LECTURE NOTES 1 FOR 254A

TERENCE TAO

1. Introduction

The aim of this course is to show how harmonic analysis techniques can be viewed in a unified manner from the perspective of the *phase plane*. This plane allows one to view a function and its Fourier transform at the same time, and it clarifies the many types of decompositions (Littlewood-Paley decomposition, wavelet decomposition, Calderón-Zygmund decomposition, etc.) which come up in harmonic analysis.

In the first week we shall review the Fourier transform on $\mathbf{R}$, and give some intuition as to what the phase plane is, although we shall keep the phase plane a fuzzy concept for the time being. Then, in the next week, we shall introduce the first of several decompositions of the phase plane, namely the Littlewood-Paley decomposition. This decomposition has many, many uses, but to begin with we shall apply this decomposition to the theory of Sobolev spaces.

2. The Fourier transform on $\mathbf{R}$; a review

We now discuss the Fourier transform on the additive group $\mathbf{R}$. Actually, there is an analogue of the Fourier transform for any abelian group, and even for a general class of non-abelian ones, but we'll begin with $\mathbf{R}$.

Let $f(t)$ be a complex-valued locally integrable function of one real variable $t \in \mathbf{R}$ (which we refer to as *time*). The *Fourier transform* $\hat f(\tau)$ is defined by the formula

$$\hat f(\tau) := \int_{\mathbf{R}} e^{-2\pi i t\tau} f(t)\ dt.$$

(This is really the best place to put the 2π; listen to the harmonic analysts and representation theorists on this one, and ignore the PDE people, physicists, and engineers!)

The variable τ also lives[1] on $\mathbf{R}$.

The *phase plane* is just the plane $\{(t,\tau) : t \in \mathbf{R}, \tau \in \mathbf{R}\}$ combining both the physical space variable t and the frequency space variable τ. Informally, one can think of the

[1] If one wants to be utterly pedantic, τ should live on the dual space $\mathbf{R}^*$ of $\mathbf{R}$ - e.g. if t has units of *seconds*, then τ should have units of *seconds*$^{-1}$, or Hertz. The quantity $t\tau$ is dimensionless, as it must in order for $\exp(-2\pi i t\tau)$ to make sense.

function f as not necessarily being a function of physical space t, or of frequency space τ, but just as a single object lurking within the phase plane which can be written in terms of one variable or the other. (This is similar to how an element of a vector space can be viewed as a string of numbers in some basis, or a string of numbers in some other basis, or just as an arrow lurking in the vector space). In analogy to the musical notation, we shall non-rigorously shade the areas (t, τ) of the phase plane in which we expect the function f to be concentrated. Such a shading is known as a *phase space portrait* of the function f (musicians would call it a *score* for f, but this notation appears to have not caught on among mathematicians).

As an example, consider the *Gaussian*

$$f(t) := e^{-\pi t^2}.$$

The Fourier transform of the Gaussian is again the Gaussian:

$$\begin{aligned}
\hat{f}(\tau) &= \int e^{-2\pi i t\tau} e^{-\pi t^2}\, dt \\
&= e^{-\pi\tau^2} \int e^{-\pi(t - i\tau)^2}\, dt \\
&= e^{-\pi\tau^2} \int e^{-\pi t^2}\, dt \quad \text{(Shifting the contour)} \\
&= e^{-\pi\tau^2} \quad \text{(Standard integral)}
\end{aligned}$$

This function should have a phase space portrait concentrated around the unit square $[-1, 1] \times [-1, 1]$.

The Fourier transform has many algebraic properties[2] (easy exercises[3]!):

- It is linear: $\widehat{f+g} = \hat{f} + \hat{g}$ and $\widehat{cf} = c\hat{f}$.
- Translations in physical space correspond to modulations in frequency space: if $g(t) := f(t - t_0)$ is the shift of f in physical space by t_0, then $\hat{g}(\tau) = e^{-2\pi i t_0 \tau} \hat{f}(\tau)$.
- Modulations in physical space correspond to translations in frequency space: if $g(t) := e^{2\pi i \tau_0 t} f(t)$, then $\hat{g}(\tau) = \hat{f}(\tau - \tau_0)$.
- Scaling in physical space corresponds to a dual scaling in frequency space: if λ is a non-zero real and $g(t) := f(t/\lambda)$ is the dilation of f by λ, then $\hat{g}(\tau) = \lambda \hat{f}(\tau\lambda)$. Note that the physical space scaling preserves height (L^∞ norm), whereas the frequency space scaling preserves mass (L^1 norm). One can use an alternate scaling $g(t) = \lambda^{-1} f(t/\lambda)$ in which the reverse happens, or one can be more balanced and use an L^2 scaling $g(t) = \lambda^{-1/2} f(t/\lambda)$ which scales both physical and frequency space the same way.
- Conjugation in physical space corresponds to conjugation and reflection in frequency space: if $g(t) := \overline{f(t)}$, then $\hat{g}(\tau) := \overline{g(-\tau)}$.

[2]To begin with we may assume that the functions f, g are Schwartz. Eventually we may use limiting arguments or duality arguments to extend to the case when f, g are (for instance) L^2 functions or even (in many cases) distributions.

[3]The exercises given in the body of the notes are not part of the assigned homework. Only the questions at the end of the notes form the homework.

