SOME TRANSFORMS IN FUNCTIONAL ANALYSIS

VIPUL NAIK

Abstract. This article describes some of the ideas and concerns that one needs to keep in mind when performing transforms in functional analysis.

1. The what and why of transforms

1.1. **Function spaces, linear operators and transforms.** A **function space**(defined) is a normed vector space (over $\mathbb{R}$ or $\mathbb{C}$) that arises as *functions* on some set. In other words, it is a normed vector space that, *as a vector space*, is a subquotient of the space of all functions (to $\mathbb{R}$ or to $\mathbb{C}$) on some set.

Some points deserve mention:

- The *subspace* part is because not all functions are being considered.
- The *quotient space* part is because functions may be considered upto some equivalence. For instance, if X is a measure space, we may be interested in the space of functions upto the equivalence relation of being equal "almost everywhere". In fact, we shall adopt the convention that when the sets in question are measure spaces, our "functions" will be functions upto almost everywhere equivalence, and moreover, we shall assume that the functions are finite almost everywhere (so they live in a vector space). When we are interested in functions honestly rather than upto equivalence, we shall use the term "specific function".
- The norm, however, could be defined in a different way for different function spaces on the same set (even on the same set with the same measure).

A **transform**(defined) is an *important* linear operator from one function space to another (possibly, the same function space). The key conditions that transforms must satisfy is *linearity*. Further things like continuity, boundedness, closedness may or may not be satisfied.

The *properties* of a transform are the way it converts operations within one function space, to operations within the other function space. Here, we need to be very careful, because some of the properties that are preserved involve operations that may not always be well-defined on either side.[1]

1.2. **Transforms as formulae.** There is a key distinction between a *formula* and the *function* it defines. A formula is a string of symbols, and can be reinterpreted in many different domains, whereas a function has a specific domain and a specific range. For instance, the formula:

$$x \mapsto x^2$$

can be viewed as giving a function from $\mathbb{Z}$ to $\mathbb{Z}$, $\mathbb{R}$ to $\mathbb{R}$, $\mathbb{C}$ to $\mathbb{C}$; it can be viewed as an operator on a function space, it can be viewed as a map in any Abelian group, any group, or any monoid.

The advantage of working around with formulae at a formal level is that we do not have to keep track constantly of issues like where things live. The disadvantage is that when we want to reinterpret the formal manipulation in terms of actual functions, we need to make sure that the "world" where we interpret the manipulations, actually permits those manipulations.

The importance of this is that a transform is often viewed primarily as a "formula" and secondarily as a linear operator. Unfortunately, I am not aware of an abstract formalism to manipulate these formulae so the manipulation has to be part-formal, and part-real-world.[2] Thus (as far as I can see) the formal symbol-pushing that one does at the formula level cannot completely forget the specific context.

[1] When I say we "need" to be careful, I'm speaking from the viewpoint of a person who does not have the intuition that makes it possible to dispense with rigour

[2] This is in contrast with some of the "purely algebraic" formulae that can be interpreted in an abstract formalism of rings or fields; or the "pure integrals" that can be interpreted over an arbitrary measure space.

1.3. **Notation for transforms.** The transform is a linear operator that takes as input a function, and for notational convenience, we distinguish function inputs to transforms from value inputs to functions. Function inputs to transforms are encoded in square brackets, or (as for the Fourier transform) the transform symbol is put as a superscript or subscript over the function input.

2. Transforming by a kernel

2.1. **Definition of transform by a kernel.** Suppose X and Y are measure spaces, and $K(x,y)$ is a function (upto measure zero) from $X \times Y$ to $\mathbb{C}$. Then, the **transform with kernel**(defined) or an **integral operator**(defined) with kernel K is the map:

$$f \mapsto \left(x \mapsto \int_Y K(x,y) f(y)\, dy \right)$$

This is a "formula"; the concrete meaning of this is if we specify where f lives. Since $f(y)$ appears in the formula, f should be a function (upto equivalence) from Y to $\mathbb{C}$, and for the formula to make sense, we need that for every x, the function:

$$y \mapsto K(x,y)f(y)$$

is in $L^1(Y)$. If instead we insist on the above condition for *almost every* x, then we get the transform defined almost everywhere. Thus, the transform is well-defined from functions upto equivalence to functions upto equivalence, and it also makes sense pointwise, at almost all points.

By the distributivity law, and the fact that $L^1(Y)$ is a vector space, we see that the set of fs for which the above is in L^1, is a vector space. Also, the range of the transform is a vector space, and in general one cannot comment on whether the transform will be injective.

To say more about the transform, we need to say more about K. The first reasonable thing to hope is that K itself is measurable with respect to the product measure on $X \times Y$, so that we can be sure that whatever is the output of the transform, is a measurable function. But we would ideally like more: we'd want to say that boundedness properties of the kernel can guarantee that "good inputs yield good outputs".

2.2. **Schur's lemma.** Schur's lemma gives sufficient conditions on the kernel of the transform, for the transform itself to be well-defined as a map from $L^2(Y)$ to $L^2(X)$. At first, L^2 might seem a strange choice, since integrability is a L^1-subject. What's really happening is that the kernel has such strong decay properties that multiplying it with a L^2-function yields a L^1-function.

Let's get a little more of the intuition behind this. It *is* true that L^2 is not contained in L^1 for infinite measure spaces, but all the L^ps are contained in the space L^1_{loc} of those functions that are L^1 when restricted to subsets of finite measure (this isn't strictly L^1_{loc}, but is close enough). The point is that the non-containment of L^2 inside L^1 is a phenomenon that happens at infinity, and if the kernel decays sufficiently fast, it would at least be possible to integrate L^2 functions against it.

What Schur's lemma tells us is more; it guarantees that under certain conditions the output is in L^2, and in fact guarantees that its L^2-norm is not more than a certain amount.

Theorem 1 (Schur's lemma)**.** Suppose $K : X \times Y \to \mathbb{C}$ satisfies the following properties:

$$\int_Y |K(x,y)|\, dy \leq C \text{ a.e. for } x \in X$$
$$\int_X |K(x,y)|\, dx \leq C \text{ a.e. for } y \in Y$$

Then the transform by K, namely:

$$f \mapsto \left(x \mapsto \int_Y K(x,y) f(y)\, dy \right)$$

defines a bounded linear operator from $L^2(Y)$ to $L^2(X)$ whose operator norm is bounded by C.

The proof is illustrative, so we give it here:

Proof. Explicitly, we need to show that:

$$\int_X \left| \left(\int_Y K(x,y) f(y)\, dy \right) \right|^2 dx \leq C^2 \|f\|_2^2$$

If one directly tries to apply Cauchy-Schwarz, one ends up with a piece for which one has no guarantee of integrability.

The trick is to write $|K(x,y)|$ as $\sqrt{|K(x,y)|}\sqrt{|K(x,y)|}$ and then use the Cauchy-Schwarz inequality on the functions $y \mapsto \sqrt{|K(x,y)|}$ and $y \mapsto \sqrt{|K(x,y)|}\,|f(y)|$. We get:

$$\int_X \left| \left(\int_Y K(x,y) f(y)\, dy \right) \right|^2 dx \leq \int_X \left(\int_Y |K(x,y)|\,|f(y)|\, dy \right)^2 dx$$

Using Cauchy-Schwarz as described above:

$$\implies \int_X \left| \left(\int_Y K(x,y) f(y)\, dy \right) \right|^2 dx \leq \int_X \left(\int_Y |K(x,y)|\, dy \right) \left(\int_Y |K(x,y)|\,|f(y)|^2\, dy \right) dx$$

Bounding first integral:

$$\implies \int_X \left| \left(\int_Y K(x,y) f(y)\, dy \right) \right|^2 dx \leq \int_X C \left(\int_Y |K(x,y)|\,|f(y)|^2\, dy \right) dx$$

Using Fubini's theorem:

$$\implies \int_X \left| \left(\int_Y K(x,y) f(y)\, dy \right) \right|^2 dx \leq C \int_Y |f(y)|^2 \left(\int_X |K(x,y)|\, dx \right) dy$$

Bounding inner integral: $$\implies \int_X \left| \left(\int_Y K(x,y) f(y)\, dy \right) \right|^2 dx \leq C^2 \int_Y |f(y)|^2\, dy$$

That's exactly the right side we want. □

A few comments are in order here. A transform involving a kernel involves a single integral, but proving integrability properties of the output function inevitably gets us into a double integral. Thus, the kernel function must satisfy good integrability properties with respect to *both* variables, rather than one. Also, Fubini's theorem, as well as some tricky manipulations, are typically needed to simplify the expressions.

