4. Fourier Transform

The Fourier transform on $\mathbb{R}^n$ takes (sufficiently nice) functions on $\mathbb{R}^n$ and maps them to functions on $\mathbb{R}^n$. It is defined by

$$\mathcal{F}f(\xi) \equiv \hat{f}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\,dx. \tag{4.1}$$

Here, 'sufficiently nice' means that the integral converges, i.e. f decays sufficiently fast at infinity. Shortly we shall define a class of rapidly decreasing functions on which all our calculations are valid.

The Fourier transform

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\,dx.$$

can be thought of as a continuum limit of Fourier series on an interval $[-L, L]$ as $L \to \infty$. The Fourier coefficients here would be defined by (restricting to one dimension for simplicity)

$$f_n = \int_{-L}^{L} e^{-in\pi x/L} f(x)\,dx. \tag{4.2}$$

Now let $\hat{f}_L(\xi)$ be a function taking the value f_n at $\xi = n\pi/L$, i.e. only defined on a lattice of points distance π/L apart.

In the limit $L \to \infty$, at least heuristically $\hat{f}_L$ becomes a function $\hat{f}$ of a continuous variable ξ and (4.2) tends to the expression (4.1).

For the Fourier series we have Bessel's identity

$$\frac{1}{2L}\sum_n |f_n|^2 = \int_{-L}^{L} |f(x)|^2\,dx$$

and there is a reconstruction formula for $f \in L^2([-L, L])$

$$f(x) = \frac{1}{2L}\sum_n f_n e^{in\pi x/L}$$

where the sum converges in L^2 since $(f_n) \in l^2(\mathbb{Z})$ is a square-summable sequence.

Bessel's identity has a continuum limit, since

$$\frac{1}{2L}\sum_n |f_n|^2 = \frac{1}{2\pi}\frac{\pi}{L}\sum_n |\hat{f}_L(\frac{n\pi}{L})|^2 \text{ ``}\to\text{''} \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{f}(\xi)|^2\,d\xi.$$

Also the reconstruction formula has a continuum limit:

$$\frac{1}{2L}\sum_n f_n e^{in\pi x/L} = \frac{1}{2\pi}\frac{\pi}{L}\sum_n e^{ixn\pi/L}\hat{f}_L(\frac{n\pi}{L}) \text{ ``}\to\text{''} \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\xi x}\hat{f}(\xi)\,d\xi.$$

Therefore we can conjecture the following formulae:

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{f}(\xi)|^2\,d\xi \tag{4.3}$$

and

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\xi x}\hat{f}(\xi)\,d\xi. \tag{4.4}$$

We shall prove these formulae for the Fourier transform shortly.

The basic property of the Fourier Transform which makes it useful for analyzing constant coefficient PDE is that it transforms derivative operators (in x) into polynomials (in ξ). Let D_j stand for the operator $-i\partial/\partial x_j$. Then integrating by parts in the integral (4.1) shows that

$$\mathcal{F}(D_j f) = \xi_j \mathcal{F} f.$$

Also, the 'opposite' is true (with a change of sign). Consider the Fourier transform of the function $x_j f$. By writing $x_j e^{-ix\cdot\xi} = i\partial/\partial\xi_j e^{-ix\cdot\xi}$, we see that

$$\mathcal{F}(x_j f) = -D_j(\mathcal{F} f).$$

More generally, let $\alpha = (\alpha_1, \ldots, \alpha_n)$, α_i nonnegative integers, be a multi-index and let D^α denote $D_{x_1}^{\alpha_1} \ldots D_{x_n}^{\alpha_n}$ and ξ^α denote $\xi_1^{\alpha_1} \ldots \xi_n^{\alpha_n}$. Then we have (assuming that f is small enough at infinity so that the integrals are all convergent, and hence integrations by parts may be performed)

$$\begin{aligned} \mathcal{F}(D^\alpha f) &= \xi^\alpha \mathcal{F} f, \\ \mathcal{F}(x^\alpha f) &= (-1)^{|\alpha|} D^\alpha \mathcal{F} f. \end{aligned} \tag{4.5}$$

Schwartz functions

What is a class of functions on which the Fourier transform acts nicely?

If we want the integral to be convergent, then we should ask that f is in $L^1(\mathbb{R}^n)$. However, it is very difficult to characterize the *range* of $\mathcal{F}$ on $L^1(\mathbb{R}^n)$, so we shall deal with a much nicer (i.e. more restrictive) class of functions, namely *Schwartz functions* or *functions of rapid decrease.* Let $\mathcal{S}(\mathbb{R}^n)$ denote the class of Schwartz functions on $\mathbb{R}^n$.

By definition, a Schwartz function on $\mathbb{R}^n$ is one such that for any two multi-indices α, β,

$$\|f\|_{\alpha,\beta} \stackrel{\text{def}}{=} \sup_{x\in\mathbb{R}^n} |x^\beta D^\alpha f(x)| < \infty.$$

Then f decreases rapidly at infinity in the sense that it is smaller than the reciprocal of any polynomial, and this is true not just for f but all its derivatives as well. Then

$$|f(x)| \le C(1 + |x|^2)^{-n} \implies f \in L^1(\mathbb{R}^n),$$

and this is also true for all derivatives of f. So the Fourier Transform is well-defined on Schwartz functions and (4.5) holds for all Schwartz f.

