the case where most of $S(\mathcal{L})$ lies on Z_p . Note that the algebraic surface Z_p now has *very low degree* (a large constant, independent of $\#S(\mathcal{L})$). This reveals very strong algebraic structure in our picture. In some sense, it is as if Z_p is just a union of very few planes, or anyway slightly curved surfaces.

We encode this information in the following lemma (the notation refers to the full generality in the statement of the **Problem** above).

LEMMA 6.4.2. *Let that $q > \frac{n}{n-1}$. When trying to show that*

$$\#S(\mathcal{L}) \lesssim_{n,q,(P)} L^q,$$

for any family $\mathcal{L}$ of L lines in $\mathbb{R}^n$, it suffices to show the estimate when $L \lesssim_{n,q,(P)} 1$, and to then assume that $S(\mathcal{L})$ lies in the zero set Z_p of a polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ with $\deg p \lesssim_{n,q} 1$.

PROOF. We are aiming for the estimate

$$(60) \quad \#S(\mathcal{L}) \leq C_{n,q,(P)} L^q,$$

for an appropriate constant $C_{n,q,(P)}$ that will be determined later. We proceed by induction on L :

- **The base case.** We first show (60) for all families $\mathcal{L}$ of lines in $\mathbb{R}^n$ with $L \leq A_{n,q,(P)}$, for some constant $A_{n,q,(P)}$.

- **The inductive hypothesis.** Let $\mathcal{L}$ be an arbitrary family of L lines in $\mathbb{R}^n$, with $L > A_{n,q,(P)}$. We will aim to show (60) for $\mathcal{L}$, assuming that (60) holds for any family of lines with strictly smaller cardinality.

- **The start of the argument: polynomial partitioning.** We apply polynomial partitioning to $S(\mathcal{L})$, using an arbitrary degree d . We will then show that we can solve the problem in the cellular case, as long as d is a large enough constant $A_{n,q}$. This will allow us to reduce the problem to the algebraic case, where a large proportion of $S(\mathcal{L})$ lies on the zero set of the partitioning polynomial (which has degree $\leq A_{n,q}$).

Indeed, let any $d \geq 1$. We apply polynomial partitioning to $S(\mathcal{L})$ with degree d . This yields a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, whose zero set Z_p splits $\mathbb{R}^n$ into $\sim d^n$ cells, each containing $\lesssim \frac{\#S(\mathcal{L})}{d^n}$ points of $S(\mathcal{L})$.

We now have two cases.

→ **The cellular case.** $> 50\%$ of $S(\mathcal{L})$ lies in the cells.

→ **The algebraic case.** $> 50\%$ of $S(\mathcal{L})$ lies in the zero set Z_p .

The point here is to find the lowest d for which we can get rid of the cellular case simply via induction inside a cell of average behaviour.

Suppose that we are in the cellular case. We know that there exists a cell C of average behaviour, meaning that

$$C \text{ contains } \sim \frac{\#S(\mathcal{L})}{d^n} \text{ points of } S(\mathcal{L}),$$

and

$$C \text{ is crossed by } \lesssim \frac{L}{d^2} \text{ lines of } \mathcal{L}.$$

Denoting by S_{cell} the set of points of $S(\mathcal{L})$ inside the cell C , by $\mathcal{L}_{\text{cell}}$ the set of lines in $\mathcal{L}$ entering C , and by $L_{\text{cell}} := \#\mathcal{L}_{\text{cell}}$, the above means that

$$\#S_{\text{cell}} \geq C_1 \frac{\#S(\mathcal{L})}{d^n}$$

and

$$L_{\text{cell}} \leq C_2 \frac{L}{d^{n-1}},$$

for some constants C_1, C_2 depending only on the dimension n . Now, two observations allow us to apply our inductive hypothesis to S_{cell} and $\mathcal{L}_{\text{cell}}$:

- (1) $S_{\text{cell}} \subseteq S(\mathcal{L}_{\text{cell}})$, as the lines of $\mathcal{L}_{\text{cell}}$ are the only lines in $\mathcal{L}$ that enter the cell, and thus they must satisfy the property (P) whenever they meet at points of S_{cell} . This further implies that

$$\#S(\mathcal{L}_{\text{cell}}) \geq C_1 \frac{\#S(\mathcal{L})}{d^n}.$$

- (2) $L_{\text{cell}} < L$, as long as d is larger than some constant depending only on n . Indeed,

$$L_{\text{cell}} \leq C_2 \frac{L}{d^{n-1}} < L,$$

as long as $d^{n-1} > C_2$, or equivalently

$$(61) \quad d > C_2^{\frac{1}{n-1}}.$$

From now on, we assume that the d we are using has this property.

This means that we can apply the inductive hypothesis to $\mathcal{L}_{\text{cell}}$, to obtain that

$$\#S(\mathcal{L}_{\text{cell}}) \leq C_{n,q,(P)} L_{\text{cell}}^q.$$

Replacing in this estimate the above bounds for $\#S(\mathcal{L}_{\text{cell}})$ and L_{cell} , we deduce that

$$C_1 \frac{\#S(\mathcal{L})}{d^n} \leq C_{n,q,(P)} \left(C_2 \frac{L}{d^{n-1}} \right)^q,$$

or equivalently that

$$\#S(\mathcal{L}) \leq \frac{C_2^q C_1^{-1}}{d^{q(n-1)-n}} \cdot C_{n,q,(P)} L^q.$$

Note that the power $q(n-1)-n$ in the denominator is *strictly positive*, exactly because $q > \frac{n}{n-1}$. Therefore, as long as d is a large enough constant, the factor $\frac{C_2^q C_1^{-1}}{d^{q(n-1)-n}}$ is smaller than 1, and the desired estimate

$$\#S(\mathcal{L}) \leq C_{n,q,(P)} L^q$$

is obtained in the cellular case. All we need is

$$d^{q(n-1)-n} \geq C_2^q C_1^{-1},$$

or equivalently

$$(62) \quad d \geq (C_2^q C_1^{-1})^{\frac{1}{q(n-1)-n}}.$$

Combining (61) and (66), we see that, as long as our d was any integer d satisfying

$$d \geq A_{n,q} := \max \left\{ C_2^{\frac{1}{n-1}}, (C_2^q C_1^{-1})^{\frac{1}{q(n-1)-n}} \right\},$$

the desired estimate

$$\#S(\mathcal{L}) \leq C_{n,q,(P)} L^q$$

has been proved in the cellular case.

This works for any d in the above range, and thus for $d \sim A_{n,q}$. Therefore, for such d , the cellular case is resolved, and we may thus assume that we are in the algebraic case: $> 50\%$ of S lies in Z_p , where now $\deg p \lesssim A_{n,q}$.

□

In the next sections, we will see how to apply polynomial partitioning in a specific example, the *joints problem*. This may be viewed as a discrete analogue of the Kakeya problem.

6.5. The joints problem

Recall that the Kakeya maximal operator conjecture claims that, for any $n \geq 2$, for any $\delta > 0$, and any family $\mathbb{T}$ of δ -separated δ -tubes in $\mathbb{R}^n$,

$$\int_{x \in \mathbb{R}^n} \#\{\text{tubes in } \mathbb{T} \text{ through } x\}^{\frac{n}{n-1}} dx \text{ is small.}$$

This is known only for $n = 2$: in this case, the proof is quite simple, as $\frac{n}{n-1} = 2$, so the multiplicity in the integrand represents the number of pairs of tubes through each x - in other words, it has a combinatorial meaning. For all $n \geq 3$, this is no longer the case.

The Kakeya maximal operator conjecture is difficult because the tubes have volume. This means that their overlap depends on the angle at which they meet - and things become more complicated the more tubes meet at a point. It thus makes sense to create a discrete analogue of the problem, where instead of tubes we have lines in $\mathbb{R}^n$. Things are simpler then: if two lines intersect, they do so at a single point.

In the Kakeya maximal operator conjecture, the separation of the tubes is due to their thickness - it is necessary so that the tubes look distinct (even when they intersect). Lines have no thickness, so, in our discrete analogue, it makes sense to not impose any separation conditions on their directions.

The discrete analogue we will now discuss is the *joints problem*. This appeared completely independently in the 1980s, in the world of computer science. It was Thomas Wolff who noticed that it is connected to the Kakeya problem.

DEFINITION 6.5.1. (Joints) Let $\mathcal{L}$ be a set of lines in $\mathbb{R}^n$. We say that a point $x \in \mathbb{R}^n$ is a joint formed by $\mathcal{L}$ if the directions of the lines in $\mathcal{L}$ that pass through x span $\mathbb{R}^n$. We denote by $J(\mathcal{L})$ the set of joints formed by $\mathcal{L}$.