Note that Fubini's theorem is justified here because all functions on the right side are taking nonnegative real values.

Finally, since the hypotheses of the lemma have "almost everywhere" conditions, the "function" that we get after applying the transform need not be finite everywhere. In general, one could expect it to blow up at those points x where the integral is infinite. However, since the output function is in L^2, it is infinite only on a set of measure zero.

Thus, if K satisfies the conditions for Schur's lemma, the transform works "honestly": if we are given f as a specific function (as opposed to just an equivalence class) we can get the transform of f as a specific function defined at points (though it may be ∞ at some points). This may seem unremarkable to those people who haven't seen the Fourier transform and the shady way it is defined on L^2.

The advantage of proving Schur's lemma is that it's a "once-and-for-all" proof; now, if we are somehow able to establish the desired properties for our kernel function, we see that it defines a bounded operator.

2.3. **A dual spaces result.** We now state another result, using the fact that L^p defines bounded linear functionals on L^q where p and q are Holder conjugates. What we're stating is more like a sequence of observations than a particular result.

The setup is same as before: X and Y are measure spaces, $K : X \times Y \to \mathbb{C}$ is a kernel function, and we are trying to use K to define a transform by the same formula:

$$f \mapsto \left(x \mapsto \int_Y K(x,y) f(y)\, dy \right)$$

Here are some observations:

- If, for every x, the function $y \mapsto K(x,y)$ is in L^q, then the transform is well-defined on L^p. This is based on the observation made earlier that for the transform to be well-defined, we need that for every x, the map $y \mapsto K(x,y)f(y)$ is in L^1.

- $L^p \to L^\infty$ condition: If, moreover, there is a *uniform* bound (say C) on the L^q-norms of the map $y \mapsto K(x,y)$, then it defines a bounded linear operator from L^p to L^∞, with operator norm at most C. This is because for each $x \in X$, Holder's inequality yields that the value at x is at most C times the L^p-norm of the original function.
- $L^1 \to L^\infty$ condition: In particular, if $K : X \times Y \to \mathbb{C}$ is a bounded function (i.e. $K \in L^\infty(X \times Y)$) then K defines a bounded linear operator from $L^1(Y)$ to $L^\infty(X)$. (this uses the fact that ∞ is the Holder conjugate of 1).
- $L^\infty \to L^\infty$ condition: Similarly, if for $K : X \times Y \to \mathbb{C}$, the function $y \mapsto K(x,y)$ is in L^1 for every x, and there is a uniform bound on the L^1 norms, K gives a well-defined map from L^∞ to L^∞.
- $L^p \to L^1$ condition: For this, we require the map $y \mapsto K(x,y)$ to be in L^q for every x, and the function $x \mapsto \|y \mapsto K(x,y)\|_q$ to be a L^1-function.
- $L^\infty \to L^1$ condition: Putting $p = 1$ and using Fubini's yields that $K \in L^1(X \times Y)$ is sufficient.
- $L^1 \to L^1$ condition: We require that the map $y \mapsto K(x,y)$ be bounded for every x, and the map $x \mapsto \|y \mapsto K(x,y)\|_\infty$ is itself a L^1 function.

In next round of edit, I'll make it clear which of these conditions is necessary (as opposed to just sufficient). Suffice it to say right now that:

- All of these define the transform pointwise, and honestly.
- All of these are certainly sufficient conditions.

3. Fourier transform

3.1. **Definition by formula.** Let's first look at the Fourier transform. The "goal" of doing a Fourier transform is to pass from the space of functions on $\mathbb{R}$, to the space of functions on $\mathbb{R}$, in a nice way. Here's the definition:

$$\hat{f} = x \mapsto \int_{\mathbb{R}} e^{-ixy} f(y)\, dy$$

This transform comes from the kernel function:

$$K(x,y) = e^{-ixy}$$

The expression e^{-ixy} makes sense only for real x, y (it makes sense for complex x, y as well, but the next few comments won't apply), so the "formula" tells us that we're trying to define an operator from "a" space of functions on $\mathbb{R}$ to a space of functions on $\mathbb{R}$.

There is also a version of the Fourier transform for functions on $\mathbb{R}^n$. The key difference in the formula is that instead of looking at e^{-ixy} (which doesn't make sense in $\mathbb{R}^n$) we look instead at $e^{ix \cdot y}$ where $\cdot$ is the standard inner product:

$$\hat{f} = x \mapsto \int_{\mathbb{R}^n} e^{-ix \cdot y} f(y)\, dy$$

How well-defined this operator is, is the next question.

3.2. **As an operator on L^1.** K is a bounded kernel: the absolute value for any pair $(x,y) \in \mathbb{R}^2$ is 1. Thus, by the dual spaces result mentioned earlier, K yields a well-defined map $L^1(\mathbb{R}) \to L^\infty(\mathbb{R})$. Thus, the Fourier transform makes sense for any L^1 function and outputs a L^∞ function. The interpretation is "honest" in the sense that if we take a "specific" L^1 function (as opposed to an equivalence class) we get a specific L^∞ function (in fact, even better, the output function is actually bounded and not just essentially bounded)

3.3. **As an operator on L^2.** A little thought shows that this operator does *not* satisfy the conditions for Schur's lemma: the problem is that the absolute value of the kernel is always 1, and 1 does not decay off at infinity. Specifically, there are functions in L^2 (and outside L^1) for which the map $y \mapsto e^{-ixy} f(y)$ has an "oscillatory" integral so we cannot define a value at these points.

Thus, we cannot hope to get an honest, pointwise map from L^2 to L^2 using the formula of the Fourier transform. Rather, the approach we take is more roundabout: we construct a dense subspace of L^2, called the Schwarz space, and prove that the Fourier transform is an isometry from the Schwarz space to itself. The isometry then extends to a completion of the Schwarz space, which is the whole of L^2.

Since this is *not* a pointwise approach, one cannot, for $f \in L^2$, make *any* sense of the expression $\hat{f}(x)$. That's because $\hat{f}$ is only an equivalence class of functions, even if one took care to define f as a specific function, and for an equivalence class of functions, the value at a point isn't at all defined.

What turns out to be true is that this circumspect approach to defining the Fourier transform on L^2, is equivalent to the honest definition on those functions that live in $L^1 \cap L^2$.

3.4. **Multiplication, differentiation, and the Fourier transform.** The Fourier transform converts multiplication to differentiation. Precise sense can be made of the statement but the crux of the idea is the fact that the derivative of x^n is nx^{n-1}, so it involves a "multiplication" by n.