- Multiplication in physical space corresponds to convolution in frequency space: $\widehat{fg}(\tau) = \hat{f} * \hat{g}(\tau)$, where
$$\hat{f} * \hat{g}(\tau) := \int_{\tau_1+\tau_2=\tau} \hat{f}(\tau_1)\hat{g}(\tau_2).$$
- Convolution in physical space corresponds to multiplication in frequency space: $\widehat{f * g}(\tau) = \hat{f}\hat{g}(\tau)$.

Using the above laws, we can compute the Fourier transform of shifted, modulated, rescaled Gaussians:
$$f(t) := \lambda^{-1/2} \exp(2\pi i \tau_0 t) \exp(-\pi(\frac{t-t_0}{\lambda})^2) \tag{1}$$
$$\hat{f}(\tau) := \lambda^{1/2} e^{2\pi i t_0 \tau_0} e(-2\pi i t_0 \tau) \exp(-\pi((\tau - t_0)\lambda)^2) \tag{2}$$
for $\tau_0 \in \mathbf{R}$, $t_0 \in \mathbf{R}$, $\lambda > 0$. Thus this is a function localized at spatial position t_0 with spatial uncertainty $\Delta t \approx \lambda$, and at frequency position τ_0 with frequency uncertainty $\Delta\tau \approx \lambda^{-1}$. (The ugly phase factor $e^{2\pi i t_0 \tau_0}$ arises as a consequence of translations and modulations not quite commuting properly; this non-commutativity is only a nuisance at present but is the root cause of the *Uncertainty principle*, which is a basic fact of life in time-frequency analysis.).

The above Gaussians are examples of *wave packets*, which is a rather informal term denoting functions which are highly localized in both physical and frequency space (in the sense that $\Delta t \Delta\tau \approx 1$). Their phase space portrait is basically a rectangle with sides parallel to the axes and area 1; such a rectangle is sometimes called a *Heisenberg tile*.

Limiting cases include *Dirac delta at* t_0 $f(t) = \delta(t - t_0)$, for which $\hat{f}(\tau) = e^{-2\pi i t_0 \tau}$, and the *plane wave of frequency* τ_0 $f(t) = \exp^{2\pi i t \tau_0}$, for which $\hat{f}(\tau) = \delta(\tau_0)$. Their phase space portraits are basically a vertical line and horizontal line respectively.

Another easy consequence of the above laws is the *Plancherel identity*
$$\begin{aligned}\langle f, g\rangle &= \int f(t)\overline{g(t)}\ dt\\ &= (\hat{f} * \overline{\hat{g}})(0)\\ &= \langle \hat{f}, \hat{g}\rangle;\end{aligned}$$
in other words, the Fourier transform $f \mapsto \hat{f}$ is an isometry on L^2. (It will in fact to be an invertible isometry, i.e. a unitary transformation).

Applying Plancherel with $f = g$ we obtain the all-important *Parseval's theorem*
$$\|f\|_2 = \|\hat{f}\|_2.$$
This theorem is also known sometimes as Plancherel's theorem.

In particular, the densities $|f(t)|^2\ dt$ and $|\hat{f}(\tau)|^2\ d\tau$ both have the same total mass. One should think of these densities as the marginal distributions of the phase space portrait (which is kind of like a density on the phase plane).

Another application of Plancherel is that if f and g have disjoint spatial supports or disjoint frequency supports, then they are orthogonal. Informally, this implies that if two functions f, g have disjoint phase space portraits, then they should be orthogonal.

The Fourier transform
$$\hat{f}(\tau) = \int e^{-2\pi i t\tau} f(t)\, dt$$
can be inverted via the *inverse Fourier transform*
$$f(t) = \int e^{2\pi i t\tau} \hat{f}(\tau)\, d\tau.$$
To see this, we first observe that this formula is true in the special case of the standard Gaussian $f(t) = e^{-\pi t^2}$ (in which case $\hat{f}(\tau)$ is also the standard Gaussian $\hat{f}(\tau) = e^{-\pi\tau^2}$). Then observe that if the Fourier inversion formula is true for f, then it is true for all translations, modulations, and scalings of f by the above laws. In particular it is true for all functions of the form
$$f(t) = \delta^{-1} e^{-\pi(\frac{t-t_0}{\delta})^2}$$
where $\delta > 0$ and $t_0 \in \mathbf{R}$. Averaging this, we see the inversion formula is true for all functions of the form
$$f(t) = \int g(t_0)\delta^{-1} e^{-\pi(\frac{t-t_0}{\delta})^2}\, dt,$$
i.e. for all functions of the form
$$g * \delta^{-1} e^{-\pi(\frac{\cdot}{\delta})^2}.$$
This is g convolved with an approximation to the identity. By taking the limit in L^2 as $\delta \to 0$ (using Plancherel's theorem) we see that the Fourier inversion formula is true for all f.

In particular, if $g := \hat{f}$, then $\hat{g}(\tau) = f(-\tau)$. On the phase plane, this means that the phase space portrait of g is the phase space portrait of f rotated clockwise by 90 degrees.

Note that all the operations on phase space have preserved area. This is because phase space has a natural symplectic structure (which in two dimensions is just a code word for "area"). Roughly speaking, symplectic geometry is to harmonic analysis (and representation theory) as classical mechanics is to quantum mechanics, but a deeper pursuit of this connection is beyond the scope of this course.

