Harmonic Analysis: from Fourier to Haar

María Cristina Pereyra

Lesley A. Ward

Department of Mathematics and Statistics, MSC03 2150, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA
E-mail address: crisp@math.unm.edu

School of Mathematics and Statistics, University of South Australia, Mawson Lakes SA 5095, Australia
E-mail address: Lesley.Ward@unisa.edu.au

2000 *Mathematics Subject Classification.* Primary 42-xx

CHAPTER 8

Beyond paradise

The Fourier transform and most of its properties can be established in a more general setting than the class $\mathcal{S}(\mathbb{R})$ of Schwartz functions. In this section we introduce first the class of *functions of moderate decrease*, which contains $\mathcal{S}(\mathbb{R})$, and then the still larger class of *tempered distributions*, which contains the $L^p(\mathbb{R})$ functions for all real p with $1 \leq p \leq \infty$, the delta distribution, the Borel measures, and more.

Strictly speaking, the tempered distributions are not functions at all, as we can see from the example of the delta distribution. However, the distributions still have a Fourier theory, meaning that we can extend to them the fundamental operations discussed above such as translation, dilation, the Fourier transform, and convolution with Schwartz functions. We clarify these statements below.

8.1. Functions of moderate decrease

A function f is said to be *of moderate decrease* if f is continuous and if there is a constant $A > 0$ such that

$$|f(x)| \leq \frac{A}{1+x^2} \qquad \text{for all } x \in \mathbb{R}. \tag{8.1}$$

The 1 in the denominator means that $|f|$ cannot be too big for x near 0, while the x^2 in the denominator means that $|f|$ cannot be too big as $x \to \pm\infty$. We note in passing that there is nothing special about the exponent 2 in the defining inequality (8.1). Every continuous function f such that $|f(x)| \leq A/(1+|x|^{1+\varepsilon})$ for all $x \in \mathbb{R}$, where $\varepsilon > 0$ and $A > 0$ are constants independent of x, has all the properties discussed in this subsection; in particular such an f is both integrable and square integrable.

Example 8.1. (*Functions of Moderate Decrease Need Not Be Schwartz*) All Schwartz functions are functions of moderate decrease. However there are functions such as

$$\frac{1}{1+|x|^n} \quad \text{for } n \geq 2, \qquad \text{and} \qquad e^{-a|x|} \quad \text{for } a > 0,$$

that are of moderate decrease, although they are not in $\mathcal{S}(\mathbb{R})$. (Check!) ◊

Example 8.2. Neither the Poisson kernel (7.20) nor the Fejér kernel (7.19) belong to $\mathcal{S}(\mathbb{R})$. However, they are functions of moderate decrease. ◊

Exercise 8.3. Verify that for each R and each $y > 0$, the Poisson kernel P_y and the Fejér kernel F_R are functions of moderate decrease. Show that the conjugate Poisson kernel Q_y is not a function of moderate decrease. ◊

We summarize the integrability and boundedness properties of functions of moderate decrease. We define the improper integral over $\mathbb{R}$ of a function f of moderate decrease as the limit as $N \to \infty$ of the Riemann integrals of f over the compact intervals $[-N, N]$:

$$\int_{-\infty}^{\infty} f(x)\,dx := \lim_{N\to\infty} \int_{-N}^{N} f(x)\,dx < \infty. \tag{8.2}$$

EXERCISE 8.4. Verify that functions of moderate decrease are closed under multiplication by bounded and continuous functions. ◊

In particular, if the function f is of moderate decrease, then the function $e^{-2\pi ix\xi}f(x)$ is also of moderate decrease, and its integral over $\mathbb{R}$ is well defined and finite for each real number ξ. That is, the Fourier transform for functions of moderate decrease is well defined by the formula

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} f(x)e^{-2\pi ix\xi}\,dx.$$

EXERCISE 8.5. Show that the Fourier transform of the Poisson kernel is $e^{-2\pi|\xi|y}$. Show that the Fourier transform of the Fejér kernel is $\left(1 - \frac{|\xi|}{R}\right)$ if $|\xi| \leq R$, and zero otherwise. ◊

EXERCISE 8.6. Show that the convolution of two functions f and g of moderate decrease is also a function of moderate decrease. Show that

$$\widehat{f * g} = \widehat{f}\widehat{g}.$$

◊

Notice that functions of moderate decrease are always bounded, since $|f(x)| \leq A$ for all $x \in \mathbb{R}$. It follows that they are in the space $L^\infty(\mathbb{R})$ of *essentially bounded* functions. (A function f is said to be essentially bounded, written $f \in L^\infty(\mathbb{R})$, if f is bounded by some fixed constant except on a set of measure zero. The space $L^\infty(\mathbb{R})$ is a Banach space with norm given by $\|f\|_\infty := \operatorname{ess\,sup}_{x\in\mathbb{R}} |f(x)|$. Here the *essential supremum* of $|f(x)|$, denoted by $\operatorname{ess\,sup}_{x\in\mathbb{R}} |f(x)|$, is defined as the smallest positive number M such that $|f(x)| \leq M$ for all $x \in \mathbb{R}$ except for a set of measure zero. When the function f is actually *bounded*, then $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. Think of $L^\infty(\mathbb{R})$ as the space of bounded functions on the line.) It also follows that functions of moderate decrease are *locally integrable*, written $f \in L^1_{\text{loc}}$, meaning that for each compact interval $[a, b]$ the integral $\int_a^b |f(x)|dx$ is finite. Finally, all functions of moderate decrease are in $L^1(\mathbb{R})$ and in $L^2(\mathbb{R})$. First, the decay condition (8.1) guarantees that every function f of moderate decrease is integrable ($f \in L^1(\mathbb{R})$). Since f is also bounded, the following norm estimate holds:

$$\|f\|_2^2 = \int_{-\infty}^{\infty} |f(x)|^2\,dx \leq \left(\sup_{x\in\mathbb{R}} |f(x)|\right) \int_{-\infty}^{\infty} |f(x)|\,dx = \|f\|_\infty \|f\|_1.$$

Thus f is also square integrable ($f \in L^2(\mathbb{R})$).

EXERCISE 8.7. (*Functions of Moderate Decrease are in $L^p(\mathbb{R})$*) Check that functions of moderate decrease are in $L^1(\mathbb{R})$ and indeed in every space $L^p(\mathbb{R})$. In other words, check that

$$\int_{-\infty}^{\infty} |f(x)|^p\,dx < \infty, \qquad \text{for } 1 \leq p < \infty.$$

◊

The Fourier transform is well defined for functions of moderate decrease. The conclusion of the Approximation of the Identity Theorem (Theorem 7.40) holds if we assume that the kernels and the function are functions of moderate decrease. Namely, if f and the good kernels K_δ are functions of moderate decrease, then the convolutions $f * K_\delta$ converge uniformly to f, and they converge in L^p to f for $1 \leq p < \infty$. Therefore, the inversion formula and Plancherel's Identity hold when f and $\widehat{f}$ are both of moderate decrease. See [**SS1**, Sections 5.1.1 and 5.1.7] for more details.

In fact, if one follows closely the proofs presented for the Schwartz class, one realizes that the full power of the Schwartz class is not being used. In the arguments what is really needed is the fast decay of the function (integrability) to ensure that integrals outside a large interval ($|x| > M$) can be made arbitrarily small for M sufficiently large, and inside the compact interval $|x| \leq M$, one could now use uniform convergence arguments to interchange limit and integral, and take advantage of the integrand being small to guarantee that the integral over the compact interval is arbitrarily small. We can repeat the arguments if we assume the functions involved are functions of moderate decrease instead of Schwartz functions. By similar arguments, one can also verify the inversion formula when f and $\widehat{f}$ are both of in $L^1(\mathbb{R})$. That requires understanding of the Lebesgue integration theory, and the theorems for interchanging limits and integrals; see for example [**SS2**, Section 2.4].

Exercise 8.8. Verify that the Approximation of the Identity Theorem (Theorem 7.40) holds when the kernels and the functions convolved are functions of moderate decrease. $\Diamond$

Exercise 8.9. Verify that the inversion formula and Plancherel's Identity hold when f and $\widehat{f}$ are both of moderate decrease. $\Diamond$

We mentioned in the interlude on kernels (Section 7.6) that the *partial Fourier integrals* $S_R f(x) := \int_{-R}^{R} \widehat{f}(\xi) e^{2\pi i \xi x}\, d\xi$ are given by convolution with the Dirichlet kernel D_R on $\mathbb{R}$. It was left as an exercise to verify that the Fejér kernel on $\mathbb{R}$ can be written as an integral mean of Dirichlet kernels,

$$F_R(x) = \frac{1}{R}\int_0^R D_t(x)\, dt,$$

and that the Fejér kernels define an approximation of the identity on $\mathbb{R}$ as $R \to \infty$. Therefore the *integral Cesàro means* of a function f of moderate decrease converge to f as $R \to \infty$; in other words

$$\sigma_R f(x) := \frac{1}{R}\int_0^R S_t f(x)\, dt = F_R * f(x) \to f(x).$$

The convergence is both uniform and in L^p.

Notice the parallel with the results obtained for Fourier series when we discussed summability methods.

In Chapter 11, we give a proof that the partial Fourier sums of f converge in $L^p(\mathbb{T})$, this time as a consequence of the boundedness of the periodic Hilbert transform in $L^p(\mathbb{T})$. The same is true in $\mathbb{R}$. Here the case $p = 2$ is taken care of by Plancherel's Identity.

8.2. Tempered distributions

We begin with some definitions. Let V be a vector field over some scalar field F; F is usually the real numbers or the complex numbers. A *linear functional* on F, or more briefly a *functional* on F, is a linear mapping from V to the scalar field F. If the vector space V has a topology (more below on the subject of topology), then a functional on V may or may not be *continuous* with respect to that topology.