In $\mathbb{R}^2$, any two (distinct) lines that intersect form a joint at their intersection. In $\mathbb{R}^3$, a family $\mathcal{L}$ of lines forms a joint at a point x if the lines in $\mathcal{L}$ that pass through x are not coplanar. As n gets larger, it is harder and harder for a joint to be formed.

We can now create a discrete analogue of the Kakeya maximal operator conjecture:

Question. Let $\mathcal{L}$ be a finite set of lines in $\mathbb{R}^n$. Is it true that

$$\sum_{x \in J(\mathcal{L})} \#\{\text{lines in } \mathcal{L} \text{ through } x\}^{\frac{n}{n-1}} \text{ is small?}$$

Ideally, like in the Kakeya maximal operator conjecture, we would like to control the above sum simply in terms of $\#\mathcal{L}$ (and independently of how the lines are distributed in $\mathbb{R}^n$). Of course, if we want to answer this question, to begin with we should be able to simply control $\#J(\mathcal{L})$ (which is the same sum, but with power 0 in place of $\frac{n}{n-1}$).

The question of controlling the number of joints formed by L lines in $\mathbb{R}^n$ appeared in 1983 in computer science, and it was solved in 2009 using the polynomial method (Dvir's vanishing lemma).

THEOREM 6.5.2. (Quilodrán, Kaplan-Sharir-Shustin, 2009) *Let $n \geq 2$. There exists a constant C_n , depending only on n , so that, for any family $\mathcal{L}$ of L lines in $\mathbb{R}^n$,*

$$\#J(\mathcal{L}) \leq C_n L^{\frac{n}{n-1}}.$$

→ **Observations.**

- The joints estimate is trivial for $n = 2$. Indeed, for any family $\mathcal{L}$ of L lines in $\mathbb{R}^2$, every joint lies in at least 2 lines of $\mathcal{L}$ (of course this is also true in all dimensions). Therefore,

$$\#J(\mathcal{L}) \leq \#\{\text{pairs of lines in } \mathcal{L}\} = \binom{L}{2} \sim L^2.$$

In other words, just as in the Kakeya maximal operator conjecture in two dimensions (where the exponent $\frac{n}{n-1} = 2$ leads us to consider the overlap of pairs of tubes), in the joints problem as well the same exponent leads us to consider intersections of pairs of lines - so the problem is easy.

In all higher dimensions, however, the joints problem was open for about 30 years.

- The bound in the joints theorem is optimal in all dimensions, in the sense that, for every $n \geq 2$, and every $L \in \mathbb{N}$, there exists a collection $\mathcal{L}$ of $\sim_n L$ lines in $\mathbb{R}^n$, so that

$$\#J(\mathcal{L}) \sim_n L^{\frac{n}{n-1}}.$$

This is easy to see; for example, consider axis-parallel lines, and the grid of points where they intersect.

□

In the upcoming sections, we will see:

- the proof of the joints theorem using Dvir's vanishing lemma,
- the proof of the slightly weaker estimate

$$\#J(\mathcal{L}) \lesssim_{n,\epsilon} L^{\frac{n}{n-1} + \epsilon}, \quad \text{for all } \epsilon > 0$$

using polynomial partitioning (here we will get rid of the cellular case using induction in the average cell, as in the previous section),

- a stronger estimate, in particular the discrete analogue

$$\sum_{x \in J(\mathcal{L})} \#\{\text{lines in } \mathcal{L} \text{ through } x\}^{\frac{3}{2}} \lesssim L^{\frac{3}{2}}$$

of the Kakeya maximal operator conjecture in $\mathbb{R}^3$, using polynomial partitioning (and a known theorem in the average cell, instead of induction, to get rid of the cellular case). In the above, the $\lesssim$ notation hides multiplicative factors of the form $(\log L)^{c_n}$.

6.6. Counting joints using Dvir's vanishing lemma (Quilodrán's proof)

We first recall the joints theorem. The rest of the section is a presentation of Quilodrán's proof.

THEOREM 6.6.1. (Quilodrán, Kaplan-Sharir-Shustin, 2009) *Let $n \geq 2$. There exists a constant C_n , depending only on n , so that, for any family $\mathcal{L}$ of L lines in $\mathbb{R}^n$,*

$$\#J(\mathcal{L}) \leq C_n L^{\frac{n}{n-1}}.$$

Quilodrán's proof is based on the following proposition, which **demonstrates the algebraic nature of the joints problem**.

PROPOSITION 6.6.2. *Let $\mathcal{L}$ be a collection of L lines in $\mathbb{R}^n$, and let J be a subset of $J(\mathcal{L})$. Suppose that a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ vanishes on J . Then,*

$$\#J \leq L \cdot \deg p.$$

PROOF. The proof is accomplished in two steps. The idea is that the estimate, roughly, claims that each line contains $\leq \deg p$ joints. This is not actually true, however it is true in terms of an iterative process: we will first find one line that contains $\leq \deg p$ joints, in the *remaining* collection of lines there will exist another line containing $\leq \deg p$ joints, and so on. In the first step we find *some* line with few joints, and in the second step we will count the joints by iteratively applying the first step.

Step 1. This step involves the core of the argument, i.e. the following.

CLAIM 6.6.3. *There exists a line in $\mathcal{L}$ containing $\leq \deg p$ joints of J .*

We prove the claim by contradiction.

- Suppose that each line $\ell \in \mathcal{L}$ contains $> \deg p$ joints of J .
- We know that p vanishes on J (as it vanishes on the whole of $J(\mathcal{L})$). Therefore, for each $\ell \in \mathcal{L}$, p vanishes along ℓ more times than its degree. This means that $\ell \subseteq Z_p$, for all $\ell \in \mathcal{L}$.

Observe how **the vanishing property of p has propagated from J to all the lines in $\mathcal{L}$** . This is a key point in any polynomial method argument.

- The fact that all lines in $\mathcal{L}$ lie in Z_p implies that, **for each joint $x \in J$, there cannot exist a tangent space to Z_p at x .**

Indeed, let $x \in J$. We know that x is a joint formed by $\mathcal{L}$, so the lines in $\mathcal{L}$ through x have directions that span $\mathbb{R}^n$. At the same time, these lines all live inside Z_p . So, if there existed a tangent space to Z_p at x , it would have to be tangent to all these lines, and this would force all these lines to live inside this tangent space. Then, however, they would not be able to form a joint at x : their directions would span something at most $(n - 1)$ -dimensional.

Observe how **the fact that all our lines are in Z_p implies further structural information**, specifically between the relationship of our joints with Z_p . This is a huge advantage of the polynomial method here: the revelation of the algebraic surface Z_p inside our picture reveals structure that was originally hidden from us.

- The above implies that, **for each joint $x \in J$, $\nabla p(x) = 0$.**

Indeed, let $x \in J$. If we had $\nabla p(x) \neq 0$, then there would exist a tangent space to Z_p at x (and $\nabla p(x)$ would be its normal).

- Therefore, some **there exists some non-zero polynomial g that vanishes on J and has lower-degree than p .**

Indeed, p is a non-zero function which vanishes somewhere (e.g. on J). Therefore, one of its partial derivatives $\partial_{x_1} p, \dots, \partial_{x_n} p$ must be a non-zero polynomial (as if all these partial derivatives were zero, p would be constant, and thus zero everywhere on $\mathbb{R}^n$, which is not true).

- The above would be a contradiction if we had started with p a non-zero polynomial of **minimal degree** vanishing on J (which we may well have). So, we run our argument above

with such a non-zero polynomial g in place of p (which of course satisfies $\deg g \leq \deg p$), and we show by contradiction that some $\ell \in \mathcal{L}$ contains $\leq \deg g$ (and thus $\leq \deg p$) points of J .

Step 2. We iteratively apply Step 1. We find a line in $\mathcal{L}$ containing $\leq \deg p$ joints of J , and we remove it (and the joints that it carries) from our collection. Amongst the remaining lines, we find another containing $\leq \deg p$ of the remaining joints of J , and we remove it (and the joints that it carries) from our collection. And so on! Eventually we have removed all the lines (and thus all the joints of J) in $\leq L$ steps, having removed $\leq \deg p$ joints of J at each step. So, it must be true that $\#J \leq L \cdot \deg p$.

Now for the details:

- p vanishes on J , so, by Claim 6.6.3, there exists $\ell_1 \in \mathcal{L}$ such that

$$\#(J \cap \ell_1) \leq \deg p.$$

We remove ℓ_1 from our collection of lines, and the joints in $J \cap \ell_1$ from our collection of joints. In this step, we have removed $\leq \deg p$ joints of J .

The remaining joints of J are those in $J_1 := J \setminus (J \cap \ell_1)$, and they are joints formed by the remaining lines, i.e. those in $\mathcal{L}_1 := \mathcal{L} \setminus \{\ell_1\}$ (since they are joints formed by $\mathcal{L}$, but ℓ_1 doesn't pass through them).

