The Fourier Transform and Distribution Theory

Minicourse by Dr. Jared Wunsch
Transcribed by Collin Kofroth
Email: ckofroth@live.unc.edu

Contents

1 Opening Remarks

These notes were taken when attending the three-week semiclassical analysis summer school at Northwestern University, as part of SNAP ("Summer Northwestern Analysis Program"). These notes comprise the series of lectures on Fourier analysis taught by Dr. Jared Wunsch. Each section represents a lecture.

Any mistakes are my own. I've added in various clarifying remarks, additional background (some of which came from Chapter 3 of Michael Taylor's first PDE book, see here), and extra details which may generate further errors. Shoot me an email if/when you catch them!

2 An Introduction to the Fourier Transform on the Schwartz Space

We will start with a rather "dry" description of the Fourier transform, as motivating it later is actually easier after describing its key properties. To start, we want to define a particularly nice class of functions called the Schwartz space, denoted $\mathcal{S}(\mathbb{R}^n)$. Informally, these functions are nice in the sense that they are smooth, with all orders of derivatives decaying faster than any polynomial. Formally, we say

$$\varphi \in \mathcal{S}(\mathbb{R}^n) \iff \varphi \in C^\infty(\mathbb{R}^n) \text{ and } \forall \alpha, \beta \in \mathbb{N}^n \quad \sup_x |x^\alpha D^\beta \varphi| = C_{\alpha\beta} < \infty.$$

Here,

$$D^\beta = \frac{1}{i^{|\beta|}} \partial_1^{\beta_1} \partial_2^{\beta_2} \cdots \partial_n^{\beta_n}.$$

We also write $D_j = -i\partial_j$.

Example:

$$e^{-x^2} \in \mathcal{S}(\mathbb{R}^n), \qquad P(x)e^{-x^2} \in \mathcal{S}(\mathbb{R}^n), \qquad \frac{1}{1+x^2} \notin \mathcal{S}(\mathbb{R}),$$

where P is any polynomial.

Equivalently,

$$\begin{aligned} \varphi \in \mathcal{S} &\iff \forall \alpha, \beta \in \mathbb{N}^n \quad D^\beta x^\alpha \varphi \in L^\infty \\ &\iff \forall k \in \mathbb{N} \quad \sum_{|\alpha| \le k} \sup_x |\langle x \rangle^k D^\alpha \varphi| < \infty, \end{aligned}$$

where $\langle x \rangle := (1+x^2)^{1/2}$. The function $\langle x \rangle$ is called the *Japanese bracket*, and it is a smoothed-out, strictly positive version of $|x|$. We will call

$$\|\varphi\|_k = \sum_{|\alpha| \le k} \sup_x |\langle x \rangle^k D^\alpha \varphi|.$$

Note that this is a countable family of semi-norms, and one can utilize them to endow $\mathcal{S}(\mathbb{R}^n)$ with a metric space topology (in fact, a Fréchet topology) via the metric

$$d(\varphi, \psi) = \sum_{k=0}^{\infty} 2^{-k} \frac{\|\varphi - \psi\|_k}{1 + \|\varphi - \psi\|_k}.$$

Thus, we can say that

$$\varphi_n \to \varphi \text{ in } \mathcal{S} \iff \|\varphi_n - \varphi\|_k \to 0 \ \ \forall k.$$

The hardest part of showing that $\mathcal{S}$ is Frechét is showing that it is complete when endowed with the previously-described metric.

Proposition 2.1.

1. *$\mathcal{S}(\mathbb{R}^n)$ is complete*
2. *$C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$ is dense*

The first statement is proved similarly to the Arzela-Ascoli theorem. The latter can be proved by approximating any Schwartz function by a smoothly-cut-off version of itself.

Note: $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$. To see this, note that

$$\varphi \in \mathcal{S} \implies |\varphi| \leq C \langle x \rangle^{-1-n},$$

then integrate in polar coordinates. In particular, we write

$$\int |\varphi(x)| \, dx = \int \left(\langle x \rangle^{n+1} |\varphi(x)| \right) \frac{1}{\langle x \rangle^{n+1}} \, dx,$$

then use the Schwartzness of φ on the first terms in parentheses and integrability of the other term. Through a similar argument, one can show that

$$\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n), \qquad p \in [1, \infty].$$

Now, we will define the Fourier transform for an L^1 function (with the Schwartz space as a subspace). There are many conventions for the multiple of 2π in front of the integral, but we adopt the one which makes the Fourier transform unitary.

Definition 2.2. If $f \in L^1(\mathbb{R}^n)$, we define the Fourier transform as

$$(\mathcal{F}f)(\xi) = \hat{f}(\xi) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-ix\xi} \, dx.$$

Note: The Fourier transform is a continuous operator from L^1 to L^∞ (in fact, even better than L^∞ by the Riemann-Lebesgue lemma).

Note that the Schwartz space is closed under application by x_j and D_j. This leads to the following proposition.

Proposition 2.3. *If $\varphi \in \mathcal{S}(\mathbb{R}^n)$, then*

$$\mathcal{F}(D_j\varphi)(\xi) = \xi_j(\mathcal{F}\varphi)(\xi)$$
$$\mathcal{F}(x_j\varphi)(\xi) = -D_j(\mathcal{F}\varphi)(\xi).$$

We sometimes say that the Fourier transform *intertwines* differentiation and multiplication (by a polynomial). If we define the operator M_j by $M_jf(x) = x_jf(x)$, then the above can be re-stated as

Proposition 2.4.

$$\mathcal{F}D_j = M_j\mathcal{F}$$
$$\mathcal{F}M_j = -D_j\mathcal{F}.$$

The proof of the first part is a straightforward application of integration by parts, after using Fubini's theorem to write our integral over $\mathbb{R}^n$ as an iterated integral, and the proof of the second follows from the Leibniz integral rule.

Proposition 2.5. *$\mathcal{F} : \mathcal{S} \to \mathcal{S}$, and it is continuous.*

Proof. Roughly,

$$\xi^\alpha D^\beta \mathcal{F}\varphi(\xi) = \mathcal{F}(D^\beta(-x)^\beta\varphi)(\xi) \in L^\infty,$$

since

$$D^\beta(-x)^\beta\varphi \in \mathcal{S} \subset L^1.$$

Continuity is similar, using the sequential characterization. □

Let $T = J \circ \mathcal{F}^2$, where $Jf(x) = f(-x)$ is coordinate reflection.

