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 7

The Fourier transform in paradise

The key idea of harmonic analysis is to express a function or signal as a superposition of simpler functions that are well understood. As we have seen, traditional Fourier series represent a periodic function as a sum of pure harmonics (sines and cosines):

$$f(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n) e^{2\pi i n x}. \tag{7.1}$$

A technical point here is that we are now using functions of period 1 instead of 2π; see Section 1.3.2. For aesthetic reasons we prefer to place the factor 2π in the exponent, where it will also be in the case of the Fourier transform on $\mathbb{R}$.

In the early 1800s, mathematics was revolutionized by Fourier's assertion that *every periodic function* could be expanded in such a series, where the coefficients (amplitudes) for each frequency n are calculated from f via the formula

$$\widehat{f}(n) := \int_0^1 f(x) e^{-2\pi i n x}\, dx, \qquad \text{for } n \in \mathbb{Z}. \tag{7.2}$$

It took almost 150 years to resolve exactly what this meant. In a remarkable paper [**Car**] that appeared in 1966, Lennart Carleson showed that for square-integrable functions on $[0, 1]$ the Fourier partial sums converge pointwise a.e. As a consequence the same holds for continuous functions (this was unknown until then).

We have already discussed these facts, and more, in Chapters 1–5. Among other things, in Chapter 5 we established the following fact.

> *The exponential functions* $\{e^{2\pi i n x}\}_{n\in\mathbb{Z}}$ *form an orthonormal basis for* $L^2([0, 1])$.

It follows immediately that equation (7.1) for the inverse Fourier transform holds in the L^2-sense.

We also discussed in Chapter 6 the discrete Fourier transform in the finite dimensional vector space $\mathbb{C}^N$. In that setting the trigonometric vectors $\{e_l\}_{l=0}^{N-1}$, with k^{th}-entries given by $e_l(k) = \frac{1}{\sqrt{N}} e^{2\pi k l/N}$, form an orthonormal basis for $\mathbb{C}^N$. Therefore the discrete inverse Fourier transform is given by

$$v(k) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} \widehat{v}(l)\, e^{2\pi k l/N},$$

and the l^{th} Fourier coefficient is given by the inner product in $\mathbb{C}^N$ of the vector $v \in \mathbb{C}^N$ and the vector e_l,

$$\widehat{v}(l) = \langle v, e_l\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k)\, e^{-2\pi k l/N}.$$

In the non-periodic setting, the *Fourier transform* $\widehat{f}$ of an integrable function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x}\, dx. \tag{7.3}$$

The *inverse Fourier transform* $(g)^{\vee}$ of an integrable function $g : \mathbb{R} \to \mathbb{R}$ is defined by

$$(g)^{\vee}(x) = \int_{\mathbb{R}} g(\xi) e^{2\pi i \xi x}\, d\xi. \tag{7.4}$$

This formulae make sense for integrable functions. There is much to be said about for which integrable functions, and in what sense, the following *Fourier inversion formula* holds:

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\xi) e^{2\pi i \xi x}\, d\xi,$$

or more concisely

$$(\widehat{f}\,)^{\vee} = f.$$

We discuss these issues and more in this chapter and the following one. We will see that the above formulas work for functions in the *Schwartz class*: infinitely often differentiable functions such that they and all their derivatives decay faster than any polynomial. In the Schwartz class the Fourier theory is perfect: we are in paradise. We will see in Chapter 8 that there is life beyond paradise, in the sense that we also have a very nice Fourier theory in larger classes of functions, such as $L^2(\mathbb{R})$ and even the class of generalized functions or *tempered distributions*.

7.1. From Fourier series to Fourier integrals

Heuristically, we could arrive at the integral formulae (7.3) and (7.4) by calculating the Fourier series on larger and larger intervals, until we cover the whole line, so to speak. This was Fourier's approach: first calculate the Fourier series of a "nice" function f on the symmetric interval $[-L/2, L/2)$ (see Section 1.3.2). For nice enough functions the L-Fourier series converges uniformly to the function f on $[-L/2, L/2)$:

$$f(x) = \sum_{n\in\mathbb{Z}} a_L(n) e^{2\pi i n x/L} \qquad \text{for all } x \in [-L/2, L/2), \tag{7.5}$$

where the L-Fourier coefficients are given by the formula

$$a_L(n) := \frac{1}{L} \int_{-L/2}^{L/2} f(x) e^{-2\pi i n x/L}\, dx.$$

Let

$$\xi_n = n/L,$$

so that

$$\Delta\xi := \xi_{n+1} - \xi_n = 1/L.$$

Thus we can now rewrite the Fourier series in equation (7.5) as

$$\sum_{n\in\mathbb{Z}} F_L(\xi_n)\Delta\xi. \tag{7.6}$$

where

$$F_L(\xi) := e^{2\pi i\xi x}\int_{-L/2}^{L/2} f(x)e^{-2\pi i\xi x}\,dx.$$

The expression (7.6) returns $f(x)$ for $x \in [-L/2, L/2]$ and resembles a Riemann sum for the improper integral $\int_{-\infty}^{\infty} F_L(\xi)\,d\xi$, except that the parameter L appears in the function to be integrated, as well as in the partition of the real line in steps of length $1/L$. Pretend now that the function f is zero outside the interval $[-M, M]$, for some $M > 0$ (i.e. the function f has *compact support*), then for L large enough so that $M < L/2$, the function F_L is now a function independent of L,

$$F_L(\xi) = e^{2\pi i\xi x}\int_R f(x)e^{-2\pi i\xi x}\,dx = e^{2\pi i\xi x}\,\widehat{f}\,(\xi),$$

where $\widehat{f}\,(\xi)$ is the Fourier transform of the function f defined on $\mathbb{R}$. In this case we could argüe that as $L \to \infty$, the sum in (7.6) should go to $\int_{-\infty}^{\infty} e^{2\pi i\xi x}\,\widehat{f}\,(\xi)\,d\xi$, and that this integral returns the value of $f(x)$ for all $x \in \mathbb{R}$.

We could perhaps accept that we need *all* frequencies to recover a function in $\mathbb{R}$, and that the following integrals will play the same role that equations (7.1) and (7.2) play in the Fourier series theory:

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}\,(\xi)e^{2\pi i\xi x}\,d\xi, \qquad \text{where} \qquad \widehat{f}\,(\xi) := \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}\,dx. \tag{7.7}$$

To make this argument rigorous, we have to specify what *nice* means, so that we have some type of convergence that entitles us to perform all these manipulations. We will not pursue this line of argument here, other than to motivate the appearance of the integral formulas.