- Since p vanishes on J_1 (as it vanishes on the whole of J), by Claim 6.6.3 there exists $\ell_2 \in \mathcal{L}_1$ (i.e., a line in $\mathcal{L}$ different from ℓ_1), so that

$$\#(J_1 \cap \ell_2) \leq \deg p.$$

We remove ℓ_2 from our collection of lines, and the joints in $J_1 \cap \ell_2$ from our collection of joints. In this step, we have again removed $\leq \deg p$ joints of J .

The remaining joints of J are those in $J \setminus (J \cap (\ell_1 \cup \ell_2))$, and they are joints formed by the remaining lines, i.e. those in $\mathcal{L}_2 := \mathcal{L} \setminus \{\ell_1, \ell_2\}$.

- We keep repeating this argument, until there are no lines (and thus no joints) left. This means that our iteration has L steps, at each of which we have removed $\leq \deg p$ joints of J . And by the end of the final step, all joints of J have been removed. Therefore,

$$\#J \leq \#(\text{steps}) \cdot \#(\text{joints of } J \text{ removed at each step}) \leq L \cdot \deg p.$$

□

The proof of the joints theorem (Theorem 6.6.1) is now a simple combination of Proposition 6.6.2 (which tells us that we can count joints *if* some polynomial vanishes on them) with Dvir's vanishing lemma (which asserts that *some polynomial*, of quite low degree, *indeed vanishes* on the joints).

PROOF OF THEOREM 6.6.1. Let $\mathcal{L}$ be a family of L lines in $\mathbb{R}^n$.

By Dvir's vanishing lemma, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \lesssim \#J(\mathcal{L})^{1/n}$, which vanishes on $J(\mathcal{L})$.

By Proposition 6.6.2,

$$\#J(\mathcal{L}) \leq \deg p \cdot L \lesssim \#J(\mathcal{L})^{1/n} \cdot L.$$

Rearranging, this gives

$$\#J(\mathcal{L})^{\frac{n-1}{n}} \lesssim L,$$

or equivalently

$$\#J(\mathcal{L}) \lesssim L^{\frac{n}{n-1}}.$$

□

Something to remember. By Proposition 6.6.2, as soon as we find *some* non-zero polynomial p vanishing on some joints (formed by a collection $\mathcal{L}$ of L lines), then we know that the joints are $\leq \deg p \cdot L$. So, the lower the degree, the better bounds we get for those joints.

There exist of course collections $\mathcal{L}$ of lines for which $\#J(\mathcal{L}) \sim L^{\frac{n}{n-1}}$, so Dvir's degree bound is optimal, if we look at the whole of $J(\mathcal{L})$ (for arbitrary collections $\mathcal{L}$). However, if we look at appropriate subcollections of $J(\mathcal{L})$, with particular structure in space, then maybe we can find *lower-degree* polynomials that vanish on those joints.

This idea will be useful later on, when we will show the discrete analogue

$$\sum_{x \in J(\mathcal{L})} \#\{\text{lines in } \mathcal{L} \text{ through } x\}^{3/2} \lesssim L^{3/2}$$

of the Kakeya maximal operator conjecture in $\mathbb{R}^3$ (a stronger estimate), using polynomial partitioning.

6.7. Counting joints using polynomial partitioning

In this section we will prove, for practice, the following slightly weaker version of the joints theorem, using polynomial partitioning (and specifically resolving the cellular case using induction inside a cell of average behaviour).

THEOREM 6.7.1. *Let $n \geq 2$. For every $\epsilon > 0$, there exists a constant $C_{n,\epsilon}$, depending only on n and ϵ , so that, for every family $\mathcal{L}$ of L lines in $\mathbb{R}^n$,*

$$\#J(\mathcal{L}) \leq C_{n,\epsilon} L^{\frac{n}{n-1} + \epsilon}.$$

PROOF. Let $\epsilon > 0$.

The idea is the following: Since the exponent $\frac{n}{n-1} + \epsilon$ is strictly larger than $\frac{n}{n-1}$, we know that, if we apply polynomial partitioning to $J(\mathcal{L})$, using as d simply a large constant $D_{n,\epsilon}$, then the cellular case will be resolved using induction inside a cell of average behaviour. So, we will reduce the problem to the algebraic case, in which the majority of $J(\mathcal{L})$ lies in the zero set of the partitioning polynomial. In this case, the partitioning polynomial p , which has $\deg p \leq d = D_{n,\epsilon}$, will vanish on a large proportion of $J(\mathcal{L})$. By Proposition 6.6.2, it will follow that $\#J(\mathcal{L}) \leq L \cdot \deg p \sim_{n,\epsilon} L$, which is even better than the desired estimate. So, the algebraic case will also be resolved.

Let us see the details. We prove the desired estimate by induction on L .

• **Base case.** For all collections $\mathcal{L}$ of lines in $\mathbb{R}^n$, with cardinality $L \leq 100$, it holds that

$$\#J(\mathcal{L}) \leq \binom{L}{2} \sim L^2 \leq L^{\frac{n}{n-1}} 100^{2 - \frac{n}{n-1}} \leq C_n L^{\frac{n}{n-1} + \epsilon},$$

for an appropriately large constant C_n .

• **Inductive hypothesis.** Let $L \in \mathbb{N}$, and suppose that, for any family $\mathcal{L}'$ of lines in $\mathbb{R}^n$, with $\#\mathcal{L}' < L$, we have shown that

$$(63) \quad \#J(\mathcal{L}') \leq C_{n,\epsilon} (\#\mathcal{L}')^{\frac{n}{n-1}+\epsilon},$$

for an appropriate constant $C_{n,\epsilon}$ depending only on n and ϵ , which is $> C_n$ and will be specified later (so that the induction will close).

• **Inductive step.** Now, let $\mathcal{L}$ be a collection of L lines in $\mathbb{R}^n$. We will show that

$$(64) \quad \#J(\mathcal{L}) \leq C_{n,\epsilon} L^{\frac{n}{n-1}+\epsilon}$$

as well (for the *same* constant $C_{n,\epsilon}$ that appears in the inductive hypothesis).

We perform polynomial partitioning on $J(\mathcal{L})$, using some degree d which will be specified later. Independently of the choice of d , this yields a non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, whose zero set Z_p splits $\mathbb{R}^n$ in $\sim d^n$ cells, each containing $\lesssim \frac{\#J(\mathcal{L})}{d^n}$ points of $J(\mathcal{L})$.

There are now two cases:

→ **The cellular case:** $> 50\%$ of the points of $J(\mathcal{L})$ lie inside the cells.

→ **The algebraic case:** $> 50\%$ of the points of $J(\mathcal{L})$ lie on Z_p .

• Suppose we are in the **cellular case**. Then, there exists a cell C demonstrating average behaviour, meaning that

$$C \text{ contains } \sim \frac{\#J(\mathcal{L})}{d^n} \text{ points of } J(\mathcal{L}),$$

and

$$C \text{ is intersected by } \lesssim \frac{L}{d^{n-1}} \text{ lines of } \mathcal{L}.$$

The implicit constants hiding above depend only on n . Therefore, denoting by J_{cell} the points of $J(\mathcal{L}) \cap C$, by $\mathcal{L}_{\text{cell}}$ the set of lines in $\mathcal{L}$ that intersect C , and $L_{\text{cell}} := \#\mathcal{L}_{\text{cell}}$, the above means that

$$\#J_{\text{cell}} \geq C_1 \frac{\#J(\mathcal{L})}{d^n}$$

and

$$L_{\text{cell}} \leq C_2 \frac{L}{d^{n-1}},$$

for some constants C_1 and C_2 that depend only on n .