Exercise. Show that, if $\phi \in \mathcal{S}(\mathbb{R}^n)$, then $x^\alpha\phi$ and $D^\beta\phi$ are also in $\mathcal{S}(\mathbb{R}^n)$.

Exercise. Show that, if $\phi \in \mathcal{S}(\mathbb{R}^n)$ and f is a C^∞ function on $\mathbb{R}^n$ with all derivatives bounded, then $f\phi \in \mathcal{S}(\mathbb{R}^n)$.

Three major theorems about $\mathcal{F}$.

Now we come to one of the main properties of the Fourier transform — the inversion formula. Let us define the map $\mathcal{G}$, which is almost the same as $\mathcal{F}$ but with two slight changes:

$$\mathcal{G} f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} f(\xi)\, d\xi. \tag{4.6}$$

The differences are the factor of $(2\pi)^{-n}$ and the change of sign in the phase of the exponential.

Theorem 4.1. *$\mathcal{F}$ maps $\mathcal{S}(\mathbb{R}^n)$ to $\mathcal{S}(\mathbb{R}^n)$.*

Theorem 4.2. *The map $\mathcal{G}$ on $\mathcal{S}(\mathbb{R}^n)$ is a two-sided inverse to $\mathcal{F}$.*

Theorem 4.3. *The Fourier transform preserves the L^2 norm of Schwartz functions in $\mathbb{R}^n$ (up to a factor $(2\pi)^{n/2}$):*

$$\|\mathcal{F} f\|_{L^2} = (2\pi)^{n/2} \|f\|_{L^2}, \quad f \in \mathcal{S}(\mathbb{R}^n).$$

Proof of Theorem 4.1 ($\mathcal{F}$ preserves the Schwartz functions): Let ϕ be a Schwartz function. Then we need to show that $\mathcal{F}\phi \in \mathcal{S}(\mathbb{R}^n)$, which amounts to showing that $\mathcal{F}\phi$ is smooth, and

$$\xi^\alpha D_\xi^\beta \hat{\phi}(\xi) \in L^\infty(\mathbb{R}^n)$$

for all multi-indices α, β. To check smoothness, we just verify that we are allowed to differentiate under the integral.

Because $\mathcal{F}$ exchanges multiplication by polynomials and partial differentiation, the second condition is equivalent to

$$\mathcal{F}\Big(D^\alpha x^\beta \phi\Big) \in L^\infty(\mathbb{R}^n).$$

According to the exercise above, $g = D^\alpha x^\beta \phi$ is in $\mathcal{S}(\mathbb{R}^n)$. Certainly then $g \in L^1$. Finally we see

$$\left|\int_{\mathbb{R}^n} e^{-ix\cdot\xi} g(x) dx\right| \le \int_{\mathbb{R}^n} |e^{-ix\cdot\xi} g(x)| dx$$
$$= \int_{\mathbb{R}^n} |g(x)| dx < \infty$$

and so $||\mathcal{F}(g)||_\infty$, which is just the supremum over ξ of this fixed quantity, is finite. □

Sketch of Theorem 4.2 ($\mathcal{G}$ is the inverse of $\mathcal{F}$):

• Step 1: compute the Fourier transform of the Schwartz function $e^{-ax^2/2}$ (in one dimension).

• Step 2: compute the Fourier transforms of Gaussians of several variables.

• Step 3: to treat the general case, take a Schwartz function $\phi(x)$, $x \in \mathbb{R}^n$. The composition $\mathcal{G} \circ \mathcal{F}\phi$ is given by the integral

$$(2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(\int_{\mathbb{R}^n} e^{-iy\cdot\xi} \phi(y)\, dy\right) d\xi.$$

This cannot be regarded as an integral over $\mathbb{R}^{2n}$ because the integrand is not convergent there; there is no decay in the ξ directions at all, only in the y direction, due to rapid decrease of the function ϕ. To remedy this we introduce artificially a function decaying as $|\xi| \to \infty$:

$$\lim_{\epsilon\to 0} (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(e^{-\epsilon|\xi|^2/2} \int_{\mathbb{R}^n} e^{-iy\cdot\xi} \phi(y)\, dy\right) d\xi.$$

The limit as $\epsilon \to 0$ gives the original integral we want. □

Proof:

• We first compute the Fourier transform of the Schwartz function $e^{-ax^2/2}$ (in one dimension). This function satisfies the differential equation

$$\frac{\partial u}{\partial x} = -axu.$$

Since, by (4.5), $\mathcal{F}(D_j f) = \xi_j \mathcal{F} f$ and $\mathcal{F}(x_j f) = -D_j \mathcal{F} f$, and remembering $D_j = -i\partial_j$, the Fourier transform of $e^{-ax^2/2}$ satisfies the differential equation

$$i\xi\hat{u} = -a(i\frac{\partial\hat{u}}{\partial\xi}) \implies \frac{\partial\hat{u}}{\partial\xi} = -\frac{\xi\hat{u}}{a}.$$

This implies that $\hat{u}$ is a multiple of $e^{-\xi^2/2a}$. To see which multiple, we compute

$$\hat{u}(0) = \int_{-\infty}^{\infty} e^{-ax^2/2}\, dx = \sqrt{\frac{2\pi}{a}}.$$

(Here we use the standard result $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$.) Hence

$$\mathcal{F}(e^{-ax^2/2}) = \sqrt{\frac{2\pi}{a}} e^{-\xi^2/2a}.$$

Thus the more peaked the original Gaussian u is, the more spread out the Fourier transform is, and vice versa.