4. Schwarz class, and some miscellanea

4.1. **The Schwarz class.** Unlike L^p (that makes sense for all measure spaces) and C_c, C_0 (that make sense for topological spaces) the Schwarz class makes sense only for spaces like $\mathbb{R}$ and $\mathbb{R}^n$ (essentially, a real-analytic structure is needed on the space). A lot of the arguments and proofs that we do depend on the Schwarz class (an example is extending the domain of definition of the Fourier transform). The two good things about the Schwarz class are:

- Functions in the Schwarz class decay "fast", so it is easier to integrate against them to get finite values. This makes it possible to define linear functionals and linear operators on the Schwarz class. Essentially, the Schwarz class is "small" and "well-behaved".
- The Schwarz class is "dense" in a number of other bigger function spaces (with their natural function space topologies). It is dense in C_0 (functions that decay to 0) with the uniform norm, and it is dense in all the L^p with their L^p-norms.
- The Schwarz class is "closed under the Fourier transform". A smaller class is C_c^∞: compactly supported smooth functions. This is smaller, and is dense in the same way that the Schwarz class is, but this class is not closed under taking Fourier transforms.

Let us now define the Schwarz class, and study the integrability properties enjoyed by it.

Definition (Schwarz class for Euclidean space)**.** The Schwarz space of $\mathbb{R}^n$, denoted $\mathcal{S}(\mathbb{R}^n)$, is the set of functions $f : \mathbb{R}^n \to \mathbb{R}$ satisfying all the following conditions (if we are considering complex-valued functions, we impose the condition on the real and imaginary parts):

- f is infinitely differentiable, i.e. $f \in C^\infty(\mathbb{R}^n)$.
- For any polynomial function $p : \mathbb{R}^n \to \mathbb{R}$, and any function g obtained as a (possibly mixed) partial derivative of any order of f, the function $x \mapsto p(x)g(x)$ goes to 0 as $\|x\| \to \infty$. Note that "derivative of any order" includes f itself, i.e. we also require that for any polynomial function $p : \mathbb{R}^n \to \mathbb{R}$, the function $x \mapsto p(x)f(x)$ goes to 0 as $\|x\| \to \infty$.

It actually suffices to say that for any polynomial p, the function $x \mapsto p(x)f(x)$ goes to 0 (I'm not sure, need to check this).

We have the following obvious inclusions:

$$C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n) \subset C_0(\mathbb{R}^n) \subset L^\infty(\mathbb{R}^n)$$

It requires a bit more work to show that Schwarz functions are in L^p for every $1 \leq p \leq \infty$. Note that C_c^∞ is clearly in L^p for every p, while $C_0(\mathbb{R}^n)$ is not, so the result says something nontrivial. Let's illustrate by explaining why $\mathcal{S}(\mathbb{R}^n)$ is contained inside L^1.

We need to show that if f is a Schwarz function, then f is integrable. We may assume $f \geq 0$. The point is that since f is a Schwarz function, any polynomial times f goes to 0, so outside a compact set, f can be dominated by a function of the form $1/(x_1^2 x_2^2 \dots x_n^2)$. This function is integrable on the outside of a cube containing the origin, and hence f has finite integral outside a compact region. Within the compact region, f is integrable because it is continuous, and hence bounded.

To show that $f \in L^p$ for $p > 1$, we can use a similar strategy; we dominate f by the function $1/(x_1^{2p} x_2^{2p} \dots x_n^{2p})$ (p not being an integer is not a serious problem) outside a cube, and use the fact that the integral is finite inside the cube.

Thus, $\mathcal{S}(\mathbb{R}^n)$ is contained inside the intersection of L^p for all p. Since a smaller subset than $\mathcal{S}(\mathbb{R}^n)$ (namely C_C^∞) is dense in L^p for all $1 \leq p < \infty$, $\mathcal{S}(\mathbb{R}^n)$ is clearly dense in L^p for $1 \leq p < \infty$. Let's quickly recall what we know about L^p-spaces, and see how $\mathcal{S}(\mathbb{R}^n)$ can be fit into our scheme of knowledge:

- All the L^ps are subspaces of the space M. M denotes the space of finite-valued measurable functions on the measure space (finite almost everywhere suffices). M itself is a $\mathbb{R}$-algebra under pointwise addition and multiplication of functions.
- Each of them has the structure of a normed vector space, but the norms are topologically different, so they do not come from any topology on M. In fact, for any p, L^p can be defined as the connected component of 0 under the p-norm viewed as a distance (each coset of L^p is a collection of functions such that the distance between any two has finite p-norm).
- Although M is a $\mathbb{R}$-algebra, none of the L^ps (for finite p) is a subalgebra, if the measure space has subsets of arbitrarily small measure. L^∞ *is* a subalgebra (under pointwise addition and multiplication).
- For infinite measure spaces like $\mathbb{R}^n$, L^r is not contained in L^s if $r \neq s$. However, L^p is contained in $L^r \cap L^s$ if p is between r and s.
- The intersection of all the L^ps, for *finite* p, is a subalgebra without multiplicative identity, but this subalgebra neither contains, nor is contained in, L^∞.
- If the measure space is *separable*, then the L^ps are all separable metric spaces for finite p. L^∞ is *not* separable if the space is expressible has infinitely many disjoint pieces of positive measure.
- The polynomials (and similarly, the trigonometric polynomials) form not just a subspace, but a *subalgebra* of the algebra M of measurable functions on $\mathbb{R}$ (or $\mathbb{R}^n$). On any compact subset of $\mathbb{R}$ (respectively, $\mathbb{R}^n$) the polynomial, or trigonometric polynomials restricted to that subset are dense in L^p for every p.
- Sitting between the subalgebra of polynomials, and the algebra of all measurable functions, are many other subalgebras: the algebra $C(\mathbb{R}^n)$ of all continuous functions, the algebra $C^r(\mathbb{R}^n)$ of r times differentiable functions, and the algebra $C^\infty(\mathbb{R}^n)$ of infinitely differentiable functions.
- For $\mathbb{R}^n$, the space C_c (continuous, compactly supported functions) is dense in L^p for every finite p, and its closure in L^∞ is C_0 (functions that decay to 0). C_c and C_0 are subalgebras without unit (because $\mathbb{R}^n$ is noncompact). In fact, they give examples of proper ideals in the algebra $C(\mathbb{R}^n)$ that do not have a common vanishing point.[3]
- In fact, the space C_c^∞ (smooth, compactly supported functions) is dense in L^p for every finite p, and its closure in L^∞ is C_0. The Schwarz space $\mathcal{S}(\mathbb{R}^n)$, which sits between C_c^∞ and C_0^∞, enjoys the same density properties. Both $\mathcal{S}(\mathbb{R}^n)$ and C_c^∞ are ideals inside the algebra C.

The Schwarz space is multiplicatively closed, it is in L^∞ and it is dense in L^p for finite p and in C_0. Thus, multiplying a function in L^p by any number of Schwarz functions keeps it in L^p. In fact, if we multiply a Schwarz function by a function in L^p, the new function we get is in L^r for all $r \leq p$ (basically, using the fact that the Schwarz space is contained in L^s for every s).

There is another subtle point to which I'd like to draw attention. When we are considering the space $L^p(\mathbb{R}^n)$, there is no *natural* representative for the class of functions. In other words, if $\tilde{L}^p(\mathbb{R}^n)$ denotes all *specific* functions that have finite p-norms, and $N_0(\mathbb{R}^n)$ denotes the space of all specific functions that are 0 almost everywhere, we have a short exact sequence of vector spaces:

$$0 \to N_0 \to \tilde{L}^p \to L^p \to 0$$

Vector space theory would tell us that this sequence splits, but there is no natural way to split the sequence in general. However, for the space of continuous functions, there is a very specific representative: the function that is honestly continuous.[4]

Thus, we shall talk of the space of continuous functions as a subspace of the space M which is a space of equivalence classes of functions; what we're using is the natural splitting to identify a subspace with a quotient. Thus, there is substance to the frequent abuse of notation.