The points in J_{cell} are joints formed by $\mathcal{L}_{\text{cell}}$ (since they are joints formed by $\mathcal{L}$, but the rest of the lines in $\mathcal{L}$ do not enter the cell). Moreover, using d so that

$$d^{n-1} \geq 2C_2,$$

or equivalently

$$(65) \quad d \geq (2C_2)^{\frac{1}{n-1}},$$

it follows that

$$L_{\text{cell}} \leq C_2 \frac{L}{d^{n-1}} \leq C_2 \frac{L}{2C_2} = \frac{L}{2} < L.$$

Therefore, the inductive hypothesis (63) holds for $\mathcal{L}_{\text{cell}}$, and gives

$$\#J_{\text{cell}} \leq C_{n,\epsilon} (L_{\text{cell}})^{\frac{n}{n-1}+\epsilon}$$

(for the constant $C_{n,\epsilon}$ that appears in the inductive hypothesis). Replacing in this estimate the earlier bounds for $\#J_{\text{cell}}$ and L_{cell} , we obtain

$$C_1 \frac{\#J(\mathcal{L})}{d^n} \leq C_{n,\epsilon} \left(C_2 \frac{L}{d^{n-1}} \right)^{\frac{n}{n-1} + \epsilon},$$

or equivalently

$$\#J(\mathcal{L}) \leq \left(\frac{C_2^{\frac{n}{n-1} + \epsilon} C_1^{-1}}{d^{(n-1)\epsilon}} \right) \cdot C_{n,\epsilon} L^{\frac{n}{n-1} + \epsilon},$$

which in turn is $\leq C_{n,\epsilon} L^{\frac{n}{n-1} + \epsilon}$ if

$$\frac{C_2^{\frac{n}{n-1} + \epsilon} C_1^{-1}}{d^{(n-1)\epsilon}} \leq 1,$$

or equivalently

$$(66) \quad d \geq \left(C_2^{\frac{n}{n-1} + \epsilon} C_1^{-1} \right)^{\frac{1}{(n-1)\epsilon}}.$$

Combining the two conditions (65) and (66) on the degree d , the desired estimate

$$\#J(\mathcal{L}) \leq C_{n,\epsilon} L^{\frac{n}{n-1} + \epsilon}$$

indeed holds (that is, the induction closes) in the cellular case, as long as we have used

$$d \geq \max \left\{ (2C_2)^{\frac{1}{n-1}}, \left(C_2^{\frac{n}{n-1} + \epsilon} C_1^{-1} \right)^{\frac{1}{(n-1)\epsilon}} \right\}$$

We fix $d := D_{n,\epsilon}$ from now on, the smallest integer satisfying the above. The cellular case has been resolved for this d .

• It remains to show the desired estimate (64) in the **algebraic case** (for $d = D_{n,\epsilon}$).

In this case, we know that $> 50\%$ of the points of $J(\mathcal{L})$ live in Z_p . We denote by J this set of joints. We have thus found a subset J of $J(\mathcal{L})$, with

$$\#J \geq \frac{1}{2} \#J(\mathcal{L}),$$

so that the partitioning polynomial p vanishes on J . And $\deg p \leq d = D_{n,\epsilon}$.

By Proposition 6.6.2, it immediately follows that

$$\#J \leq \deg p \cdot L \leq D_{n,\epsilon} L \leq D_{n,\epsilon} L^{\frac{n}{n-1} + \epsilon},$$

and thus

$$\#J(\mathcal{L}) \leq (2 \cdot D_{n,\epsilon}) L^{\frac{n}{n-1} + \epsilon}.$$

So, the induction closes in this case too, as long as the multiplicative constant $C_{n,\epsilon}$ that we chose in our inductive hypothesis satisfies

$$C_{n,\epsilon} \geq 2D_{n,\epsilon}.$$

Therefore, fixing

$$C_{n,\epsilon} := \max\{C_n, 2D_{n,\epsilon}\}$$

(where, recall, C_n is the multiplicative constant that appeared in the base case of our induction process), the inductive argument closes in the algebraic case too, and the proof is complete. □

6.8. Counting joints with multiplicities in $\mathbb{R}^3$ (using polynomial partitioning)

In this section we will prove a stronger estimate for $n = 3$, specifically the discrete analogue of the Kakeya maximal operator conjecture in $\mathbb{R}^3$, using polynomial partitioning.

PROPOSITION 6.8.1. *For every collection $\mathcal{L}$ of L lines in $\mathbb{R}^3$, it holds that*

$$\sum_{x \in J(\mathcal{L})} \#\{\text{lines in } \mathcal{L} \text{ through } x\}^{\frac{3}{2}} \lesssim \log L \cdot L^{\frac{3}{2}}$$

The idea here is to sort the joints in $J(\mathcal{L})$ in $\sim \log L$ groups, according to how many lines pass through each joint. In particular, the group J_k will contain the joints in $J(\mathcal{L})$, through each of which $\sim k$ lines of $\mathcal{L}$ are passing (specifically, $\geq k$ and $< 2k$ lines of $\mathcal{L}$). Then, we will control each $\#J_k$ independently: we will show that

$$\#J_k k^{3/2} \lesssim L^{3/2}.$$

Adding over the logarithmically many k will give the desired estimate.

To control each $\#J_k$ as above, we will apply polynomial partitioning to J_k . However, since the exponent in the RHS is exactly $3/2$, we know that induction inside a cell of average behaviour will *not* resolve the cellular case. Instead:

- We will apply partitioning with a polynomial of *more interesting degree* (than just a large constant),
- We will resolve the cellular case by applying a *known theorem* (which counts points) inside the average cell.

In other words, this approach does not involve induction.

The theorem that will produce the interesting degree, and which we will also apply inside the average cell, is the *Szemerédi-Trotter theorem*. There are some simple proofs of the Szemerédi-Trotter theorem (compared to the original one), e.g. a proof using the crossing number inequality from graph theory, and another using polynomial partitioning. We will not prove the Szemerédi-Trotter here, but we will discuss it, before proceeding to the proof of Proposition 6.8.1. In fact, we present here a slightly weaker version of the original theorem, since it is a bit more intuitive and serves our purposes.

THEOREM 6.8.2. (**Szemerédi-Trotter, 1983**) *For any family $\mathcal{L}$ of L lines in $\mathbb{R}^2$, and every $k \geq 2$, denote*

$$S_k := \{x \in \mathbb{R}^2 : \geq k \text{ and } < 2k \text{ lines of } \mathcal{L} \text{ pass through } x\}.$$

Then,

$$\#S_k \lesssim \frac{L}{k} + \frac{L^2}{k^3}.$$

The implicit constant above is independent of $\mathcal{L}$ and k .

—→ **Observations.**

- While the Szemerédi-Trotter theorem is phrased for $\mathbb{R}^2$, it actually holds inside any $\mathbb{R}^n$. Indeed, any finite collection $\mathcal{L}$ of lines inside $\mathbb{R}^n$ can be projected to $\mathbb{R}^2$, in a way that the number of projected lines remains the same, and so that the number of projected lines that meet at each projected point of S_k remains the same (in fact, a generic projection will work).

- There are two terms that control $\#S_k$: $\frac{L}{k}$ and $\frac{L^2}{k^3}$. According to the range of k , one of them is dominant. In particular,

$$(67) \quad \#S_k \lesssim \begin{cases} \frac{L^2}{k^3}, & \text{if } k \lesssim L^{1/2}, \\ \frac{L}{k}, & \text{if } k \gtrsim L^{1/2}. \end{cases}$$

- Each of the above two bounds is *non-trivial*. Indeed, since $k \geq 2$, we know that at least 2 lines of $\mathcal{L}$ pass through each point of S_k . So, a trivial bound would come from counting the pairs of (distinct) lines in $\mathcal{L}$ that pass through the points of S_k :

$$\begin{aligned} \#S_k k^2 &\sim \sum_{x \in S_k} \binom{k}{2} \\ &= \sum_{x \in S_k} \#\{\text{pairs of distinct lines in } \mathcal{L} \text{ passing through } x\} \\ &\leq \#\{\text{pairs of distinct lines in } \mathcal{L}\} \\ &= \binom{L}{2} \sim L^2. \end{aligned}$$

In other words, a trivial estimate is

$$\#S_k \lesssim \frac{L^2}{k^2}.$$

Both bounds in (67) are *much stronger* than this:

- The bound $\frac{L^2}{k^3}$ (which is dominant for small k) tells us that we can gain a whole extra power of k in the bound's denominator.
- The bound $\frac{L}{k}$ (which is dominant for large k) tells us that, at least numerologically, the S_k behave like roots of distinct bushes, out of which $\sim k$ lines of $\mathcal{L}$ emanate.

We are now ready to prove Proposition 6.8.1.

PROOF OF PROPOSITION 6.8.1. Let $\mathcal{L}$ be a collection of L lines in $\mathbb{R}^3$. For each *dyadic* $k \geq 2$, i.e. any k of the form 2^m , for some $m \in \mathbb{N}$, denote

$$J_k := \{x \in J(\mathcal{L}) : \geq k \text{ and } < 2k \text{ lines of } \mathcal{L} \text{ pass through } x\}.$$

Observe that

$$J(\mathcal{L}) = \bigsqcup_{\text{dyadic } k \in [2, L]} J_k,$$

as at least 3 and at most L lines of $\mathcal{L}$ pass through each joint in $J(\mathcal{L})$. Therefore,

$$\sum_{x \in J(\mathcal{L})} \#\{\text{lines in } \mathcal{L} \text{ through } x\}^{3/2} \sim \sum_{\text{dyadic } k \in [2, L]} J_k k^{3/2}.$$

Now, the number of dyadic k in $[2, L]$ is $\sim \log L$. Indeed, it equals $\sim M$, where M is the smallest integer satisfying $2^M \geq L$, in which case $M \sim \log L$.