Theorem 2.6. *$T : \mathcal{S} \to \mathcal{S}$ is the identity operator.*

Corollary 2.7. *$\mathcal{F}^{-1} = J \circ \mathcal{F}$. In particular, we have the Fourier inversion formula*

$$f(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \hat{f}(\xi)e^{ix\xi}\, d\xi.$$

(Proof of theorem). Due to time constraints, some of the proof will only be on $\mathbb{R}$, but the general proof is similar.

Step 1: $TD_j = D_jT$ and $TM_j = M_jT$

Check that, for example,

$$TD_j = J\mathcal{F}^2D_j = J\mathcal{F}M_j\mathcal{F} = -JD_j\mathcal{F}^2 = D_jJ\mathcal{F}^2 = D_jT.$$

Step 2: Fix $\gamma \in \mathcal{S}$ strictly positive. For any $\varphi \in \mathcal{S}$ and $y \in \mathbb{R}$, we can write, via Taylor's theorem,

$$\varphi(x) = \varphi(y)\frac{\gamma(x)}{\gamma(y)} + (x - y)r_y(x),$$

with $r_y \in C^\infty$. Note that the first term agrees with φ at $x = y$, and it is in $\mathcal{S}$. To see that $r_y \in \mathcal{S}$, note that away from $x = y$, we can write

$$r_y(x) = \frac{\varphi(x) - \varphi(y)\frac{\gamma(x)}{\gamma(y)}}{x - y}.$$

Away from $x = y$, $\frac{1}{x-y} \in L^\infty$, as are all of its derivatives, and $\varphi, \gamma \in \mathcal{S}$, so the $r_y \in \mathcal{S}$. Thus, we have written φ as a sum of two Schwartz functions, the first of which matches φ at $x = y$. Applying T to φ and using the previous lemma, we get that

$$T\varphi(x) = \frac{\varphi(y)}{\gamma(y)} T\gamma(x) + (x - y)Tr_y(x),$$

and evaluating at $x = y$ shows that

$$T\varphi(y) = \varphi(y)\frac{T\gamma(y)}{\gamma(y)} =: \varphi(y)h(y).$$

Hence, h exists so that for all $\varphi \in \mathcal{S}$,

$$T\varphi(x) = h(x)\varphi(x).$$

Since ∂_x commutes with T, we get that $h\varphi' = h'\varphi + h\varphi'$, implying that h is constant. So, $T\varphi = c\varphi$, for some $c \in \mathbb{C}$. The result now follows from the following proposition. □

Proposition 2.8.

$$\mathcal{F}(e^{-x^2/2}) = e^{-\xi^2/2}$$

Proof. If $u = e^{-x^2/2}$, then

$$(D - iM_x)u = 0.$$

Taking the Fourier transform yields the ODE

$$(M - i(-D))\hat{u} = 0,$$

and so

$$\hat{u} = \hat{u}(0)e^{-\xi^2/2},$$

with

$$\hat{u}(0) = \frac{1}{\sqrt{2\pi}} \int\limits_{\mathbb{R}} e^{-x^2/2} e^{-ix(0)}\, dx = 1.$$

□

3 Interpretations of the Fourier Transform and an Introduction to Tempered Distributions

We will write $(\cdot,\cdot)$ to denote the real L^2 inner product, and $\langle\cdot,\cdot\rangle$ to denote the sesquilinear inner product on L^2. One can readily use Fubini to compute that

$$(\mathcal{F}\varphi,\psi) = (\varphi,\mathcal{F}\psi),$$

and

$$\langle\mathcal{F}\varphi,\mathcal{F}\psi\rangle = (\mathcal{F}\varphi,\mathcal{F}^{-1}\bar{\psi}) = (\varphi,\bar{\psi}) = \langle\varphi,\psi\rangle\,.$$

Lemma 3.1. *$\mathcal{F}$ extends from $\mathcal{S}\subset L^2$ (densely) to a continuous linear map $\mathcal{F}: L^2\to L^2$, and it is unitary.*

To prove this, take an element of L^2 and an approximating sequence. This sequence is Cauchy in L^2 and by the above work, the Fourier transform of the Cauchy sequence converges in L^2, which implies that it converges. Define this limit as the Fourier transform, which is independent of choice of approximating sequence. The fact that $\mathcal{F}$ is unitary is called the *Plancheral theorem.*

Now, we will discuss motivations of the Fourier transform.

1. $\mathcal{F}$ exchanges information about smoothness and decay: For example, let $f\in L^1$ decrease quickly at ∞ (i.e. $x^\alpha f\in L^1$ for all α). Then,

$$(-D)^\alpha(\mathcal{F}f) = \mathcal{F}(x^\alpha f)\in L^\infty(\mathbb{R})\cap C(\mathbb{R}) \implies \mathcal{F}f\in C^\infty.$$

Conversely, if $D^\alpha f\in L^1$ for all α, then

$$\xi^\alpha\mathcal{F}f\in L^\infty \implies \mathcal{F}f = \mathcal{O}(\langle\xi\rangle^{-k})\ \ \forall k.$$

2. (A more physical interpretation) $\mathcal{F}$ decomposes "signals" into components with different frequencies: In the Fourier inversion formula, we can interpret $e^{ix\xi}$ as a plane wave with amplitude $\hat{f}(\xi)$, where large ξ yields rapid oscillations. This allows us to re-construct f via its frequencies. For example, say we have a violinist who plays a note at a frequency λ. We will assume that the rise and fall of the tone is Gaussian. For small positive ε, we will model this (interpret x as time) via

$$f(x) = e^{-\varepsilon x^2/2}e^{i\lambda x}.$$

So, this is a complex oscillation with frequency λ and envelope $e^{-\varepsilon x^2/2}$, decaying very slowly. Then,

$$\hat{f}(\xi) = ce^{-(\xi-\lambda)^2/2\varepsilon}.$$

This is a Gaussian centered at $\xi=\lambda$ which is highly peaked and rapidly decaying (small ε gives a narrow peak here, as opposed to a wide envelope for f). So, we get a huge spike at one frequency, namely λ.