3. The Fourier transform on other groups

There are analogues of the Fourier transform for all locally compact abelian groups G, and even for non-abelian groups provided that they satisfy some nice properties (such as amenability). In the abelian cases the Fourier transform lives on the dual group G^* (the space of characters on G); in the non-abelian case the Fourier transform lives on the space of irreducible representations of G. We won't go into the general theory, but just give some important examples.

For us, the most important example is Euclidean space $G := \mathbf{R}^n$, in which case the dual group is also $G^* = \mathbf{R}^n$, the Fourier transform is

$$\hat{f}(\xi) := \int_{\mathbf{R}^n} e^{-2\pi i (x\cdot\xi)} f(x)\ dx$$

and the inverse is

$$f(x) = \int_{\mathbf{R}^n} e^{2\pi i (x\cdot\xi)} \hat{f}(\xi)\ d\xi.$$

For future reference we record the following scaling property: if L is an invertible linear transformation on $\mathbf{R}^n$ and $g(x) := f(L^{-1}x)$, then $\hat{g}(\xi) = (\det L) g(L^*\xi)$. (Compare this with the one-dimensional scaling property).

Another example is the circle $\mathbf{T} := \mathbf{R}/\mathbf{Z}$, in which case the dual group is the integers $G^* = \mathbf{Z}$, Fourier transform is

$$\hat{f}(n) := \int_{\mathbf{T}} e^{-2\pi i n x} f(x)\ dx$$

and the inverse is

$$f(x) = \sum_{n\in\mathbf{Z}} e^{2\pi i n x} \hat{f}(n).$$

If we take a circle $\mathbf{T}_\lambda := \mathbf{R}/\lambda\mathbf{Z}$ of length λ, the dual group is $\mathbf{Z}/\lambda$, the Fourier transform is

$$\hat{f}(n) := \int_{\mathbf{T}/\lambda} e^{-2\pi i n x} f(x)\ dx$$

and the inverse is

$$f(x) = \frac{1}{\lambda} \sum_{n\in\mathbf{Z}/\lambda} e^{2\pi i n x} \hat{f}(n).$$

Conversely, if G is the integers $\mathbf{Z}$, the dual group is $G^* = \mathbf{T}$, the Fourier transform is given by

$$\hat{f}(\xi) := \sum_n e^{-2\pi i n\xi} f(n)$$

and the inverse is

$$f(n) = \int_{\mathbf{T}} e^{2\pi i n \xi} \hat{f}(\xi)\ d\xi.$$

If G is the finite additive group $\mathbf{Z}_n$ with counting measure, then the dual group is also $\mathbf{Z}_n$ but with normalized counting measure, the Fourier transform is given by

$$\hat{f}(b) := \sum_{a\in\mathbf{Z}_n} e^{-2\pi i ab/n} f(a)$$

and the inverse is

$$f(a) := \frac{1}{n} \sum_{b\in\mathbf{Z}_n} e^{2\pi i ab/n} f(b).$$

One has analogues of the above results for all of these groups, except for the fact that Gaussians do not always have nice counterparts in all groups. In all cases phase space is now the region $G \times G^*$.

One can spend a lot of time deciding the rules for the Fourier transform for a direct sum of two groups, or the quotient of one group by another, etc., but we won't do so here. We will give one example though: if one takes G to be the group $G = \mathbf{Z}_p^n$ - the additive group of p-ary digit strings of length n with no "carry" operation - then the dual group is also $G^* = \mathbf{Z}_p^n$, the Fourier transform is given by

$$\hat{f}(b) := \sum_{a \in \mathbf{Z}_p^n} e^{-2\pi i a \cdot b/p} f(a)$$

and the inverse is

$$f(a) := \frac{1}{p^n} \sum_{b \in \mathbf{Z}_p^n} e^{2\pi i a \cdot b/p} f(b).$$

This example will eventually give rise to the Walsh-Fourier transform.

4. The uncertainty principle

In all the phase space portraits drawn to date, the portrait always occupies a region of area at least 1. Certainly it can occupy a larger region, but the question arises of whether it can occupy a smaller region.

In the finite case, e.g. $G = \mathbf{Z}_5$, there is a clear problem, at least if one reasons heuristically. In this case phase space is a 5×5 square. A "dirac mass" δ_{a,a_0} occupies a vertical line in phase space (5 points), while a "plane wave" $\exp(2\pi i a b_0/5)$ occupies a horizontal line in phase space. One can ask whether smaller regions of space can be occupied, however the space of functions on G is only five-dimensional. If one could find, for instance, a function with phase space support on a single point, then by translation and modulation one could make 25 orthogonal functions on G, a contradiction. From this we expect all functions to occupy at least 5 points of phase space.

In the case $G = \mathbf{R}$ there is a more precise statement, because we have a notion of infinitesimal translation and modulation.

We begin by recalling that translations in physical space correspond to modulations in frequency space. In particular, the Fourier transform of $f(t+\varepsilon)$ is $\exp(2\pi i \varepsilon \tau)\hat{f}(\tau)$. Differentiating this with respect to ε, we see that the Fourier transform of $f'(t)$ is $2\pi i \tau \hat{f}(\tau)$. Thus differentiation in physical space is an infinitesimal translation, and corresponds to multiplication by $2\pi i \tau$ in frequency space, which is an infinitesimal modulation[4]. Similarly multiplication by $-2\pi i t$ in physical space corresponds to differentiation in frequency space.