4.2. **Separate versus joint.** One of the theoretical issues that we should discuss at this stage is that of *separate* versus *joint*. In "category-theoretic language", the question is as follows: we have two objects X and Y in some Cartesian-closed monoidal concrete category (basically sets with additional structure). Suppose further that for every $y \in Y$, the map $X \to X \times Y$ given by $x \mapsto (x, y)$ is a morphism in the

[3]Such ideals do not exist in compact spaces; an elementary exercise in topology that has important applications to algebra; but we're doing analysis here

[4]In fact, by an observation due to B. Werness, the class of functions for which a natural section exists is somewhat bigger than the class of continuous functions for $\mathbb{R}^n$, it is the class of functions for which the metric density of the induced measure, is well-defined at every point.

category, and vice versa. Then the question: if a map $X \times Y \to Z$ restricts to a morphism on each of the X-fibers and also on each of the Y-fibers, is it a morphism?

In general, the answer is *no*. A map that restricts to a morphism on each of the fibers is sometimes said to be *separately* a morphism. For instance:

- *Topological spaces*: There exist maps from a product space $X \times Y$ to Z that are separately continuous (i.e. continuous on each X-fiber and each Y-fiber) but not jointly continuous (i.e. not continuous from $X \times Y$ with the product topology). Specifically, the map $\sin(4\text{Arg}(x+iy))$ is a separately continuous map from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$, that is not jointly continuous.
- *Smooth manifolds*: There exist maps from a product of manifolds that are jointly continuous, and smooth on each fiber, but not jointly smooth. In particular, there are maps from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$ that are smooth on each vertical and horizontal line, but not jointly smooth.
- *Measure spaces*: There exist maps from a product of measure sapces that are measurable on each fiber but not jointly measurable.

In general, joint continuity, smoothness, measurability, differentiability imply separate continuity, et al. but the converse isn't true. Thus, if we are given the datum that a particular function is jointly smooth, and the only way we're using it is "separate" smoothness, we are missing out on some part of the problem data (or equivalently, are trying to prove something more general). Schur's lemma is a "separate" lemma in the sense that it says something about the restrictions to the fibers, which is one of the sources of its power. However, applying this in a larger problem that involves iterated integrations may require that we use the joint behaviour.

An example of particular interest is the Schwarz space. If $f \in \mathcal{S}\left(\mathbb{R}^2\right)$, then we may be interested in what we can say about the fiber-wise integrals of f. Note first that since f is jointly Schwarz, it is separately Schwarz, so for every value of x, the map $y \mapsto f(x,y)$ is in $\mathcal{S}\left(\mathbb{R}^1\right)$. But the joint Schwarz property tells us that the function:

$$g = x \mapsto \int_{\mathbb{R}} f(x,y)\, dy$$

is again in $\mathcal{S}\left(\mathbb{R}^1\right)$. In other words, the joint Schwarz property tells us that the function we get by fiber-wise integration is again Schwarz. We couldn't have made such guarantees if we were merely given that f is separately Schwarz.

4.3. **Closure of Schwarz space under Fourier transforms.** The importance of the following fact cannot be overestimated: the Fourier transform defines a L^2-isometry from the Schwarz space to itself. This is an honest isometry of specific functions, since, as mentioned earlier, Schwarz functions are their own natural representatives in the equivalence class.

Something more deserves mention regarding Fourier transform. Given a Schwarz function (or any function) on $\mathbb{R}^m \times \mathbb{R}^n$, one can choose to take the Fourier transform on the last n variables, which basically means that we think of the function as a function only on the last n variables, with the first m variables serving as parameters. The great thing is that if we start with a function that is jointly Schwarz, then taking the Fourier transform in only some of the variables, again gives a function that is *jointly Schwarz*. (This requires more proof or explanation).

5. Quantization

The goal for now is not to get into the deeper physical meanings of the transforms, but to understand the mechanics of the integration process, and what they tell us about where things live and how valid certain manipulations are.

A *quantization* is a very "meta" object in this sense: it is a rule to convert a function into a transform. If we start with a function a it gives a outputs a quantized version $A = Q[a]$ that takes as input a function f and outputs a function $A[f]$.

Now the quantization rules that we have are linear, so we can pull the a into the argument as well, and get a "two-input transform". Rather than viewing Q as something that takes one function and then gives a transform that acts on another function, we could view Q as taking in *two* functions, a and f, and acting on both of them to output the function $(Q[a])[f]$. This is the way we'll be viewing it in practice, as we shall soon see.

To complicate things further, the quantization is not viewed as a single quantization, but rather as a family of quantizations Q_ε parametrized by $\varepsilon \in \mathbb{R}$, such that when $\varepsilon = 0$, we get the operation of "multiplication" of the two functions.

5.1. **Ordinary quantization.** The "ordinary quantization" has a beautiful interpretation in terms of noncommutative algebra, but that would take us too far afield, so instead we think of the formula as "God-given" and try only to study it. Since quantization wasn't made by functional programmers or people to whom set-theoretic clarity was at a premium, the notation may appear confusing at first. What I would probably call $Q_\varepsilon[a]$ is denoted as $a(x, \varepsilon D)$. This is "not", in any precise sense, the evaluation of a at x and εD.

Here's the setup, $a \in \mathcal{S}(\mathbb{R}^n \times \mathbb{R}^n)$. The **ordinary quantization**(defined) of a at ε, denoted $a(x, \varepsilon D)$, takes as input a function $f \in \mathcal{S}(\mathbb{R}^n)$ and outputs the function $a(x, \varepsilon D)[f] \in \mathcal{S}(\mathbb{R}^n)$, by the following rule:

$$a(x, \varepsilon D)[f] = x \mapsto \frac{1}{(2\pi)^n} \int e^{ix.\xi} a(x, \varepsilon\xi) \hat{f}(\xi)\, d\xi$$

When $\varepsilon = 0$, the operator boils down to multiplication by the function $x \mapsto a(x, 0)$.

5.2. **Ordinary quantization in terms of kernels.** If we think of a as fixed, and view $a(x, \varepsilon D)$ as an operator on $\mathcal{S}(\mathbb{R}^n)$, then we can think of it as a composite of two kernel transforms. The first kernel is the kernel for the Fourier transform, namely $(x, \xi) \mapsto e^{-ix.\xi}$. The second kernel is given as:

$$(x, \xi) \to \frac{1}{(2\pi)^n} e^{ix.\xi} a(x, \varepsilon\xi)$$

In terms of the deeper meaning of ordinary quantization, ordinary quantization corresponds to doing things sequentially (as opposed to the Weyl quantization, where everything's done at once). This makes ordinary quantization both easier and harder to handle than Weyl quantization.

However, it turns out that we can simplify the "composite" of these two transforms into a single transform. This requires a use of Fubini and integral interchange. We shall discuss this after a few subsections.

5.3. **Weyl quantization.** Before going further into proving properties of ordinary quantizations (like the fact that it is well-defined, and gives a bounded operator) we define the Weyl quantization. The **Weyl quantization**(defined) has the same setup but a different formula: starting with $a \in C^\infty(\mathbb{R}^n \times \mathbb{R}^n)$ and a $\varepsilon > 0$, we get an operator $a^w(x, \varepsilon D)$ from $\mathcal{S}(\mathbb{R}^n)$ to $\mathcal{S}(\mathbb{R}^n)$, as follows:

$$a^w(x, \varepsilon D)[f] = x \mapsto \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} a\left(\frac{x+y}{2}, \varepsilon\xi\right) f(y) e^{i(x-y).\xi}\, d\xi\, dy$$

The Weyl quantization can directly be viewed as a transform with a kernel. The kernel function is given by:

$$K(x, y) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^d} a\left(\frac{x+y}{2}, \varepsilon\xi\right) e^{i(x-y).\xi}\, d\xi$$

On the positive side, the quantization is described by integration against a kernel; on the negative side, the kernel here is more complicated than the kernels for ordinary quantization. Philosophically, the Weyl quantization is about "doing it all at once" as opposed to separately.