7.2. The Fourier transform on the Schwartz class $\mathcal{S}(\mathbb{R})$

It turns out that formulae (7.7) hold when f belongs to the *Schwartz space* $\mathcal{S}(\mathbb{R})$, also known as the *Schwartz class*. First, $C^\infty(\mathbb{R})$ denotes the space of infinitely often differentiable functions $f : \mathbb{R} \to \mathbb{C}$. Next, the Schwartz space $\mathcal{S}(\mathbb{R})$ consists of those functions $f \in C^\infty(\mathbb{R})$ such that f and all its derivatives f', f'', ..., $f^{(\ell)}$, ... are *rapidly decreasing*, meaning that as $|x| \to \infty$ they decrease faster than any polynomial increases. Rephrasing, we obtain the following definition.

DEFINITION 7.1. The *Schwartz space* $\mathcal{S}(\mathbb{R})$ is defined by

$$\mathcal{S}(\mathbb{R}) := \Big\{ f \in C^\infty(\mathbb{R}) : \lim_{|x|\to\infty} |x|^k |f^{(\ell)}(x)| = 0 \text{ for all integers } k,\ \ell \geq 0 \Big\}. \tag{7.8}$$

◊

One can replace the limiting property with the statement that the products $|x|^k|f^{(\ell)}(x)|$ are bounded functions for all k and ℓ.

EXERCISE 7.2. Let $f \in C^\infty(\mathbb{R})$ verify that

$$\lim_{|x|\to\infty} |x|^k |f^{(\ell)}(x)| = 0 \text{ for all integers } k,\ \ell \geq 0$$

if and only if

$$\sup_{x\in\mathbb{R}} |x|^k |f^{(\ell)}(x)| < \infty \text{ for all integers } k,\ \ell \geq 0$$

. ◊

The Schwartz space is a vector space over the complex numbers. We invite you to verify that it is closed under the operations of multiplication by polynomials, multiplications by trigonometric functions, and differentiation.

EXERCISE 7.3. Show that if $f \in \mathcal{S}(\mathbb{R})$, then the products $x^k f(x) \in \mathcal{S}(\mathbb{R})$ for all $k \geq 0$. Show that the product $e^{-2\pi i x\xi} f(x) \in \mathcal{S}(\mathbb{R})$ for all $\xi \in \mathbb{R}$. Finally show that the derivatives $f^{(\ell)} \in \mathcal{S}(\mathbb{R})$ for all $\ell \geq 0$. ◊

Our first examples of Schwartz functions are the compactly supported functions in $C^\infty(\mathbb{R})$.

DEFINITION 7.4. A function f defined on $\mathbb{R}$ is *compactly supported* if there is a closed interval $[a,b] \subset \mathbb{R}$ such that $f(x) = 0$ for all $x \notin [a,b]$. We say that such a function f *has compact support*, and informally that f *lives on the interval* $[a,b]$. ◊

EXAMPLE 7.5. (*Compactly Supported* $C^\infty(\mathbb{R})$ *Functions are Schwartz*) If $f \in C^\infty(\mathbb{R})$ is compactly supported, then *a fortiori* f and all its derivatives decay faster as $|x| \to \infty$ than the reciprocals of all polynomials, and so $f \in \mathcal{S}(\mathbb{R})$. ◊

Compactly supported C^∞ functions are examples of Schwartz functions, but can we find a compactly supported C^∞ function that is not identically equal to zero?

EXAMPLE 7.6. (*A Bump Function*) Here is an example of a *compactly supported Schwartz function* that is supported on the interval $[a,b]$ and non-zero on the interval (a,b):

$$B(x) = \begin{cases} \exp\left(\dfrac{-1}{x-a}\right)\exp\left(\dfrac{-1}{b-x}\right), & \text{if } a < x < b; \\ 0, & \text{otherwise.} \end{cases}$$

◊

EXERCISE 7.7. Show that the bump function in Example 7.6 is in the Schwartz class. Notice that it suffices to verify that $B(x)$ is C^∞ at $x = a$ and at $x = b$. ◊

Here is an example of a function that is in the Schwartz class but is not compactly supported.

EXAMPLE 7.8. (*The Gaussian Function*) The canonical example of a function in $\mathcal{S}(\mathbb{R})$ is the *Gaussian function* $G(x)$, defined by

$$G(x) := e^{-\pi x^2}.$$

◊

We define the integral over $\mathbb{R}$ of a Schwartz function f to be the limit of the integrals of f over larger and larger intervals (notice that those integrals can be defined in the Riemann or in the Lebesgue sense, and they coincide):

$$\int_{\mathbb{R}} f(x)\,dx := \lim_{T\to\infty} \int_{-T}^{T} f(x)\,dx. \tag{7.9}$$

EXERCISE 7.9. Justify the existence of the limit of $\int_{-T}^{T} f(x)\,dx$ as $T \to \infty$, for functions $f \in \mathcal{S}(\mathbb{R})$. ◊

We then define the Fourier transform $\widehat{f}:\mathbb{R}\to\mathbb{C}$ of a Schwartz function f as follows:

$$\widehat{f}(\xi)=\int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\,dx := \lim_{T\to\infty}\int_{-T}^{T} f(x)e^{-2\pi i\xi x}\,dx. \tag{7.10}$$

Notice that the integrand is in $\mathcal{S}(\mathbb{R})$ (see Exercise 7.3).

The Gaussian, introduced in Example 7.8, plays a very important role in Fourier analysis, probability theory and physics. It has the unusual property of being equal to its own Fourier transform:

$$\widehat{G}(\xi)=e^{-\pi\xi^2}.$$

One can prove this fact using the observation that both the Gaussian and its Fourier transform satisfy the ordinary differential equation $f'(x)=-2\pi x f(x)$, with initial condition $f(0)=1$. Therefore, by uniqueness of the solution, they must be the same function.

Exercise 7.10. (*The Gaussian is its own Fourier transform*) Convince yourself that the Gaussian belongs to $\mathcal{S}(\mathbb{R})$. Find the Fourier transform of the Gaussian, either by filling in the details in the previous paragraph, or by a direct calculation. $\Diamond$

Convolution is a very important operation in harmonic analysis. In the context of Fourier series, we have already seen how to convolve two periodic functions on $\mathbb{T}$. Periodic convolution was extensively discussed in Section 4.3, and used informally before that when convolving with the Dirichlet kernel. Section 4.3.2 described the smoothing properties of convolution. In Chapter 6 we touched on circular convolution for vectors. On the line, the convolution of two functions f, $g\in\mathcal{S}(\mathbb{R})$ is defined as follows:

$$f*g(x):=\int_{\mathbb{R}} f(x-y)g(y)\,dy. \tag{7.11}$$

This integral is well defined for each $x\in\mathbb{R}$ and for f, $g\in\mathcal{S}(\mathbb{R})$. As we will see, the Schwartz class is closed under convolutions. In Chapter 8 we discuss convolution of a Schwartz function with a distribution, and in Chapter 11 convolution of an L^p function with an L^q function.