Definition 8.10. A *tempered distribution* T is a continuous linear functional from the space $\mathcal{S}(\mathbb{R})$ of Schwartz functions to the complex numbers:

$$T : \mathcal{S}(\mathbb{R}) \to \mathbb{C}.$$

The space of tempered distributions is denoted by $\mathcal{S}'(\mathbb{R})$. Further, two tempered distributions T_a and T_b are said to *coincide in the sense of distributions* if $T_a(\phi) = T_b(\phi)$ for all $\phi \in \mathcal{S}(\mathbb{R})$. A sequence of tempered distributions $\{T_n\}_{n\in\mathbb{N}}$ is said to converge in the distribution sense to $T \in \mathcal{S}'(\mathbb{R})$ if for all $\phi \in \mathcal{S}(\mathbb{R})$

$$\lim_{n\to\infty} T_n(\phi) = T(\phi).$$

$\Diamond$

The notation $\mathcal{S}'(\mathbb{R})$ is used because, more generally, the collection of all continuous linear functionals $T : V \to F$ on a topological vector space V is usually denoted by V'. This collection V' is called the *dual* of V. Thus the space of tempered distributions is the dual of the space of Schwartz functions.

To say that T is linear means that

$$T(a\phi + b\psi) = aT(\phi) + bT(\psi) \qquad \text{for all } a,\, b \in \mathbb{C} \text{ and all } \phi,\, \psi \in \mathcal{S}(\mathbb{R}).$$

Continuity requires a bit more explanation. Very schematically, we note that a *topology* on a set X is a collection of subsets of X; they are known as the *open* subsets of X. The collection has to satisfy certain conditions that we won't go into here. Convergence of a sequence of elements x_n of X can be defined in terms of the topology on X. A function from a set X that has a topology to another set Y that also has a topology is said to be *continuous* if the inverse image of every open subset of Y is an open subset of X. In the familiar cases where X and Y are $\mathbb{R}$ or $\mathbb{R}^n$, there is a standard topology that arises from the notion of Euclidean distance, and the above definition of continuity for functions from X to Y coincides with the usual ε–δ definition of continuity. If X is a normed vector space, with the natural topology that arises from the norm, it turns out that a linear functional on X is continuous if and only if it is bounded. The topology of $\mathcal{S}(\mathbb{R})$ is not induced by a norm but by a *family of seminorms*. In the following aside we describe very briefly what that means. See for example [**Mun**] for more about topology, and [**Fol**, Chapter 5] for more about continuous functionals.

Aside 8.11. *We say a little more about the topology on* $\mathcal{S}(\mathbb{R})$. *If* $\phi \in \mathcal{S}(\mathbb{R})$, *then the quantities* $\rho_{k,\ell}(\phi) = \sup_{x\in\mathbb{R}} |x|^k |\phi^{(\ell)}(x)|$ *are finite for each* $k,\ \ell \in \mathbb{N}$, *see Exercise 7.2. These quantities are* seminorms, *meaning that they are positive (although each one of them is not necessarily positive definite, that is* $\rho_{k,\ell}(\phi) = 0$ *does not imply* $\phi = 0$, *but if for all* $k,\ \ell \in \mathbb{N}$ *the corresponding seminorm is zero, then the function must be zero), they are homogeneous (*$\rho_{k,\ell}(\lambda\phi) = |\lambda|\rho_{k,\ell}(\phi)$*), and they satisfy the triangle inequality. These seminorms generate a topology on* $\mathcal{S}(\mathbb{R})$.

With respect to that topology, **a sequence $\{\phi_n\}$ converges to ϕ in $\mathcal{S}(\mathbb{R})$ if and only if $\lim_{n\to\infty} \rho_{k,\ell}(\phi_n - \phi) = 0$ for all k, $\ell \in \mathbb{N}$.**

A linear functional T on $\mathcal{S}(\mathbb{R})$ is continuous if and only if whenever $\{\phi_n\}$ is a sequence in $\mathcal{S}(\mathbb{R})$ that converges in the topology of $\mathcal{S}(\mathbb{R})$ to some $\phi \in \mathcal{S}(\mathbb{R})$, the sequence $\{T(\phi_n)\}$ of complex numbers converges to $T(\phi)$ in $\mathbb{C}$.

We systematically check the continuity of only a few of the tempered distributions introduced in this chapter. In particular, we verify that the delta distribution is continuous. Other cases are left as exercises for the reader.

The canonical example of a tempered distribution is the functional T_f given by integration against a reasonable function f:

$$T_f(\phi) := \int_{\mathbb{R}} f(x)\phi(x)\,dx \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}). \tag{8.3}$$

How reasonable must the function f be? For the integral to be well defined for all $\phi \in \mathcal{S}(\mathbb{R})$, f cannot grow too fast at infinity. For example, exponential growth would overwhelm the decay of the Schwartz functions ϕ as $x \to \pm\infty$. However, polynomials or continuous functions that increase no faster than a polynomial are acceptable. Functions f in the Schwartz class are as reasonable as they could possibly be since they are rapidly decreasing, and functions of moderate decrease are also reasonable. Both Schwartz functions and functions of moderate decrease are bounded.

We say for short that a distribution T is a function f, when we mean that T is the tempered distribution induced by f, in other words $T = T_f$.

EXAMPLE 8.12. (*Bounded Functions Induce Tempered Distributions*) As an example, we show that if f is a bounded function then the formula (8.3) for T_f defines a continuous linear functional on $\mathcal{S}(\mathbb{R})$. So T_f defines a tempered distribution: $T_f \in \mathcal{S}'(\mathbb{R})$. In particular T_f is a tempered distribution if $f \in \mathcal{S}(\mathbb{R})$ or if f is a function of moderate decrease.

Linearity is a consequence of the linearity of the integral. To check continuity, consider $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, this means in particular that $\lim_{n\to\infty} \rho_{2,0}(\phi_n) = 0$, that is for all $n > N_0$, $|x|^2|\phi_n(x)| \leq 1$. By choosing $M > 0$ large enough we can guarantee that for all $n > N_0$

$$\int_{|x|>M} |\,\phi_n(x)|\,dx \leq \int_{|x|>M} \frac{1}{x^2}\,dx < \frac{\varepsilon}{2}.$$

On the compact interval $[-M, M]$ the functions ϕ_n converge uniformly to zero (this is the fact that $\lim_{n\to\infty} \rho_{0,0}(\phi_n) = 0$), and we can interchange the limit and the integral:

$$\lim_{n\to\infty} \int_{|x|\leq M} |\phi_n(x)|\,dx = \int_{|x|\leq M} \lim_{n\to\infty} |\phi_n(x)|\,dx = 0.$$

So given $\varepsilon > 0$ there exists $N_1 > 0$ so that for all $n > N_1$,

$$\int_{|x|\leq M} |\phi_n(x)|\,dx < \frac{\varepsilon}{2}.$$

Thus for $n > N := \max\{N_0, N_1\}$ we can ensure that

$$\int |\phi_n(x)|\,dx \leq \varepsilon.$$

Finally, T_f is continuous, since for $n > N$,

$$|T_f(\phi_n)| \leq \int |f(x)||\phi_n(x)|\,dx \leq \sup_{x\in\mathbb{R}} |f(x)| \int |\phi_n(x)|\,dx \leq \sup_{x\in\mathbb{R}} |f(x)|\varepsilon,$$

and therefore $\lim_{n\to\infty} T_f(\phi_n) = 0$. ◊

In particular the distribution T_1 induced by the bounded function $f = 1$ is a well defined tempered distribution.

EXERCISE 8.13. (*Polynomials Induce Tempered Distributions*) Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_kx^k$ be a polynomial of degree k. Show that the linear functional defined by integration against the polynomial f is a tempered distribution. ◊

Example 8.12 tells us that we can view $\mathcal{S}(\mathbb{R})$ as sitting inside $\mathcal{S}'(\mathbb{R})$, via the mapping $\varepsilon : \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ given by

$$\varepsilon(f) := T_f \qquad \text{for } f \in \mathcal{S}(\mathbb{R}).$$

The mapping ε is itself continuous, in the sense that if $f_n,\ f \in \mathcal{S}(\mathbb{R})$ and $f_n \to f$ in the topology of $\mathcal{S}(\mathbb{R})$, then $T_{f_n}(\phi) \to T_f(\phi)$ in $\mathbb{C}$ for all $\phi \in \mathcal{S}(\mathbb{R})$, that is, T_{f_n} converges to T_f in the sense of distributions. Such continuous mapping are sometimes called *embeddings*. This nomenclature will be clearer in the next section, where we will define some operations for distributions that generalize ordinary operations on functions, in the sense that if $\mathcal{O}f$ is the action of an operation on a function (for example differentiation, translation, etc.), then the corresponding operation for the distribution is $\mathcal{O}\big(\varepsilon(f)\big) := \varepsilon(\mathcal{O}f)$.

EXERCISE 8.14. Prove that if $f_n,\, f \in \mathcal{S}(\mathbb{R})$ and $f_n \to f$ in $\mathcal{S}(\mathbb{R})$, then $T_{f_n}(\phi) \to T_f(\phi)$ in $\mathbb{C}$ for all $\phi \in \mathcal{S}(\mathbb{R})$. ◊

Like bounded functions, unbounded functions that have some decay at infinity and some integrability properties can also induce tempered distributions via integration, as we show in Section 8.6 for L^p functions. Some tempered distributions are not induced by any function, however, as we show in Section 8.4 for the example of the delta distribution, and in Section 8.6 for the example of the *principal value distribution* of $1/x$.

In Example 8.12 we used the fact that if $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then $\|\phi_n\|_{L^1} \to 0$. A similar argument allows one to show that the L^p-norms will also go to zero.