So, it suffices to show that

$$(68) \quad \#J_k k^{3/2} \lesssim L^{3/2}$$

for each dyadic $k \in [2, L]$. We will prove (68) for all $k \in [2, L]$ (k being specifically dyadic doesn't matter in terms of this).

So, let $k \in [2, L]$.

If $k \geq L^{1/2}$: Then, by the Szemerédi-Trotter theorem,

$$\#J_k \lesssim \frac{L}{k} \stackrel{\frac{L}{k} \geq 1}{\leq} \left(\frac{L}{k}\right)^{3/2},$$

hence $\#J_k k^{3/2} \lesssim L^{3/2}$, as desired. In other words, (68) holds for ‘large’ k .

If $k < L^{1/2}$: Then, by the Szemerédi-Trotter theorem,

$$\#J_k \lesssim \frac{L^2}{k^3}.$$

Therefore, for an appropriate absolute constant A (independent of $\mathcal{L}$ and k), $\#J_k \leq A \cdot \frac{L^2}{k^3}$, so the quantity

$$d := A \cdot \frac{L^2}{\#J_k k^3}$$

is *larger than 1*, and can thus play the role of the upper bound for the degree of a partitioning polynomial.

In particular, for this d , we apply polynomial partitioning to J_k . This yields a non-zero polynomial $p \in \mathbb{R}[x_1, x_2, x_3]$, with $\deg p \leq d$, whose zero set Z_p splits $\mathbb{R}^3$ in $\sim d^3$ cells, each containing $\lesssim \frac{\#J_k}{d^3}$ points of J_k .

There are now two cases:

→ **The cellular case.** $> 50\%$ of the points of J_k lie inside the cells.

→ **The algebraic case.** $> 50\%$ of the points of J_k lie on Z_p .

• Suppose that we are in the **cellular case**. Then, there exists a cell C , so that, denoting

$$J_{k,\text{cell}} := J_k \cap C, \quad \mathcal{L}_{\text{cell}} := \{\text{lines in } \mathcal{L} \text{ that enter } C\}, \quad \text{and} \quad L_{\text{cell}} := \#\mathcal{L}_{\text{cell}},$$

we have

$$\#J_{k,\text{cell}} \sim \frac{\#J_k}{d^3} \quad \text{and} \quad L_{\text{cell}} \lesssim \frac{L}{d^2}.$$

Each point in $J_{k,\text{cell}}$ lies in $\geq k$ and $< 2k$ lines of $\mathcal{L}_{\text{cell}}$ (as the rest of the lines in $\mathcal{L}$ do not enter the cell). Therefore, by the Szemerédi-Trotter theorem,

$$\#J_{k,\text{cell}} \lesssim \frac{L_{\text{cell}}}{k} + \frac{L_{\text{cell}}^2}{k^3},$$

so

$$\#J_{k,\text{cell}} \lesssim \frac{L_{\text{cell}}^2}{k^3} \quad \text{or} \quad \#J_{k,\text{cell}} \lesssim \frac{L_{\text{cell}}}{k}.$$

The estimate $\#J_{k,\text{cell}} \lesssim \frac{L_{\text{cell}}^2}{k^3}$ is impossible, as long as the large constant A in the definition of d has been chosen to be appropriately large. Indeed, $\#J_{k,\text{cell}} \lesssim \frac{L_{\text{cell}}^2}{k^3}$ would imply that

$$\frac{\#J_k}{d^3} \lesssim \frac{(L/d^2)^2}{k^3},$$

or equivalently

$$d \lesssim \frac{L^2}{\#J_k k^3},$$

a contradiction for A larger than the multiplicative constant hiding above. So, making sure that our A is this large, $\#J_{k,\text{cell}} \lesssim \frac{L_{\text{cell}}^2}{k^3}$ is impossible, and therefore

$$\#J_{k,\text{cell}} \lesssim \frac{L_{\text{cell}}}{k}.$$

This implies that

$$\frac{\#J_k}{d^3} \lesssim \frac{(L/d^2)}{k},$$

or equivalently

$$\#J_k k \lesssim L d^{\frac{L^2}{\#J_k k^3}} \sim L \cdot \frac{L^2}{\#J_k k^3},$$

whence

$$\#J_k^2 k^4 \lesssim L^3.$$

Rearranging, we obtain the estimate

$$\#J_k \lesssim \frac{L^{3/2}}{k^2},$$

which is even better than the desired estimate $\#J_k \lesssim \frac{L^{3/2}}{k^{3/2}}$.

Therefore, (68) holds in the cellular case.

• It remains to deal with the **algebraic case**. In this case, denote by J the set of joints in J_k which lie on Z_p ; we know that $\#J \gtrsim \#J_k$.

Since p vanishes on J , Proposition 6.8.1 implies that

$$\#J \leq L \cdot \deg p.$$

Therefore, recalling that $d \sim \frac{L^2}{\#J_k k^3}$, we obtain

$$\#J_k \lesssim L \cdot \frac{L^2}{\#J_k k^3},$$

or equivalently

$$\#J_k \lesssim \frac{L^{3/2}}{k^{3/2}},$$

as desired. So, (68) has also been proved in the algebraic case. □

6.9. Polynomial partitioning in Fourier restriction theory: a taste

In this section, we explain the basics of how polynomial partitioning can be used in the Fourier restriction problem².

Thanks to the wave-packet decomposition of functions supported close to Σ , polynomial partitioning works well for L^2 -based estimates, e.g. estimates of the form

$$(69) \quad \|\widehat{f d\sigma}\|_{L^p(B_R)} \leq C_\epsilon R^\epsilon \|f\|_2, \quad f \in L^2(\Sigma).$$

Moreover, perhaps it will not be a surprise (given our applications of polynomial partitioning to discrete examples, as well as from the way Khinchin's inequality turns Fourier restriction estimates into Kakeya maximal operator estimates) that polynomial partitioning also works well when $p > \frac{2n}{n-1}$ (the Fourier restriction exponent).

Of course, we have seen the proof of (69) for the best possible p ($p = \frac{2(n+1)}{n-1}$), and without any R^ϵ -loss (the Stein-Tomas theorem). Even so, (69) is still a very good toy model to test polynomial partitioning, because:

²The ideas here are based on L. Guth's article 'A restriction estimate using polynomial partitioning' (JAMS, 2016, <https://arxiv.org/pdf/1407.1916>).

- Even when people are trying to show restriction estimates that are not L^2 -based, they instead first tackle other estimates with polynomial partitioning, such as *multilinear inequalities*, which *are* L^2 -based (and then they transform these estimates into the Fourier restriction estimates they want).

- The way they tackle these different L^2 -based estimates is very similar to the approach we will see now for (69).

Here we go...

Polynomial partitioning approach to (69).

→ **Considering functions on $N_{R^{-1}}\Sigma$.** To begin with, recall that we can consider, instead of $f \in L^2(\Sigma)$, an appropriately fattened function f_{fat} , for any $R \geq 1$, supported on $N_{R^{-1}}\Sigma$, so that

$$|\widehat{f d\sigma}| \sim R |\check{f}| \quad \text{on } B_R$$

and

$$\|f\|_{L^2(\Sigma)} \sim R^{1/2} \|f_{\text{fat}}\|_{L^2(\mathbb{R}^n)}.$$

Given this, it suffices to show that, for all $\epsilon > 0$, there exists $C_\epsilon > 0$ so that

$$(70) \quad \|\check{g}\|_{L^p(B_R)} \lesssim C_\epsilon R^\epsilon R^{-1/2} \|g\|_2, \quad g \in L^2(\mathbb{R}^n), \quad \text{supp } g \subseteq N_{R^{-1}}\Sigma$$

for all such fattened versions g .

We now fix $R \geq 1$ and $\epsilon > 0$, and we aim for (70). For simplicity, we denote by $\mathcal{D}$ the class of functions g that feature in (70), for this given R .

→ **Inductive hypothesis on the radius.** We may assume that, for all $1 \leq \rho < \frac{R}{2}$, it holds that

$$(71) \quad \|\check{g}\|_{L^p(B_\rho)} \lesssim C_\epsilon \rho^\epsilon R^{-1/2} \|g\|_2, \quad \text{for all } g \in \mathcal{D}.$$

The reason is that (71) trivially holds for $1 \leq \rho \leq 100$ (for C_ϵ large enough), as

$$(\|\check{g}\|_{L^p(B_{100})})^{1/p} = \left(\int_{B_{100}} |\check{g}|^p \right)^{1/p} \leq |B_{100}|^{\frac{1}{p}} \cdot \|g\|_{L^1(N_{R^{-1}}\Sigma)} \lesssim |N_{R^{-1}}\Sigma|^{1/2} \|g\|_2 \sim R^{-1/2} \|g\|_2,$$

for all $g \in \mathcal{D}$.