• We can compute the Fourier transforms of Gaussians of several variables in the same way (one variable at a time), and we get in $\mathbb{R}^n$

$$\mathcal{F}(e^{-a|x|^2/2}) = \left(\frac{2\pi}{a}\right)^{n/2} e^{-|\xi|^2/2a}. \tag{4.7}$$

• Now to treat the general case, take a Schwartz function ϕ. The composition $\mathcal{G} \circ \mathcal{F}\phi$ is given by the integral

$$(2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(\int_{\mathbb{R}^n} e^{-iy\cdot\xi} \phi(y)\, dy\right) d\xi.$$

This cannot be regarded as an integral over $\mathbb{R}^{2n}$ because the integrand is not convergent there; there is no decay in the ξ directions at all, only in the y direction, due to rapid decrease of the function ϕ. To remedy this we introduce artificially a function decaying as $|\xi| \to \infty$:

$$\lim_{\epsilon\to 0} (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(e^{-\epsilon|\xi|^2/2} \int_{\mathbb{R}^n} e^{-iy\cdot\xi} \phi(y)\, dy\right) d\xi.$$

The limit as $\epsilon \to 0$ gives the original integral we want. (Why? To see this, note that $\hat{\phi} \in \mathcal{S}(\mathbb{R}^n)$, hence certainly in L^1. Then, since $e^{-\epsilon|\xi|^2/2} \le 1$ everywhere, and converges pointwise to 1, we have

$$\lim_{\epsilon\to 0} \int_{\mathbb{R}^n} e^{ix\cdot\xi} e^{-\epsilon|\xi|^2/2} \hat{\phi}(\xi)\, d\xi = \int_{\mathbb{R}^n} e^{ix\cdot\xi} \hat{\phi}(\xi)\, d\xi$$

by the dominated convergence theorem.)

Now, for each $\epsilon > 0$, the integrand is an L^1 function in $\mathbb{R}^{2n}$. (One nice way to see this is that the 'external' product $f(x)g(y)$ of Schwartz functions on $\mathbb{R}^n$ is always Schwartz on $\mathbb{R}^{2n}$.) This entitles us to apply Fubini's theorem and switch the order of integration; i.e. we can do the ξ integral first. This produces for us the Fourier transform of the Gaussian, which fortunately we just computed:

$$\lim_{\epsilon\to 0}(2\pi)^{-n}\left(\frac{2\pi}{\epsilon}\right)^{n/2}\int_{\mathbb{R}^n} e^{-|x-y|^2/2\epsilon}\phi(y)\,dy.$$

We change variable to $z=(x-y)/\sqrt{\epsilon}$:

$$=(2\pi)^{-n/2}\lim_{\epsilon\to 0}\int_{\mathbb{R}^n} e^{-|z|^2/2}\phi(x+\sqrt{\epsilon}z)\,dz.$$

The integrand is uniformly bounded by $e^{-|z|^2/2}\|\phi\|_\infty$ as ϵ goes to zero. This is an L^1 function, so by the dominated convergence theorem, we can take the pointwise limit in the integrand:

$$=\phi(x)(2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{-|z|^2/2}\,dz=\phi(x).$$

Thus, we have shown that

$$(\mathcal{G}\mathcal{F}\phi)(x)=\phi(x),$$

which proves that $\mathcal{G}\circ\mathcal{F}$ is the identity on $\mathcal{S}(\mathbb{R}^n)$. An essentially identical argument proves that $\mathcal{F}\circ\mathcal{G}$ is the identity on $\mathcal{S}(\mathbb{R}^n)$, so this proves that $\mathcal{F}$ is a bijection (one-to-one and onto) on $\mathcal{S}(\mathbb{R}^n)$ with inverse $\mathcal{G}$. □

Proof of Theorem 4.3: If we write out the square of the L^2 norm of $\hat{f}$ longhand we get

$$\int\left(\int e^{-iy\cdot\xi}f(y)\,dy\right)\left(\int e^{ix\cdot\xi}\overline{f(x)}\,dx\right)\,d\xi.$$

Similar to before, this integrand is not L^1 on $\mathbb{R}^{3n}$ because we have no decay at infinity in the ξ variable. So we introduce a decaying Gaussian factor as before and take a limit:

$$\lim_{\epsilon\to 0}\int e^{-\epsilon|\xi|^2/2}\left(\int e^{-iy\cdot\xi}f(y)\,dy\right)\left(\int e^{ix\cdot\xi}\overline{f(x)}\,dx\right)\,d\xi.$$