EXERCISE 8.15. Verify that if $\phi_n \in \mathcal{S}(\mathbb{R})$ and $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then for all $1 \leq p \leq \infty$,

$$\|\phi_n\|_p^p := \int |\phi_n(x)|^p\,dx \to 0.$$

◊

The collection of continuous linear functionals that acts on the space of compactly supported C^∞ functions (a subspace of the Schwartz class) are the so-called *distributions*. If we denote by $\mathcal{D}(\mathbb{R})$ the space of compactly supported C^∞ functions, the space of distributions $\mathcal{D}'(\mathbb{R})$ is its dual. Tempered distributions are distributions, the two spaces are nested, that is

$$\mathcal{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R}).$$

This is no accident but a fundamental fact in functional analysis. Given V, W topological vector spaces, denote their duals by V', W', then if $V \subset W$ (the topology of V is inherited from the topology in W) then $W' \subset V'$. Although we sometimes write just distributions for short, in this textbook we always mean tempered distributions. One could also consider the larger class of infinitely differentiable functions and its dual which will be a subset of $\mathcal{S}'(\mathbb{R})$, the so-called *compactly supported distributions*. If we think about what functions will give a continuous linear functional on C^∞, you realize the functions have to be compactly supported, hence the name (there is also a notion of support for distributions, and of compact support for distributions). The book by Robert Strichartz [**Str2**] is a delightful introduction to distribution theory at a level comparable to that of this book; it includes applications.

8.3. The time–frequency dictionary for $\mathcal{S}'(\mathbb{R})$

We show how to extend from functions to distributions the fundamental operations discussed so far: translation, dilation, differentiation, the Fourier and inverse Fourier transforms, multiplication by Schwartz functions and by polynomials, and convolution by Schwartz functions. We will take it for granted that these extensions of the fundamental operations, and the Fourier transform, are nicely behaved with respect to the corresponding topology.

Given a tempered distribution $T : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$, we define nine new tempered distributions as follows.

(i) The *translate* $\tau_h T$ of T by $h \in \mathbb{R}$, defined by

$$(\tau_h T)(\phi) := T(\tau_{-h}\phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(ii) The *dilation* $D_s T$ of T by $s > 0$, defined by

$$(D_s T)(\phi) := T\big(sD_{1/s}(\phi)\big) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}),$$

where $D_s\phi(x) = s\phi(sx)$.

(iii) The *derivative* T' of T, defined by

$$(T')(\phi) := -T(\phi') \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(iv) The *Fourier transform* $\widehat{T}$ of T, defined by

$$(\widehat{T}\,)(\phi) := T(\widehat{\phi}\,) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(v) The *inverse Fourier transform* $(T)^\vee$ of T, defined by

$$(T)^\vee(\phi) := T(\phi^\vee) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(vi) The *product* $\mathcal{M}_g T$ of T with a function g such that if $\phi \in \mathcal{S}(\mathbb{R})$ then $g\phi \in \mathcal{S}(\mathbb{R})$ (g could be a function on the Schwartz class, a bounded function like the trigonometric functions $e_\xi(x) = e^{2\pi i x\xi}$, or an unbounded function that does not increase too fast, such as a polynomial). In the case of multiplication the trigonometric function e_ξ we called this *modulation* and denoted it $M_\xi = \mathcal{M}_{e_\xi}$.

$$(\mathcal{M}_g T)(\phi) := T(g\phi) = T(\mathcal{M}_g\phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(vii) The *reflection* $\widetilde{T}$ of T defined by,

$$(\widetilde{T})(\phi) := T(\widetilde{\phi}) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(vii) The *conjugate* $\overline{T}$ of T defined by,

$$(\overline{T})(\phi) := T(\overline{\phi}) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(ix) The *convolution* $\psi * T$ of T with $\psi \in \mathcal{S}(\mathbb{R})$, defined by

$$(\psi * T)(\phi) := T(\widetilde{\psi} * \phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

The new functionals defined in (i)–(ix) are linear. It can be shown that they are also continuous. So we have indeed defined new tempered distributions, starting from a given T.

EXERCISE 8.16. Verify that if $T \in \mathcal{S}'(\mathbb{R})$ then the nine linear functionals defined in (i)-(ix) are in $\mathcal{S}'(\mathbb{R})$. $\Diamond$

It takes some time to get used to working with these definitions. The key idea is that since tempered distributions are defined by the way they act on the space of Schwartz functions, it makes sense to define a modification of a tempered distribution by pushing the modification across to the Schwartz function. But we should be careful to which Schwartz function we pass it on, the argument ϕ? NO.

Let us try to understand with an example, differentiation. Since $\mathcal{S}(\mathbb{R})$ is embedded in $\mathcal{S}'(\mathbb{R})$ via the mapping $f \to T_f$, $f \in \mathcal{S}(\mathbb{R})$ we would like to be able to identify a new distribution, T_f', the *derivative* of T_f, with $T_{f'}$, so that the differentiation defined for distributions reduces to good old differentiation of functions when the distribution under consideration is given by integration by the function f. With that in mind, we can attempt to give a meaning to T' for general tempered distributions, not just the ones given by integration. For instance, to understand the derivative T' of a tempered distribution T, we need to know what T' does to each $\phi \in \mathcal{S}(\mathbb{R})$, how are we going to figure it out? We compute $T_{f'}(\phi)$, and attempt to write the output as T_f acting on some modification of ϕ involving, hopefully, ordinary differentiation. In this case, this is nothing more than integration by parts, where the boundary terms disappear because the functions decay at infinity:

$$T_{f'}(\phi) = \int f'(x)\phi(x)\,dx = -\int f(x)\phi'(x)\,dx = -T_f(\phi').$$

Now we keep the left and right ends of these string of equalities and use them to define T' for any tempered distribution T:

$$T'(\phi) := -T(\phi').$$

The process of moving from a function f to its induced tempered distribution T_f interacts as one would expect with the operations of translation of f, dilation of f, and so on. For example, the translate $\tau_h T_f$ of the tempered distribution T_f induced by f is the same as the tempered distribution $T_{\tau_h f}$ induced by the translate $\tau_h f$ of f:

$$\tau_h T_f = T_{\tau_h f}.$$

Note that, as for differentiation, $\tau_h T(\phi) \neq T(\tau_h \phi)$. In a few instances, like Fourier transform and multiplication by functions, the operation "migrates" unchanged to the argument. This is only because those operations are "self-adjoint', meaning that $\int (\mathcal{O}f)\, g = \int f\, (\mathcal{O}g)$.

Example 8.17 and Exercise 8.18 offer some practice in working with the definition of tempered distributions, as we work out the details of this interaction for each of the nine new distributions.

EXAMPLE 8.17. (*Forming the Tempered Distributions induced by Functions Respect Convolution.*) We verify that for Schwartz functions f and ψ, the convolution of ψ with the tempered distribution T_f induced by f is the same as the tempered distribution induced by the convolution of ψ with f:

$$\psi * T_f = T_{\psi * f} \qquad \text{for all } f \in \mathcal{S}(\mathbb{R}),\ \psi \in \mathcal{S}(\mathbb{R}).$$

For f, ψ, and $\phi \in \mathcal{S}(\mathbb{R})$,

$$\begin{aligned} T_{\psi*f}(\phi) &= \int_{\mathbb{R}} \phi(y)(\psi * f)(y)\,dy \\ &= \int_{\mathbb{R}} \phi(y)\left(\int_{\mathbb{R}} f(x)\psi(y-x)\,dx\right) dy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x)\psi(y-x)\phi(y)\,dy\,dx \\ &= \int_{\mathbb{R}} f(x)\left(\int_{\mathbb{R}} \widetilde{\psi}(x-y)\phi(y)\,dy\right) dx \\ &= \int_{\mathbb{R}} f(x)(\widetilde{\psi} * \phi)(x)\,dx \\ &= T_f(\widetilde{\psi} * \phi) \;=\; (\psi * T_f)(\phi), \end{aligned}$$

as required. ◊

Notice that the above calculation leads us to the correct definition of $\psi * T(\phi) = T(\widetilde{\psi} * \phi)$.

It turns out that the object obtained by convolution of $\psi \in \mathcal{S}(\mathbb{R})$ and $T \in \mathcal{S}'(\mathbb{R})$ can be identified with an infinitely differentiable function that grows slower than some polynomial. Specifically, there exists a function $f : \mathbb{R} \to \mathbb{C}$, $f \in C^\infty(\mathbb{R})$, and a polynomial $P(x)$ such that $|f(x)| \le P(|x|)$ for all $x \in \mathbb{R}$, and such that $\psi * T = T_f$, where f depends of course on ψ and T. In fact, f is the function defined by $f(y) = T(\tau_y\widetilde{\psi})$. See [**Graf**, pp.116–117] for details.

Exercise 8.18. (*The Tempered Distributions induced by Functions Respect the Fundamental Operations*) Check that for all $f \in \mathcal{S}(\mathbb{R})$, $h \in \mathbb{R}$, $s > 0$, and $g \in C^\infty$ such that $gf \in \mathcal{S}(\mathbb{R})$ for all $f \in \mathcal{S}(\mathbb{R})$, we have

$$\tau_h(T_f) = T_{\tau_h f}, \quad D_s(T_f) = T_{D_s f}, \quad \widehat{T_f} = T_{\widehat{f}}, \quad (T_f)^\vee = T_{(f)^\vee},$$

$$\widetilde{T_f} = T_{\widetilde{f}}, \quad \overline{T_f} = T_{\overline{f}}, \quad \text{and } (\mathcal{M}_g T_f) = T_{\mathcal{M}_g f}.$$

◊

Definitions (iv) and (v) show that the Fourier transform can be extended to be a bijection on $\mathcal{S}'(\mathbb{R})$, because it is a bijection on $\mathcal{S}(\mathbb{R})$. Just observe that

$$[\widehat{T}]^\vee(\phi) = T([\widehat{\phi}]^\vee) = T(\phi).$$

Similarly $\widehat{[T]^\vee} = T$.