Define the operators D and X by

$$Xf(t) := tf(t); \quad Df(t) = \frac{1}{2\pi i} f'(t).$$

[4]This fact is the principal reason why the Fourier transform plays such an important role in the theory of differential equations

From the preceding discussion we have

$$\widehat{Df}(\tau) = \tau \hat{f}(\tau).$$

In other words, X multiplies by the physical variable, while D multiplies by the frequency variable. The quantity

$$\frac{\|Xf\|_2}{\|f\|_2} = (\frac{\int t^2 |f(t)|^2 \; dt}{\int |f(t)|^2 \; dt})^{1/2}$$

is a measure of the average value of $|t|$, or in other words the average deviation of the physical variable t from the origin. Similarly

$$\frac{\|Df\|_2}{\|f\|_2} = (\frac{\int \tau^2 |f(\tau)|^2 \; d\tau}{\int |f(\tau)|^2 \; d\tau})^{1/2}$$

measures the average deviation of the frequency variable τ from the origin.

Note that X and D are both self-adjoint operators.

If it were possible for a non-zero function f to have a phase space portrait at the origin $(0, 0)$, then the above considerations would give $Xf = 0$ and $Df = 0$, thus f is a joint eigenvector of X and D. However X and D do not commute:

$$[D, X]f = DXf - XDf = \frac{1}{2\pi i} f$$

which would give a contradiction. A more quantitative version is

Proposition 4.1 (Heisenberg uncertainty principle)**.** *We have*

$$\frac{\|Xf\|_2}{\|f\|_2} \frac{\|Df\|_2}{\|f\|_2} \geq \frac{1}{4\pi}.$$

Proof We consider the quantity $\|(aX + ibD)f\|_2^2$, where a, b are real numbers to be chosen later. Clearly this quantity is non-negative. On the other hand, we have

$$\begin{aligned}
\|(aX + ibD)f\|_2^2 &= \langle (aX + ibD)f, (aX + ibD)f \rangle \\
&= \langle (aX - ibD)(aX + ibD)f, f \rangle \\
&= \langle a^2 X^2 + b^2 D^2 + abi(XD - DX)f, f \rangle \\
&= a^2 \langle Xf, Xf \rangle + b^2 \langle Df, Df \rangle - \langle \frac{ab}{2\pi} f, f \rangle \\
&= a^2 \|Xf\|_2^2 + b^2 \|Df\|_2^2 - \frac{ab}{2\pi} \|f\|_2^2.
\end{aligned}$$

We therefore have

$$a^2 \|Xf\|_2^2 + b^2 \|Df\|_2^2 \geq \frac{ab}{2\pi} \|f\|_2^2.$$

Now if we pick $a := \|Df\|_2$ and $b := \|Xf\|_2$, we thus have

$$2\|Xf\|_2^2 \|Df\|_2^2 \geq \frac{\|Df\|_2 \|Xf\|_2}{2\pi} \|f\|_2^2$$

and the claim follows. ■

By modulating f by a plane wave we can generalize the above proposition to

$$\frac{\|Xf\|_2}{\|f\|_2}\frac{\|(D-\tau_0)f\|_2}{\|f\|_2} \geq \frac{1}{4\pi}$$

for any $\tau_0 \in \mathbf{R}$, and then by translating f we can generalize further to

$$\frac{\|(X-t_0)f\|_2}{\|f\|_2}\frac{\|(D-\tau_0)f\|_2}{\|f\|_2} \geq \frac{1}{4\pi}. \tag{3}$$

If one sets t_0 equal to the mean position

$$t_0 := \frac{\langle Xf, f\rangle}{\langle f, f\rangle} = \frac{\int |f(t)|^2 t\ dt}{\int |f(t)|^2\ dt} \tag{4}$$

and τ_0 equal to the mean frequency

$$\tau_0 := \frac{\langle Df, f\rangle}{\langle f, f\rangle} = \frac{\int |\hat{f}(\tau)|^2 \tau\ d\tau}{\int |f(\tau)|^2\ d\tau} \tag{5}$$

then $\frac{\|(X-t_0)f\|_2}{\|f\|_2}$ can be viewed as the standard deviation of the position variable t (using the density $|f(t)|^2\ dt$ and similarly $\frac{\|(D-\tau_0)f\|_2}{\|f\|_2}$ is the standard deviation of the frequency variable τ (using $|\hat{f}(\tau)|^2\ d\tau$). (Exercise: show that this choice of t_0, τ_0 minimizes these two expressions, and is thus in some sense the optimal way to use (3)). This can be expressed more succinctly as

$$(\Delta t)(\Delta \tau) \geq \frac{1}{4\pi}.$$

We will not use the Heisenberg uncertainty principle much in its most precise form, and will instead use it as a heuristic principle $(\Delta t)(\Delta\tau) \gtrsim 1$ to guide us. The principle places a limit as to how much we may meaningfully localize both physical space and frequency space simultaneously. (Of course, one can always define some sort of operator which claims to localize both further, but such an object is unlikely to have much application).

5. Phase space localization

A fundamental notion in analysis is that of *localization* - taking a function and restricting it to a small region in physical space, or frequency space, or more exotic spaces (spectra, phase space, etc.). One can then recover the original function from the localized pieces by using a partition of unity. There are similar notions in algebra - for instance, the concept of localizing to a single characteristic p - but the concept of localization in analysis is special in that it is "fuzzy" - there is no canonical "best" way to localize, but instead one has a wide freedom in exactly what recipe to use. (This fuzzy flavor seems to be a distinguishing trait of analysis compared against other branches of mathematics).