Now the integrand is L^1 on $\mathbb{R}^{3n}$ and we may invoke Fubini's theorem and do the ξ integral first. We get

$$\lim_{\epsilon\to 0}\left(\frac{2\pi}{\epsilon}\right)^{n/2}\iint e^{-|x-y|^2/2\epsilon}f(y)\overline{f(x)}\,dx\,dy.$$

Using similar reasoning as above, the limit is equal to

$$(2\pi)^n\int|f(x)|^2\,dx.$$

□

Convolutions

The convolution of two functions f,g, say in $\mathcal{S}(\mathbb{R}^n)$, is the function $f*g$ given by

$$f*g(x)=\int_{\mathbb{R}^n} f(x-y)g(y)\,dy.$$

Notice that $f*g=g*f$, since

$$\int_{\mathbb{R}^n} f(x-y)g(y)\,dy=\int_{\mathbb{R}^n} f(z)g(x-z)\,dz.$$

This turns up frequently in analysis. There are two main ways in which they show up. One is from translation invariance. Suppose we have an operator K on functions defined on $\mathbb{R}^n$, which is translation invariant: $T_a(Kf)=K(T_af)$, where T_a is translation by a, i.e. $T_ag(x)=g(x-a)$. Also suppose that K has an integral representation:

$$(Kf)(x)=\int K(x,y)f(y)\,dy.$$

Then $K(x,y)$ is a function of $x-y$ and the integral above is then a convolution.

The other way convolution turns up is in smoothing and approximation. Suppose that we have a function $f(x)$, $x\in\mathbb{R}^n$ which is not very smooth — say we only have $f\in L^1$, but with no differentiability. We might want to find a smoothed version of f, i.e. a function g such that g is close to f in L^1 (i.e. $\|f-g\|_1$ is small), but such that g is smooth (i.e. C^∞). One way to do this is to average f against a smooth function. Let ϕ be a smooth, nonnegative, compactly supported function on $\mathbb{R}^n$ with integral 1. Then

$$f*\phi(x)=\int f(x-y)\phi(y)\,dy$$

is a averaged version of f, which is smooth since we can write, with $z=x-y$,

$$(f*\phi)(x)=\int f(z)\phi(x-z)\,dz \tag{4.8}$$

and then we may differentiate under the integral sign arbitrarily many times. We can scale ϕ as follows:

$$\phi_\epsilon(x)=\epsilon^{-n}\phi(x/\epsilon). \tag{4.9}$$

This preserves all conditions on ϕ, making it narrower and steeper. Then it turns out that

$$f*\phi_\epsilon\to f \text{ in } L^1(\mathbb{R}^n). \tag{4.10}$$

You should think here that $(f * \phi_\epsilon)(x)$ is an average of some of the values of f near x; as $\epsilon \to 0$, we are only averaging over a very small region, so $(f*\phi_\epsilon)(x)$ should approach $f(x)$. The impressive thing about this claim is that it also converges in the L^1 sense.

Unfortunately there is no function which is an identity element under convolution, i.e. no function g such that $f * g = f$ for all f.

We can think of (ϕ_ϵ) as forming an 'approximate identity', which is often almost as good. (Another way to remedy this lack of identity element is to pass to distributions.)

Theorem 4.4. *Let $f, g \in \mathcal{S}(\mathbb{R}^n)$. Then the Fourier transform of $f * g$ is the pointwise product of $\hat{f}\hat{g}$, and the Fourier transform of fg is $(2\pi)^{-n}\hat{f} * \hat{g}$.*

This is a straightforward calculation: Write the Fourier transform of the convolution

$$\int e^{-ix\cdot\xi}\left(\int_{\mathbb{R}^n} f(x-y)g(y)\,dy\right)\,dx.$$

Now write $e^{-ix\cdot\xi} = e^{-i(x-y)\cdot\xi}e^{-iy\cdot\xi}$ and change variables from y to $z = x - y$.

To prove the second result apply $\mathcal{G}$ to $(2\pi)^{-n}\hat{f} * \hat{g}$.

Density results

Our next goal is to show that the Fourier transform makes sense on $L^2(\mathbb{R}^n)$ and satisfies Theorems 4.1, 4.2 and 4.3 with $\mathcal{S}(\mathbb{R}^n)$ replaced by $L^2(\mathbb{R}^n)$.

This is not as difficult as it may seem in view of the following:

Theorem 4.5 (**B**ounded **L**inear **T**ransformation Thm)**.** *Suppose that $S \subset H_1$ is a dense subspace and $T : S \to H_2$ is a bounded linear transformation:*

$$\|Tf\|_{H_2} \leq M\|f\|_{H_1}, \quad f \in S.$$

Then T extends uniquely to a bounded linear transformation from H_1 to H_2.