The Fourier transform interacts very nicely with a number of operations. In particular, *differentiation is transformed into polynomial multiplication* and vice versa, which sheds light on the immense success of Fourier transform techniques in the study of differential equations (certainly in the linear case). Also, *convolutions are transformed into products* and vice versa, which explains the success of Fourier transform techniques in signal processing, since *filtering* (= convolution) is one of the most important signal-processing tools. In Section 7.3, we give a time–frequency dictionary that lists all these useful interactions.

We show in Section 7.4 that as a consequence of the time–frequency dictionary, the Schwartz class is closed under the Fourier transform. Furthermore we show in Section 7.7 that the Fourier inversion formula holds in $\mathcal{S}(\mathbb{R})$, and as a consequence, the Fourier transform is a bijection from $\mathcal{S}(\mathbb{R})$ into $\mathcal{S}(\mathbb{R})$. Moreover, the Fourier transform on $\mathcal{S}(\mathbb{R})$ preserves energy or L^2-norms (Plancherel).

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

The Fourier transform is a *linear transformation*, meaning that the Fourier transform of a linear combination of Schwartz functions f and g is equal to the same linear combination of the Fourier transforms of f and g:

$$[af + b\,g]^{\wedge} = a\widehat{f} + b\,\widehat{g}.$$

We have already mentioned that the Fourier transform converts differentiation into polynomial multiplication, and convolution into multiplication. It also interacts very nicely with *translations*, *modulations*, and *dilations*. In Table 7.1 we list ten of these extremely useful fundamental properties of the Fourier transform.

We refer to the list in Table 7.1 as the *time–frequency dictionary*. Here the word *time* is used because the variable x in $f(x)$ often stands for time, for instance when $f(x)$ represents a voice signal taking place over some interval of time, as in Example 1.1. Similarly, the word *frequency* refers to the variable ξ in $\widehat{f}(\xi)$. Operations on the function $f(x)$ are often said to be happening *in the time domain* or *on the time side*, while operations on the Fourier transform $\widehat{f}(\xi)$ are said to be *in the frequency domain* or *on the frequency side*, or, less precisely, *on the Fourier side*.

As in a dictionary used for converting between two languages, an entry in the right-hand column of the time–frequency dictionary gives the equivalent on the frequency side of the corresponding entry, expressed on the time side, in the left-hand column.

For example, the expressions $e^{2\pi ihx}f(x)$ and $\widehat{f}(\xi - h)$ convey in two different ways the same idea, that of shifting by the amount h all the frequencies present in the signal $f(x)$. To give some terminology, *modulation* of a function $f(x)$ means multiplying the function by a term of the form $e^{2\pi ihx}$ for some real number h, as shown in the term $e^{2\pi ihx}f(x)$ on the left-hand side of property (c) in Table 7.1. As usual, horizontal *translation* of a function means adding a constant, say $-h$, to its argument, as shown in the term $\widehat{f}(\xi - h)$ on the right-hand side of property (c).

Thus, row (c) in the table should be read as saying that

the Fourier transform of $e^{2\pi ihx}f(x)$ *is equal to* $\widehat{f}(\xi - h)$.

In other words,

modulation by h *on the time side is transformed into translation by* h *on the frequency side.*

Notice that if $f, g \in \mathcal{S}(\mathbb{R})$, then all the functions in the time column of the table belong to the Schwartz class as well. In other words, the Schwartz class is closed not only under products, multiplication by polynomials and trigonometric functions (modulations), differentiation, and convolution, but also under simpler operations such as linear operations, translations, dilations, and conjugations. Hence we are entitled to compute the Fourier transforms of the functions in the time column of the table.

Property (a) says that the Fourier transform is linear. Properties (b)–(e), known as *symmetry* properties or *group invariance* properties, explain how the Fourier transform interacts with the symmetries (translation, dilation, reflection) of the domain of the function f. For us this domain is either $\mathbb{R}$ or $\mathbb{T}$; both are groups. Properties (f)–(j) explain how the Fourier transform interacts with conjugation, and differentiation of functions f, and with multiplication of functions f by polynomials, as well as products and convolution of functions f with other functions in the Schwartz class.

TABLE 7.1. The time–frequency dictionary in $\mathcal{S}(\mathbb{R})$.

	Time	**Frequency**
(a)	linear properties $af + bg$	linear properties $a\widehat{f} + b\widehat{g}$
(b)	translation $\tau_h f(x) := f(x-h)$	modulation $\widehat{\tau_h f}(\xi) = e^{-2\pi i h\xi}\widehat{f}(\xi) = M_{-h}\widehat{f}(\xi)$
(c)	modulation $M_h f(x) := e^{2\pi i h x} f(x)$	translation $\widehat{M_h f}(\xi) = \widehat{f}(\xi - h) = \tau_h \widehat{f}(\xi)$
(d)	dilation $D_s f(x) := s f(sx)$	inverse dilation $\widehat{D_s f}(\xi) = \widehat{f}(s^{-1}\xi) = s D_{s^{-1}} \widehat{f}(\xi)$
(e)	reflection $\widetilde{f}(x) := f(-x)$	reflection $\widehat{\widetilde{f}}(\xi) = -\widetilde{\widehat{f}}(\xi)$
(f)	conjugation $\overline{f}(x) := \overline{f(x)}$	conjugate reflection $\widehat{\overline{f}}(\xi) = \overline{\widehat{f}(-\xi)} = \overline{\widetilde{\widehat{f}}}(\xi)$
(g)	differentiation $f'(x)$	multiplication by polynomial $\widehat{f'}(\xi) = 2\pi i \xi \widehat{f}(\xi)$
(h)	multiplication by polynomial $-2\pi i x f(x)$	differentiation $[-2\pi i x f(x)]^\wedge(\xi) = \frac{d}{d\xi}\widehat{f}(\xi)$
(i)	convolution $f * g(x) := \int f(x-y) g(y)\, dy$	product $\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi)$
(j)	product $f(x)g(x)$	convolution $\widehat{fg}(\xi) = \widehat{f} * \widehat{g}(\xi)$

We leave it as an exercise to prove most of the properties listed in the time–frequency dictionary. We have already proved most of these properties in the Fourier-series setting, and the proofs in the present setting are almost identical.

EXERCISE 7.11. Verify properties (a)–(g) of the time–frequency dictionary. $\Diamond$

Let us prove property (h), which gives the connection between taking a derivative on the Fourier side and multiplying f by an appropriate polynomial on the time side. Namely, multiplying a function f by the polynomial $-2\pi i x$ yields a function whose Fourier transform is the derivative with respect to ξ of the Fourier transform $\widehat{f}$ of f. Property (h) does not have an immediate analogue in the Fourier-series setting, although we did prove in Chapter 3 the analogue of its companion property (g), about the Fourier coefficients of the derivative of a given periodic and continuously differentiable function. Convolution in $\mathcal{S}(\mathbb{R})$ as well as properties (i) and (j) will be discussed in Section 7.5.