Physical space localization is the most familiar. If one wants to localize a function $f(t)$ to an interval I in physical space, the easiest thing to do is to apply the rough cutoff function χ_I, thus yielding the restriction $f(t)\chi_I(t)$ to I. Or, one can use a bump function adapted to I, e.g. a C_0^∞ function $\tilde{\chi}_I(t)$ supported on I which equals 1 on the inner third of I and obeys the derivative bounds[5]

$$|\tilde{\chi}_I^{(j)}(t)| \lesssim |I|^{-j}$$

for all $j \geq 0$, with the implicit constant depending on j, and with $|I|$ denoting the length of I. The cutoff is then given by $f(t)\tilde{\chi}_I(t)$. In many cases the smooth cutoff is better than the rough cutoff (for instance, it preserves differentiability) although the exact choice of cutoff function is almost never important. (Occasionally one cares about the total mass of $\tilde{\chi}_I$ or some moment or symmetry conditions, but there is always a large amount of freedom left in how to choose $\tilde{\chi}_I$).

Equally important is the notion of *frequency space localization* (which engineers call band pass filters). Given an interval I in frequency space, one can apply a rough cutoff π_I defined by

$$\widehat{\pi_I f}(\tau) := \chi_I \hat{f}(\tau)$$

(i.e. one takes the Fourier transform, cuts off to I, and then undoes the Fourier transform) and or perhaps a smooth cutoff $\tilde{\pi}_I$ defined by

$$\widehat{\tilde{\pi}_I f}(\tau) := \tilde{\chi}_I \hat{f}(\tau).$$

These operators are examples of *Fourier multipliers.* A Fourier multiplier T is an operator defined by a formula of the form

$$\widehat{Tf}(\tau) := m(\tau)\hat{f}(\tau)$$

where m is some fixed function, known as the *symbol* of the multiplier. (A *spatial multiplier* is given by the more mundane formula $Tf(t) := m(t)f(t)$). Using our basic facts about the Fourier transform, we can write T as a convolution operator

$$Tf(t) = f * K = \int f(t-s)K(s)\, ds$$

where $K := \check{m}$ is the inverse Fourier transform of m:

$$K(s) = \int e^{2\pi i s\tau} m(\tau)\, d\tau.$$

The identity operator is of course a Fourier multiplier with symbol 1. The differentiation operator $\frac{d}{dt}$is a multiplier with symbol $2\pi i\tau$. More generally, any differentiation operator $P(\frac{d}{dt})$, with P a polynomial, is a multiplier with symbol $P(2\pi i\tau)$. (One can therefore think of the cutoff π_I as $\chi_I(\frac{1}{2\pi i}\frac{d}{dt})$, but this is rarely a useful perspective).

It is clear what physical space cutoffs do in physical space, and what frequency space cutoffs do in frequency space. Now let's look at what frequency cutoffs do in physical space (the corresponding situations for physical cutoffs in frequency space is, of course, virtually identical).

[5]Throughout the course we use $A \lesssim B$ to denote the estimate $A \leq CB$ for some constant C which only depends on unimportant quantities - in this case, j. The derivative bounds here are natural - basically, these are the best possible bounds one can place on such a $\tilde{\chi}$ without being inconsistent with $\tilde{\chi}$ being supported on I and equal to 1 somewhere. This can be seen by Taylor's theorem with remainder.

For sake of argument let I be the interval $[-1,1]$. The rough cutoff π_I is then a Fourier multiplier with the kernel given by the sinc function:

$$K(s) = \int e^{2\pi i s\tau} \chi_{[-1,1]}(\tau)\ d\tau = \frac{\sin(2\pi s)}{\pi s}.$$

This kernel (also known as the *Dirichlet kernel*) decays as $s \to \pm\infty$, but only very slowly. Because of this, the operator π_I tends to spread things around in physical space quite a bit. For instance, if f is an integrable function supported on a set Ω, then $\pi_I(f)$ will obey bounds of the form

$$|\pi_I(f)(x)| \lesssim \|f\|_1 \mathrm{dist}(x,\Omega)^{-1}$$

whenever $\mathrm{dist}(x,\Omega) \gtrsim 1$ (why?). Thus, even though f vanishes outside Ω, $\pi_I(f)$ does not quite vanish outside Ω, although it does decay as one moves further away from Ω. This illustrates the annoying (but usually controllable) phenomenon of *leakage*: a localization in Fourier space inevitably causes a loss of localization in physical space, and vice versa.