→ **Recalling the wave-packet decomposition of g .** Consider $\mathbb{T}_R = \cup_{\theta \in \Theta} \mathbb{T}(\theta)$, a collection of $R^{1/2} \times \dots \times R^{1/2} \times R$ tubes inside B_R , so that Θ consists of $O(1)$ -overlapping, $R^{-1/2}$ -separated caps covering Σ , and the tubes in $\mathbb{T}(\theta)$ are all pointing to the normal to θ . Then, we can write

$$g = \sum_{T \in \mathbb{T}_R} g_T,$$

where each $\widehat{g_T}$ is essentially supported in T (in the sense that it decays quickly outside $R^\epsilon T$). Importantly, the g_T are essentially pairwise orthogonal, so

$$\|g\|_2^2 \sim \sum_{T \in \mathbb{T}_R} \|g_T\|_2^2,$$

so it suffices to show that

$$\left\| \sum_{T \in \mathbb{T}_R} \check{g_T} \right\|_{L^p(B_R)} \leq C_\epsilon R^\epsilon R^{-1/2} \left(\sum_{T \in \mathbb{T}} \|g_T\|_2^2 \right)^{1/2}$$

(with perhaps a slightly altered constant C_ϵ).

→ **Inductive hypothesis on the number of tubes in the wave-packet decomposition of g .** Let $M \in \mathbb{N}$, $M \leq \#\mathbb{T}_R$. We may assume that, for any subcollection $\mathbb{W} \subseteq \mathbb{T}_R$, with $\#\mathbb{W} < M$, the estimate

$$(72) \quad \left\| \sum_{T \in \mathbb{W}} \widetilde{g}_T \right\|_{L^p(B_R)} \leq C_\epsilon R^\epsilon R^{-1/2} \left(\sum_{T \in \mathbb{W}} \|g_T\|_2^2 \right)^{1/2}$$

holds. The reason is that (72) holds when $\#\mathbb{W} \lesssim 1$. Indeed, let $T \in \mathbb{T}_R$, and denote by $\theta(T)$ the smallest $R^{-1/2} \times \dots \times R^{-1/2} \times R^{-1}$ -rectangle along Σ , with normal the direction of T . Then, recall from the wave-packet decomposition section that

$$g_T = c_k \exp(k) \psi_\theta$$

for some k and some exponential function $\exp(k)$ of unit length pointwise, where ψ_θ is a smooth bump function adapted to θ . Therefore,

$$\widetilde{g}_T = c_k |\theta| \psi_T,$$

where ψ_T is a Schwartz function adapted to T (it is a smooth version of χ_T that decays quickly outside $R^\epsilon T$). Therefore,

$$\|g_T\|_2 = |c_k| \cdot |\theta|^{1/2},$$

while

$$\|\widetilde{g}_T\|_{L^p(B_R)} \sim |c_k| \cdot |\theta| \cdot |T|^{1/p}.$$

So, indeed

$$\|\widetilde{g}_T\|_{L^p(B_R)} \leq C_\epsilon R^\epsilon R^{-1/2} \|g_T\|_2$$

if and only if

$$|\theta| \cdot |T|^{1/p} \leq C_\epsilon R^\epsilon R^{-1/2} \cdot |\theta|^{1/2}.$$

or equivalently (since $|\theta| = |T|^{-1}$)

$$1 \leq C_\epsilon R^\epsilon R^{-1/2} \cdot |\theta|^{-1/2} \cdot |T|^{-1/p} = C_\epsilon R^\epsilon R^{-1/2} \cdot |T|^{\frac{1}{2} - \frac{1}{p}} = C_\epsilon R^\epsilon R^{-1/2} \left(R^{\frac{n+1}{2}} \right)^{\frac{1}{2} - \frac{1}{p}},$$

which holds for all $p \geq \frac{2(n+1)}{n-1}$ (as then $\frac{1}{2} - \frac{1}{p} \geq \frac{1}{n+1}$).

→ **The inductive step.** We now fix $g \in \mathcal{D}$. Our goal is to show that, for any $\mathbb{T} \subseteq \mathbb{T}_R$ with $\#\mathbb{T} = M$,

$$(73) \quad \left\| \sum_{T \in \mathbb{T}} \widetilde{g}_T \right\|_{L^p(B_R)} \leq C_\epsilon R^\epsilon R^{-1/2} \left(\sum_{T \in \mathbb{T}} \|g_T\|_2^2 \right)^{1/2}.$$

Fix such $\mathbb{T}$.

• **Polynomial partitioning.** Setting

$$f := \sum_{T \in \mathbb{T}} g_T,$$

in which case

$$\check{f} = \sum_{T \in \mathbb{T}} \widetilde{g}_T,$$

we see that our goal is to show that

$$\int_{B_R} |\check{f}|^p \leq C_\epsilon^p R^{p\epsilon} R^{-p/2} \|f\|_2^p.$$

Since we wish to control $\int |\check{f}|^p \chi_{B_R}$, we apply polynomial partitioning to $|\check{f}|^p \chi_{B_R}$, with some degree d that will be decided later. This yields some non-zero polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$, with $\deg p \leq d$, whose zero set Z splits $\mathbb{R}^n$ into $\sim d^n$ cells C , each satisfying

$$\int_{C \cap B_R} |\check{f}|^p = \frac{\int_{B_R} |\check{f}|^p}{\#\text{cells}} \sim \frac{\int_{B_R} |\check{f}|^p}{d^n}.$$

• **An issue: a tube can enter a huge number of cells.** Our situation, at first glance, appears nicer than the corresponding one in discrete problems:

- we only have a cellular case (the zero set Z contributes nothing to the total mass, since it has measure 0), while also
- every cell contains exactly the same amount of mass.

To be able to run the cellular case as in the discrete setting, we would also like to know that the average cell is intersected by $\lesssim \frac{\#\mathbb{T}}{d^{n-1}}$ tubes in $\mathbb{T}$. If that was the case, then we would choose such an average cell C , we would apply our inductive hypothesis (72) inside C , and we would hope for our induction process to close.

We need to be careful though. In the discrete setting, given a family $\mathcal{L}$ of lines in $\mathbb{R}^n$, we proved that the average cell is intersected by $\lesssim \frac{\#\mathcal{L}}{d^{n-1}}$ lines of $\mathcal{L}$ using that *every* $\ell \in \mathcal{L}$ enters $\leq d + 1$ cells.

In our continuous setting, *this is no longer the case*. Let $T \in \mathbb{T}$. While every line ℓ inside T and parallel to T (such as its central axis) enters $\leq d + 1$ cells, the tube T itself can enter as many cells as it wishes (even all of them!). So, we cannot deduce that every tube enters few cells, and thus we cannot conclude that the average cell is intersected by few tubes.

• **An adjustment to partitioning, to fix the issue.** To fix the above problem, we make an adjustment to our partitioning. Motivated by the fact that each $\widetilde{g}_T$ is negligible outside $R^\epsilon T$, we thicken Z by $R^{1/2+\epsilon}$, the width of the tubes $R^\epsilon T$. So, instead of having a partitioning surface Z , we now have a partitioning ‘wall’

$$W := N_{\frac{1}{2}R^{1/2+\epsilon}} Z \cap B_R,$$

and shrunken cells

$$C' := (C \setminus W) \cap B_R,$$

(over all original cells C). Therefore, while the partitioning polynomial has not changed, we now use that

$$B_R := W \sqcup \bigsqcup_{\text{shrunken cells } C'} C'.$$

The number of new cells has also not changed: they are still $\sim d^n$ in total.

The advantage of this partitioning is that the following holds.

CLAIM 6.9.1. *Every $R^{1/2+\epsilon} \times \dots \times R^{1/2+\epsilon} \times R$ -tube in $\mathbb{R}^n$ enters $\leq d + 1$ shrunken cells C' .*

PROOF. Whenever T enters a shrunken cell C' , then its central axis ℓ enters the original cell C (as either ℓ either enters $C' \subseteq C$, or its distance from C' is smaller than the width $R^{1/2+\epsilon}$ of T , and thus the width of the wall W).

Since ℓ enters $\leq d + 1$ of the original cells, T enters $\leq d + 1$ of the shrunken cells.

□

Thanks to this, we will be able to deal with the cellular case as in discrete settings, using induction inside a cell of average behaviour (both in terms of the mass it contains, and in terms of the number of tubes entering it).

Let us see the details. It holds that

$$\int_{B_R} |\check{f}|^p = \int_W |\check{f}|^p + \sum_{\text{shrunk cells } C'} \int_{C'} |\check{f}|^p,$$

so we split in two cases.