Proof of property (h). Differentiating (formally) under the integral sign leads us immediately to the right formula. It only remains to justify the interchange of the limit and the integral.

We will check that $\widehat{f}$ is differentiable (as a bonus we get continuity of f), and that its derivative coincides with the Fourier transform of $-2\pi ixf(x)$, both in one stroke. By the definition derivative we need to show that,

$$\lim_{h\to 0} \frac{\widehat{f}(\xi+h)-\widehat{f}(\xi)}{h} = [-2\pi ixf(x)]^\wedge(\xi).$$

This is equivalent to showing that the limit of the difference is zero,

$$\lim_{h\to 0} \left[\frac{\widehat{f}(\xi+h)-\widehat{f}(\xi)}{h} - [-2\pi ixf(x)]^\wedge(\xi)\right] = 0.$$

Consider the difference

$$\frac{\widehat{f}(\xi+h)-\widehat{f}(\xi)}{h} - [-2\pi ixf(x)]^\wedge(\xi)$$

$$= \int_{\mathbb{R}} f(x)e^{-2\pi ix\xi}\left(\frac{e^{-2\pi ixh}-1}{h}+2\pi ix\right)dx. \tag{7.12}$$

It suffices to check that for each $\varepsilon > 0$ we can make the absolute value of the integral on the right-hand side of equation (7.12) smaller than a constant multiple of ε, provided h is small enough, say for $|h| < h_0$. We bound separately the contributions for large x and for small x. For all $N > 0$,

$$|\text{RHS of (7.12)}| \le \underbrace{\int_{|x|\le N} |f(x)|\left|\frac{e^{-2\pi ixh}-1}{h}+2\pi ix\right| dx}_{(A)}$$

$$+ \underbrace{2\pi\int_{|x|>N} |x||f(x)|\left|\frac{e^{-2\pi ixh}-1}{2\pi xh}+i\right| dx}_{(B)}.$$

Since $f \in \mathcal{S}(\mathbb{R})$, f is bounded, say by $M > 0$, and f is rapidly decreasing. In particular, there is some $C > 0$ such that $|x|^3|f(x)| \le C$ for all $|x| > 1$, and so

$$|xf(x)| \le Cx^{-2} \quad \text{for all } |x| > 1.$$

Since $\int_{|x|>1} x^{-2}\,dx < \infty$, the tails of this integral must be going to zero, meaning that $\lim_{N\to\infty}\int_{|x|>N} x^{-2}\,dx = 0$. Therefore, given $\varepsilon > 0$, there is an $N > 0$ such that

$$\int_{|x|\ge N} |xf(x)|\,dx \le \int_{|x|>N} \frac{C}{x^2}\,dx \le \varepsilon.$$

Moreover, for $|x| \le N$, there is some h_0 such that for all $|h| < h_0$,

$$\left|\frac{e^{-2\pi ixh}-1}{h}+2\pi ix\right| \le \frac{\varepsilon}{N},$$

because $\frac{d}{dy}e^{-2\pi ixy}\Big|_{y=0} = -2\pi ix$, and $e^{-2\pi ixy}\big|_{y=0} = 1$, and on the compact interval $[-N,N]$, the continuous function $\frac{d}{dy}e^{-2\pi ixy}$ is necessarily uniformly continuous, and this ensures that the sequence indexed in the parameter h, $F_h(x) = \frac{e^{-2\pi ixh}-1}{h}+2\pi ix$ converges uniformly to zero as $h \to 0$ for $|x| \le N$. These imply that $(A) \le 2M\varepsilon$.

Also remember that for all real θ, $|e^{i\theta} - 1| \leq |\theta|$, and so $\left|\frac{e^{-i\theta}-1}{\theta} + i\right| \leq 2$. All these facts together imply that $(B) \leq 4\pi\varepsilon$. Hence for each $\varepsilon > 0$ there exists $h_0 > 0$ such that for all $|h| < h_0$

$$\left|\frac{\widehat{f}(\xi + h) - \widehat{f}(\xi)}{h} - [-2\pi i x f(x)]^\wedge(\xi)\right| \leq (2M + 4\pi)\varepsilon.$$

Therefore, by the definition of the limit as $h \to 0$,

$$\lim_{h\to 0} \frac{\widehat{f}(\xi + h) - \widehat{f}(\xi)}{h} = [-2\pi i x f(x)]^\wedge(\xi).$$

We conclude that $\frac{d}{d\xi}\widehat{f}(\xi) = [-2\pi i x f(x)]^\wedge(\xi)$, which establishes property (h). □

Notice that in the proof of property (h) we only used decay properties of f. More precisely we used the fact that $\int |x||f(x)|\, dx < \infty$, which is the minimum necessary to ensure that we can calculate the Fourier transform of $-2\pi i x f(x)$.

Exercise 7.12. Use an induction argument to show that if

$$\int_{\mathbb{R}} (1 + |x|)^k |f(x)|\, dx < \infty,$$

then $\widehat{f}$ is k times differentiable. Moreover, in that case

$$\frac{d^k}{d\xi^k}\widehat{f}(\xi) = [(-2\pi i x)^k f(x)]^\wedge(\xi).$$

Hint: Do the case $k = 1$ first, and then use an induction argument to get all other ks. For the case $k = 1$, assume that $\int_{\mathbb{R}} (1 + |x|)|f(x)|\, dx < \infty$, in particular $\int_{\mathbb{R}} |f(x)|\, dx < \infty$ and $\int_{\mathbb{R}} |x||f(x)|\, dx < \infty$. The proof is very similar to the proof of property (h), there are two places where properties of f were used: first to estimate term (A) boundedness of f was used to show that $\int_{|x|\leq N} |f(x)|\, dx < M$, and second to estimate term (B) the integrability of $|x|\,|f(x)|$ was used to ensure that the tail of the integral (for $|x| > N$) could be made small enough. ♢

7.4. The Schwartz class is closed under the Fourier transform

In this section we verify that the Fourier transform of a Schwartz function f is a Schwartz function. We must check that $\widehat{f}$ is infinitely differentiable, and that $\widehat{f}$ and all its derivatives are rapidly decaying.

Exercise 7.12 shows that $f \in \mathcal{S}(\mathbb{R})$ implies that $\widehat{f} \in C^\infty$, and gives a formula for the k^{th} derivative of $\widehat{f}$.

As a second step toward the proof that the Schwartz class is closed under the Fourier transform, we prove a version of the Riemann–Lebesgue Lemma in $\mathcal{S}(\mathbb{R})$, showing that the Fourier transform $\widehat{f}$ of a Schwartz function tends to zero as $|\xi| \to \infty$. Of course, this result must hold if $\widehat{f}$ is to be in the Schwartz class $\mathcal{S}(\mathbb{R})$.