Having seen what the rough cutoff did, now let's try the smooth cutoff $\tilde\pi_I$. The kernel is now

$$K(s) = \int e^{2\pi i s\tau} \tilde\chi_{[-1,1]}(\tau)\ d\tau.$$

This kernel cannot be worked out explicitly because we have not specified $\tilde\chi_{[-1,1]}$ explicitly, but we can still get very good estimates on what K looks like. First of all, since $\tilde\chi_{[-1,1]}$ is bounded and supported on $[-1,1]$, we can obtain the bound $|K(s)| \lesssim 1$. When $|s| \lesssim 1$ this is pretty much the best bound available, but for $|s| \gg 1$ we can do a lot better by exploiting the rapid oscillation[6] of the phase $e^{2\pi i s\tau}$. Indeed, by integration by parts we have

$$K(s) = -\int \frac{1}{2\pi i s} e^{2\pi i s\tau} \frac{d}{d\tau} \tilde\chi_{[-1,1]}(\tau)\ d\tau$$

and more generally

$$K(s) = \int (\frac{-1}{2\pi i s})^j e^{2\pi i s\tau} (\frac{d}{d\tau})^j \tilde\chi_{[-1,1]}(\tau)\ d\tau$$

for all $j > 0$. Since the smooth cutoff $\tilde\chi_{[-1,1]}$ has all derivatives bounded, we thus have

$$|K(s)| \lesssim |s|^{-j}$$

for all j (with the implicit constant depending on j). In other words, K is *rapidly decreasing*. This implies, for instance, that if f is an integrable function supported on Ω as before, then

$$|\tilde\pi_{[-1,1]}(f)(x)| \lesssim \|f\|_1 \mathrm{dist}(x,\Omega)^{-j}$$

for all j and $\mathrm{dist}(x,\Omega) \gtrsim 1$. Informally, $\tilde\pi_{[-1,1]}$ spreads f by a distance of about $O(1)$, plus a rapidly decreasing "tail". This tail is of course a nuisance, but in most applications[7] it causes no real damage (beyond the damage already caused by the

[6]This is a simple example of the *principle of non-stationary phase*: the more rapidly varying a phase in an integrand, the smaller the integral gets. The theory of stationary phase is a topic in itself but we do not have time to discuss it in detail here.

[7]At least, in applications where one does not care about the exact constants which occur in bounds. Many applications in analysis, especially local analysis, are of this form.

initial spreading by $O(1)$). This is analogous to the fact that a geometric series like $\sum_{n=0}^{\infty} 2^{-n}$ is not too much larger than just the initial term of 1.

Clearly the smooth cutoff is better behaved than the rough cutoff, and it is almost always preferable to use the smooth one.

One may hope to eliminate the tail by choosing $\tilde{\chi}_{[-1,1]}$ so that it's inverse Fourier transform $K(s)$ is compactly supported, but unfortunately this is impossible:

Lemma 5.1. *There does not exist a non-zero integrable function f such that both f and $\hat{f}$ are compactly supported.*

Proof If f has compact support, then the function

$$\hat{f}(\tau) = \int e^{-2\pi i t\tau} f(t)\, dt$$

extends to an entire function

$$\hat{f}(z) = \int e^{-2\pi i t z} f(t)\, dt.$$

But the zero set of an entire function must be discrete, contradicting the compact support of $\hat{f}(\tau)$. ∎

This no-go lemma has been the bane of many a harmonic analyst. However, there is one special species of Fourier analysis in which Lemma 5.1 is not an issue, namely the *Walsh-Fourier* case when the abelian group is $G = \mathbf{Z}_2^{\mathbf{Z}}$ (or more generally $\mathbf{Z}_p^{\mathbf{Z}}$ for some prime p) rather than $\mathbf{R}$. In other words the group consists of doubly infinite 2-adic strings instead of $\mathbf{R}$ (the main difference being the lack of a carry operation). We will discuss this model in a later week.

For more general intervals I, the frequency space localization operator $\tilde{\pi}_I$ spreads things in physical space by about $O(|I|^{-1})$, plus a tail. Note that this is consistent with the Uncertainty principle.

So far we have only looked at the costs of localization. Now we look at the benefits of localization. One benefit is orthogonality. If one localizes a function in physical space to two disjoint intervals I and J, then the resulting localized functions are orthogonal (regardless of whether one uses the rough or smooth cutoff). Similarly for localization in frequency space.

Another benefit has to do with the operators X and D mentioned earlier. Physical space localization has the effect of localizing the operator X to a constant: if f has physical space support in an interval $[t_0 - r, t_0 + r]$, then

$$Xf = (t_0 + O(r))f.$$

Thus, if r is small compared to t_0, we have

$$Xf \approx t_0 f.$$

Similarly, if f has frequency space support[8] in an interval $[\tau_0 - s, \tau_0 + s]$ with $s \ll |\tau_0|$, then we heuristically have

$$Df \approx \tau_0 f$$

(where we do not make the notion $\approx$ rigorous). Thus frequency space localization becomes very useful when studying expressions involving derivatives (which, of course, happens all the time in applications).

A third benefit is another spin off of the uncertainty principle - a localization in frequency to a frequency interval I destroys all the physical space information at scales $|I|^{-1}$ and lower. To make this clearer we begin again with the example of $I = [-1, 1]$.

An algebraic manifestation of this principle is the well-known

Lemma 5.2 (Shannon-Nyquist sampling formula). *If f has Fourier support on $[-1, 1]$, then*

$$f(x) := \sum_{n \in \mathbf{Z}/2} f(n) \frac{\sin(2\pi(x - n))}{2\pi(x - n)}.$$

In particular, the function f can be reconstructed from its values on the half-integers $\mathbf{Z}/2$.

Proof We will not need the Shannon-Nyquist sampling formula anywhere in this course, and so we leave this as an exercise. (It is an easy consequence of the Poisson summation formula, for instance). ■

The Shannon-Nyquist sampling theorem suggests that a function with Fourier support near the origin is more like a function on $\mathbf{Z}$ than on $\mathbf{R}$ (but blurred out by a spatial uncertainty of $O(1)$).