- This is a special case of a theorem about metric spaces: we can replace H_1 by any metric space and H_2 by any complete metric space, S any dense subset of H_1, and the condition by

$$d(Tx_1, Tx_2) \leq Md(x_1, x_2).$$

- The idea is to define T for a general $f \in H_1$ by taking a sequence of elements of S converging to f and observing that their images under T converge in H_2.

Sketch of the BLT theorem: To define Tf for an arbitrary $f \in H_1$, we will choose some sequence $\{g_n\} \subset S$ with $g_n \to f$. Now $\|Tg_n - Tg_m\| \leq M\,\|g_n - g_m\|$, so the sequence $\{Tg_n\}$ is Cauchy and hence converges to some $g \in H_2$. We want to define $Tf = g$.

Still todo:

- Show that this makes sense, i.e. that our definition of Tf didn't depend on which sequence $\{g_n\}$ we choose.
- Show that this extension is linear.
- Show that this extension is bounded.
- Show that any other extension is the same as this one. □

The BLT theorem and Theorem 4.3 implies that the Fourier transform extends from $\mathcal{S}(\mathbb{R}^n)$ to its closure in $L^2(\mathbb{R}^n)$. Next we show that this closure is all of L^2.

Theorem 4.6. *Let $1 \le q < \infty$. Then $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^q(\mathbb{R}^n)$.*

Proof: First we prove this for $q = 1$. We use the fact, proved in MATH 3320 (and in SS, Chapter 2 Theorem 2.4) that continuous functions of compact support are dense in $L^1(\mathbb{R}^n)$.

So, let $f \in L^1(\mathbb{R}^n)$ and $\epsilon > 0$ be given. We need to show that there is a Schwartz function g such that $\|f - g\|_1 < \epsilon$. First select a continuous h of compact support, with L^1 distance less than $\epsilon/2$ from f. It suffices to find a Schwartz function g such that $\|h - g\|_1 \le \epsilon/2$.

We use a convolution to smooth g. Select a nonnegative $C^\infty(\mathbb{R}^n)$ function ϕ with compact support, and such that $\int \phi = 1$. Without loss of generality we can assume that the support of ϕ is the unit ball B. Now define

$$\phi_\delta(x) = \delta^{-n}\phi(x/\delta).$$

Then $\phi_\delta(x)$ is in $C^\infty(\mathbb{R}^n)$, supported in the ball of radius δ and also has integral 1. Now consider

$$h_\delta = h * \phi_\delta, \text{ i.e. } h_\delta(x) = \int h(x-y)\phi_\delta(y)\, dy.$$

This is C^∞ since ϕ_δ is C^∞. It is compactly supported since the support of h_δ is contained in the set of points of distance not more than δ from the support of h. To make things concrete, let A be the set of points of distance not more than 1 from the support of h. Then A is compact and has a finite measure $|A|$. If we take $\delta < 1$ then h_δ is supported in the set A.

Now h is uniformly continuous on A, so we can find δ so that $|h(x) - h(y)| < \epsilon/2|A|$ whenever $|x - y| < 2\delta$. Then we can estimate

$$|h(x) - h_\delta(x)| \le \int_{B(0,\delta)} \Big|h(x) - h(x-z)\Big|\phi_\delta(z)\, dz.$$

Since $|x - (x-z)| = |z| < 2\delta$ when z is in $B(0,\delta)$, we can estimate this by

$$|h(x) - h_\delta(x)| \le \int_{B(0,\delta)} \frac{\epsilon}{2|A|}\phi_\delta(z)\, dz = \frac{\epsilon}{2|A|}.$$

Integrating this over A we find that

$$\|h - h_\delta\|_{L^1} \le \frac{\epsilon}{2},$$

as required.

For $q > 1$, we proceed as follows. Given $f \in L^q(\mathbb{R}^n)$, we can find a bounded function of compact support $\tilde{f}$ such that $\|f - \tilde{f}\|_q < \epsilon/2$. (We know simple functions are dense in L^q; in fact compactly supported simple functions are also dense, since given any measurable set A there is a compact $K \subset A$ so $\mu(A \setminus K)$ is arbitrarily small. This ensures that bounded functions of compact support are dense too. Alternatively, define $E_n = \{|x| \le N, |f(x)| \le N\}$ and use the dominated convergence theorem to show that $f\chi_{E_N}$ converges in the L^q sense to f.)

Now suppose that $|\tilde{f}| \le M$ and that $\tilde{f}$ is supported in $B(0,R)$. Then $\tilde{f} \in L^1(\mathbb{R}^n)$. Choose g, constructed as above via convolution with a smooth function, so that it approximates $\tilde{f}$ in L^1 to within $\tilde{\epsilon}$. We have also $|g| \le M$. We estimate

$$\|\tilde{f} - g\|_{L^q}^q \le \|\tilde{f} - g\|_\infty^{q-1}\|\tilde{f} - g\|_1.$$

This implies that

$$\|\tilde{f} - g\|_{L^q} \le (2M)^{(q-1)/q}\tilde{\epsilon}^{1/q}$$

and to complete the proof we simply choose $\tilde{\epsilon}$ so the right hand side is smaller than $\epsilon/2$. □

Exercise. Using ideas from this proof, prove the closely related result (4.10).