Lemma 7.13 (Riemann–Lebesgue Lemma in $\mathcal{S}(\mathbb{R})$). *If $f \in \mathcal{S}(\mathbb{R})$, then*

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

PROOF. Using a symmetrization trick that we have already encountered when verifying the Riemann–Lebesgue Lemma for periodic functions, we see that

$$\begin{aligned}\widehat{f}(\xi) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}\,dx \\ &= -\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}e^{\pi i}\,dx \\ &= -\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi\left(x-\frac{1}{2\xi}\right)}\,dx \\ &= -\int_{-\infty}^{\infty} f\left(y+\frac{1}{2\xi}\right)e^{-2\pi i\xi y}\,dy.\end{aligned}$$

The last equality uses the change of variable $y = x - \frac{1}{2\xi}$. It follows that

$$\widehat{f}(\xi) = \frac{1}{2}\int_{-\infty}^{\infty}\left[f(x) - f\left(x+\frac{1}{2\xi}\right)\right]e^{-2\pi i\xi x}\,dx.$$

Hence

$$|\widehat{f}(\xi)| \le \frac{1}{2}\int_{-\infty}^{\infty}\left|f(x) - f\left(x+\frac{1}{2\xi}\right)\right|\,dx.$$

Because $f \in \mathcal{S}(\mathbb{R})$, we know there exists $M > 0$ such that $|f(x)| \le M/x^2$. Also, given $\epsilon > 0$, there exists $N > 0$ large enough such that for $|\xi| > 1$

$$\int_{|x|>N}\left|f(x) - f\left(x+\frac{1}{2\xi}\right)\right|\,dx \le \int_{|x|>N}\left[\frac{M}{x^2} + \frac{M}{(x+\frac{1}{2\xi})^2}\right]dx \le \epsilon.$$

As for the integral over $|x| \le N$, since $g_\xi(x) = f(x) - f\left(x+\frac{1}{2\xi}\right)$ converges uniformly to zero as $|\xi| \to \infty$ on $|x| \le N$ (because f is uniformly continuous on $[-N, N]$), we can make it as small as we wish. More precisely, given $\varepsilon > 0$ there exists $K > 0$ such that for all $|\xi| > K$, and for all $|x| \le N$, $|g_\xi(x)| < \frac{\varepsilon}{4N}$. Therefore

$$|\widehat{f}(\xi)| \le \frac{\varepsilon}{2} + \int_{|x|\le N}|g_\xi(x)|\,dx \le \frac{\varepsilon}{2} + 2N\frac{\varepsilon}{4N} = \epsilon.$$

We conclude that given $\varepsilon > 0$, there exists $K > 0$ such that for all $|\xi| > K$, $|\widehat{f}(\xi)| \le \varepsilon$, that is $\lim_{|\xi|\to\infty}|\widehat{f}(\xi)| = 0$, as required. □

In fact this argument works, essentially unchanged, for functions f that are only assumed to be continuous and integrable. Integrability takes care of the integral for $|x| > N$, while continuity controls the integral for $|x| \le N$. A density argument shows that if $f \in L^1(\mathbb{R})$ then the same result holds. More is actually true. If $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous (we showed under the condition that $f \in \mathcal{S}(\mathbb{R})$, or just that $\int |x||f(x)| < \infty$, that $\widehat{f}$ was actually differentiable, and hence continuous). The statement of the full *Riemann–Lebesgue Lemma* is as follows.

LEMMA 7.14 (Riemann–Lebesgue Lemma). *If $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous and $\lim_{|\xi|\to\infty}|\widehat{f}(\xi)| = 0$. That is, $\widehat{f} \in C_0(\mathbb{R})$.*

EXERCISE 7.15. Show that if $F \in L^1(\mathbb{R})$ then $\widehat{f}(\xi)$ is a continuous function of ξ.

The final ingredient needed to prove that the Fourier transform takes $\mathcal{S}(\mathbb{R})$ to itself (Theorem 7.17) is that the Fourier transforms of Schwartz functions decrease rapidly.

LEMMA 7.16. *If $f \in \mathcal{S}(\mathbb{R})$ then $\widehat{f}$ is rapidly decreasing.*

PROOF. To prove the Lemma, we must check that for $f \in \mathcal{S}(\mathbb{R})$,

$$\sup_{\xi\in\mathbb{R}} |\xi|^n |\widehat{f}(\xi)| < \infty \qquad \text{for all integers } n \geq 0. \tag{7.13}$$

We already know that $\widehat{f}$ is continuous, and in fact that $\widehat{f}$ is C^∞. Moreover, the Riemann–Lebesgue Lemma (Lemma 7.13) shows that $\widehat{f}$ vanishes at infinity. We must show that as $|\xi| \to \infty$, $\widehat{f}$ vanishes faster than the reciprocals of all polynomials. We proved the analogue of property (g) in the Fourier series case, and (Exercise 7.11) that property (g) itself holds: $\widehat{f'}(\xi) = 2\pi i\xi\widehat{f}(\xi)$. Iterating, we obtain the formula

$$\widehat{f^{(n)}}(\xi) = (2\pi i\xi)^n \widehat{f}(\xi) \qquad \text{for } f \in \mathcal{S}(\mathbb{R}),\ n \geq 0. \tag{7.14}$$

Therefore, to verify that $\widehat{f}$ is rapidly decreasing it suffices to check that for each $n \geq 0$,

$$\sup_{\xi\in\mathbb{R}} |\widehat{f^{(n)}}(\xi)| < \infty. \tag{7.15}$$

Since $f \in \mathcal{S}(\mathbb{R})$, $f^{(n)}$ is also a Schwartz function. Applying the Riemann–Lebesgue Lemma to $f^{(n)}$, we conclude that for all $n \geq 0$,

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

Therefore

$$\lim_{|\xi|\to\infty} (2\pi i\xi)^n \widehat{f}(\xi) = 0. \tag{7.16}$$

Since we have already shown that $\widehat{f}$ is a C^∞ function, and in particular that $\widehat{f}$ is continuous and hence necessarily bounded on every closed interval, the decay of $\widehat{f}$ at infinity shown in equation (7.16) implies inequality (7.13) for all $n \geq 0$. Hence $\widehat{f}$ is rapidly decreasing, as required. □

After much preparation, we can now prove that the Fourier transform of any Schwartz function is also a Schwartz function (Theorem 7.17).