3. Solving PDEs: We have the nice property that $\mathcal{F}D^\alpha = \xi^\alpha \mathcal{F}$. Suppose that $p(\xi)$ is a polynomial in n variables, say $p(\xi) = \sum_{|\alpha|\le m} a_\alpha \xi^\alpha$. Given this polynomial, we can define the object

$$p(D) = \sum_{|\alpha|\le m} a_\alpha D^\alpha,$$

which is a constant-coefficient differential operator. Then,

$$\mathcal{F}(p(D)\varphi) = p(\xi)\mathcal{F}\varphi.$$

This can turn PDE problems to, in a way, problems in real algebraic geometry. This is a pretty easy problem to solve when the polynomial has no real zeros. For example, let

$$\Delta = -\sum_j D_j^2 = \sum_j \partial_j^2,$$

and say we want to solve $(I - \Delta)\varphi = \psi$, where $\varphi, \psi \in \mathcal{S}$. The above holds

$$\iff (1 + |\xi|^2)\hat{\varphi} = \hat{\psi} \iff \hat{\varphi} = \frac{1}{1+|\xi|^2}\hat{\psi} \iff \varphi = \mathcal{F}^{-1}\left(\frac{1}{1+|\xi|^2}\mathcal{F}\psi\right).$$

That is, there must exist a unique solution in $\mathcal{S}$. We call the above a *Fourier multiplier*. This allows us to define a functional calculus:

$$f(\Delta)\varphi = \mathcal{F}^{-1} f(-|\xi|^2)\mathcal{F}\varphi,$$

provided f is nice (e.g. bounded with bounded derivatives, or even with polynomial growth).

Next, we will begin distribution theory. The continuous dual space of $\mathcal{S}$ is called the space of tempered distributions, denoted

$$\mathcal{S}' = \{u : \mathcal{S} \to \mathbb{C} : \ u \text{ is linear and continuous}\},$$

where continuity is equivalent to the statement

$$\exists k \in \mathbb{N}\, \exists C \in \mathbb{R}^+\, \forall \varphi \in \mathcal{S} \ \ |u(\varphi)| \le C \|\varphi\|_k,$$

where (as before)

$$\|\varphi\|_k = \sum_{|\alpha|\le k} \sup_x \langle x\rangle^k |D^\alpha \varphi|.$$

The order of a distribution is the number of derivatives needed to ensure continuity. One can, equivalently, use the sequential criterion:

$$u \in \mathcal{S}' \iff u : \mathcal{S} \to \mathbb{C} \text{ is linear and } u(\varphi_j) \to u(\varphi) \text{ whenever } \varphi_j \to \varphi \text{ in } \mathcal{S}.$$

There are more general notions of distributions (the dual space of test functions), but the topology is messy, and tempered distributions are more fundamental to the Fourier transform. One cannot take the Fourier transform of a generic non-tempered distribution since the Fourier transform of a C_c^∞ function cannot be C_c^∞. For this reason, we will not discuss them further.

Example: Suppose that $f \in L^1(\mathbb{R}^n)$, and define

$$u_f(\varphi) = \int f\varphi\, dx.$$

This defines a tempered distribution, as

$$|u_f(\varphi)| \leq \|f\|_{L^1} \|\varphi\|_0 .$$

We often do away with the above notation and use the same symbol for a function and the distribution that it induces. We could have also take $f \in L^1_{loc}$ with polynomial growth.

Example: Let μ be a finite, positive (or even complex-valued) Borel measure. Then, it induces a tempered distribution via the map

$$\varphi \mapsto \int \varphi\, d\mu,$$

with a similar estimate to the previous example. A particular example is the Dirac delta distribution: $\delta(\varphi) = \varphi(0)$, which is induced by the Dirac measure:

$$\delta_x(A) = \chi_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases},$$

and we have that

$$\int f\, d\delta_x = f(x).$$

We write the distributional pairing $(u, \varphi) = u(\varphi)$ for $u \in \mathcal{S}'$ and $\varphi \in \mathcal{S}$. This has many conventions, such as a being a right action or using angled brackets. $\mathcal{S}'$ is endowed with the weak* topology, meaning that

$$u_n \to u \text{ in } \mathcal{S}' \iff (u_n, \varphi) \to (u, \varphi) \quad \forall \varphi \in \mathcal{S}.$$

We can continuously extend many operations from $\mathcal{S}$ to $\mathcal{S}'$. We first consider the pairing of two Schwartz functions. For example, for all $\varphi, \psi \in \mathcal{S}$,

$$(\varphi', \psi) = -(\varphi, \psi')$$

via an integration by parts. This motivates the definition of the distributional derivative as

$$(u', \varphi) = -(u, \varphi')$$

for $u \in \mathcal{S}', \varphi \in \mathcal{S}$.

Exercise: Check that this defines an element of $\mathcal{S}'$ if u does.

Example: Let H be the Heaviside function. Then, H induces a distribution via acting by integration:

$$(H, \varphi) = \int_0^\infty \varphi(x)\, dx.$$

Compute directly that

$$(H', \varphi) = -(H, \varphi') = -\int_0^\infty \varphi'(x)\, dx = \varphi(0) = (\delta, \varphi) \implies H' = \delta.$$

We can similarly extend many other operations, such as multiplication by x, multiplication by a function (smooth with polynomial growth in all derivatives), translation, and scaling ($\varphi \mapsto \varphi(\lambda \cdot)$). We can also define the Fourier transform of tempered distributions via duality:

$$(\mathcal{F}u, \varphi) = (u, \mathcal{F}\varphi).$$

The inverse Fourier transform extends in the same way (and it is still the genuine inverse), and these maps are continuous in the weak* topology.