We have now proved that *the Fourier transform $\mathcal{F}$ is, up to a numerical factor, an isometry on $L^2(\mathbb{R}^n)$, with inverse $\mathcal{G}$.*

Parseval's formula

Proposition 4.7. *If $f, g \in L^2(\mathbb{R}^n)$, then*

$$\int f(x)\overline{g(x)}\,dx = (2\pi)^{-n}\int \hat{f}(\xi)\overline{\hat{g}(\xi)}\,d\xi.$$

This follows immediately from the fact the Fourier transform preserves norms up to a scalar factor, and the polarization identity

$$2\langle f, g\rangle = \|f+g\|^2 + i\,\|f+ig\|^2 - (1+i)\,\|f\|^2 - (1+i)\,\|g\|^2.$$

To remember the polarization identity, just expand out $\|f+g\|^2$, and then $\|f+ig\|^2$ (to get a different coefficient in front of $\langle g, f\rangle$, then take the appropriate linear combination to cancel out the $\langle g, f\rangle$ terms.

We could also prove it directly by proving it for Schwartz functions f, g and then use the density of Schwartz space in L^2 to get the general case.

If $f, g \in \mathcal{S}(\mathbb{R}^n)$, then we compute

$$\begin{aligned}\int f(x)\overline{g(x)}\,dx &= \int \left((2\pi)^{-n}\int \hat{f}(\xi)e^{ix\cdot\xi}\,d\xi\right)\overline{g(x)}\,dx \\ &= (2\pi)^{-n}\iint \hat{f}(\xi)e^{ix\cdot\xi}\overline{g(x)}\,dx\,d\xi \\ &= (2\pi)^{-n}\int \hat{f}(\xi)\overline{\hat{g}(\xi)}\,d\xi.\end{aligned}$$

- This result also holds if, say, $f \in \mathcal{S}(\mathbb{R}^n)$ and $g \in L^1(\mathbb{R}^n)$.

Central limit theorem

The central limit theorem is about the limiting behaviour of the repeated convolution of a probability measure with itself. To simplify things we will consider just nonnegative L^1 functions on $\mathbb{R}$, with integral 1 (particular examples of probability measures).

First we motivate the construction. Let X and Y be real-valued random variables with probability density functions f and g, respectively. This means that, for all $t \in \mathbb{R}$,

$$P(X < t) = \int_{-\infty}^{t} f(s)\,ds,$$

and similarly for (Y, g). What is the probability density function h for $X+Y$? Let's assume X and Y are independent. Heuristically,

$$\begin{aligned}&P(X+Y<t) \\ &\quad= \lim_{n\to\infty}\frac{1}{n}\sum_{k\in\mathbb{Z}} nP\left(X \in [k/n, (k+1)/n]\right)P(Y < t-k/n) \\ &\quad\to \int_{-\infty}^{\infty} f(s)P(Y<t-s)\,ds.\end{aligned}$$

Slightly more carefully, but still at a heuristic level, we have

$$P(X+Y<t) \geq \sum_k P\left(X \in \left[\frac{k}{n}, \frac{k+1}{n}\right]\right)P\left(Y < t - \frac{k+1}{n}\right)$$

and

$$P(X+Y<t) \leq \sum_k P\left(X \in \left[\frac{k}{n}, \frac{k+1}{n}\right]\right)P\left(Y < t - \frac{k}{n}\right)$$

and then as $n \to \infty$ we expect both of these to converge to the given integral.

Now differentiate with respect to t and we get

$$h(t) = \int_{-\infty}^{\infty} f(s)g(t-s)\,ds,$$

or equivalently, $h = f * g$.

In scientific experiments one often runs an experiment many times and averages the results to obtain greater accuracy. Mathematically, we can ask what is the pdf of $(X_1+\cdots+X_n)/n$ in terms of the pdf f, if X_i are all independent with pdf f. Clearly the pdf of $X_1+\ldots X_n$ is the n-fold convolution of f, denoted f^{*n}. The scaling has the effect of changing this to $nf^{*n}(nx)$. To see this, compute

$$P(X/n < t) = P(X < nt) = \int_{-\infty}^{nt} f(s)d(s)$$

and changing variables $s = ns'$

$$= \int_{-\infty}^{t} nf(ns')ds'.$$

Now it turns out that, provided the pdf f satisfies

$$\int_{-\infty}^{\infty} x^2 f(x)\,dx < \infty, \tag{4.11}$$

i.e. has finite variance, that the n-fold convolution looks more and more like a Gaussian which concentrates more and more at the mean value. To make this precise, assume that the mean value of X is 0, or equivalently,

$$\int_{-\infty}^{\infty} x f(x)\,dx = 0.$$

(Otherwise, we can just shift out probability distribution by subtracting off the mean.) Then (4.11) is precisely the variance σ. To obtain a mathematical formulation of this statement we use a different scaling, one that keeps the variance fixed. It is not hard to compute that the variance of f^{*n} is $n\sigma$, the variance of af is $a\sigma$, and the variance of $f(b\cdot)$ is $b^{-3}\sigma$. We will consider $g_n(x) = n^{1/2} f^{*n}(n^{1/2}x)$; this has fixed variance σ.