We can now build a time–frequency dictionary on $\mathcal{S}'(\mathbb{R})$, displayed in Table 8.1 that is in a one-to-one correspondence with the time–frequency dictionary we built for $\mathcal{S}(\mathbb{R})$ (Table 7.1). In the time column we recall the definitions of the basic operations with tempered distributions; in the frequency column, how they interact with the Fourier transform.

TABLE 8.1. The time–frequency dictionary in $\mathcal{S}'(\mathbb{R})$.

	Time	**Frequency**
(a)	linear properties $aT_1 + bT_2$	linear properties $a\widehat{T_1} + b\widehat{T_2}$
(b)	translation $\tau_h T(\phi) := T(\tau_{-h}\phi)$	modulation $\widehat{\tau_h T} = M_{-h}\widehat{T}$
(c)	modulation by $e_h(x) = e^{2\pi i h x}$ $M_h T(\phi) := T(M_h\phi)$	translation $\widehat{M_h T} = \tau_h \widehat{T}$
(d)	dilation $D_s T(\phi) := T(sD_{s^{-1}}\phi)$	inverse dilation $\widehat{D_s T} = sD_{s^{-1}}\widehat{T}$
(e)	reflection $\widetilde{T}(\phi) := T(\widetilde{\phi})$	reflection $\widehat{\widetilde{T}} = -\widetilde{\widehat{T}}$
(f)	conjugate $\overline{T}(\phi) := T(\overline{\phi})$	conjugate reflection $\widehat{\overline{T}} = \overline{[T]^{\vee}}$
(g)	derivative $T'(\phi) := -T(\phi)$	multiply by polynomial $\widehat{T'} = [2\pi i\xi]\widehat{T} = \mathcal{M}_{2\pi i x}\widehat{T}$
(h)	multiply by polynomial $\mathcal{M}_{-2\pi i x}T(\phi) := T(-2\pi i x\phi(x))$	derivative $[\mathcal{M}_{-2\pi i x}T]^{\wedge} = \widehat{T}'$
(i)	convolution $\psi * T(\phi) := T(\widetilde{\psi} * \phi)$	product $\widehat{\psi * T} = M_{\widehat{\psi}}\widehat{T}$

EXAMPLE 8.19. As an example of how to derive the time–frequency dictionary encoded in Table 8.1, we derive the multiplication formula for the convolution. Consider $\psi, \phi \in \mathcal{S}(\mathbb{R})$, apply the definitions of Fourier transform of a distribution, and of convolution with a distribution,

$$\widehat{\psi * T}(\phi) = \psi * T(\widehat{\psi}) = T(\widetilde{\psi} * \widehat{\phi}). \tag{8.4}$$

Let us look more closely at the argument on the right hand side of (8.4), and we will see that it is the Fourier transform of $\widehat{\psi}\phi$,

$$\begin{aligned}
\widetilde{\psi} * \widehat{\phi}(\xi) &= \int \widetilde{\psi}(s)\widehat{\phi}(\xi - s)\, ds \\
&= \int \widetilde{\psi}(s)\left(\int \phi(x)e^{-2\pi i x(\xi - s)}\, dx\right) ds \\
&= \int \left(\int \psi(-s)e^{-2\pi i x(-s)}\, ds\right) \phi(x)e^{-2\pi i x\xi}\, dx \\
&= \int \widehat{\psi}(x)\phi(x)e^{-2\pi i x\xi}\, dx \\
&= \widehat{\widehat{\psi}\phi}.
\end{aligned}$$

Now we can continue, using again the definition of the Fourier transform of a distribution and the definition of multiplication of a distribution by the function $\widehat{\psi}$, we get that for all $\psi \in \mathcal{S}(\mathbb{R})$,

$$\widehat{\psi * T}(\phi) = T(\widehat{\widehat{\psi}\phi}) = \widehat{T}(\widehat{\psi}\phi) = \mathcal{M}_{\widehat{\psi}}\widehat{T}(\phi).$$

This is exactly the multiplication formula we were seeking:

$$\widehat{\psi * T} = \mathcal{M}_{\widehat{\psi}}\widehat{T}.$$

◊

Exercise 8.20. Derive the formulas in the frequency column of Table 8.1. It will be useful to consult Table 7.1, the corresponding time–frequency dictionary for the Schwartz class. ◊

8.4. The delta distribution

There are objects in $\mathcal{S}(\mathbb{R})$ other than functions. The *delta distribution* is the canonical example of a tempered distribution that is not a function. Informally, the delta distribution "takes the value infinity at $x = 0$ and zero everywhere else". Rigorously, the delta distribution is the Fourier transform of the tempered distribution induced by the function that is identically equal to one. Computationally, the delta distribution has a very simple effect on Schwartz functions: it evaluates them at the point 0, as we now show.

By Exercise 8.12, the bounded function $f(x) \equiv 1$ induces a tempered distribution T_1 via integration against the function 1:

$$T_1(\phi) := \int_{\mathbb{R}} \phi(x)\,dx \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

The Fourier transform of T_1 defines a new tempered distribution $\widehat{T_1}$ by

$$\widehat{T_1}(\phi) := T_1(\widehat{\phi}\,) = \int_{\mathbb{R}} \widehat{\phi}\,(\xi)\,d\xi = \int_{\mathbb{R}} \widehat{\phi}\,(\xi)e^{2\pi i\xi 0}\,d\xi = (\widehat{\phi}\,)^{\vee}(0) = \phi(0)$$

for $\phi \in \mathcal{S}(\mathbb{R})$. The last equality holds by of the Fourier Inversion Formula for Schwartz functions (Theorem 7.33).

Definition 8.21. The *delta distribution*, $\delta : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$, is defined by

$$\delta(\phi) := \widehat{T_1}(\phi\,) = \phi(0) \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

◊

It is clear from the definition that δ is linear. It can be shown that δ is also continuous, and so δ is indeed a tempered distribution. In fact, suppose $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then we want to verify that the complex numbers $\delta(\phi_n) = \phi_n(0) \to 0$ as $n \to \infty$. But $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$ implies that in particular $\rho_{0,0}(\phi_n) = \sup_{x\in\mathbb{R}} |\phi_n(x)| \to 0$, clearly $|\phi_n(0)| \le \rho_{0,0}(\phi_n)$ hence it must go to zero as desired.

What is its Fourier transform $\widehat{\delta}$? We compute that

$$\widehat{\delta}\,(\phi) = \delta(\widehat{\phi}\,) = \widehat{\phi}\,(0) = \int_{\mathbb{R}} \phi(x)\,dx = T_1(\phi) \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

That is, the Fourier transform $\widehat{\delta}$ of δ is the tempered distribution T_1 induced by the function $f(x) \equiv 1$. We write $\widehat{\delta} = T_1$, or sometimes $\widehat{\delta} = 1$ for short.

EXERCISE 8.22. Show that the *inverse* Fourier transform of δ is also given by T_1,

$$[\delta]^{\vee}(\phi) = T_1(\phi) = \widehat{\delta}\,(\phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

$\Diamond$

Why isn't the delta distribution a true function? For a moment let's try to think of the delta distribution as a function, written $\delta(x)$ say, and called the *delta function*. Then the distribution induced by $\delta(x)$ should give us back the delta *distribution* again, and so we would have

$$\int_{\mathbb{R}} \delta(x)\phi(x)\,dx = \phi(0) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}). \tag{8.5}$$

In particular the integral on the left must be zero for all $\phi \in \mathcal{S}(\mathbb{R})$ such that $\phi(0) = 0$. So $\delta(x)$ must be zero for 'most' x; if for instance $\delta(x)$ were positive on some interval not including the point $x = 0$, we could build a Schwartz function supported on that interval whose integral against $\delta(x)$ would not be zero, contradicting equation (8.5). One can use measure theory to extend this idea, concluding that $\delta(x)$ can only be non-zero at a small collection of points x, specifically on a set E of measure zero, and further that we can assume $\delta(x) = 0$ at all points except $x = 0$, since even if δ is non-zero on E, the set is so small that the values of δ on E do not contribute anything to the integral. (To see the idea, think of E as being just a single point x_0.) Thus $\delta(x)$ would have to be a function whose mass is concentrated at zero. In other words, the support[1] of the delta function $\delta(x)$ should be the singleton set $\{0\}$. To reiterate, the delta function should take the value zero at all $x \neq 0$. But what value can we assign to it at the special point $x = 0$ so that integration against $\delta(x)$ will pick out the value of ϕ at 0? If we assign a finite value there, then $\int_{\mathbb{R}} \delta\phi = 0$ rather than 1. Therefore $\delta(0)$ cannot be a finite number, and so $\delta(0)$ must be infinity. Thus the delta function is not a function in the usual sense.

Fortunately, all is not lost: one can interpret δ as a distribution as we do in this section, or as a *point-mass measure* in the language of measure theory [**Fol**].

It is interesting to note that the support of the delta distribution is as small as it could possibly be without being the empty set, while the support of its Fourier transform is as large as it could possibly be. This observation illustrates a time–frequency principle.

> *The smaller the support of the function, the larger the support of its Fourier transform, and vice versa.*

This idea will be made precise by the *Heisenberg Uncertainty Principle* in Section 8.5.3.

EXERCISE 8.23. (*The Derivatives and the Antiderivative of the Delta Distribution.*) Find the derivatives $D^k\delta$, $k \geq 1$, of the delta distribution. Check that their Fourier transforms $\widehat{D^k\delta}$ can be identified with the polynomials $(2\pi ix)^k$. Check that δ coincides, in the sense of distributions, with the derivative H' of the *Heaviside function* H defined by $H(x) = 0$ if $x \leq 0$, $H(x) = 1$ if $x > 0$. $\Diamond$

---

[1]One can define a notion of support of a distribution [**Str2**, Section 6.1] that extends the usual notion of the support of a function (Definition 7.4). With that notion the support of the delta distribution is the singleton set $\{0\}$. In fact, all distributions supported on $\{0\}$ are finite linear combinations of the delta distribution and its derivatives. See [**Str2**, Sections 6.2, 6.3] and also [**Graf**, p.122].