— **The cellular case.** The LHS mass is mostly contained in the cells, meaning that the term involving the cells is the dominant one:

$$\sum_{\text{shrunk cells } C'} \int_{C'} |\check{f}|^p \geq \int_W |\check{f}|^p.$$

In this case,

$$\int_{B_R} |\check{f}|^p \leq 2 \cdot \sum_{\text{shrunk cells } C'} \int_{C'} |\check{f}|^p.$$

— **The algebraic case.** The LHS mass is mostly contained in the wall W , meaning that the term involving the wall is the dominant one:

$$\int_W |\check{f}|^p \geq \sum_{\text{shrunk cells } C'} \int_{C'} |\check{f}|^p.$$

In this case,

$$\int_{B_R} |\check{f}|^p \leq 2 \cdot \int_W |\check{f}|^p.$$

• **Dealing with the cellular case.** In this case, we know that more than 50% of the mass of $|\check{f}|^p$ in B_R lies inside the cells C' . At the same time, each cell carries at most the average of the mass. Therefore, similarly to the discrete setting, a definite proportion of the cells each contribute roughly the average.

CLAIM 6.9.2. *In the cellular case, $> 0.5\%$ of the shrunk cells C' satisfy*

$$\int_{C' \cap B_R} |\check{f}|^p \sim \frac{\int_{B_R} |\check{f}|^p}{d^n}.$$

Moreover, again as in the discrete setting, it can be proved that, since each $R^\epsilon T$ enters a small number of cells (for all $T \in \mathbb{T}$), the average cell will be entered by few of the tubes $R^\epsilon T$. For simplicity, denote

$$R^\epsilon \mathbb{T} := \{R^\epsilon T : T \in \mathbb{T}\}.$$

CLAIM 6.9.3. *Let $\alpha \in (0, 100)$. Then, $> \alpha\%$ of the cells C' have the property that*

$$C' \text{ is intersected by } \leq A \frac{\#\mathbb{T}}{d^{n-1}} \text{ tubes of } R^\epsilon \mathbb{T},$$

where $A := \frac{10^{10}}{(100-\alpha)\%c_0}$ (for c_0 a constant depending only on the dimension, and for which $\#\text{cells} \geq c_0 d^n$).

We leave the proofs of both claims as exercises. However, we prove the following version of Claim 6.9.3, which states that the tubes entering the average cell will also carry little mass in total.

CLAIM 6.9.4. Let $\alpha \in (0, 100)$. Then, $> \alpha\%$ of the cells C' have the property that

$$\sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} \|g_T\|_2^2 \leq A \cdot \frac{1}{d^{n-1}} \sum_{T \in \mathbb{T}} \|g_T\|_2^2,$$

where A is as in Claim 6.9.3.

PROOF. On the one hand, since each $R^\epsilon T \in R^\epsilon \mathbb{T}$ enters $\leq d+1$ cells C' , it follows that

$$\sum_{T \in \mathbb{T}} \sum_{\text{cells } C' \text{ that } R^\epsilon T \text{ enters}} \|g_T\|_2^2 \leq \sum_{T \in \mathbb{T}} (d+1) \|g_T\|_2^2 \leq 2d \sum_{T \in \mathbb{T}} \|g_T\|_2^2.$$

Changing the order of summation on the LHS above, we obtain

$$(74) \quad \sum_{\text{cells } C'} \sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} \|g_T\|_2^2 \leq 2d \sum_{T \in \mathbb{T}} \|g_T\|_2^2.$$

Now, let

$$C_{\text{good}} := \left\{ \text{cells } C' \text{ for which } \sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} \|g_T\|_2^2 \leq A \cdot \frac{1}{d^{n-1}} \sum_{T \in \mathbb{T}} \|g_T\|_2^2 \right\}$$

and

$$C_{\text{bad}} := \{\text{the remaining cells } C'\}.$$

Our goal is to show that $\#C_{\text{good}} > \alpha\% \# \text{cells}$.

Suppose for contradiction that $\#C_{\text{good}} \leq \alpha\% \# \text{cells}$. Then, $\#C_{\text{bad}} \geq (100 - \alpha)\% \# \text{cells}$, hence

$$\begin{aligned} \sum_{C' \in C_{\text{bad}}} \sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} \|g_T\|_2^2 &\geq \#C_{\text{bad}} \cdot A \cdot \frac{1}{d^{n-1}} \sum_{T \in \mathbb{T}} \|g_T\|_2^2 \\ &\geq (100 - \alpha)\% \# \text{cells} \cdot A \cdot \frac{1}{d^{n-1}} \sum_{T \in \mathbb{T}} \|g_T\|_2^2 \\ &\geq (100 - \alpha)\% c_0 d^n \cdot \frac{10^{10}}{(100 - \alpha)\% c_0} \cdot \frac{1}{d^{n-1}} \sum_{T \in \mathbb{T}} \|g_T\|_2^2 \\ &\geq 10^{10} d \sum_{T \in \mathbb{T}} \|g_T\|_2^2, \end{aligned}$$

which contradicts (74). Therefore, $\#C_{\text{good}} > \alpha\% \# \text{cells}$. □

Choosing sufficiently large α , and combining Claims 6.9.2, 6.9.3 and 6.9.4, we find a cell of average behaviour in all three senses³.

COROLLARY 6.9.5. In the cellular case, there exists a cell C' so that

$$\int_{C'} |\check{f}|^p \sim \frac{\int_{B_R} |\check{f}|^p}{d^n},$$

$$C' \text{ is intersected by } \lesssim \frac{\#\mathbb{T}}{d^{n-1}} \text{ tubes of } R^\epsilon \mathbb{T}$$

and

$$\sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} \|g_T\|_2^2 \lesssim \frac{1}{d^{n-1}} \sum_{T \in \mathbb{T}} \|g_T\|_2^2.$$

³To be precise, the second and third property of the corollary involve the tubes $T \in \mathbb{T}$ so that $R^\epsilon T$ enters C' .

The crucial observation now is that, since each $\widetilde{g}_T$ is negligible outside $R^\epsilon T$, within the cell C' the only tubes $T \in \mathbb{T}$ that contribute to the mass are those for which $R^\epsilon T$ enters C' . That is, modulo a negligible error,

$$\check{f} = \sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} \widetilde{g}_T \quad \text{on } C'.$$

Therefore, defining

$$f_{C'} := \sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} g_T,$$

we see that

$$\check{f} = \widetilde{f}_{C'} \quad \text{on } C'$$

(while also $\widetilde{f}_{C'}$ is negligible outside C'). Therefore, the first property in the above corollary can be rewritten as

$$\int_{B_R} |\widetilde{f}_{C'}|^p \sim \frac{\int |\check{f}|^p}{d^n},$$

Moreover, due to the orthogonality of the wave packets g_T ,

$$\sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} \|g_T\|_2^2 \sim \left\| \sum_{T \in \mathbb{T}: R^\epsilon T \text{ enters } C'} g_T \right\|_2^2 \sim \|f_{C'}\|_2^2,$$

while also

$$\sum_{T \in \mathbb{T}} \|g_T\|_2^2 \sim \left\| \sum_{T \in \mathbb{T}} g_T \right\|_2^2 \sim \|f\|_2^2.$$

Therefore, the third property in the above corollary can be rewritten as

$$\|f_{C'}\|_2^2 \lesssim \frac{1}{d^{n-1}} \|f\|_2^2.$$

So, we rewrite the above corollary in this language, which reveals a direct analogy with the cellular case numerology from discrete settings.

COROLLARY 6.9.6. (Cellular case numerology) *For every cell C' , we define*

$$\mathbb{T}_{C'} := \{T \in \mathbb{T} : R^\epsilon T \text{ intersects } C'\}$$

and

$$f_{C'} := \sum_{T \in \mathbb{T}_{C'}} g_T.$$

Then, in the cellular case there exists a cell C' satisfying

$$\int_{B_R} |\widetilde{f}_{C'}|^p \sim \frac{\int |\check{f}|^p}{d^n},$$

$$\#\mathbb{T}_{C'} \lesssim \frac{\#\mathbb{T}}{d^{n-1}},$$

and

$$\|f_{C'}\|_2^2 \lesssim \frac{1}{d^{n-1}} \|f\|_2^2.$$

We now fix a cell C' satisfying these three properties (i.e., exhibiting average behaviour in all three senses). For d a sufficiently large constant,

$$\#\mathbb{T}_{C'} < \#\mathbb{T},$$

so we can apply our inductive hypothesis (72) with $\mathbb{W} := \mathbb{T}_{C'}$. This yields

$$\left\| \sum_{T \in \mathbb{T}_{C'}} \widetilde{g}_T \right\|_{L^p(B_R)} \leq C_\epsilon R^\epsilon R^{-1/2} \left(\sum_{T \in \mathbb{T}_{C'}} \|g_T\|_2^2 \right)^{1/2},$$

i.e.

$$\|\widetilde{f}_{C'}\|_{L^p(B_R)} \leq C \cdot C_\epsilon R^\epsilon R^{-1/2} \|f_{C'}\|_2,$$

for some constant C depending only on the dimension n . Equivalently,

$$\int_{B_R} |\widetilde{f}_{C'}|^p \leq C^p C_\epsilon^p R^{p\epsilon} R^{-p/2} (\|f_{C'}\|_2^2)^{p/2}.$$