Theorem 4.8 (Central Limit Theorem)**.** *Assume that f satisfies* (4.11)*. Then the pdf g_n converges weakly to the normal distribution with variance σ, namely*

$$(2\pi\sigma)^{-1/2} e^{-x^2/2\sigma}.$$

• Here, 'weakly' means that for every continuous function k on $\mathbb{R}$ that converges to 0 at infinity, we have

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g_n(x)k(x)\,dx = \int_{-\infty}^{\infty} (2\pi\sigma)^{-1/2} e^{-x^2/2\sigma} k(x)\,dx.$$

This is convergence 'in the weak-$*$ topology of measures', as we will discuss later in the course.

This is equivalent to, by a change of variable

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left(nf^{*n}(nx) - \frac{1}{\sqrt{2\pi\sigma}} e^{-nx^2/2s} \right) k(x)dx = 0$$

for all such k.

We first prove a lemma.

Lemma 4.9. *(i) (Riemann-Lebesgue lemma) Suppose that $f \in L^1(\mathbb{R})$. Then $\hat{f}$ is uniformly continuous and tends to zero at infinity.*

(ii) Suppose that f has finite variance (i.e. satisfies (4.11)*). Then $\hat{f}$ is C^2.*

Proof: (i) Hölder's inequality shows that $\hat{f}$ is bounded, with $\|\hat{f}\|_\infty \le \|f\|_1$. To show continuity, given $\epsilon > 0$, we choose $R > 0$ large enough that $\int_{|x|>R} |f| < \epsilon$. Then, we choose δ small enough so that $|1 - e^{iy}| < \epsilon \|f\|_1^{-1}$ for all $|y| \le \delta R$. Then, if $|\xi_1 - \xi_2| < \delta$, we have

$$\begin{aligned} &|\hat{f}(\xi_1) - \hat{f}(\xi_2)| \\ &\quad \le \Big| \int_{-R}^{R} (e^{-ix\xi_1} - e^{-ix\xi_2}) f(x)\,dx + \int_{|x|>R} (e^{-ix\xi_1} - e^{-ix\xi_2}) f(x)\,dx \Big| \\ &\quad \le \epsilon \|f\|_1^{-1} \|f\|_1 + 2\|f\|_{L^1(\{|x|>R\})} \\ &\quad < 3\epsilon \end{aligned}$$

so $\hat{f}$ is uniformly continuous.

To prove that $\hat{f}(\xi) \to 0$ as $|\xi| \to \infty$, we use the density of $\mathcal{S}(\mathbb{R})$ in $L^1(\mathbb{R})$. Given $\epsilon > 0$, and $f \in L^1(\mathbb{R})$, choose $\phi \in \mathcal{S}(\mathbb{R})$ such that $\|f - \phi\|_1 < \epsilon$. Since $\hat{\phi} \in \mathcal{S}(\mathbb{R})$ it tends to zero at infinity. On the other hand, $\widehat{f - \phi}$ is bounded by ϵ in L^∞. Therefore, $\limsup_{|\xi|\to\infty} |\hat{f}(\xi)| \le \epsilon$. Since this is true for all $\epsilon > 0$, we have $\limsup_{|\xi|\to\infty} |\hat{f}(\xi)| = 0$, which is equivalent to saying that $\hat{f}(\xi) \to 0$ as $|\xi| \to \infty$.

Part (ii) follows readily from (i). Observe xf and $x^2 f$ are both in L^1, and use the fact that Fourier transform intertwines multiplication by x and differentiation. □

Proof of CLT: We use the Fourier transform. The Fourier transform of $g_n(x)$ is $(\hat{f}(\xi/n^{1/2}))^n$. Now, the condition (4.11) implies that $\hat{f}$ is C^2, and that the second derivative of $\hat{f}$ at $\xi = 0$ is $-\sigma$. (Since $-D^2\hat{f}(0) = \widehat{-x^2 f}(0) = -\int_{-\infty}^{\infty} e^0 x^2 f(x)dx$.) Also, the mean value of X being zero implies that the first derivative of $\hat{f}$ at $\xi = 0$ is zero. That means, at $\xi = 0$, that

$$\hat{f}(\xi) = 1 - \sigma\xi^2/2 + o(\xi^2).$$

(Recall that $o(\xi^2)$ here means that for every $\epsilon > 0$, there is a $\delta > 0$ so that if $|\xi| < \delta$ then

$$|\hat{f}(\xi) - (1 - \sigma\xi^2/2)| < \epsilon\xi^2$$

or

$$1 - (\sigma/2 + \epsilon)\xi^2 \le \hat{f}(\xi) \le 1 - (\sigma/2 - \epsilon)\xi^2.)$$

Now scaling ξ and raising to the nth power, we get

$$\Big(1 - (\sigma/2+\epsilon)\xi^2/n\Big)^n \le \hat{g}_n(\xi) \le \Big(1 - (\sigma/2-\epsilon)\xi^2/n\Big)^n.$$