Example:

$$(\mathcal{F}\delta, \varphi) = (\delta, \mathcal{F}\varphi) = \hat{\varphi}(0) = (2\pi)^{-n/2} \int \varphi(x)\, dx = ((2\pi)^{-n/2}, \varphi) \implies \hat{\delta} = (2\pi)^{-n/2}.$$

Then, via the Fourier inversion formula,

$$\mathcal{F}1 = (2\pi)^{n/2}\delta.$$

More generally,

$$(\mathcal{F}e^{i\lambda x}, \varphi) = (e^{i\lambda x}, \hat{\varphi}) = \int e^{i\lambda x}\hat{\varphi}(x)\, dx = (2\pi)^{n/2}\mathcal{F}^{-1}(\hat{\varphi})(\lambda) = (2\pi)^{n/2}\varphi(\lambda).$$

So,

$$\mathcal{F}e^{i\lambda x}(\xi) = (2\pi)^{n/2}\delta(\xi - \lambda).$$

4 Schwartz Kernels and Sobolev Spaces I

If we take a smooth cut-off function, then it is easy to see that

$$\chi(\varepsilon^{-1}\cdot)\varphi \to \varphi$$

as $\varepsilon \to 0$ in $\mathcal{S}$, giving density of C_c^∞ in $\mathcal{S}$. By duality, one can check that

$$\chi(\varepsilon^{-1}\cdot)u \to u$$

in $\mathcal{S}'$ for $u \in \mathcal{S}'$. Of course, this tells us something about the Fourier transform, as now

$$\mathcal{F}^{-1}\chi(\varepsilon^{-1}\cdot)\mathcal{F}u \to u$$

in $\mathcal{S}'$ by continuity of $\mathcal{F}$ on $\mathcal{S}'$. To check that this makes sense,

$$u \in \mathcal{S}' \implies \mathcal{F}u \in \mathcal{S}' \implies \chi(\varepsilon^{-1}\cdot)\mathcal{F}u \in \mathcal{S}' \implies \mathcal{F}^{-1}\chi(\varepsilon^{-1}\cdot)\mathcal{F}u \in \mathcal{S}'.$$

In particular, $\chi(\varepsilon^{-1}\cdot)\mathcal{F}u$ is a compactly-supported tempered distribution. Recall that a distribution is compactly supported if its support is compact, where the support of a distribution is defined as the complement of the union of all open sets on which the distribution vanishes (and we say that a distribution vanishes on an open set if its action on all functions supported on that set is zero). That is, a point is in the support if the distribution is not the zero distribution when acting on test functions supported on any neighborhood of the point. We call

$$\mathcal{E}' = \{u \in \mathcal{S}' : \operatorname{supp} u \text{ is compact}\}.$$

In fact, we can pair with any C^∞ function (it is actually the dual of C^∞). That is, $\mathcal{E}' = (C^\infty)^*$, where C^∞ is endowed with the Frechét topology generated by the semi-norms

$$p_{R,k}(u) = \sup_{|x|\le R} \sum_{|\alpha|=1}^{k} |D^\alpha u|.$$

For $f \in C^\infty$, then we define $u \in \mathcal{E}'$ via the pairing

$$(u, f) = (u, \chi f),$$

where χ is a smooth cut-off that is 1 on $\operatorname{supp} u$. This definition is independent of the choice of cut-off, since all χf agree on $\operatorname{supp} u$. This actually allows us to extend $u \in \mathcal{S}'$ to a linear functional on C^∞. Since $\mathcal{E}' \subset \mathcal{S}'$, we can take the Fourier transform of elements of $\mathcal{E}'$. We can actually say more.

Lemma 4.1. *If $u \in \mathcal{E}'$, then we can make sense of the Fourier transform as a function:*

$$\mathcal{F}u(\xi) = (u, (2\pi)^{-n/2}e^{-i\xi x}) = (2\pi)^{-n/2}(u, \chi(x)e^{-ix\xi}) \in C^\infty(\mathbb{R}^n).$$

In fact, this is analytic, and it extends to an entire function on $\mathbb{C}^n$ with an exponential bound (see the Paley-Wiener-Schwartz theorem).

Corollary 4.2.

$$\mathcal{F} : \mathcal{E}' \to C^\infty$$

So, $\mathcal{F}^{-1}\chi(\varepsilon^{-1}\cdot)\mathcal{F}u \to u$, as discussed earlier, and the left is C^∞. Throwing in another cut-off, we get that

$$\chi((\varepsilon')^{-1}\cdot)\mathcal{F}^{-1}\chi(\varepsilon^{-1}\cdot)\mathcal{F}u \to u$$

in $\mathcal{S}'$ as $\varepsilon, \varepsilon' \to 0$. Hence, C_c^∞ is dense in $\mathcal{S}'$. Since $C_c^\infty \subset \mathcal{S}$, this tells us that $\mathcal{S}$ is dense in $\mathcal{S}'$. This, in turn, implies that the extension of operations from $\mathcal{S}$ to $\mathcal{S}'$ is unique. This can also tell us, for example, that the Fourier transform of an L^1 function f defined by integration is the same as when we consider f as a distribution T_f.

Back to PDE's: Consider

$$\begin{cases} (iD_t - \Delta)u = 0 \\ \qquad u(0, x) = f \end{cases}$$

In the Schwartz space, we have the solution

$$u = \mathcal{F}^{-1} e^{-t|\xi|^2} \mathcal{F} f.$$

This makes sense now if $f \in \mathcal{S}'$. Indeed, $f \in \mathcal{S}' \implies \mathcal{F}f \in \mathcal{S}'$, and the growth of the exponential and its derivatives guarantees that its product with $\mathcal{F}f$ is tempered, then the inverse transform continues to keep us in the space. We would like to understand this operator better. Define the *heat propagator*

$$e^{t\Delta} : f \mapsto \mathcal{F}^{-1} e^{-t|\xi|^2} \mathcal{F} f.$$

To understand this, we need to understand the inverse Fourier transform of a product.

Definition 4.3. If $\varphi, \psi \in \mathcal{S}$, then we define their *convolution* as

$$(\varphi * \psi)(x) = \int \varphi(x-y)\psi(y)\, dy.$$

This has a nice physical interpretation as a moving average of the two functions. It is a bilinear, symmetric operator $\mathcal{S} \times \mathcal{S} \to \mathcal{S}$. It can be extended continuous to $\mathcal{S} \times \mathcal{S}' \to \mathcal{S}'$, by duality (in fact, one can extend it further to a map $\mathcal{S} \times \mathcal{E}' \to \mathcal{S}$ and $\mathcal{E}' \times \mathcal{S}' \to \mathcal{S}'$). Typically, one defines $(\varphi * u, \psi) = (u, \tilde{\varphi} * \psi)$, where $\tilde{\varphi}(x) = \varphi(-x)$. One can readily check that this is consistent with the definition given for Schwartz functions. For symmetry, we merely record

Lemma 4.4. $\varphi * \psi = \psi * \varphi$

To prove this, simply change variables. A quick application of Fubini yields the key property of convolution.