5.4. **Extending to L^2.** Although the ordinary and Weyl quantizations are initially defined (and make direct sense) only for the Schwarz space, we can use the density of the Schwarz space in L^2 to *attempt* to extend it to a map from L^2 to L^2. (Can we do this? How well-defined would it be?)

Note that a continues to remain a C^∞ function.

5.5. **Boundedness of the Weyl quantization.** Let us try to show that the Weyl quantization is *uniformly bounded* as a linear operator from $\mathcal{S}(\mathbb{R}^n)$ to $\mathcal{S}(\mathbb{R}^n)$ *with respect to* the L^2-norm. We need to figure out whether the kernel function satisfies the conditions of Schur's lemma. It turns out that the answer is *yes*, though this is not immediately obvious. Let's inspect the kernel more closely:

$$K(x, y) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^d} a\left(\frac{x+y}{2}, \varepsilon\xi\right) e^{i(x-y).\xi}\, d\xi$$

We first observe that the expression inside the integration symbol looks like a Fourier transform. In addition to normalization considerations, we need to make linear changes of variables, and view things as living in the correct space. For this, view a as a function on the second coordinate (the ξ coordinate) parametrized by the first coordinate – so we're really thinking of the first coordinate as fixed, and seeing the Fourier transform as happening in the second coordinate. Then, $\varepsilon\xi$ should be our new ξ, (the variable

against which we are integrating) and the new x thus becomes $(y-x)/\varepsilon$ (the sign changes because in the Fourier transform we have a negative sign, and the ε is to cancel the positive factor of ε that comes up when we replace ξ by $\varepsilon\xi$.

We now use two facts, mentioned earlier:

- Taking the Fourier transform in some of the variables, treating the others as parameters, preserves the (joint) Schwarz property.
- For a jointly Schwarz function, we can do a linear change of variables and still get a jointly Schwarz function. Moreover, any jointly Schwarz function is in L^1 for each variable. This part is needed to show that the kernel satisfies the conditions for Schur's lemma.

Modulo these two basic facts, the proof is a mere formal manipulation, that we (do not) give below.

5.6. **Boundedness of the ordinary quantization.** Ordinary quantization is trickier to handle than Weyl quantization because it is not described as a "single-stroke" integration against a kernel. However, the good news is that we can reduce it to such a thing with a little manipulation, and from that point onwards, the proof ideas are precisely the same as those for the Weyl quantization.

$$\begin{aligned} a(x,\varepsilon D)[f](x) &= \frac{1}{(2\pi)^n}\int e^{ix\cdot\xi}a(x,\varepsilon\xi)\hat{f}(\xi)\,d\xi \\ \implies a(x,\varepsilon D)[f](x) &= \frac{1}{(2\pi)^n}\int e^{ix\cdot\xi}a(x,\varepsilon\xi)\int f(y)e^{-i\xi\cdot y}\,dy\,d\xi \\ \implies a(x,\varepsilon D)[f](x) &= \frac{1}{(2\pi)^n}\int f(y)\left(\int e^{ix\cdot\xi}a(x,\varepsilon\xi)\,d\xi\right)\,dy \end{aligned}$$

We thus get:

$$K(x,y) = \int e^{ix\cdot\xi}a(x,\varepsilon\xi)\,d\xi$$

5.7. **Continuity in the ε parameter.** We have defined the ordinary and Weyl quantizations with a parameter: ε. One natural question is: how smoothly do the quantizations change with ε? To answer this, we again inspect the original definitions. We see that the dependence on ε stems in how much "weight" we give to the ξ coordinate. Since a has a smooth dependence on ξ, we see that the dependence of the transform on ε is a smooth dependence. The family of quantizations thus varies smoothly. The same is true for the Weyl quantization.

5.8. **Asymptotic equivalence of the quantizations.** It is clear that at $\varepsilon = 0$, the ordinary quantization and Weyl quantization are both multiplication operators, by the function $x \mapsto a(x,0)$. Thus, the operators a and a^w converge "pointwise" as $\varepsilon \to 0$. Here, we show that they converge to each other in another sense: the L^2 sense. In other words, we show that the L^2 norm of the operator:

$$a(x,\varepsilon D) - a^w(x,\varepsilon D)$$

goes to 0 as $\varepsilon \to 0$. It turns out that this follows again from Schur's lemma. I outline the important steps:

- First, we use the fact that both of them are expressible as integral operators i.e. as transforms with kernels.
- Thus, the difference is again expressible as an integral operator whose kernel is the difference of the kernels of the two quantizations.
- We now show that for this difference, we can compute a Schur's lemma bound on the fiber-wise integrals $C(\varepsilon)$, such that $\lim_{\varepsilon\to 0} C(\varepsilon) = 0$.
- Combining this with Schur's lemma, we see that the operator norm, which is bounded from above by $C(\varepsilon)$, goes to zero as ε goes to 0.

5.9. **Self-adjointness property.** If a is real-valued, a^w is self-adjoint.

The proof of this relies on a more general fact. Namely, go back to the general situation of a kernel function $K : X \times Y \to \mathbb{C}$. We could try to use K to define an integral operator from a function space on X to a function space on Y, or we could use it to define an integral operator from a function space on Y to a function space on X. The two integral operators are closely related by an application of Fubini's theorem: they are "transpose" to each other.

Now, in the case of the Weyl quantization, the kernel function, namely:

$$K(x,y) = \int e^{i(x-y)\cdot\xi} a\left(\frac{x+y}{2}, \varepsilon\xi\right) d\xi$$

is particularly nice: interchanging the role of x and y yields the conjugate function when a is real (basically, the a part is unchanged and the other part gets conjugated). In symbols:

$$K(y,x) = \overline{K(x,y)}$$

5.10. **Other properties of the Weyl quantization.** We state here one important property of the Weyl quantization: the so-called product rule.

Let a and b be two elements of $\mathcal{S}(\mathbb{R}^n \times \mathbb{R}^n)$. We then have:

$$b(x,\varepsilon D) \circ a(x,\varepsilon D) = (ab)(x,\varepsilon D) + \frac{\varepsilon}{i}\left(\nabla_\xi b \cdot \nabla_x a\right)(x,\varepsilon D) + \varepsilon^2 R_\varepsilon$$

The crucial corollary of the product rule, that we shall apply in many subsequent situations, is that $b \circ a$, in the limit, approaches the product ab. The precise limiting formula that we use is:

$$\lim_{\varepsilon\to 0} \langle\ \varphi\ ,\ (b\circ a)^w(\psi)\ \rangle = \lim_{\varepsilon\to 0} \langle\ \varphi\ ,\ (ab)^w\psi\ \rangle$$

In other words, we can, in the limit, substitute a product for a composition.

5.11. **The upshot.** The product rule and self-adjointness properties of the Weyl quantization, along with the basic fact that the Weyl quantization gives well-defined, uniformly bounded operators, give us a toolkit for manipulating expressions using the Weyl quantizations in inner products. We shall see that this toolkit gives us some surprising flexibility with proving results about the Wigner transform. Manipulatively, we have the following:

- The self-adjointness property of the Weyl quantization allows us to move a Weyl quantization from one side of an inner product to another.
- The product rule allows us to simplify a composite of two Weyl quantizations, in terms of a single Weyl quantization by their product. This is done only when we are passing to the limit as $\varepsilon \to 0$.