THEOREM 7.17 (The Schwartz Space is Closed Under the Fourier Transform). *If $f \in \mathcal{S}(\mathbb{R})$ then $\widehat{f} \in \mathcal{S}(\mathbb{R})$.*

PROOF. Suppose $f \in \mathcal{S}(\mathbb{R})$. We want to show that $\widehat{f} \in \mathcal{S}(\mathbb{R})$. We have already shown that $\widehat{f} \in C^\infty(\mathbb{R})$, and that

$$\frac{d^\ell}{d\xi^\ell}\widehat{f}(\xi) = [(-2\pi i x)^\ell f(x)]^\wedge(\xi).$$

In other words, the ℓ^{th} derivative of $\widehat{f}$ is the Fourier transform of the Schwartz function $g_\ell(x) = (-2\pi i x)^\ell f(x)$. It remains to show that $\widehat{f}$ and all its derivatives are rapidly decreasing. Applying Lemma 7.16 to each $g_\ell \in \mathcal{S}(\mathbb{R})$, we find that $\frac{d^\ell}{d\xi^\ell}\widehat{f} = \widehat{g_\ell}$ is rapidly decreasing for each $\ell \geq 0$. We conclude that $\widehat{f} \in \mathcal{S}(\mathbb{R})$, as required. □

Equations (7.14) and (7.16) illustrate an important principle in Fourier analysis that we have already discussed in the context of Fourier series in Chapter 3.

> *Smoothness of a function is correlated with fast decay of its Fourier transform at infinity.*

In other words, the more times a function f can be differentiated, the faster its Fourier transform $\widehat{f}(\xi)$ goes to zero as $\xi \to \pm\infty$. This principle does not seem all that interesting when we are in paradise. After all, functions in the Schwartz class and their Fourier transforms are infinitely differentiable and rapidly decaying.

A *band-limited function* is one whose Fourier transform is compactly supported. Thus the Fourier transform of a band-limited function shows the most dramatic form possible of decay at infinity. In accordance with the previous principle, it is not surprising that band-limited functions turn out to be analytic, which is an even stronger property than being infinitely often differentiable. Likewise, smoothness of the Fourier transform is correlated with fast decay of the function transformed at infinity. If the function has compact support, then its Fourier transform will be analytic. Not all analytic functions can be reached since they must satisfy some at most exponential growth conditions (the code word *analytic functions of exponential type*).

Much is known about the decay properties of a function at infinity, and the corresponding analyticity properties of its Fourier transform. Such results go under the generic name of Paley–Wiener theorems. To learn more about them, see for example [**Str2**, Section 7.2] and [**RS**, Chapter IX.3].

7.5. Convolution and approximations of the identity on $\mathbb{R}$

The convolution of two functions f, $g \in \mathcal{S}(\mathbb{R})$ is defined as follows:

$$f * g(x) := \int_{\mathbb{R}} f(x-y)g(y)\,dy. \tag{7.17}$$

This integral is well defined for each $x \in \mathbb{R}$ and for f, $g \in \mathcal{S}(\mathbb{R})$. Notice that by change of variables, $f * g = g * f$: convolution is commutative.

The new function $f * g$ defined by the integral in equation (7.17) is rapidly decreasing, because of the smoothness and fast decay of the functions f and g. It turns out that the derivatives of $f * g$ are rapidly decreasing as well, and so the convolution of two Schwartz functions is again a Schwartz function. The key observation is that

$$\frac{d}{dx}\big(f * g\big)(x) = f' * g(x) = f * g'(x), \qquad \text{for } f,\ g \in \mathcal{S}(\mathbb{R}).$$

The first identity above is true because in this setting, we are entitled to interchange differentiation and integration. The formal justification of this fact involves breaking the integral into two pieces, one where the variable of integration is bounded ($|y| \leq M$), the other where is not ($|y| > M$): on the bounded part Theorem 2.49 about interchange of limits and integration can be used, on the unbounded part the decay of the function can be used to guarantee that the integral goes to zero as $M \to \infty$. This is exactly the argument we used when proving Property (h) in the time–frequency dictionary.

The second identity holds because convolution is commutative; it can also be verified by integration by parts.

EXERCISE 7.18. (*The Schwartz Space Is Closed Under Convolution*) Verify that if f, $g \in \mathcal{S}(\mathbb{R})$, then $f * g \in \mathcal{S}(\mathbb{R})$. $\Diamond$

Convolution is a *smoothing* operation. The output keeps the best of each input. This heuristic is not completely clear when thinking about convolution of Schwartz functions, because we are already in paradise. One can convolve much less regular functions (e.g. just an integrable function, or even a generalized function or distribution) with a smooth function; then the smoothness will be inherited by the convolution. In fact when convolving Schwartz functions with distributions (as we will do in Chapter 8), we will see that the resulting distribution can be identified with a C^∞ function.

Convolutions correspond to *translation-invariant bounded linear transformations*, and to *Fourier multipliers*. These, and their generalizations, have been and are the object of intense study, we discussed the discrete analogue of this statement in Chapter 6.

EXERCISE 7.19. Verify properties (i) and (j) of the time–frequency dictionary: if f, $g \in \mathcal{S}(\mathbb{R})$ then

$$\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi), \qquad \widehat{fg}(\xi) = \widehat{f} * \widehat{g}(\xi).$$

$\Diamond$

Convolution can be viewed as a binary operation on the Schwartz class $\mathcal{S}(\mathbb{R})$. It is a commutative operation. However, the Schwartz class with convolution does not form a group, since there is no *identity element* in $\mathcal{S}(\mathbb{R})$. This is very simple to verify once we have established property (i) in the time–frequency dictionary. Suppose there exists $e \in \mathcal{S}(\mathbb{R})$ such that for all $f \in \mathcal{S}(\mathbb{R})$ we have that $f * e = f$, Fourier transform on both sides and obtain that $\widehat{f}(\xi)\widehat{e}(\xi) = \widehat{f}(\xi)$. In particular you can set $f(x) = G(x) = e^{-\pi x^2}$, the Gaussian, which is its own Fourier transform and never vanishes. Hence we can cancel G from the identity $G(\xi)\widehat{e}(\xi) = G(\xi)$, and conclude that $\widehat{e} = 1$. But we have just shown that if $e \in \mathcal{S}(\mathbb{R})$, then $\widehat{e} \in \mathcal{S}(\mathbb{R})$, therefore it must decrease rapidly, but the function identically equal to one does not decrease. We have reached a contradiction. There cannot be an identity in $\mathcal{S}(\mathbb{R})$ for the convolution product.

There are two ways to address this lack of an identity element. The first is to introduce the *delta function*, to be defined in Section 8.4 below. The delta function can be thought of as a point-mass since, roughly speaking, it takes the value zero except at a single point. The delta function is not a Schwartz function, and in fact is not a true function at all, but rather a generalized kind of function called a *distribution*. It acts as an identity under convolution (see Exercise 8.24; one must extend the definition of convolution appropriately).