Lemma 4.5. $\mathcal{F}(\varphi * \psi) = (2\pi)^{n/2}(\mathcal{F}\varphi)(\mathcal{F}\psi)$

Hence, we can write

$$e^{t\Delta} f = \mathcal{F}^{-1} e^{-t|\xi|^2} \mathcal{F} f = K * f,$$

where

$$K = (2\pi)^{-n/2} \mathcal{F}^{-1} e^{-t|\xi|^2}.$$

We call K is the *heat kernel*. In particular,

$$K(x,t) = (4\pi t)^{-n/2} e^{-|x|^2/4t} \implies e^{t\Delta} f = \int (4\pi t)^{-n/2} e^{-|x-y|^2/4t} f(y)\, dy,$$

for a function f. Many answers in PDE's have solutions in this form. So, our solution has the form

$$e^{t\Delta} f = \int K(x-y) f(y)\, dy$$

In greater generality, we may have a solution

$$\int K(x,y) f(y)\, dy,$$

where K is a function on $\mathbb{R}^{2n}$. For example, if $K \in L^2(\mathbb{R}^{2n})$, then this gives an operator on L^2 functions, which is an example of a Hilbert-Schmidt operator (this is actually exhaustive, which is the content of the so-called Hilbert-Schmidt kernel theorem).

Now, let us take $K \in \mathcal{S}'(\mathbb{R}^{2n})$. Define

$$T_K(\varphi) = \text{``}\int K(x,y)\varphi(y)\,dy\text{''}$$

by

$$(T_K\varphi, \psi)_{\mathbb{R}^n} = (K, \psi \otimes \varphi), \qquad (\psi \otimes \varphi)(x,y) = \varphi(x)\psi(y).$$

Note that this is consistent with what would happen if K were some nice function. So, each $K \in \mathcal{S}'(\mathbb{R}^{2n})$ defines a continuous, linear operator $T_K : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$. This turns out to be exhaustive.

Theorem 4.6 (Schwartz Kernel Theorem)**.** *Let $T : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$. If T is linear and continuous, then there exists $K \in \mathcal{S}'(\mathbb{R}^{2n})$ so that $T = T_K$. i.e.*

$$\text{``}T\varphi(x) = \int K(x,y)\varphi(y)\,dy\text{''}.$$

We often do not distinguish between an operator and its Schwartz kernel, since they are in one-to-one correspondence. The proof of this result is quite non-trivial, and we will omit it.

Now, we move on to a discussion on Sobolev spaces. We are especially interested in L^2-based Sobolev spaces.

Definition 4.7. We define the Sobolev space $H^k(\mathbb{R}^n)$ for $k \in \mathbb{N} \cup \{0\}$ as

$$H^k(\mathbb{R}^n) = \{u \in \mathcal{S}'(\mathbb{R}^n) : D^\alpha u \in L^2 \quad \forall |\alpha| \le k\}.$$

When we say $D^\alpha u \in L^2$, we mean that the distributional pairing is represented by integration against and L^2 function. Note that

$$u \in H^k \iff \hat{u} \in L^2 \text{ and } \xi^\alpha \hat{u} \in L^2 \ \forall\ |\alpha| \le k \iff \langle\xi\rangle^k \hat{u} \in L^2,$$

and the last statement makes sense for any $s \in \mathbb{R}$.

Definition 4.8. We define the Sobolev space $H^s(\mathbb{R}^n)$ for $s \in \mathbb{R}$ as

$$H^s(\mathbb{R}^n) = \{u \in \mathcal{S}'(\mathbb{R}^n) : \langle\xi\rangle^s \hat{u} \in L^2\}.$$

We can see that $s > 0 \implies H^s \subset L^2 = H^0$, $s < 0 \implies L^2 \subset H^s$, which further implies that $H^s \subset H^t$ for all $s > t$ (this subset notation should be understood as meaning that the inclusion map is injective and continuous) These are all Hilbert spaces, as they are L^2-based. They are endowed with the inner product

$$\langle u, v\rangle_{H^s} = \int \langle\xi\rangle^{2s}\,\hat{u}(\xi)\overline{\hat{v}}(\xi)\,d\xi.$$

We also remark that for $s > 2n$, H^s forms a Banach algebra.

Consider the Fourier multiplier

$$\Lambda_s = (1 - \Delta)^{s/2},$$

which acts via the Fourier multiplier functional calculus as

$$\Lambda_s f = \mathcal{F}^{-1}\left(\langle\xi\rangle^s \hat{f}(\xi)\right).$$

Since $\mathcal{F}$ is an isometric isomorphism, one could equivalently define H^s as the tempered distributions u so that $\Lambda_s u \in L^2$, or $H^s = \Lambda_{-s}L^2$. In particular, this is an isometric isomorphism

$$\Lambda_s : H^s \to L^2,$$

and

$$\Lambda_s : H^t \to H^{t-s}$$

in the same manner.

One can readily prove the following proposition.

Proposition 4.9. *For all $s \in \mathbb{R}$, $D^\alpha : H^s \to H^{s-|\alpha|}$*

Furthermore,

Lemma 4.10. *Suppose that $f \in H^1$, so that $\partial_{x_j} f \in L^2$ for each j (this denotes the distributional derivative). Then,*

$$\lim_{k\to 0} \frac{f(x + ke_j) - f(x)}{k} \to \partial_{x_j} f(x)$$

in L^2.

So, the difference quotients converge in L^2 to the distributional derivative. Next, we record a result on *duality*.

Proposition 4.11. $(H^s)^* = H^{-s}$, *where the dual is taken with respect to the L^2 inner product. Hence, any $\alpha : H^s \to \mathbb{C}$ continuous and linear has the form*

$$\alpha(u) = \langle u, v\rangle_* := \langle \hat{u}, \hat{v}\rangle,$$

for a unique $v \in H^{-s}$.