We see that although the delta distribution is not a function, it can be viewed as the derivative of a function. The delta distribution is not an anomaly in this respect. It turns out that for *each tempered distribution*, no matter how wild, there is a function such that the given tempered distribution is the derivative of some order (in the sense of distributions) of the function. For example, consider the distribution given by the k^{th} derivative of the delta distribution. By Exercise 8.23, $D^k\delta$ coincides with the $(k+1)^{\text{th}}$-derivative of the Heaviside function. See [**Str2**, Section 6.2].

We now return to the ideas at the start of Section 7.5, where we lamented that there is no identity element in $\mathcal{S}(\mathbb{R})$, and suggested that the delta distribution, and more generally an approximation to the identity, could act in that role.

EXERCISE 8.24. (*The Delta Distribution Acts as an Identity Under Convolution.*) Verify that the delta distribution acts as an identity for the operation of convolution. More precisely, show that for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\phi * \delta = \phi \qquad \text{(as usual, we mean } = T_\phi\text{)}.$$

$\Diamond$

EXERCISE 8.25. (*Approximations of the Identity Converge to the Delta Distribution.*) Let $\{K_t(x)\}_{x\in\Lambda}$ with some t_0 be an approximation to the identity in $\mathbb{R}$. Prove that the distribution induced by K_t converges to the delta distribution as $t \to t_0$, in the sense of distributions, that is, for all $\phi \in \mathcal{S}(\mathbb{R})$, $T_{K_t}(\phi) \to \delta(\phi) = \phi(0)$ as $t \to t_0$. $\Diamond$

For those who have seen some measure theory, we note that every Borel measure μ on $\mathbb{R}$ defines a tempered distribution T_μ, via integration, similarly to the way in which equation (8.3) defines the tempered distribution induced by a function:

$$T_\mu(\phi) := \int_{\mathbb{R}} \phi(x)\, d\mu \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

In the next section we discuss some beautiful and very classical consequences of the Fourier theory. To close the chapter, we return to tempered distributions, Fourier transform and its interplay with the Lebesgue spaces $L^p(\mathbb{R})$. We also work out the details of a tempered distribution unlike the examples that we have discussed so far. This tempered distribution is not induced by a function, nor is it a linear combination of delta distributions or their derivatives. It is the so-called principal value distribution of $1/x$, the building block for the Hilbert transform.

8.5. Some applications of the Fourier transform

Among many applications of the Fourier transform we have chosen to present some of the most classical applications. The formulae and inequalities discussed in the next pages have remarkable consequences in a range of subjects: from number theory (the Poisson Summation Formula), to signal processing (the Shannon Sampling Formula), to quantum mechanics (Heisenberg's Uncertainty Principle). The proofs are beautiful and elegant. For many more applications to physics, partial differential equations, probability, and more, see the books by Strichartz [**Str2**], by Dym and McKean [**DM**], by Körner [**?**] by Stein and Shakarchi [**SS1**], and by Reed and Simon [**RS**].

8.5.1. The Poisson Summation Formula. Given a function on the line, how can we construct a periodic function of period 1?

There are two methods. One is called *periodization*, and the other uses a Fourier series whose Fourier coefficients are given by the Fourier transform evaluated at the integers. The Poisson Summation Formula gives conditions under which these two methods coincide.

The periodization of a function ϕ on the line is constructed by summing over its integer translates,

$$P_1\phi(x) = \sum_{n\in\mathbb{Z}} \phi(x+n) = \lim_{N\to\infty} \sum_{|n|\le N} \phi(x+n),$$

where the series is supposed to converge in the sense of symmetric partial sums. Notice that when ϕ is in $\mathcal{S}(\mathbb{R})$ or is a function of moderate decrease, the sum in the periodization is absolutely convergent.

EXAMPLE 8.26. If H_t is the *heat kernel* on the line, defined by formula (7.21), then its periodization is the periodic heat kernel

$$P_1 H_t(x) = (4\pi t)^{-1/2} \sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t}.$$

◊

The second method assigns to a given function $\phi \in \mathcal{S}(\mathbb{R})$ a periodic function defined by,

$$P_2\phi(x) = \sum_{n\in\mathbb{Z}} \widehat{\phi}(n) e^{2\pi ixn}.$$

The series is well defined and convergent because $\phi \in \mathcal{S}(\mathbb{R})$.

EXAMPLE 8.27. If $f \in L^1(0,1)$ and is defined to be zero outside $(0,1)$, then both periodizations are well defined, and they coincide with its periodic extension to the whole line (in particular, $P_1 f = P_2 f$). ◊

The phenomenon displayed in the example is not an accident. The Poisson Summation Formula states precisely that.

THEOREM 8.28 (Poisson Summation Formula). *If $\phi \in \mathcal{S}(\mathbb{R})$ then*

$$P_1\phi(x) = \sum_{n\in\mathbb{Z}} \phi(x+n) = \sum_{n\in\mathbb{Z}} \widehat{\phi}\,(n) e^{2\pi inx} = P_2\phi(x). \tag{8.6}$$

In particular, setting $x = 0$,

$$\sum_{n\in\mathbb{Z}} \phi(n) = \sum_{n\in\mathbb{Z}} \widehat{\phi}\,(n). \tag{8.7}$$

PROOF. It suffices to check that the right- and left-hand sides of equation (8.6) have the same Fourier coefficients. The n^{th} Fourier coefficient on the right-hand side is $\widehat{\phi}(n)$. Computing the left-hand side we get

$$\begin{aligned}\int_0^1 \left(\sum_{m\in\mathbb{Z}} \phi(x+m)\right) e^{-2\pi inx}\,dx &= \sum_{m\in\mathbb{Z}} \int_0^1 \phi(x+m)e^{-2\pi inx}\,dx \\ &= \sum_{m\in\mathbb{Z}} \int_m^{m+1} \phi(y)e^{-2\pi iny}\,dy \\ &= \int_{-\infty}^{\infty} \phi(y)e^{-2\pi iny}\,dy \\ &= \widehat{\phi}(n).\end{aligned}$$

We may interchange the sum and the integral in the first step since f is rapidly decreasing. □

EXAMPLE 8.29. If H_t is the heat kernel, then all conditions in the Poisson Summation Formula are satisfied. Furthermore $\widehat{H_t}(n) = e^{-4\pi^2 tn^2}$, and so the periodic heat kernel has a natural Fourier series in terms of separated solutions of the heat equation:

$$(4\pi t)^{-1/2} \sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t} = \sum_{n\in\mathbb{Z}} e^{-4\pi^2 tn^2} e^{2\pi inx}.$$

♢

The above calculation can be justified for functions of moderate decrease, and so the Poisson Summation Formula also holds for that larger class of functions.

EXERCISE 8.30. Use the Poisson Formula to show that the periodizations of the Poisson and Fejér kernels on the line coincide with the 1-periodic Poisson and Fejér kernels, introduced in Sections 4.5 and 4.6. ♢

We can give an interpretation in the language of distributions of the Poisson Summation Formula. Recall that $\tau_n\delta(\phi) = \phi(n)$ for all $\phi \in \mathcal{S}(\mathbb{R})$. With this in mind, the left hand side of (8.7) can be writen as,

$$\sum_{n\in\mathbb{Z}} \phi(n) = \sum_{n\in\mathbb{Z}} \tau_n\delta(\phi).$$

Similarly for the right hand side,

$$\begin{aligned}\sum_{n\in\mathbb{Z}} \widehat{\phi}(n) &= \sum_{n\in\mathbb{Z}} \tau_n\delta(\widehat{\phi}), \\ &= \sum_{n\in\mathbb{Z}} \widehat{\tau_n\delta}(\phi),\end{aligned}$$

where in the last equality we used the definition of the Fourier transform for distributions. What the Poisson Summation Formula is saying is that the Fourier transform of the distribution $\sum_{n\in\mathbb{Z}} \tau_n\delta$ is the same distribution,

$$\widehat{\sum_{n\in\mathbb{Z}} \tau_n\delta} = \sum_{n\in\mathbb{Z}} \tau_n\delta.$$

EXERCISE 8.31. Verify that $\sum_{n\in\mathbb{Z}} \tau_n\delta$ is a tempered distribution. ♢

Notice that the distribution above is like an infinite comb with point masses at the integers. Its Fourier transform can be calculated directly, by noticing that $\widehat{\tau_n\delta}$ is the distribution induced by the function $e^{2\pi i n\xi}$, see Exercise 8.40, so the apparently very complicated distribution $\sum_{n\in\mathbb{Z}} e^{2\pi i n\xi}$ is nothing more than the infinite comb $\sum_{n\in\mathbb{Z}} \tau_n\delta$.

See [**Str2**, Section 7.3] for a beautiful application of the Poisson Summation Formula in two dimensions to crystals and quasicrystals.