Using now the bounds from Corollary 6.9.6, which relate $f_{C'}$ to f , we obtain

$$\frac{\int_{B_R} |\check{f}|^p}{d^n} \leq C C^p C_\epsilon^p R^{p\epsilon} R^{-p/2} \left(\frac{\|f\|_2^2}{d^{n-1}} \right)^{p/2},$$

where again C' is a constant depending only on the dimension. Rearranging, we obtain

$$\int_{B_R} |\check{f}|^p \leq \frac{C C^p}{d^{(n-1)\frac{p}{2}-n}} \cdot C_\epsilon^p R^{p\epsilon} R^{-p/2} \|f\|_2^p.$$

Importantly, as long as

$$p > \frac{2n}{n-1},$$

the denominator is a *positive* power of d . So, as long as d is a large enough constant $D_{n,p}$ (depending only on n and p), we have that

$$\frac{C C^p}{d^{(n-1)\frac{p}{2}-n}} \leq 1,$$

hence

$$\int_{B_R} |\check{f}|^p \leq C_\epsilon^p R^{p\epsilon} R^{-p/2} \|f\|_2^p,$$

or equivalently

$$\|\check{f}\|_{L^p(B_R)} \leq C_\epsilon R^\epsilon R^{-1/2} \|f\|_2,$$

as desired.

So, as long as $p > \frac{2n}{n-1}$ (the critical exponent for Fourier restriction) and the degree d of the partitioning polynomial is a large enough constant (depending only on n and p), *the induction closes in the cellular case*. We fix such d from now on, and we can thus reduce the problem to the algebraic case.

• **Dealing with the algebraic case: a sketch.** We give a quick overview of how the algebraic case can be approached. Recall that in this case

$$\int_{B_R} |\check{f}|^p \leq 2 \int_W |\check{f}|^p,$$

therefore it suffices to show that

$$\int_W |\check{f}|^p \leq \frac{C_\epsilon^p}{2} R^{\epsilon p} R^{-p/2} \|f\|_2^p.$$

Denoting

$$\mathbb{T}_W := \{T \in \mathbb{T} : R^\epsilon T \text{ intersects } W\},$$

we see that, modulo a negligible error,

$$\check{f} = \sum_{T \in \mathbb{T}_W} \check{g}_T \quad \text{on } W.$$

So, defining

$$h := \sum_{T \in \mathbb{T}_W} g_T,$$

it suffices to show that

$$(75) \quad \int_W |\check{h}|^p \leq c C_\epsilon^p R^{\epsilon p} R^{-p/2} \|f\|_2^p,$$

for an explicit constant c (perhaps small) that has arisen from the argument so far.

We now outline the rest of the argument (with some white lies):

- Cover W by a finitely overlapping family $\mathcal{B}$ of balls B of radius $\rho := R^{1-\epsilon}$.
- For every $B \in \mathcal{B}$, we split the tubes $T \in \mathbb{T}$ (for which $R^\epsilon T$ intersects B) into two groups:

$$\mathbb{T}_{\text{tang},B} := \{T \in \mathbb{T}_W : (R^\epsilon T) \cap B \subseteq W\}$$

(the set of tubes that are ‘tangential’ to W inside B), and

$$\mathbb{T}_{\text{tr},B} := \{T \in \mathbb{T}_W \setminus \mathbb{T}_{\text{tang},B} : R^\epsilon T \text{ intersects } B\}$$

(the set of tubes that are ‘transverse’ to W inside B). Then,

$$\check{h} = \check{h}_{\text{tang},B} + \check{h}_{\text{tr},B} \quad \text{on } B \cap W,$$

where

$$h_{\text{tang},B} := \sum_{T \in \mathbb{T}_{\text{tang},B}} g_T, \quad h_{\text{tr},B} := \sum_{T \in \mathbb{T}_{\text{tr},B}} g_T.$$

- By the above,

$$\int_W |\check{h}|^p \lesssim \sum_{B \in \mathcal{B}} \int_B |\check{h}_{\text{tang},B}|^p + \sum_{B \in \mathcal{B}} \int_B |\check{h}_{\text{tr},B}|^p.$$

So, we need to control each of the two terms above.

→ **The transverse term.** This can be treated by using the radial inductive hypothesis inside every $B \in \mathcal{B}$ (since each such B has radius $R^{1-\epsilon}$, much smaller than R), combined with the claim below⁴.

CLAIM 6.9.7. *Every $T \in \mathbb{T}$ can belong to $\mathbb{T}_{\text{tr},B}$ for at most d^{A_n} balls B .*

⁴This claim is highly non-trivial, and requires ideas from algebraic geometry. For it to hold, we actually need Z to be a smooth hypersurface (we can assume this by Sard’s theorem, which allows us to appropriately perturb the partitioning polynomial p without changing its degree too much). This means that a tangent space exists to Z at every one of its points. So, we can talk about the angle at which a tube T intersects W .

In particular, if a tube T is transverse to W inside B , then $R^\epsilon T$ must intersect W at angle $> R^{-1/2+\epsilon/2}$ somewhere inside B (as otherwise $(R^\epsilon T) \cap B$ would fully lie inside W , so T would be tangential to W inside B , a contradiction). Then, one can prove that $R^\epsilon T$ cannot cross W at such a large angle too many times across its length.

This can be viewed as an analogue of the fact that each tube $R^\epsilon T$ enters a small number of cells (only now, instead of cells, we have small balls). And, instead of using the inductive hypothesis (on the number of tubes) inside every cell, we use the inductive hypothesis (on the radius) inside each B . The argument closes very similarly to the cellular case:

For each $B \in \mathcal{B}$, by the inductive hypothesis (71) on the radius, we have that

$$\int_B |\check{h}_{\text{tr},B}|^p \leq C_\epsilon^p \rho^{p\epsilon} R^{-p/2} \|h_{\text{tr},B}\|_2^p,$$

so

$$\begin{aligned} \sum_{B \in \mathcal{B}} \int_B |\check{h}_{\text{tr},B}|^p &\leq C_\epsilon^p \rho^{p\epsilon} R^{-p/2} \sum_{B \in \mathcal{B}} \|h_{\text{tr},B}\|_2^p \\ &= C_\epsilon^p \rho^{p\epsilon} R^{-p/2} \sum_{B \in \mathcal{B}} \left(\left\| \sum_{T \in \mathbb{T}_{\text{tr},B}} g_T \right\|_2 \right)^p \\ &\stackrel{\text{orthogonality}}{\lesssim} C_\epsilon^p \rho^{p\epsilon} R^{-p/2} \sum_{B \in \mathcal{B}} \left(\sum_{T \in \mathbb{T}_{\text{tr},B}} \|g_T\|_2^2 \right)^{p/2} \\ &\stackrel{p/2 > 1}{\leq} C_\epsilon^p \rho^{p\epsilon} R^{-p/2} \left(\sum_{B \in \mathcal{B}} \sum_{T \in \mathbb{T}_{\text{tr},B}} \|g_T\|_2^2 \right)^{p/2} \\ &= C_\epsilon^p \rho^{p\epsilon} R^{-p/2} \left(\sum_{T \in \mathbb{T}} \#\{B \in \mathcal{B} : T \in \mathbb{T}_{\text{tr},B}\} \cdot \|g_T\|_2^2 \right)^{p/2} \\ &\stackrel{\text{Claim 6.9.7}}{\leq} C_\epsilon^p \rho^{p\epsilon} R^{-p/2} d^{A_n \cdot \frac{p}{2}} \left(\sum_{T \in \mathbb{T}} \|g_T\|_2^2 \right)^{p/2} \\ &\lesssim C_\epsilon^p (\rho^{p\epsilon} d^{A_n \cdot \frac{p}{2}}) R^{-p/2} \|f\|_2^p. \end{aligned}$$

The induction now closes because, thanks to the fact that $\rho = R^{1-\epsilon}$ is much smaller than R , while also d is simply a large constant, we have

$$\rho^{p\epsilon} d^{A_n \cdot \frac{p}{2}} \ll R^{p\epsilon}.$$

→ **The tangential term.** This is the hard part, and is usually handled in ways that we cannot address here. The idea is that the tubes T tangential to the wall W have *restricted directions*, so the corresponding wave packets g_T are supported in certain ‘slices’ of $N_{R^{-1}}\Sigma$, which thus have *lower dimension* than Σ .

This case is handled through different techniques than the ones we have seen so far, e.g. decoupling arguments or multilinear estimates along these slices.

In any case, we can keep the following moral out of this story:

Moral. In the Fourier restriction problem, the most interesting wave packets are those that evolve tangentially to algebraic hypersurfaces of low degree.