6. Wigner transform

6.1. **Definition of the Wigner transform.** The Wigner transform starts off with a function of "one" vector and gives a function of two vectors. Namely, for every ε, the Wigner transform is a linear operator from a space of functions on $\mathbb{R}^d$, to a space of functions on $\mathbb{R}^d \times \mathbb{R}^d$. The formula is as follows:

$$W_\varepsilon[\varphi] = (x,k) \mapsto \frac{1}{(2\pi)^d}\int e^{ik\cdot y}\varphi\left(x - \frac{\varepsilon y}{2}\right)\overline{\varphi}\left(x + \frac{\varepsilon y}{2}\right) dy$$

This certainly looks like a monstrosity, and since we've basically decided to go the route of formal mathematics minus physical motivation for now, we just have to live with it, and figure out the formal properties of this.

First of all, observe that it is not clear that the Wigner transform gives something well-defined if $\varepsilon = 0$. The problem is that we're only given that $\phi \in L^2$ so $\phi\overline{\phi} \in L^1$, but there's no reason to suspect that $e^{ik\cdot y}$ times this will still be in L^1. In fact, it never lands in L^1. Thus, we really need to use something about $\varepsilon \neq 0$ in the proof.

6.2. **Proof of boundedness.** We in fact show the bound:

$$\|W_\varepsilon\|_2 = \frac{1}{\varepsilon^{d/2}}\|\varphi\|_2^2$$

The correct way to view the Wigner transform is to think of it as the Fourier transform of the function:

$$g(x,y) = \varphi\left(x + \frac{\varepsilon y}{2}\right)\overline{\varphi}\left(x - \frac{\varepsilon y}{2}\right)$$

Note that the signs got switched because in the Fourier transform, we require a negative sign in the exponent.

Since the Fourier transform is an isometry, it boils down to showing that the function g is in L^2. The idea here is to make a change of variables, giving importance now to the variables $x + \varepsilon y/2$ and $x - \varepsilon y/2$. Note, crucially, that this change of variables goes throuh only if $\varepsilon \neq 0$, and to contributes a volume factor.

Think of it pictorially as follows: the vectors $x + \varepsilon y/2$ and $x - \varepsilon y/2$ are vectors very close to x, but they together generate x and y by taking the average and half the difference (and scaling back by ε). But the smaller ε is, the closer the vectors get, and the more we need to scale back by to get x and y. Finally, when $\varepsilon = 0$, the two vectors collapse to x, so we can no longer use them to generate y.

We basically do a change of variables: $u = x + \varepsilon y/2$ and $v = x - \varepsilon y/2$. This is a linear change of variables whose Jacobian is ε^d, so we get:

$$\|W_\varepsilon[\varphi]\|_2^2 = \frac{1}{(2\pi\varepsilon)^d} \iint |\varphi(u)|^2 \, |\varphi_\varepsilon(v)|^2 \, du \, dv$$

Separating variables and integrating yields:

$$\|W_\varepsilon[\varphi]\|_2^2 = \frac{1}{(2\pi\varepsilon)^d} \, \|\varphi\|_2^4$$

Taking squareroots yields the desired result.

6.3. **Some more observations about the Wigner transform.**

- The Wigner transform is real-valued. This is basically because we're multiply φ with $\overline{\varphi}$. The check is direct.
- The Wigner transform is not defined in an honest, pointwise fashion. That's because the last stage of the Wigner transform is a Fourier transform. So we cannot always evaluate the Wigner transform at a point (x, k). We can do honest Wigner transforms for Schwarz functions, though.

6.4. **Relating the Wigner transform to the Weyl quantization.** It turns out that there is a rather elementary way in which the Wigner transform is related to the Weyl quantization. Let's first try to get an idea of the spaces where things live.

- The Wigner transform starts off with a function on $\mathbb{R}^d$ and outputs a function on $\mathbb{R}^d \times \mathbb{R}^d$.
- The Weyl quantization takes in one function on $\mathbb{R}^d \times \mathbb{R}^d$ (the quantizing function) and one function on $\mathbb{R}^d$ and outputs a function on $\mathbb{R}^d$.
- The relation is as follows. Start with a function η on $\mathbb{R}^d \times \mathbb{R}^d$ and a function φ on $\mathbb{R}^d$. Then, we can consider $\eta^w(x, \varepsilon D)(\varphi)$ as another function on $\mathbb{R}^d$ and we can consider $W[\varphi]$ as another function on $\mathbb{R}^d \times \mathbb{R}^d$. The relation between these is as follows:

$$\langle \, \eta \, , \, W_\varepsilon \, \rangle = \langle \, \eta^w(x, \varepsilon D)\varphi \, , \, \varphi \, \rangle$$

To make precise sense of the above, we need to carefully interpret where things live. It turns out that the usual hypotheses do: $\varphi \in L^2$, and $\eta \in C_c^\infty$.

The importance of this is that we can now relate properties of the Wigner transform with properties of the Weyl quantization. More specifically, the manipulative toolkit that we developed for the Weyl quantization gives us a way of establishing results about the Wigner transform. The typical strategy is:

- Write the inner product in terms of the Wigner transform
- Convert it to an inner product in terms of the Weyl quantization
- Now, manipulate this inner product using the self-adjointness property of the Weyl quantization (can be done only for real-valued functions), and use the product rule to convert compositions to products (can be done only when sending ε to 0

6.5. **Families of functions.** Before sending things to zero in the limit, we need one more important idea. So far, we have considered the Wigner transform and Weyl quantization as ε-parametrized families of integral operators. But we have been applying only one operator at a time, and to only one function.

The more general scenario is that one has a ε-parametrized family of functions φ, and the Wigner transform W_ε-member of φ. In other words, when we now use the notation $W_\varepsilon[\varphi]$ and φ happens to be a family of functions, what we *really* mean is $W_\varepsilon[\varphi_\varepsilon]$.

6.6. **The two-input Wigner transform.** So far, we have viewed the Wigner transform as something that inputs a function on $\mathbb{R}^d$ and outputs a function on $\mathbb{R}^d \times \mathbb{R}^d$. We now describe a *two-input* Wigner transform; this takes as input *two* functions on $\mathbb{R}^d$ and outputs a function on $\mathbb{R}^d \times \mathbb{R}^d$, in other words, it is a map:

$$L^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d \times \mathbb{R}^d)$$

The definition is such that if we apply it to a pair (f, f) (i.e. a diagonal element on the product on the left side) we get the Wigner transform of f.

The two-input version is:

$$\tilde{W}_\varepsilon[f,g] = (x,k) \mapsto \int e^{ik\cdot y} f\left(x + \frac{\varepsilon y}{2}\right) \overline{g}\left(x - \frac{\varepsilon y}{2}\right)\, dy$$

With exactly the same reasoning as before, we derive a relation between the two-input Wigner transform, and Weyl quantization:

$$\left\langle\ \eta\ ,\ \tilde{W}_\varepsilon[f,g]\ \right\rangle = \langle\ \eta^w(f)\ ,\ g\ \rangle$$

6.7. **Uniform boundedness results.**

6.8. **Using the Wigner transform to obtain measures.** Let's review the setup. We have a family φ_ε of functions parametrized by ε (in practice, we would like the φ to vary smoothly (or continuously) with respect to ε, but this condition is not necessary to make sense of things we can say.

Now, in general, the φ_ε may not converge to anything, and even if they did, it is certainly not necessary that $W_\varepsilon[\varphi_\varepsilon]$ converge to anything as $\varepsilon \to 0$ (that's because, as we noted/pointed out earlier, W_0 doesn't make a whole lot of sense for arbitrary L^2 functions). But we could still look at the limit *points* of the set $W_\varepsilon[\varphi_\varepsilon]$, when suitably topologized. A somewhat surprising result (that would have been hard to prove without all the machinery that we've set up) is that all the limit points of this are nonnegative linear functionals, and thus give nonnegative measures.