This is now valid for $|\xi|/n^{1/2} < \delta$, which for a fixed ξ and δ is true for large enough n. Therefore, for fixed ϵ, and ξ, we can take the limit as $n \to \infty$ and obtain

$$e^{-(\sigma/2+\epsilon)\xi^2} \le \lim_n \hat{g}_n(\xi) \le e^{-(\sigma/2-\epsilon)\xi^2}.$$

Since this is true for all $\epsilon > 0$ we obtain

$$\lim_n \hat{g}_n(\xi) = e^{-\sigma\xi^2/2}.$$

Note also that $|\hat{g}_n(\xi)| \leq 1$ everywhere, since $\|g_n\|_1 = 1$.

Let us denote the integral $\int_{\mathbb{R}} f(x)k(x)\,dx$ by $\langle f, k\rangle$. To complete the proof we need to compute

$$\lim_{n\to\infty} \langle g_n(x) - \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/2s}, k(x)\rangle = 0.$$

To check that a sequence of functions h_n (uniformly bounded in L^1 norm) converges weakly to zero, it is only necessary to check on a subspace (of continuous functions converging to 0 at ∞) which is dense with respect to the sup norm.

Suppose $\lim_{n\to\infty} \left|\int g_n(x)k'(x)dx\right| = 0$, and $\sup|k' - k| < \epsilon$.

$$\begin{aligned}\lim_{n\to\infty}\left|\int h_n(x)k(x)dx\right| &= \lim_{n\to\infty}\left|\int h_n(x)k'(x) - h_n(x)(k'(x) - k(x))dx\right| \\ &\leq \lim_{n\to\infty}\left|\int h_n(x)k'(x)dx\right| + \sup|k' - k| \int |h_n|\,dx \\ &< 0 + \epsilon\,\|h_n\|_1\end{aligned}$$

Now the Schwartz functions are just such a dense subspace. (To see this, take an arbitrary continuous function h converging to 0 at ∞. Multiply it by a smooth compactly supported function which is identically 1 for $|x| < R$, for sufficiently large R that this is close in sup norm to h. Now convolve with some ϕ_δ to smooth it, and show that, as in the proof of Theorem 4.6, for sufficiently small δ this is close enough in sup norm.)

So assume that k is Schwartz. Using Parseval's formula, we have

$$\lim_{n\to\infty} \langle g_n(x), k(x)\rangle = (2\pi)^{-1} \lim_{n\to\infty} \langle \hat{g}_n, \hat{k}\rangle.$$

Since $\hat{k}$ is Schwartz, hence certainly L^1, and $|g_n(\xi)| \leq 1$ and converges pointwise, we can use the dominated convergence theorem to show that the limit is

$$\frac{1}{2\pi}\langle e^{-\sigma\xi^2/2}, \hat{k}\rangle.$$

Using Parseval again, this is equal to

$$\langle \mathcal{G}(e^{-\sigma\xi^2/2}), k\rangle,$$

which completes the proof since we know that

$$\mathcal{G}(e^{-\sigma\xi^2/2}) = (2\pi\sigma)^{-1/2} e^{-x^2/2\sigma}.$$

□

A counterexample

To show that the finite variance condition is necessary consider the following innocent-looking probability distribution:

$$\psi(t) = \frac{1}{\pi}\frac{1}{1+t^2}.$$

Since this is an analytic function, the Fourier transform can be computed using contour integration, and we obtain

$$\hat{\psi}(\xi) = e^{-|\xi|}.$$

The idea here is to approximate the integral defining the Fourier transform by the contour integral along a semicircular contour, with diameter $[R, -R]$. Depending on the sign of ξ, this only works for a semi-circle above the x-axis or below the x-axis (so that we can show in the limit the contribution from the semicircular piece of the contour vanishes. We see that $1/(1+t^2)$ has poles at $\pm i$, so depending on which contour we are using we evaluate the residue at the relevant enclosed pole, and use Cauchy's theorem to obtain the answer. If we take independent random variables X_1 and X_2 with distribution ψ, then $(X_1 + X_2)/2$ also has distribution ψ! To see this, note that the Fourier transform of the distribution of $(X_1+X_2)/2$ is $\hat{\psi}^2(\xi/2) = \hat{\psi}(\xi)$. Thus, this distribution does NOT become Gaussian after repeated averaging.

The problem with ψ, of course, is that it does not have finite variance (or even a well-defined mean), and consequently, its Fourier transform is, quite visibly, not C^2.

A few notes on central limit theorems:

- In our proofs we said nothing about the rate of convergence; this can be quantified, for example in the Berry-Esséen theorem.
- There's a zoo of extensions; more precisely, we've proved here the Lindeberg-Lévy CLT.
- If we don't assume the random variables are identically distributed, there's the Lindeberg-Feller theorem and Liapounov's theorem.
- It's also possible to say something about the infinite variance case, where under some conditions the averages converge to one of the "stable distributions". The counterexample above is in fact a Cauchy distribution, which is one of the stable distributions.