Note that the dual is taken with respect to L^2 (the above might have set off alarm bells, as Hilbert spaces are reflexive with respect to their inner product).

Remark 4.12. Let me elaborate on this pairing. On $\mathcal{S}$, this pairing is the same as the standard L^2 pairing via Plancherel. For $u, v \in \mathcal{S}$, we apply Cauchy-Schwarz to get

$$|\langle u, v\rangle| = |\langle \Lambda_s u, \Lambda_{-s} v\rangle| \le \|u\|_{H^s} \|v\|_{H^{-s}}.$$

So, this is a continuous bilinear form which extends by density to H^s and H^{-s}. I will not show that this pairing gives an isometric isomorphism between H^{-s} and $(H^s)^*$ via the map $v \mapsto \langle\cdot, v\rangle_*$; see page 122 of Friedlander and Joshi or page 302 in Folland.

It will be nice to know how these Sobolev spaces relate to more classical spaces of functions, especially those which record regularity.

Proposition 4.13 (Sobolev Embedding). *$H^s(\mathbb{R}^n) \hookrightarrow C^k(\mathbb{R}^n)$ if $s > n/2 + k$.*

For example, $H^s \hookrightarrow C^0$ if $s > n/2$.

Proof. Both are Banach spaces, so it suffices to show that the inclusion map is continuous on a dense subset. We only do $k = 0$. It will suffice to show that there exists $C \in \mathbb{R}^+$ so that for all $\varphi \in \mathcal{S}$,

$$\sup |\varphi| \leq C \|\varphi\|_{H^s}.$$

Write

$$\varphi(x) = (2\pi)^{-n/2} \int e^{ix\xi} \hat{\varphi}(\xi)\, d\xi = (2\pi)^{-n/2} \int (\langle \xi \rangle^{-s} e^{ix\xi})(\langle \xi \rangle^{s} \hat{\varphi}(\xi)\, d\xi.$$

Applying the Cauchy-Schwarz inequality,

$$\sup |\varphi| \leq C \|\varphi\|_s.$$

We used that

$$s > n/2 \implies \langle \xi \rangle^{-s} e^{ix\xi} \in L^2.$$

□

Remark 4.14. I prefer the following proof:

We start with $k = 0$. For any $f \in H^s$, with $s > n/2$, we note that

$$\|\hat{f}\|_{L^1} \lesssim \|f\|_{H^s}.$$

Since $\hat{f} \in L^1$, the Fourier inversion theorem and Riemann-Lebesgue lemma tell us that $f \in C_0^0$, with

$$\|f\|_{L^\infty} \lesssim \|f\|_{H^s}.$$

Now, suppose that $f \in H^s$ and $s > k + n/2$, $k \in \mathbb{N}$. Given $|\alpha| \leq k$, we know that $D^\alpha : H^s \to H^{s-|\alpha|}$. Applying the $k = 0$ case to $D^\alpha f$, we have that $D^\alpha f \in C_0^0$ and

$$\|D^\alpha f\|_{L^\infty} \lesssim \|f\|_{H^{s-|\alpha|}} \lesssim \|f\|_{H^s}.$$

We can see the continuity in the following way, as well. By density, there exists a sequence of Schwartz functions $\{\varphi_j\}$ so that $\varphi_j \to D^\alpha f$ in $H^{s-|\alpha|}$. Since $\varphi_j \in H^{s-|\alpha|}$. By the above inequality, we have that $\varphi_j \to D^\alpha f$ uniformly, which yields continuity.

Technically, this only shows that the weak derivatives of f are continuous. General weak derivative theory tells us that we are good (e.g. if $D_j f = u$ and u is continuous, then f is a.e. C^1), but we can also note the following:

By density, there exists a sequence of Schwartz function f_j so that $f_j \to f$ in H^s. By the estimate, $f_j \to g$ in C^k, due to completeness. In particular, $f_j \to f$ in L^2, and $f_j \to g$ in C_0^0; thus,

$$\int f_j \varphi \to \int f \varphi, \qquad \forall \varphi \in \mathcal{S}$$
$$\int f_j \varphi \to \int g \varphi, \qquad \forall \varphi \in \mathcal{S}$$

implying that $f = g$.

Note that the proof actually gave us an embedding into C_0^k.

Corollary 4.15.

$$\bigcap_{s \in \mathbb{R}} H^s \subset C^\infty,$$

and

$$\bigcup_{s \in \mathbb{R}} H^s \supset \mathcal{E}'.$$

Another important embedding is the compact embedding of higher-order Sobolev spaces into lower-order ones.

Theorem 4.16 (Rellich-Kondrachov)**.** *Take $s > t$, K compact, and consider*

$$S = \{u \in H^s(\mathbb{R}^n) : \operatorname{supp} u \subset K, \|u\|_s \leq 1\}.$$

Then, S is compact in H^t.

Remark 4.17. The same is true. for $\langle x \rangle^{-\alpha} H^s \hookrightarrow \langle x \rangle^{-\beta} H^t$ for $\alpha > \beta$ and $s > t$. This is nice, as it is expressed in terms of both decay and regularity.

We will prove this next time, but let us talk about it a bit. We know that the closed unit ball is never compact in an infinite-dimensional Banach space. There are two primary obstructions to compactness in L^2: functions marching off to infinity in space or in frequency (both are issues by Plancherel). If we support things in space and control the high frequencies in some way (better than L^2 in frequency), then we get compactness in a weaker space.

5 Sobolev Spaces II: Proof of Rellich's Theorem, Complex Interpolation, and Applications

To start off, we proceed with the proof of the Rellich-Kondrachov theorem.

Proof of Theorem 4.16. Let $u_j \in S$ be bounded in H^t. WLOG, take $t = 0$ (otherwise, use the Λ_s action). We want a subsequence converging in L^2. Passing to a subsequence (by the Banach-Alaoglu theorem) $u_j \to u \in S$ weakly in H^s. We want to show that $u_j \to u$ (i.e. strongly).