7. Riesz-Thorin interpolation theorem

7.1. **Dense and complete.** By now, we are familiar with the situation: a formula is used to define a transform from one L^p space to another, but in reality the formula makes honest pointwise sense only in a very limited range of cases. But that limited range of cases is enough to form a dense subspace of L^p. We first discuss a general but elementary result:

Claim. Let X and Y be normed linear spaces, and $A \subset X$ a dense linear subspace of X. Suppose $T : A \to Y$ is a map. Then, if Y is complete, and T is a bounded linear map fom A to Y, then T extends uniquely to a bounded linear map form X to Y.

The key steps of the proof are:

- Pick a point $x \in X$
- Find a sequence of points in A that converges to x
- Take the images of these points. Since the original sequence was convergent, it was Cauchy, so its image is a Cauchy sequence. But now since Y is complete, the image is convergent, and the point to which it converges is the image of x.
- Show that the image is independent of the choice of sequence (using the fact that combining two sequences converging to x still yields a sequence converging to x).

Thus, if our transforms have been defined on a space like C_c^∞, or the Schwarz space, or some other space that is dense in all the L^ps, and we have shown that the transform sends it inside L^r and the operator norm viewed in terms of the p-norm to the r-norm is bounded, then we can actually extend to a map $L^p \to L^r$ (this little detail will be implicit in whatever we do for Riesz-Thorin).

7.2. **Norm from one L^p to another.** Having actually dealt with some transforms, we can now attempt to understand more challenging ideas in the subject. Pick $1 \leq p, r \leq \infty$. We want to study situations when a transform from functions on a measure space X to a measure space Y, gives a well-defined bounded linear operator:

$$L^p(X) \to L^r(Y)$$

Moreover, we want to find upper bounds on the operator norm of the transform.

We had done some baby cases of this a while ago. In the baby cases, we had set $r = \infty, 1$ and had derived some sufficient conditions, that weren't necessarily necessary. We can now look at the more general case.

The proof of Schur's lemma actually shows something stronger than the statement itself. If we consider the integral operator $T : L^2 \to L^2$ with kernel K, then we have:

$$\|T\| \leq \sqrt{\sup_X \int_Y |K(x,y)|\, dy} \sqrt{\sup_Y \int_X |K(x,y)|\, dx}$$

A little simplification effort yields that, in fact, the first term on the right is $\sqrt{M(\infty,\infty)}$ and the second term on the right is $\sqrt{M(1,1)}$. So we can write:

$$M(2,2) \leq \sqrt{M(1,1)M(\infty,\infty)}$$

We shall see that the Riesz-Thorin interpolation theorem provides a significant generalization of this idea. Roughly speaking, it states that if $M(p_0,r_0)$ and $M(p_1,r_1)$ are both finite for the linear transform T, then $M(p(a),r(a)) < \infty$, where $(p(a),q(a))$ "interpolates" between the pair (p_0,r_0) and the pair (p_1,r_1).

7.3. **Correctly linearizing the L^ps.** Observe that the set $[1,\infty]$, under the map $x \mapsto 1/x$, goes to the set $[0,1]$, wherein we have a nice linear structure. We use convex combinations of the reciprocals as the "linearized" view on $[1,\infty]$. Thus, in this view, 2 is the midpoint of $[1,\infty]$, and conjugate numbers are equidistant from the midpoint (and on opposite sides of it).

With this view, given p_0 and p_1, we define $p(a)$, for $a \in [0,1]$, as follows: invert, take the convex linear combination, and then invert back. Formally:

$$\frac{1}{p(a)} = \frac{a}{p_1} + \frac{1-a}{p_0}$$

With this, we can now state the Riesz-Thorin interpolation theorem. Let T be a linear operator from a function space on X to a function space on Y. Then, if $M(p_0,r_0)(T) < \infty$ and $M(p_1,r_1)(T) < \infty$, we have:

$$M(p(a),r(a))(T) \leq (M(p_0,r_0)(T))^{1-a}(M(p_1,r_1)(T))^a$$

I will not sketch a complete proof here, but I'll provide the key ideas. The key idea is to somehow construct a function on $[0,1]$ and argue that the values of the function on the boundary control the value of the function inside. Unfortunately, there is no theorem of real analysis that guarantees such a result, so we extend to a holomorphic function on a complex strip, and apply the so-called "three lines theorem" from complex analysis.

7.4. **Application to Fourier transform.** The power of the Riesz-Thorin interpolation theorem is that it allows us to use proofs for some nice and cute L^ps, to provide proof for a continuous range of L^ps. One such application is the so-called Hausdorff-Young inequality.

Recall that we know two facts about the Fourier transform:

- Since the kernel is in L^∞ jointly, the Fourier transform is in $M(1,\infty)$
- By a somewhat tedious argument that everybody knows, the Fourier transform defines an isometry from L^2 to L^2.

We can apply the Riesz-Thorin interpolation theorem with $(p_0,r_0) = (1,\infty)$ and $(p_1,r_1) = (2,2)$. Pictorially, this tells us that the Fourier transform "reflects" the reciprocal about the point $1/2$, if the reciprocal is 1, or $1/2$. This means that the Fourier transform reflects the reciprocal for any p between 1 and 2. In other words, the Fourier transform is in $M(p,q)$ where q is the Holder conjugate to p and $1 \leq p \leq 2$.

Moreover, Riesz-Thorin interpolation also provides a precise bound on the norm of the Fourier transform (the bound depends on the way we choose to normalize, and has 2π floating around). This bound goes by the name of Hausdorff-Young inequality.

7.5. **Information about convolutions.** We now use Riesz-Thorin to say some very powerful things, most of which are far from obvious, and certainly very hard to prove by direct means. Suppose $f,g \in M(X)$ i.e. f and g are both measurable functions on $\mathbb{R}$ or $\mathbb{R}^n$ (or more generally a locally compact topological Abelian group with a regular translation-invariant measure). We try to define the **convolution**(defined) of f against g as follows:

$$(f*g)(x) = \int_X f(x-y)g(y)\, dy$$

Again, a few comments. We aren't trying to make sure the convolution is defined at *each* point. But we do want it to make sense at *almost every* point, and we want the result to be integrable.

If we genuinely insisted that the convolution be defined at every point, we'd basically be looking at a very narrow range of stuff: $f \in L^p$, $g \in L^q$ where p and q are conjugate exponents. But with the slight latitude, we can prove that the coonvolution sends $L^1 \times L^1$ to L^1. This is an elementary application of Fubini's theorem, and the key point to note is that it doesn't guarantee finiteness (or well-definedness) everywhere.

The most important observation about convolution is that it yields a commutative, associative operation. Though this can be viewed formally, the intuitive idea is that $(f * g)(x)$ is the integral, over all possible decompositions of x as a sum of elements w and y, of $f(w)g(y)$. Commutativity is thus essentially the Abelianness of the underlying group.

It's also clear that the convolution is well-defined and bounded by 1, as a map from $L^1 \times L^\infty$ to L^∞ (in other words, if one function is in L^1 and the other is in L^∞, the convolution is in L^∞). This is a special case of an observation made earlier, that if the kernel is in L^∞, then we get a map from L^1 to L^∞.