For our purposes we will prefer a version of the principle with quantitative estimates. Like many principles in analysis, there is no definitive statement to this principle; the instance below is just a typical example. (The *proof*, however, is ubiquitous).

Lemma 5.3. *Let f be a function with frequency space support in $[-1, 1]$ (e.g. f might be the smooth or rough frequency localization of some other function to $[-1, 1]$). Then f contains no sudden spikes (compared to nearby values):*

$$|f(t)| \lesssim \int |f(t - s)| \frac{ds}{(1 + |s|)^j} \tag{6}$$

for all $j > 0$. Indeed, we have

$$|f^{(k)}(t)| \lesssim \int |f(t - s)| \frac{ds}{(1 + |s|)^j}$$

for all $j, k > 0$, with the implicit constant depending on j, k.

[8] We shall sometimes abbreviate "physical space support" as just "support", and "frequency space support" as "Fourier support".

Proof Let $\tilde{\pi}_{[-2,2]}$ be a frequency localization operator whose symbol $\tilde{\chi}_{[-2,2]}$ equals 1 on $[-1,1]$. Then

$$\hat{f} = \tilde{\chi}_{[-2,2]}\hat{f}$$

and hence

$$f = \tilde{\pi}_{[-2,2]}f,$$

so that

$$f(t) = \int f(s)K(t-s)\ ds$$

where K is the inverse Fourier transform of $\tilde{\chi}_{[-2,2]}$. Since $\tilde{\chi}_{[-2,2]}$ is a Schwartz function, so is K. So

$$f^{(k)}(t) = \int f(s)K^{(k)}(t-s)\ ds = \int f(t-s)K^{(k)}(s)\ ds.$$

The claim then follows by taking absolute values and using the fact that K is Schwartz. ■

Another way of stating the principle is that if f has frequency space support in $[-1,1]$, then one has $|f(x)| \sim |f(y)|$ with large probability if $|x-y| \lesssim 1$. More succinctly, f is essentially constant at unit scales. For larger scales of separation there are basically no constraints between $f(x)$ and $f(y)$; for instance take

$$f(x) = \sum_n c_n K(x-n)$$

where K is the Fourier transform of a bump function adapted to $[-1,1]$ and c_n is some arbitrary sequence of scalars. The function f is then essentially c_n on the interval $[n-1/2, n+1/2]$ and so $f(x)$ and $f(y)$ are essentially unrelated for $|x-y| \gg 1$.

More generally, if f has frequency support on an interval I centered at the origin, then f is essentially constant at scales $O(|I|^{-1})$. If I is centered instead at some other frequency τ_0, then $e^{-2\pi it\tau_0}f(t)$ is essentially constant at scales $O(|I|^{-1})$, so that f behaves like a constant multiple of the plane wave $e^{2\pi it\tau_0}$ at scales $|I|^{-1}$. (This principle is also called the Uncertainty principle, and in a certain sense it is much stronger than the relationship $\Delta t\Delta\tau \gtrsim 1$. It generalizes to higher dimensions; for instance, if we are in $\mathbf{R}^2$ and the frequency space variable of f is localized to an $a \times b$ rectangle, then f is essentially constant on rectangles of size $\ll a^{-1} \times b^{-1}$.).

A typical consequence of the uncertainty principle is

Lemma 5.4 (Bernstein's inequality). *If $1 \leq p \leq q \leq \infty$ and f has frequency support in an interval I, then*

$$\|f\|_q \lesssim |I|^{\frac{1}{q}-\frac{1}{p}}\|f\|_p$$

In other words, once a function is frequency localized, then lower L^p norms control higher L^q norms. (This is analogous to the fact that on a discrete space such as the integers, lower l^p norms control higher l^q norms). The factor $|I|^{\frac{1}{q}-\frac{1}{p}}$ is sharp; to see this, think about a function f with frequency support on I and spatially localized to an interval of size $|I|^{-1}$. (In a certain sense, this is the only case in which Bernstein's inequality is sharp; if f is spread out on more than one $|I|^{-1}$-interval then one can obtain a gain. This principle is often useful in more delicate estimates).

Proof The claim is clearly true when $p = q$. By interpolation it suffices to show the case $p = 1$, $q = \infty$. By scaling and shifting I we may assume that $I = [-1, 1]$. The claim then becomes

$$\|f\|_\infty \lesssim \|f\|_1$$

which follows from (6). ■

On the phase plane, a physical space localization restricts the phase space portrait to a vertical strip, while a frequency space localization to a horizontal strip. If one wants to localize a function to a rectangle in phase space, it thus seems natural to compose a frequency space and a physical space localization. The catch is that a frequency space localization spreads things out in physical space and vice versa, so no matter how one composes operators together, there is going to be some leakage either in physical space or in frequency space. (Indeed, Lemma 5.1 tells us that this is inevitable). Furthermore, the uncertainty principle tells us we cannot localize to a rectangle of area $\ll 1$. On the other hand, we can localize to any larger rectangle:

Proposition 5.5. *Let $I \times J$ be a rectangle in the phase plane with area $\gtrsim 1$. Then for any $f \in L^2$ (say), the function $F := \tilde{\pi}_J(\tilde{\chi}_I f)$ has Fourier support in J and is localized in J in the sense that*

$$|F(x)| \lesssim \|f\|_2 |I|^{1/2} (|I| \mathrm{dist}(x, J))^{-j}$$

whenever $j \geq 0$ and $\mathrm{dist}(x, J) \gtrsim |I|^{-1}$, with the implicit constant depending on j.