Note that

$$w \in S \subset \mathcal{E}' \implies \hat{w}(\xi) = (w, (2\pi)^{-1/2} e^{ix\xi}) = (w, (2\pi)^{-1/2} \chi e^{ix\xi}), \qquad \chi \in C_c^\infty, \chi \equiv 1 \text{ on } \operatorname{supp} w.$$

Then,

$$|\hat{w}(\xi)| \leq C \|w\|_{H^s} \left\|\chi e^{i(\cdot)\xi}\right\|_{H^{-s}} = C \left\|\chi e^{i(\cdot)\xi}\right\|_{H^{-s}},$$

and

$$\begin{aligned} \left\|\chi e^{i(\cdot)\xi}\right\|_{H^{-s}} &= \int |\widehat{\chi e^{ix\xi}}(\eta)|^2 \langle\eta\rangle^{-2s} \, d\eta = \int |\hat{\chi}(\eta - \xi)|^2 \langle\eta\rangle^{-2s} \, d\eta \\ &= \mathcal{O}(\langle\xi\rangle^{-2s}). \end{aligned}$$

Thus,

$$|\hat{w}(\xi)| \leq C \langle\xi\rangle^{-s}.$$

If we apply this to u_j, then we get that $\hat{u}_j$ is uniformly bounded on compact subsets of $\mathbb{R}^n$. One can do the same for partial derivatives in ξ, in which case the Arzela-Ascoli theorem guarantees the existence of a locally uniformly convergence subsequence. Since

$$u_j \overset{w}{\rightharpoonup} u \implies (u_j, e^{ix\xi}) \to (u, e^{ix\xi}) \implies \hat{u}_j(\xi) \to \hat{u}(\xi) \qquad \forall \xi$$

passing to our subsequence gives that

$$\hat{u}_j \to \hat{u} \text{ uniformly.}$$

Fix $\varepsilon > 0$. Then, there exists $R > 0$ such that

$$\langle R\rangle^{-2s} < \frac{\varepsilon}{4}.$$

By Plancherel,

$$\begin{aligned} \|u_j - u\|_{L^2}^2 &= \int |\hat{u}_j(\xi) - \hat{u}(\xi)|^2 \, d\xi = \int\limits_{|\xi|>R} |\hat{u}_j(\xi) - \hat{u}(\xi)|^2 \, d\xi + \int\limits_{|\xi|\leq R} |\hat{u}_j(\xi) - \hat{u}(\xi)|^2 \, d\xi \\ &=: I_1 + I_2. \end{aligned}$$

Now,

$$I_2 = \int\limits_{|\xi|>R} |\hat{u}_j - \hat{u}|^2 \langle\xi\rangle^{-2s} \langle\xi\rangle^{2s} \, d\xi \leq \langle R\rangle^{-2s} \|u_j - u\|_{H^s}^2 < \frac{\varepsilon}{4} \|u_j - u\|_{H^s}^2 \leq \frac{\varepsilon}{2}.$$

For I_2, we simply use the established uniform convergence to get that there exists $N \in \mathbb{N}$ such that

$$I_2 < \frac{\varepsilon}{2}.$$

Thus, $u_j \to u$ in L^2. ☐

We will change gears a bit. Recall that if

$$p(\xi) = \sum_{|\alpha|\le m} a_\alpha \xi^\alpha, \qquad a_\alpha \in \mathbb{C},$$

then

$$p(D) : H^s \to H^{s-m}$$

for all $s \in \mathbb{R}$. We are interested in studying the action of variable-coefficient differential operators

$$P \in \mathrm{Diff}^m(\mathbb{R}^n), \qquad P = \sum_{|\alpha|\le m} a_\alpha(x) D^\alpha, \qquad a_\alpha \in C_b^\infty(\mathbb{R}^n).$$

In particular, we pose the following question:

Question: Is it still true that $P : H^s \to H^{s-m}$?

If P is the map $\varphi \mapsto a\varphi$, with $a \in C_b^\infty$, then $P : H^s \to H^s$ for any $s \in \mathbb{N}$. For example, if $s = 1$, and $u \in H^1$, then

$$\|au\|_{H^1} \le \|au\|_{L^2} + \|\nabla(au)\|_{L^2} \lesssim \|u\|_{L^2} + \|\nabla u\|_{L^2} + \|u\|_{L^2} < \infty.$$

The argument is similar for higher natural numbers (induct). Consider P to be given in the form in our stated question. Note that $a_\alpha(x)D^\alpha = M_{a_\alpha} \circ D^\alpha$ and $D^\alpha : H^s \to H^{s-|\alpha|}$ for all $s \in \mathbb{R}$. This coupled with the work for M_{a_α} gives the result for natural s, and we can see that, for general $s \in \mathbb{R}$, our question boils down to showing that $M_a : H^s \to H^s$ for all $s \in \mathbb{R}$. We will show this using the method of *complex interpolation*.

We will start with $s \in [0, 1]$. One can natural extend the work to the non-negative reals, then use a duality argument to extend to $s \in \mathbb{R}$. We already know that result when $s = 0, 1$. Notice that it will suffice to prove that $T_s := \Lambda_s M_a \Lambda_{-s} : L^2 \to L^2$ for $s \in (0, 1)$.

Key idea: T_z makes sense for all $z \in \mathbb{C}$ and is analytic in z.

Since $\|T_z\| = \sup_{\|\psi_1\|=\|\psi_2\|=1} |\langle T_z\psi_1, \psi_2\rangle|$ for $\psi_1, \psi_2 \in L^2$ and $\mathcal{S}$ is dense in L^2, it will suffice to show that for all $\psi_1, \psi_2 \in \mathcal{S}$

$$|\langle T_z\psi_1, \psi_2\rangle| \lesssim \|\psi_1\| \, \|\psi_2\|,$$

where the $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote the L^2 inner product and norm, respectively. We already have the result if $\operatorname{Re} z = 0, 1$, since $\Lambda_{x+iy} = \Lambda_x \Lambda_{iy}$, and Λ_{iy} is unitary on L^2. To ease notation, let us call

$$f = f_{\psi_1,\psi_2} = \langle T_z\psi_1, \psi_2\rangle .$$

If we call $\Omega = \{z \in \mathbb{C} : \operatorname{Re} z \in [0, 1]\}$, then we already know that $|f(z)| \lesssim \|\psi_1\| \, \|\psi_2\|$ for $z \in \partial\Omega$. So, f is scalar-valued and holomorphic on Ω (in fact, it is entire).