Thus, fixing one function from L^1, we see that convolution with that function gives bounded linear operators with bound 1:

$$L^1 \to L^1, \qquad L^\infty \to L^\infty$$

Riesz-Thorin interpolation now tells us that for $f \in L^1$ convolution with f sends L^p to L^p, and $M(p,p)$'1.[5]

We now apply Riesz-Thorin again, this time fixing $f \in L^p$. We know tha tfor $f \in L^p$, convolution with f maps L^1 to L^p (as we just proved) and maps L^q to L^∞ (by Holder's inequality, and as a special case of some general observations we made long ago). Applying Riesz-Thorin yields that convolving a function in L^p and a function in L^r yields a function in L^s, where:

$$\frac{1}{s} + 1 = \frac{1}{p} + \frac{1}{r}$$

Here is a more geometric way of viewing this. Again, go from p to $1/p$, so we are viewing the exponents as their reciprocals in $[0,1]$. Then, the result on convolutions states that when we convolve two functions, their convolution lands inside the space corresponding to the point whose distance fron 1 is the sum of their distances from 1. In other words, distances from 1 add up. In particular, when both are in L^1, the convolution is in L^1.

In fact, this gives a more geometric way of deriving the result. Suppose our goal is to find triples (p,r,s) such that $L^p * L^r \subset L^s$. Basically, Riesz-Thorin states that if we fix p, the reciprocals of r and s form a convex set, and if we fix r, the reciprocals of p and s form a convex set. And we know that $(1,1,1)$, $(1,\infty,\infty)$, and (p,q,∞) lie within this set. By taking the convex hull on both variables, we get the desired result.

8. Hilbert transform

8.1. **Idea behind the Hilbert transform.** The key idea behind the Hilbert transform is:

- Start off with a real-valued function defined on the real line
- Find a complex-valued function defined on the upper half plane, including the real line, with the property that it is anlytic in the open upper half plane and continuous on the whole upper half plane, and such that the real part of its restriction to the upper half-plane is the function we started with.
- Now take the imaginary part of that complex-valued function.

Results from complex analysis (that we do not have the space to go into here) guarantee that if such a function exists, it will be unique, so we can choose any recipe to consturct such a function. We describe here two recipes, which look somewhat different, but yield the same result: the Hilbert transform.

Note that the approach we have taken to the Hilbert transform really works only for real-valued functions on the real line. However, once we have the formulae, it will be possible to see how the formulae can be generalized to other situations.

[5]equality is easily established, but is irrelevant

8.2. **Using the Cauchy method.** We start off by assuming that our original function, h, is in $C^1(\mathbb{R})$ (real-valued) , and moreover, it goes to zero at infinity at a rate that is at least inverse quadratic. In other words, $|h(t)| = O(|t|^{-2})$ for sufficiently large t. In particular, the space of functions that we are considering includes $\mathcal{S}(\mathbb{R})$ (by this we only mean real-valued Schwarz functions. We can then define the analytic function f on the upper half-plane as follows:

$$f(\zeta) = \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{h(t)}{t-\zeta}\,dt$$

Since f comes via an integral formula, it is analytic. We need to verify that when ζ is real, then the real part of f is h, and we also need to certify that the imaginary part is reasonably nice.

Let's do this. Write $\zeta = \xi + i\eta$. The integral now becomes:

$$\begin{aligned} f(\zeta) &= \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{h(t)(t-\overline{\zeta})}{|t-\zeta|^2}\,dt \\ \implies f(\zeta) &= \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{h(t)(t-\overline{\zeta})}{(t-\xi)^2+\eta^2} \\ \implies f(\zeta) &= \frac{1}{\pi}\int_{\mathbb{R}} \frac{\eta}{(t-\xi)^2+\eta^2}h(t)\,dt + \frac{i}{\pi}\int \frac{\xi-t)h(t)}{(t-\xi)^2+\eta^2}\,dt \end{aligned}$$

It's clear now that for $\eta = 0$ the first integral vanishes (on account of being the integral of an odd function). It thus remains to compute the second integral anbd obtain the imaginary part. This looks like a good old integral operator, with kernel:

$$K(t,\xi) = \frac{1}{\pi}\frac{1}{t-\xi}$$

This kernel does not satisfy any remotely good integrability properties, because it blows up at ξ. So instead of trying to compute the integral directly on the real line, we look at the integral over the complement of a small ball aroud ξ, and then send the radius of the small ball to 0. This is called the *principal value*, and yields the well-defined function that we're seeking.

In the subsequent sections, we shall use the letter k for the function $H[h]$, i.e. for the Hilbert transform applied to the function h.

8.3. **Proving some properties of the Hilbert transform.** The Hilbert transform has so far been defined from a very restricted class of functions, namely, those in $C^1(\mathbb{R})$ that vanish sufficiently fast as we go out to ∞. To observe that it can be extended to a map from L^2 to L^2, we shall show that it defines an isometry. Then, remarks made earlier will show that we get a Hilbert transform from L^2 to L^2.

As usual, the Hilbert transform is far from honest on L^2; given a specific function in the L^2 class, there is no reason why its image should have a specific representative function.

Let's first prove boundedness properties of the Hilbert transform. For this, we do not need to know the specific form of the analytic function f, but only the fact that on the real line, the real part of f is h and the imaginary part of f is $H[h] = k$. Further, we also need to use the assumption we made that h goes to zero at a rate faster than inverse quadratic.

The fact that h decays at a rate faster than quadratic, tells us that f decays to zero at a rate faster than quadratic; in other words, $f(t) = O(|t|^{-2})$ for sufficiently large $|t|$. Thus, if we take semicircular contours strictly in the upper half plane whose diameter parts are very close to the real line, the contribution of the circular part is $O(1/r)$ for large enough r (a linear length is compensated for by an inverse quadratic modulus. We also know that the integral of f^2 along the boundary is 0 because the function is analytic on the boundary, so taking limits we see that for a horizontal line close to the real axis, the integral along the line limits to 0. Thus, we get that the integral of f^2 along the real axis is 0, so:

$$\int (h+ik)^2 = 0 \implies \int (h^2-k^2) = 0 \implies \|h\|_2 = \|k\|_2$$

Hence, the Hilbert transform is an isometry from a dense subspace of L^2, to within L^2, and hence extends to an isometry from L^2 to L^2.

8.4. **Hilbert transform versus Fourier transform.** The Hilbert transform and Fourier transform enjoy a close relation. To see this, consider the following alternative definition of Hilbert transform:

$$H[h] = x \mapsto \text{IFT}\left(\xi \mapsto \frac{1}{i}\text{sgn}(\xi)\hat{h}(\xi)\right)$$

The equivalence of this with the earlier definition follows from an alternative approach to constructing the analytic function in the upper half-plane.

Here IFT denotes the inverse Fourier transform. In other words, the Hilbert transform simply multiples the Fourier coefficients by a sign factor, and a $1/i$ factor.

It is easy to see, from either definition of the Hilbert transform, that the square of the Hilbert transform is the $-I$ map (negative identity; in other words, it sends h to $-h$. Intuitively, this is because the Hilbert transform is some sort of rotation by i in a larger, more invisible space.

8.5. **Boundedness of the Hilbert transform.** The Hilbert transform is not well-defined from L^1 to L^1 or from L^∞ to l^∞, but it turns out to be well-defined from L^p to L^p for any $1 < p < \infty$. We sketch the main proof ideas for this below:

- The Hilbert transform is an isometry from L^2 to L^2, because the square of the analytic function integrates to 0 on the real line. In fact for any even number $2j$, the expression f^{2j} integrates to 0 along the real line, and this, combined with AM-Gm inequalities yields that the Hilbert transform is a bounded linear operator from L^{2j} to L^{2j}.
- The Riesz-Thorin interpolation theorem yields that the Hilbert transform is well-defined from L^p to L^p for $2 \leq p < \infty$.
- The self-adjointness of the Hilbert transform tells us that if p and q are conjugate, $M(p,p) = M(q,q)$ for the Hilbert transform. This shows us that the Hilbert transform is well-defined and bounded as a linear operator from L^p to L^p.

Index