8.5.2. The Shannon Sampling Formula. The *Shannon Sampling Formula* states that *band-limited functions* (functions whose Fourier transform is supported on a compact interval) can be recovered from appropriate samples. We will need the so-called *sine kernel* function sinc, defined by

$$\operatorname{sinc}(x) := \frac{\sin(\pi x)}{\pi x}.$$

THEOREM 8.32 (Shannon Sampling Formula). *Suppose f is a function of moderate decrease such that*

$$\operatorname{supp} \widehat{f} \subset [-1/2, 1/2].$$

Then

$$f(x) = \sum_{n\in\mathbb{Z}} f(n)\, \operatorname{sinc}(x-n).$$

PROOF. Since the support of $\widehat{f}$ is the interval $[-1/2, 1/2]$, we can consider the periodization of $\widehat{f}$ to $\mathbb{R}$. Viewed as a 1-periodic function, we can compute its Fourier coefficients (notice that $\widehat{f}$ is the Fourier transform in $\mathbb{R}$, and in the next formula we are abusing the hat notation to denote also the Fourier coefficients of a 1-periodic function, $(g)^\wedge(n)$):

$$\begin{aligned} (\widehat{f}\,)^\wedge(n) &= \int_{-1/2}^{1/2} \widehat{f}\,(x) e^{-2\pi i n x}\, dx \\ &= \int_{-\infty}^{\infty} \widehat{f}\,(x) e^{2\pi i(-n)x}\, dx && \big(\widehat{f} \text{ is supported on } [-1/2, 1/2]\big) \\ &= f(-n). && (\text{Fourier Inversion Formula on } \mathbb{R}) \end{aligned}$$

The Fourier inversion formula was justified for functions in $\mathcal{S}(\mathbb{R})$ and for functions of moderate decrease such that $\widehat{f}$ is of moderate decrease. This entails verifying that $\widehat{f}$ is continuous, which it is true for all functions in $L^1(\mathbb{R})$ as a consequence of the Riemann–Lebesgue Lemma 8.46 (see also Exercise 8.47), and that it decreases at a certain rate. Notice that the band-limited assumption forces $\widehat{f}$ to vanish outside the interval $[-1/2, 1/2]$, hence it decays as fast as it could be. We can safely apply the Fourier inversion formula for band-limited functions of moderate decrease.

Therefore, on the interval $[-1/2, 1/2]$, $\widehat{f}$ coincides with its Fourier series:

$$\widehat{f}\,(\xi) = \sum_{n\in\mathbb{Z}} f(-n) e^{2\pi i n\xi} = \sum_{n\in\mathbb{Z}} f(n) e^{-2\pi i n\xi}, \qquad \text{for } \xi \in [-1/2, 1/2]. \tag{8.8}$$

The Fourier inversion formula on $\mathbb{R}$ also tells us that

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

Using the compact support of $\widehat{f}$ again, substituting formula (8.8), and doing a little more calculation, we can obtain the desired formula. More precisely,

$$\begin{aligned} f(x) &= \int_{-1/2}^{1/2} \widehat{f}(\xi) e^{2\pi i \xi x}\, d\xi \;=\; \int_{-1/2}^{1/2} \sum_{n\in\mathbb{Z}} f(n) e^{2\pi i n\xi} e^{-2\pi i \xi x}\, d\xi \\ &= \sum_{n\in\mathbb{Z}} f(n) \int_{-1/2}^{1/2} e^{2\pi i n\xi} e^{-2\pi i \xi x}\, d\xi \;=\; \sum_{n\in\mathbb{Z}} f(n) \int_{-1/2}^{1/2} e^{2\pi i (x-n)\xi}\, d\xi \\ &= \sum_{n\in\mathbb{Z}} f(n)\, \operatorname{sinc}(x-n). \end{aligned}$$

The interchange of sum and integral is justified because for each x there is uniform convergence as functions of $\xi \in [-1/2, 1/2]$ of the partial sums $g_N(x,\xi) := \sum_{|n|\leq N} f(n) e^{2\pi i n\xi} e^{-2\pi i \xi x}$ as $N \to \infty$ to the series, see Exercise 8.33.

The last identity holds because

$$\operatorname{sinc}(x) = \int_{-1/2}^{1/2} e^{2\pi i x\xi}\, d\xi. \qquad \square$$

EXERCISE 8.33. Verify using the M-Weirstrass test that for each x the functions $g_N(x,\xi) := \sum_{|n|\leq N} f(n) e^{2\pi i n\xi} e^{-2\pi i \xi x}$ converge uniformly in $\xi \in [-1/2, 1/2]$ as $N \to \infty$.

The assumption that the support of the Fourier transform lies in an interval of length 1 is artificial. As long as the Fourier transform is supported on a compact interval, say of length L, we can carry out similar calculations. We would use the Fourier theory for L-periodic functions; see Section 1.3.2.

EXERCISE 8.34. Suppose that f is a function of moderate decrease and supp $\widehat{f} \subset [-L/2, L/2]$. Show that

$$f(x) = \sum_{n\in\mathbb{Z}} f(n/L)\; \operatorname{sinc}\big((x-n)/L\big).$$

$\Diamond$

The larger the support of $\widehat{f}$, the more we will have to sample.

This statement is reasonable, for if only frequencies satisfying $|\xi| < L/2$ are present in f, then it is plausible that sampling at the $1/L$ rate will suffice. Sampling at a larger rate would not suffice.

This formula has many applications in signal processing. See [**Sha**].

8.5.3. Heisenberg's Uncertainty Principle. *It is impossible to find a function that is simultaneously well localized in time and in frequency.* The Heisenberg Uncertanty Principle is an inequality from which one can infer (if one reads carefully) the above principle. This idea can be carried one step further: there is no function that is both compactly supported and band-limited.

THEOREM 8.35 (Heisenberg's Uncertainty Principle). *Suppose that $\psi \in \mathcal{S}(\mathbb{R})$, and that ψ is normalized in $L^2(\mathbb{R})$: $\|\psi\|_2 = 1$. Then*

$$\left(\int_{\mathbb{R}} x^2 |\psi(x)|^2\, dx\right) \left(\int_{\mathbb{R}} \xi^2 |\widehat{\psi}(\xi)|^2\, d\xi\right) \geq \frac{1}{16\pi^2}.$$

Before we present the proof, let us use the Heisenberg Uncertainty Principle to argue that IF the support of a function $f \in \mathcal{S}(\mathbb{R})$ is on the interval $[-a, a]$, and the support of its Fourier transform is on $[-b, b]$ (by appropriate translations and dilations we can always consider supports centered at zero), and $\|f\|_2 = \|\widehat{f}\|_2$, then $ab \geq 1/4\pi$. This will imply that if a is very small, then b has to be large, and viceversa, so that their product remains bounded away from zero. Of course there is no function that has both the assumed properties, but one can then consider approximate supports where most of the L^2-norm is concentrated. Under the assumption, one can estimate,

$$\int_{\mathbb{R}} x^2 |\psi(x)|^2 \, dx \leq \int_{-a}^{a} x^2 |\psi(x)|^2 \, dx \leq a^2 \|f\|_2^2 = a^2.$$

Likewise,

$$\int_{\mathbb{R}} \xi^2 |\widehat{\psi(\xi)}|^2 \, d\xi \leq b^2.$$

Applying now the Heisenberg's Uncertainty Principle and taking square root we conclude that $ab \geq 1/4\pi$ as claimed.

Proof. By hypothesis

$$1 = \int_{\mathbb{R}} |\psi(x)|^2 \, dx.$$

We integrate by parts, setting $u = |\psi(x)|^2$ and $dv = dx$, and taking advantage of the fast decay of ψ to get rid of the boundary terms. Note that

$$\frac{d}{dx}|\psi|^2 = \frac{d}{dx}\psi\overline{\psi} = \overline{\psi}\psi' + \overline{\overline{\psi}\psi'} = 2\mathrm{Re}(\overline{\psi}\psi').$$

Also remember that $\mathrm{Re} z \leq |z|$, and that $|\int f| \leq \int |f|$. Hence

$$1 = 2\int_{\mathbb{R}} x \mathrm{Re}\big[\overline{\psi(x)}\psi'(x)\big] \, dx \leq 2 \int_{\mathbb{R}} |x||\psi(x)||\psi'(x)| \, dx.$$

Now apply the Cauchy–Schwarz inequality to the right-hand side obtaining

$$1 \leq 2 \left(\int_{\mathbb{R}} |x\psi(x)|^2 \, dx \right)^{1/2} \left(\int_{\mathbb{R}} |\psi'(x)|^2 \, dx \right)^{1/2}.$$

But by Plancherel and property (g) of the time–frequency dictionary,

$$\int_{\mathbb{R}} |\psi'(x)|^2 \, dx = 4\pi^2 \int_{\mathbb{R}} \xi^2 |\widehat{\psi}(\xi)|^2 \, d\xi.$$

This completes the proof. □

The Uncertainty Principle holds for the class of functions of moderate decrease, and for the even larger class $L^2(\mathbb{R})$ of square-integrable functions.

Exercise 8.36. Check that equality holds in Heisenberg's Uncertainty Principle if and only if

$$\psi(x) = \sqrt{2B/\pi}\, e^{-Bx^2} \qquad \text{for some } B > 0.$$

Hint: Remember that equality holds in the Cauchy–Schwarz inequality if and only if one of the functions is a constant multiple of the other. ♢

Exercise 8.37. Check that under the conditions of Theorem 8.35,

$$\left(\int_{\mathbb{R}} (x-x_0)^2 |\psi(x)|^2\, dx\right)\left(\int_{\mathbb{R}} (\xi-\xi_0)^2 |\widehat{\psi}(\xi)|^2\, d\xi\right) \geq \frac{1}{16\pi^2}$$

for all x_0, $\xi_0 \in \mathbb{R}$. **Hint:** Use Heisenberg's Uncertainty Principle and the time–frequency dictionary for $\psi_{x_0,\xi_0}(x) = e^{2\pi i x \xi_0}\psi(x+x_0) \in \mathcal{S}(\mathbb{R})$. ♢

Exercise 8.38. Prove the following inequality:

$$\int_{\mathbb{R}} x^2 |\psi(x)|^2\, dx + \int_{\mathbb{R}} \xi^2 |\widehat{\psi}(\xi)|^2\, d\xi \geq \frac{1}{2\pi}$$

for $\psi \in \mathcal{S}(\mathbb{R})$. ♢

Exercise 8.39. Show that there is no function that is compactly supported both in time and in frequency, as follows. Suppose f is continuous on $\mathbb{R}$. Show that f and $\widehat{f}$ cannot both be compactly supported, unless $f = 0$. **Hint:** Assume f is supported on $[0, 1/2]$, expand f in a Fourier series on $[0, 1]$, and observe that f must be a trigonometric polynomial, but trigonometric polynomials can not vanish on an interval. ♢

The trigonometric functions are not in $L^2(\mathbb{R})$. However we can view them as distributions, and we can compute their Fourier transforms.