The second substitute for an identity element in $\mathcal{S}(\mathbb{R})$ under convolution builds on the first. The idea is to use a parametrized *sequence* or *family* of related functions, that tends to the delta distribution as the parameter approaches some limiting value. Such families are called *approximations of the identity*. We discussed them in the context of Fourier series; see Section 4.4 and especially Definition 4.17. The functions in the family may actually be Schwartz functions, or they may be less smooth.

DEFINITION 7.20. An *approximation of the identity* in $\mathbb{R}$ is a family $\{K_t\}_{t\in\Lambda}$ of real-valued functions on $\mathbb{R}$, where Λ is an index set, together with a point t_0 that is an accumulation[1] point of Λ, with the following three properties.

(i) The functions K_t have mean value one:
$$\int_{\mathbb{R}} K_t(x)\,dx = 1 \quad \text{for all } t \in \Lambda.$$

(ii) The functions K_t are uniformly integrable in t: there is a constant $C > 0$ such that
$$\int_{\mathbb{R}} |K_t(x)|\,dx \le C \quad \text{for all } t \in \Lambda.$$

(iii) Stated informally, the mass[2] of K_t becomes increasingly concentrated at $x = 0$ as $t \to t_0$:
$$\lim_{t\to t_0} \int_{|x|>\delta} |K_t(x)|\,dx = 0 \quad \text{for each } \delta > 0.$$

$\Diamond$

In the examples it will become clear what the accumulation point t_0 should be. Sometimes $t_0 = 0 < t$ and then $t \to 0^+$. Or one can have $t_0 = \infty$ and then $t \to \infty$, or $t_0 = 1$ and $t \to 1$. In the periodic case the kernels were indexed by $N \in \mathbb{N}$ (Fejér kernel), or by $0 < r < 1$ (Poisson kernel), and we considered $N \to \infty$ and $r \to 1^-$ respectively.

Notice that in the definition there is no explicit mention of the delta function. Nevertheless the three properties in the definition do imply that as $t \to t_0$, $K_t(x)$ converges in some sense to the delta function, that is, to the identity under convolution; see Exercise 8.25. This is the origin of the term *approximation of the identity*, or in a language we already met, because of the property $f * K_t \to f$ as $t \to t_0$, which is another aspect of the same idea, see Theorem 7.23.

EXERCISE 7.21. (*Generating an Approximation of the Identity from a Kernel*) An easy way to produce an approximation of the identity is to start with a non-negative function (or 'kernel') $K(x) \ge 0$ with mean value one, $\int_{\mathbb{R}} K(x)\,dx = 1$, and define the family of dilations

$$K_t(x) := t^{-1}K(t^{-1}x), \qquad \text{for all } t > 0. \tag{7.18}$$

Verify that $\{K_t\}$ is an approximation of the identity, with $t_0 = 0$. $\Diamond$

EXAMPLE 7.22. (*Gaussian Kernels Form an Approximation of the Identity*) Gaussians are good kernels, and by scaling them we obtain an approximation of the identity. More precisely, given $G(x) = e^{-\pi x^2}$, the *Gauss kernels* $G_t(x)$ are defined by
$$G_t(x) := t^{-1}e^{-\pi(t^{-1}x)^2}.$$
Since $e^{-\pi x^2} > 0$ and $\int_{\mathbb{R}} e^{-\pi x^2}\,dx = 1$, this family is an approximation of the identity with $t_0 = 0$. As $t \to 0$, the graphs of $G_t(x)$ look more and more like the delta function; see Figure 7.1. $\Diamond$

Another facet of the approximation-of-the-identity idea is that by convolving against an approximation of the identity, one can approximate a function very well.

[1] That is we can find points $t \neq t_0$ from the index set Λ that are arbitrarily close to t_0.

[2] The mass is the integral of the absolute value of K_t.

FIGURE 7.1. Graphs of Gauss kernels G_t for $t = 1$, 3, and 7.5.

One instance of this heuristic is given in the next theorem, whose analogue on $\mathbb{T}$ we proved in Theorem 4.21.

THEOREM 7.23. *Let $\{K_t\}_{t\in\Lambda}$ be an approximation of the identity as $t \to t_0$. Let $K_t, f \in \mathcal{S}(\mathbb{R})$. Then the convolutions $f * K_t$ converge uniformly to f as $t \to t_0$.*

See also Theorem 7.40 and its proof, for convergence in L^p.

EXERCISE 7.24. Prove Theorem 7.23. Verify that Theorem 7.23 remains true if f is only assumed to be continuous and integrable, so that the convolution is still well defined. $\lozenge$

Note that because of the smoothing effect of convolution, the approximating functions $f * K_t$ may be much smoother than the original function f. For instance, this idea gives examples illustrating that the uniform limit of differentiable functions need not be differentiable.

7.6. An interlude on kernels

When we studied summability methods in Sections 4.5 and 4.6, we used the Abel and Cesàro means to regularize the convergence of Fourier series, by convolving with the Fejér and Poisson kernels on the circle $\mathbb{T}$. Those kernels happened to be approximations of the identity on $\mathbb{T}$. We now define analogous Fejér and Poisson kernels on the real line $\mathbb{R}$; they can be used to regularize the convergence of the Fourier integral. In this subsection we also define the heat kernel on $\mathbb{R}$, the Dirichlet kernel on $\mathbb{R}$, and the conjugate Poisson kernel on $\mathbb{R}$.

The Fejér kernel on $\mathbb{R}$:

$$F_R(x) := R\left(\frac{\sin(\pi R x)}{R\pi x}\right)^2, \qquad \text{for } R > 0. \tag{7.19}$$

The Poisson kernel on $\mathbb{R}$:

$$P_y(x) := \frac{y}{\pi(x^2+y^2)}, \qquad \text{for } y > 0. \tag{7.20}$$

EXERCISE 7.25. Show that the Poisson kernel is a solution of *Laplace's equation*,

$$\Delta P := \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} = 0,$$

for (x, y) in the upper half-plane. $\lozenge$

Neither the Fejér kernel nor the Poisson kernel are functions in the Schwartz class. They do belong to the class of *functions of moderate decrease*, discussed in Chapter 8 below. For functions of moderate decrease we can reproduce essentially all the proofs presented so far concerning approximations of the identity. In Chapter 8 we also discuss a much larger class of objects, the *tempered distributions*, which includes all the kernels in this section and much more.