Proof We can rescale so that $I = [-1, 1]$ (exercise!). We then have

$$F(x) := \int \tilde{\chi}_I(x-y) f(x-y) K(y)\, dy$$

where K is the inverse Fourier transform of $\tilde{\chi}_{[-1,1]}$. By Cauchy-Schwarz we thus have

$$|F(x)| \leq \|f\|_2 \|\tilde{\chi}_I(\cdot - y) K(y)\, dy\|_2.$$

Since K is rapidly decreasing, the claim then follows from a routine computation. ■

In the above proposition we localized in space first and then in frequency. If one did the reverse then one would then have perfect space localization and imperfect frequency localization instead. Other variants are possible but one cannot perfectly localize in both space and frequency simultaneously.

Suppose we have two disjoint rectangles $I \times J$ and $I' \times J'$ in phase space, and we localize a function f to both rectangles. Since the rectangles are disjoint, they are either separated in physical space (I and I' disjoint), separated in frequency space (J and J' disjoint), or both. In either case we expect the localizations to be orthogonal (although we cannot e demand perfect orthogonality in both directions because we must accept a leakage somewhere by Lemma 5.1).

These heuristics suggest that we may be able to decompose $L^2(\mathbf{R})$ into almost orthogonal subspaces by partitioning the phase plane into disjoint rectangles and applying phase space localizations to each such rectangle, although in order to respect the uncertainty principle each rectangle needs to have area at least 1.

If the rectangles have area exactly one (so the rectangle is a Heisenberg tile), then we expect the associated spaces to be one-dimensional. To motivate this let us return to the finite group case, e.g. $G = \mathbf{Z}_n$. In this case the uncertainty principle asserts that one can only hope to localize to those rectangles of area at least n. If one then partitions the $n \times n$ phase plane into n rectangles of area n each, and assumes perfect orthogonality for the sake of argument, then the portion of $l^2(\mathbf{Z}^n)$ corresponding to each such space can only have one dimension each, since the full space $l^2(\mathbf{Z}_n)$ is only n-dimensional. Another motivation comes from the uncertainty principle: if f was more or less localized to the Heisenberg tile $I \times J$ centered at the origin (say) then the frequency localization at J forces f to be essentially constant at scales $|I|$, while the spatial localization at I forces $\hat{f}$ to be essentially constant at scales $|J|$. This suggests that the phase space portrait of f is essentially a constant on the tile $I \times J$ and very small or zero elsewhere.

There are many ways to carve up the phase plane into appropriate pieces. One of the most useful is the *Littlewood-Paley decomposition*, in which the phase plane is carved up into semi-infinite rectangles

$$\{\mathbf{R} \times \pm[2^j, 2^{j+1}] : \pm \in \{+,-\}, j \in \mathbf{Z}\}$$

so that the frequency variable is chopped up into dyadic intervals around the origin, while the spatial variable is untouched. This decomposition is basic in understanding anything involving derivatives (e.g. Sobolev spaces, Hölder spaces, multipliers, differential operators, pseudo-differential operators).

Of course these rectangles are not pushing anywhere near the Heisenberg limit, as their area is infinite rather than 1. One can refine the Littlewood-Paley decomposition further to a *wavelet decomposition* or something similar, in which the Littlewood-Paley rectangles are subdivided further into Heisenberg tiles

$$\{[k2^{-j}, (k+1)2^{-j}] \times \pm[2^j, 2^{j+1}] : \pm \in \{+,-\}, j, k \in \mathbf{Z}\}$$

This decomposition is basically an enhanced Littlewood-Paley decomposition which can deal with localized spatial phenomena (such as boundary conditions) without too much trouble.

Another partition is the *Gabor partition*, in which we just take unit squares

$$\{[j, j+1] \times [k, k+1] : j, k \in \mathbf{Z}\}.$$

These squares correspond to standard Gaussians which are translated and modulated but not scaled. This partition comes up in the Gabor transform, or in such devices as the windowed Fourier transform (also known as the local cosine transform). This transformation is generally useful when the unit scale plays a prominent and distinguished role.

For completeness, we give two limiting partitions, the *physical space partition*

$$\{t \times \mathbf{R} : t \in \mathbf{R}\},$$

which basically ignores the frequency variable and analyzes a function f in terms of its physical space representation $f(t)$, or the *frequency space partition*

$$\{\mathbf{R} \times \tau : \tau \in \mathbf{R}\}$$

which is the reverse.

Next week we shall focus on the Littlewood-Paley decomposition and its applications to Sobolev spaces.

6. Homework for week 1 - due Friday, January 19

- Show that if f is equal to a modulated Gaussian (1), then the Heisenberg uncertainty inequality (3) is obeyed with equality. Conversely, if f is a non-zero Schwartz function which obeys (3) with identity for some t_0, τ_0, then f is given by (1) for some λ.
- State and prove a statement which has the flavour of "the Fourier transform on the circle $\mathbf{R}/\lambda\mathbf{Z}$ approaches the Fourier transform on the circle $\mathbf{R}$ as $\lambda \to \infty$". Give an informal description of what is happening to the phase space portraits as $\lambda \to \infty$. (Yes, this is an extremely vague question, especially given that we have not formally defined phase space portraits. Be creative.)

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: `tao@math.ucla.edu`