Exercise 8.40. Check that the Fourier transform of $e_\xi(x) = e^{2\pi i x\xi}$ in the sense of distributions is the shifted delta distribution, δ_ξ, defined by $\delta_\xi(\phi) = \phi(\xi)$. ♢

The preceding exercise also illustrates the time–frequency localization principle.

Exercise 8.41. Prove the analogues of the previous exercise in the finite-dimensional context, and in the Fourier-series context. Which vectors/sequences play the role of the delta distribution? ♢

In general, if the support of a function is essentially localized on an interval of length d, then its Fourier transform is essentially localized on an interval of length d^{-1}. This says that the support on the *phase plane* or *time–frequency plane* will be essentially a *rectangle* of area 1 and dimensions $d \times d^{-1}$.

Exercise 8.42. Draw the phase planes for the trigonometric basis, the standard basis, and the Haar basis in $\mathbb{C}^8$ introduced in Chapter 6. ♢

8.6. $L^p(\mathbb{R})$ as distributions

Functions in the classical *Lebesgue spaces* $L^p(\mathbb{R})$, for real numbers p with $1 \leq p < \infty$, also define tempered distributions. The space $L^p(\mathbb{R})$ consists of those functions such that $\int_{\mathbb{R}} |f(x)|^p\, dx < \infty$. Here the integral is in the sense of Lebesgue. In Section 2.1 we discussed the analogous spaces $L^p(\mathbb{T})$ of L^p functions on the unit circle. All the $L^p(\mathbb{R})$ spaces are *normed*, with norm given by

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p\, dx\right)^{1/p}.$$

The $L^p(\mathbb{R})$ spaces are also *complete*, meaning that every Cauchy sequence in the space converges to an element in the space. Complete normed spaces are also known as *Banach spaces*. The space $L^2(\mathbb{R})$ of square-integrable functions on the line is a *Hilbert space*, with *inner product* given by

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)\overline{g(x)}\, dx.$$

Functions in $L^p(\mathbb{R})$ are always locally integrable. In symbols we write $L^p(\mathbb{R}) \subset L^1_{\text{loc}}(\mathbb{R})$.

The Schwartz class is *dense* in each $L^p(\mathbb{R})$ (see Appendix A). In other words, we can approximate any function $f \in L^p(\mathbb{R})$ by functions $\phi_n \in \mathcal{S}(\mathbb{R})$ such that the L^p norm of the difference $f - \phi_n$ tends to zero as n tends to infinity,

$$\lim_{n\to\infty} \|f - \phi_n\|_p = 0.$$

Or equivalently, given $\epsilon > 0$ there exists a $\phi \in \mathcal{S}(\mathbb{R})$ such that

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

In fact more is true: the smaller class of compactly supported C^∞ functions (bump functions) are also dense in $L^p(\mathbb{R})$.

The idea that the Schwartz functions are dense in $L^p(\mathbb{R})$ leads to another way to think of the space $L^p(\mathbb{R})$. We have already mentioned that we can equip $\mathcal{S}(\mathbb{R})$ with the L^p-norm, making it into a normed space, although this space is not complete. If we "completed" the space by adding equivalence classes of Cauchy sequences in $\mathcal{S}(\mathbb{R})$ with respect to the L^p-norm, in the same way as one constructs the real numbers from the rational numbers, then at the end of this completion process (and there is an end to it) we obtain the Lebesgue spaces $L^p(\mathbb{R})$. [***similar comment was made in section 7.8]

We say that an identity, $f = g$, or a limit, $\lim_{t\to t_0} f_t = f$, holds *in the L^p-sense* if

$$\|f - g\|_p = 0, \qquad \text{or } \lim_{t\to t_0} \|f_t - f\|_p = 0.$$

Equality in the L^p-sense occurs if and only if $f = g$ except possibly on a set of measure zero, that is $f = g$ a.e.

Functions $f \in L^p(\mathbb{R})$ define tempered distributions T_f by integration. The first thing to worry about is that the integral in equation (8.3) must be well defined for all $f \in L^p(\mathbb{R})$ and $\phi \in \mathcal{S}(\mathbb{R})$. In fact the integral $\int_{\mathbb{R}} f(x)\phi\, dx$ is well defined (in the sense of Lebesgue) as long as $f \in L^p(\mathbb{R})$ and $\phi \in L^q(\mathbb{R})$ for p, q *dual or conjugate exponents*, that is such that $\frac{1}{p} + \frac{1}{q} = 1$ (in particular if $\psi \in \mathcal{S}(\mathbb{R}) \subset L^q(\mathbb{R})$ for all $1 \le q \le \infty$). This is a consequence of the celebrated *Hölder's Inequality* (proved in Section 11.5). Hölder's Inequality says that if $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$ with $\frac{1}{p} + \frac{1}{q} = 1$, then $fg \in L^1(\mathbb{R})$. Furthermore,

$$\|fg\|_1 = \int_{\mathbb{R}} |f(x)g(x)|\, dx \le \|f\|_p \|g\|_q.$$

This estimate implies that $\int_{\mathbb{R}} f(x)g(x)\, dx$ is well defined and finite. Furthermore, the linearity of the integral guarantees the linearity of the functional T_f. Hölder's inequality can be used to prove the continuity of the functional, once one notices that the L^p-norms of a function $\phi \in \mathcal{S}(\mathbb{R})$ can be controlled by the Schwartz seminorms, see Exercise 8.15. Notice that when $p = q = 2$ Hölder's inequality is nothing more than the Cauchy-Schwarz inequality in $L^2(\mathbb{R})$.

EXERCISE 8.43. Verify that if $f \in L^p(\mathbb{R})$ then the linear functional T_f defined by integration against f, is continuous, that is if $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$ then the sequence of complex numbers $T_f(\phi) \to 0$. $\Diamond$

With this discussion in mind, we can compute the Fourier transform of L^p-functions in the sense of distributions. Can we identify these Fourier transforms with reasonable functions? There are two cases we have already discussed, at least in the cases of Schwartz functions and functions of moderate decrease.

- $\boxed{p=1}$ If $f \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$, then
$$|\widehat{f}(\xi)| = \left| \int_{\mathbb{R}} f(x) e^{2\pi i x\xi}\, dx \right| \le \int_{\mathbb{R}} |f(x)|\, dx.$$
It follows that, in the language of norms, $\|\widehat{f}\|_\infty \le \|f\|_1$.
- $\boxed{p=2}$ If $f \in \mathcal{S}(\mathbb{R}) \subset L^2(\mathbb{R})$, then Plancherel's Identity holds:
$$\|\widehat{f}\|_2 = \|f\|_2.$$

It turns out that we can identify Fourier transforms of L^1-functions with bounded functions (actually with the subset of continuous functions vanishing at infinity), and the Fourier transform is a bijection in $L^2(\mathbb{R})$.

EXERCISE 8.44. The conjugate Poisson kernel $Q_y(x)$ (see equation (7.23)) is in $L^2(\mathbb{R})$ for each $y > 0$, therefore it can be considered as a tempered distribution by integration against it, and we can compute its Fourier transform. Check that
$$\widehat{Q_y}(\xi) = -i\ \mathrm{sgn}(\xi)\, e^{-2\pi|y\xi|}.$$
Compare with the result of Exercise 8.53. $\Diamond$

The following example shows that even in $\mathcal{S}(\mathbb{R})$, if we hope for an inequality of the type
$$\|\widehat{f}\|_q \le C\|f\|_p, \qquad \text{for all } f \in \mathcal{S}(\mathbb{R}), \tag{8.9}$$
then p and q must be conjugate exponents: $\frac{1}{p} + \frac{1}{q} = 1$. Although it is not obvious from the example, it is also true that such an inequality will only hold for $1 \le p \le 2$.

EXAMPLE 8.45. Consider the following one-parameter family of functions in $\mathcal{S}(\mathbb{R})$, together with their Fourier transforms,
$$g_t(x) = e^{-\pi t x^2}, \qquad \widehat{g_t}(\xi) = \frac{1}{\sqrt{t}} e^{-\pi x^2/t}.$$
Recall that $\int_{\mathbb{R}} e^{-\pi x^2}\, dx = 1$, and so $\int_{\mathbb{R}} e^{-\pi A x^2}\, dx = 1/\sqrt{A}$. It follows that the L^p-norm of g_t and the L^q-norm of $\widehat{g_t}$ are given by
$$\|g_t\|_p^p = 1/\sqrt{tp}, \qquad \|\widehat{g_t}\|_q^q = (\sqrt{t})^{1-q}/\sqrt{q}.$$
According to Plancherel, when $p = q = 2$ the previous quantities should coincide, and indeed, they do. If we are hoping to obtain an inequality like (8.9), then we must be able to bound the ratios $\|\widehat{g_t}\|_q/\|g_t\|_p$ independently of t, and that is only possible when p and q are conjugate exponents. In fact,
$$\frac{\|\widehat{g_t}\|_q}{\|g_t\|_p} = (\sqrt{t})^{\frac{1}{p}+\frac{1}{q}-1} \frac{(\sqrt{p})^{1/p}}{(\sqrt{q})^{1/q}}.$$

This ratio is bounded independently of t if and only if the exponent of $\sqrt{t}$ is zero. $\Diamond$

Table 8.2 summarizes how the Fourier transform acts on the L^p-spaces.

TABLE 8.2. The effect of the Fourier transform on L^p-spaces.