Theorem 5.1 (Hadamard 3-lines theorem)**.** *Suppose that f is analytic in $\operatorname{int}\Omega$ and both continuous and bounded on Ω. Set $M(x) = \sup_{y\in\mathbb{R}} |f(x + iy)|$. Then, $\log M(x)$ is convex. Equivalently,*

$$M(x) \le (M(0))^{1-x}(M(1))^x.$$

Proof. Set
$$g(z) = f(z)((M(0))^{-1+z}(M(1))^{-z}).$$
We must show that $|g(z)| \leq 1$, and we already know that $|g(iy)| \leq 1$ and $|g(1+iy)| \leq 1$. For each $n \in \mathbb{N}$, set
$$g_n(z) = g(z)e^{(z^2-1)/n}.$$
Since $|g_n(z)| = |g(z)|e^{(x^2-y^2-1)/n}$, it follows that $|g_n(z)| \to 0$ as $|z| \to \infty$ for all $z \in \partial\Omega$. By the maximum modulus principle,
$$\sup_{z\in\Omega} |g_n(z)| = \sup_{z\in\partial\Omega} |g_n(z)| \leq \sup_{z\in\partial\Omega} |g(z)| \leq 1.$$
Since $g_n(z) \to g(z)$ as $n \to \infty$ for all $z \in \Omega$, it follows that $|g(z)| \leq 1$ for all $z \in \Omega$. □

In order to apply this to f, we must check that f is bounded on $\partial\Omega$. But, this is easy: fix $z \in \partial\Omega$, and notice that
$$|f(z)| \leq \|T_z\psi_1\| \, \|\psi_2\| = \|\Lambda_z M_a \Lambda_{-z}\psi_1\| \, \|\psi_2\|$$
If $\operatorname{Re} z = 0$, then
$$|f(z)| \leq \|M_a\Lambda_0\psi_1\| \, \|\psi_2\| \, ,$$
and if $\operatorname{Re} z = 1$,
$$|f(z)| \leq \|M_a\Lambda_{-1}\psi_1\|_{H^1} \, \|\psi_2\| \, .$$
Recall that $M_a : H^s \to H^s$ if $s = 0$ and or $s = 1$. If $\operatorname{Re} z = 0$, then we use that $s = 0$ case to get that
$$|f(z)| \leq C_0 \, \|\Lambda_0\psi_1\| \, \|\psi_2\| = C_0 \, \|\psi_1\| \, \|\psi_2\| \, ,$$
and if $\operatorname{Re} z = 1$, then we use the $s = 1$ case to get that
$$|f(z)| \leq C_1 \, \|\Lambda_{-1}\psi_1\|_{H^1} \, \|\psi_2\| = C_1 \, \|\psi_1\| \, \|\psi_2\| \, .$$
By Theorem 5.1, we conclude via the least upper bound property that
$$|f(x+iy)| \leq C_0^{1-x} C_1^x \, \|\psi_1\| \, \|\psi_2\|$$
for all $x \in [0,1]$. In particular,
$$|f(s)| \leq C \, \|\psi_1\| \, \|\psi_2\|$$
for all $s \in [0,1]$.

For $s \in \mathbb{R}_+$, one can employ the same type of argument (using boundedness on H^k for an appropriate $k \in \mathbb{N}$). If $-s \geq 0$ and $u \in H^{-s}$, then
$$\|M_a u\|_{-s} = \sup_{\substack{v\in H^s \\ \|v\|_s=1}} | \langle M_a u, v\rangle_{L^2} | = \sup_{\substack{v\in H^s \\ \|v\|_s=1}} | \langle u, M_a v\rangle_{L^2} | \lesssim \|u\|_{H^{-s}} \, \|v\|_{H^s} = \|u\|_{H^{-s}} \, .$$
Thus, we have answered our question in the affirmative: variable-coefficient differentiable operators of order m (with smooth coefficients bounded in all derivatives) send H^s to H^{s-m}, just as in the constant coefficient case. We will record this in the following theorem.

Theorem 5.2. *Let $P \in \mathit{Diff}^m(\mathbb{R}^n)$ be given by $P = \sum_{|\alpha|\le m} a_\alpha(x) D^\alpha$, where $a_\alpha \in C_b^\infty(\mathbb{R}^n)$. Then,*

$$P : H^s \to H^{s-m}.$$

Another application of complex interpolation is the following: recall that the Fourier transform $\mathcal{F}$ sends L^2 to L^2 and L^1 to L^∞. Using complex interpolation, one can prove the result for all p in between.

Theorem 5.3 (Hausdorff-Young)**.** *For any $p \in [1,2]$,*

$$\mathcal{F} : L^p(\mathbb{R}^n) \to L^{p'}(\mathbb{R}^n), \qquad \frac{1}{p} + \frac{1}{p'} = 1.$$

For a final application, we claim that Sobolev spaces are invariant under coordinate changes. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism, and let us say that $\Phi \equiv 1$ outside of a ball of radius R. If $u \in L^2$, then

$$\|\Phi^* u\|_{L^2}^2 = \int |u(\Phi(x))|^2 \, dx = \int |u(y)|^2 |\det D\Phi^{-1}(y)| \, dy < \infty,$$

since $(D\Phi^{-1})_{ij} \in C_b^\infty$. If $u \in H^1$, then we note that

$$\int \left| \frac{\partial (u \circ \Phi)}{\partial x_j} \right|^2 dx = \int \sum_{k=1}^n \left| \frac{\partial u}{\partial y_k} \frac{\partial \Phi_k}{\partial x_j} \right|^2 \det D\Phi^{-1}(y)| \, dy < \infty,$$

since

$$\frac{\partial \Phi_k}{\partial x_j} \in L^\infty.$$

Thus, $\Phi^* : L^2 \to L^2$ and $\Phi^* : H^1 \to H^1$. Using complex interpolation and duality, it follows that $\Phi^* : H^s \to H^s$ for all $s \in \mathbb{R}$. This is the content of the following theorem.

Theorem 5.4. *Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism, and suppose $\Phi \equiv 1$ outside of a ball of radius R. Then,*

$$\Phi^* : H^s \to H^s.$$