ϕ	$\to$	$\widehat{\phi}$
$\mathcal{S}(\mathbb{R})$	unitary bijection	$\mathcal{S}(\mathbb{R})$
$L^1(\mathbb{R})$	bounded map $\|\widehat{f}\|_\infty \leq \|f\|_1$ (Riemann–Lebesgue Lemma)	$C_0(\mathbb{R}) \subset L^\infty(\mathbb{R})$
$L^p(\mathbb{R})$ $1 < p < 2$	bounded map $\|\widehat{f}\|_q \leq C_p\|f\|_p$ (Hausdorff–Young inequality)	$L^q(\mathbb{R})$ $\frac{1}{p} + \frac{1}{q} = 1$
$L^2(\mathbb{R})$	isometry $\|\widehat{f}\|_2 = \|f\|_2$ (Plancherel)	$L^2(\mathbb{R})$
$\mathcal{S}'(\mathbb{R})$	bijection	$\mathcal{S}'(\mathbb{R})$

The space of continuous functions that vanish at infinity is denoted by $C_0(\mathbb{R})$. It is a subset of $L^\infty(\mathbb{R})$, the space of *essentially bounded* functions on the line. The Riemann–Lebesgue Lemma is a strong statement. It says that the Fourier transform of an integrable function ($f \in L^1(\mathbb{R})$) is a *continuous* function, such that $\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0$. However, the Fourier transform is not a surjective map from $L^1(\mathbb{R})$ to $C_0(\mathbb{R})$; there is a function in $C_0(\mathbb{R})$ that is not the Fourier transform of any function in $L^1(\mathbb{R})$. See [**Kra**, p. 112, Prop 2.3.15].

LEMMA 8.46 (Riemann–Lebesgue Lemma). *Suppose $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous and*

$$\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0.$$

EXERCISE 8.47. Prove the Riemann–Lebesgue Lemma for functions f of moderate decrease (remember that functions of moderate decrease are continuous). More precisely, prove that if f is of moderate decrease, then $\widehat{f}$ is continuous, and $\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0$. In particular $\widehat{f}$ is bounded. (We verified the inequality $\|\widehat{f}\|_\infty \leq \|f\|_1$ for $f \in \mathcal{S}(\mathbb{R})$. The same proof goes through for functions of moderate decrease, and also for functions in $L^1(\mathbb{R})$, using the theory of Lebesgue integration.) $\Diamond$

Warning: the class of functions of moderate decrease is *not* closed under the Fourier transform. In fact, there is a familiar function of moderate decrease whose Fourier transform is not integrable [****careful moderate decrease implies continuous by definition, should be acareful with the next example]

EXERCISE 8.48. Verify that the Fourier transform of $\chi_{[-1/2,1/2]}$ is not integrable, but that it is in $C_0(\mathbb{R})$. $\Diamond$

The Fourier transform is well defined on $L^1(\mathbb{R})$, but since $\widehat{f}$ is not necessarily integrable, one has difficulties with the inversion formula. This is why we have chosen to present the theory on $\mathcal{S}(\mathbb{R})$. Since $\mathcal{S}(\mathbb{R})$ is dense in both $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ (see Appendix A), one can extend the Fourier transform to these spaces by continuity. In $L^2(\mathbb{R})$, Plancherel will be transferred by continuity as well. Are these extensions the same on $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$? The answer is yes, and they also coincide with the definition of the Fourier transform in those spaces in the sense of distributions. Notice that in $\mathbb{R}$, we do not have the nice ladder structure for L^p-spaces that we had on the circle; see Figure 2.4 in Section 2.1. In particular $L^2(\mathbb{R})$ is not a subset of $L^1(\mathbb{R})$, nor is $L^1(\mathbb{R})$ a subset of $L^2(\mathbb{R})$.

For $1 < p < 2$, the Fourier transform is a bounded map from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$, for p, q dual exponents $\frac{1}{p} + \frac{1}{q} = 1$. This is a consequence of the Hausdorff–Young inequality, which we will prove later, in Corollary 11.54. The Fourier transform does not map $L^p(\mathbb{R})$ for $p > 2$ into a nice functional space. In particular, for each $p > 2$, there exists a function $f \in L^p(\mathbb{R})$ whose Fourier transform $\widehat{f}$ is not locally integrable, but functions in L^q are always locally integrable, hence $\widehat{f}$ is not in L^q for any $1 \leq q \leq \infty$. Fourier transforms of generic L^p-functions for $p > 2$ are tempered distributions, but unless more information is given about the function we cannot say more. What is so special about $L^p(\mathbb{R})$, $1 < p < 2$? It turns out that in that range it is true that $L^p(\mathbb{R}) \subset L^1(\mathbb{R}) + L^2(\mathbb{R})$, and one can then use the theory in $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. We will come back to this issue when we discuss interpolation in Section 11.5. [*****reference for further reading Grafakos?]

EXERCISE 8.49. ($L^p(\mathbb{R})$ *and* $L^q(\mathbb{R})$ *are Not Nested*) Suppose $1 \leq p < q \leq \infty$. Find a function in $L^q(\mathbb{R})$ that is not in $L^p(\mathbb{R})$. Find a function that is in $L^p(\mathbb{R})$ but is not in $L^q(\mathbb{R})$. $\Diamond$

To finish we will present an example of a distribution that is given neither by a function in L^p, nor by a point mass (like the delta distribution), nor by a measure. It is the *principal value* distribution of $1/x$, the building block of the Hilbert transform, discussed in more detail in Chapter 11.

EXAMPLE 8.50. (*The Principal Value Distribution* $1/x$.) Define a linear functional on $\mathcal{S}(\mathbb{R})$ by,

$$H_0(\phi) = \lim_{\epsilon \to 0} \int_{|y|>\epsilon} \frac{\phi(y)}{y}\, dy. \tag{8.10}$$

To verify the continuity of this functional, we need to show that if $\phi_k \to 0$ in $\mathcal{S}(\mathbb{R})$ as $k \to \infty$, then the complex numbers $H_0(\phi_k) \to 0$ as $k \to \infty$. We need to take advantage of some "hidden cancellation". For each $\epsilon > 0$,

$$\int_{\epsilon \leq |y| < 1} \frac{1}{y}\, dy = 0.$$

We can add zero to (8.10),

$$\begin{aligned}
\left|\int_{|y|>\epsilon} \frac{\phi_k(y)}{y}\,dy\right| &= \left|\int_{|y|>1} \frac{\phi_k(y)}{y}\,dy + \int_{\epsilon\le|y|<1} \frac{\phi_k(y)-\phi_k(0)}{y}\,dy\right| \\
&\le \int_{|y|>1} \frac{|\phi_k(y)|}{|y|}\,dy + \int_{\epsilon\le|y|<1} \frac{|\phi_k(y)-\phi_k(0)|}{|y|}\,dy \\
&\le \|y\phi_k\|_\infty \int_{|y|>1} y^{-2}dy + \|\phi_k'\|_\infty \int_{1\ge|y|>\epsilon} 1\,dy \\
&\le 2\big(\rho_{1,0}(\phi_k) + \rho_{0,1}(\phi_k)\big)
\end{aligned}$$

To verify that the right hand side can be made arbitrarily small for k large enough, we use the fact that $\rho_{0,1}(\phi_k) \to 0$ and $\rho_{1,0}(\phi_k) \to 0$ as $k \to \infty$ since $\phi_k \to 0$ in $\mathcal{S}(\mathbb{R})$. Therefore H_0 is a tempered distribution, the *principal value distribution*, $H_0 = \text{p.v.}\frac{1}{y}$. For each $x \in \mathbb{R}$ define a new tempered distribution by translation, $\tau_x H_0(\phi) = H_0(\tau_{-x}\phi)$. We can now define the so-called *Hilbert transform* of $\phi \in \mathcal{S}(\mathbb{R})$:

$$\begin{aligned}
H\phi(x) &:= \frac{1}{\pi} H_0(\tau_x\tilde{\phi}) \\
&= \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|y|>\epsilon} \frac{\tilde{\phi}(y-x)}{y}\,dy \\
&= \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|y|>\epsilon} \frac{\phi(x-y)}{y}\,dy \\
&= \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|x-t|>\epsilon} \frac{\phi(t)}{x-t}\,dt.
\end{aligned}$$

$\lozenge$

Exercise 8.51. The functions $x^{-1}\chi_{\{|x|>\epsilon\}}$ are bounded and define tempered distributions for each $\epsilon > 0$, that we will denote H_ϵ. Check that the limit in the sense of distributions as $\epsilon \to 0$ of H_ϵ is H_0. This means that for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{\epsilon\to 0} H_\epsilon(\phi) = H_0(\phi).$$

(This is just a consequence of the definition of H_0.) $\lozenge$

Now we will compare the conjugate Poisson kernel to the principal value distribution $\frac{1}{x}$.

Exercise 8.52. Check that the limit as $y \to 0$ of the conjugate Poisson kernel $Q_y(x)$ in $\mathcal{S}'(\mathbb{R})$ is $\frac{1}{\pi}H_0$. This means that for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{y\to 0}\int_{\mathbb{R}} Q_y(x)\phi(x)\,dx = \frac{1}{\pi}H_0(\phi).$$

Hint: Notice that by Exercise 8.51 it will suffice to show that $(\pi Q_t - H_t) \to 0$ in the sense of distributions as $t \to 0$, that is, for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{t\to 0}\left(\int_{\mathbb{R}} Q_t(x)\phi(x)\,dx - \int_{|x|>t} \frac{\phi(x)}{x}\,dx\right) = 0.$$

$\lozenge$

EXERCISE 8.53. Using Exercises 8.52 and 8.44 calculate the Fourier transform of the tempered distribution $\frac{1}{\pi}H_0$. Use this calculation to justify the following formula for the Fourier transform of the Hilbert transform:

$$\widehat{H(\phi)}(\xi) = i\text{sgn}\xi\widehat{\phi}(\xi).$$

◊