The heat kernel on $\mathbb{R}$ (a modified Gaussian):

$$H_t(x) := \frac{1}{\sqrt{4\pi t}}\, e^{-\frac{|x|^2}{4t}}, \qquad \text{for } t > 0. \tag{7.21}$$

EXERCISE 7.26. Show that the heat kernel is a solution of the *heat equation* on the line:

$$\frac{\partial H}{\partial t} = \frac{\partial^2 H}{\partial x^2}.$$

◊

EXERCISE 7.27. Show that the Gaussian kernels $\{G_\delta\}$, the Fejér kernels $\{F_R\}$, the Poisson kernels $\{P_y\}$, and the heat kernels $\{H_t\}$ generate approximations of the identity on $\mathbb{R}$, as $\delta \to 0$, $R \to \infty$, $y \to 0$, and $t \to 0$ respectively. The identities $\int_{\mathbb{R}} \frac{1-\cos x}{x^2}\, dx = \pi$ and $\int_{\mathbb{R}} \frac{1}{1+x^2}\, dx = \pi$ may be helpful. ◊

The Dirichlet kernel on $\mathbb{R}$:

$$D_R(x) := \int_{-R}^{R} e^{2\pi i x\xi}\, d\xi = \frac{\sin(2\pi R x)}{\pi x}, \qquad \text{for } R > 0. \tag{7.22}$$

EXERCISE 7.28. Let $S_R f$ denote the *partial Fourier integral* of f, defined by

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

Show that $S_R f = D_R * f$. ◊

The Dirichlet kernels are not uniformly integrable, and so the family $\{D_R\}$ does not generate an approximation of the identity. (This obstacle is the origin of many of the convergence problems for Fourier series and integrals.) However, we can get around the problem by averaging, as we did in the periodic context. The integral averages of the Dirichlet kernels are the Fejèr kernels, which are an approximation of the identity.

EXERCISE 7.29. Show that

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

◊

The conjugate Poisson kernel on $\mathbb{R}$:

$$Q_y(x) := \frac{x}{\pi(x^2+y^2)}, \qquad \text{for } y > 0. \tag{7.23}$$

EXERCISE 7.30. Show that the conjugate Poisson kernels are not integrable, but that they are square integrable. In other words, $Q_y \notin L^1(\mathbb{R})$, but $Q_y \in L^2(\mathbb{R})$. Also show that the conjugate Poisson kernel satisfies Laplace's equation $\Delta Q = 0$. Finally, show that $P_y(x) + iQ_y(x) = 1/(\pi i z)$, where $z = x + iy$. The function $1/(\pi i z)$ is analytic on the upper half-plane; it is known as the *Cauchy kernel*. Thus the Poisson kernel and the conjugate Poisson kernel are the real and imaginary parts, respectively, of the Cauchy kernel $1/(\pi i z)$. ◊

7.7. The Fourier Inversion Formula, and Plancherel's Identity

In this subsection we present several fundamental ingredients of the theory of Fourier transforms: the multiplication formula, the Fourier Inversion Formula, Plancherel's Identity, and the polarization identity. Then we use all these tools in an application to linear differential equations.

It follows as a consequence of differentiation being transformed into polynomial multiplication, and vice versa, that *the Fourier transform maps $\mathcal{S}(\mathbb{R})$ into itself.* See Theorem 7.17.

We turn to the *multiplication formula*, relating the inner products of two Schwartz functions with each other's Fourier transforms.

EXERCISE 7.31. (*The Multiplication Formula*) Verify that for functions $f, g \in \mathcal{S}(\mathbb{R})$,

$$\int_{\mathbb{R}} f(s)\widehat{g}(s)\,ds = \int_{\mathbb{R}} \widehat{f}(s)g(s)\,ds. \tag{7.24}$$

Hint: Use the Fubini–Tonelli Theorem (Appendix A) to interchange the order of the integrals in an appropriate double integral. ♢

We have deliberately used the variable name s in the terms $f(s)$, $\widehat{g}(s)$, $g(s)$, and $\widehat{f}(s)$, rather than the more usual variable names x for f and g and ξ for $\widehat{g}$ and $\widehat{f}$, since we have to use the same variable of integration in each of the integrals.

It is interesting to note that there is no analogue of the multiplication formula in the Fourier series setting. The Fourier coefficients of a periodic function are not a function but a sequence of numbers, and so the pairing of say f and the Fourier coefficients of g (the analogue of the inner product of f and $\widehat{g}$ in equation (7.24)) would not make sense. However it will make sense in the discrete case, where both the input and the output of the discrete Fourier transform are vectors in $\mathbb{C}^N$.

EXERCISE 7.32. State and prove a multiplication formula, analogous to formula (7.24), that is valid in $\mathbb{C}^N$. ♢

Several of the preceding facts can be woven together to prove the important Fourier inversion formula in $\mathcal{S}(\mathbb{R})$.

THEOREM 7.33 (Fourier Inversion Formula in $\mathcal{S}(\mathbb{R})$). *If $f \in \mathcal{S}(\mathbb{R})$, then for all $x \in \mathbb{R}$,*

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\xi)e^{2\pi i\xi x}\,d\xi = (\widehat{f}\,)^{\vee}(x). \tag{7.25}$$

Notice the symmetry between the definition in formula (7.10) of the Fourier transform $\widehat{f}(\xi)$ of the function $f(x)$, and the formula (7.25) for $f(x)$ in terms of its Fourier transform $\widehat{f}(\xi)$. Formally the only differences are the exchange of the symbols x and ξ and the exchange of the negative or positive signs that appear in the exponent. Whatever conditions are required on f so that formula (7.10) makes sense must also be required on $\widehat{f}$ so that formula (7.25) makes sense. In other words, the functions f and $\widehat{f}$ must enjoy the same status, which is the case in the Schwartz space. This is why the Fourier theory on $\mathcal{S}(\mathbb{R})$ is so beautiful and self-contained.

PROOF OF THEOREM 7.33. Fix $x \in \mathbb{R}$. The idea is to apply the multiplication formula (7.24) to the functions f and g, where f is an arbitrary Schwartz function and g is the specific Schwartz function given by

$$g(s) := e^{2\pi isx}e^{-\pi|ts|^2} \in \mathcal{S}(\mathbb{R}).$$

Compute directly, or use properties (c) and (d) from the time–frequency dictionary in Section 7.3, to verify that

$$\widehat{g}\,(s) = t^{-1}e^{-\pi|t^{-1}(x-s)|^2} = G_t(x-s), \tag{7.26}$$

which is a translate of the approximation of the identity generated by the Gaussian. Now equation (7.24) can be rewritten as

$$\int_{\mathbb{R}} f(s)G_t(x-s)\,ds = \int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}e^{-\pi|ts|^2}\,ds.$$

Let $t \to 0$. The left-hand side is $f * G_t(x)$, which converges to $f(x)$ (uniformly in x) as $t \to 0$ by Exercise 7.24. As for the right-hand side, we can interchange the limit as $t \to 0$, and the integral by the Dominated Convergence Theorem **??**, since the integrand, a function indexed by $t > 0$, is dominated by the absolute value of $\widehat{f} \in \mathcal{S}(\mathbb{R})$ by Theorem 7.17 hence integrable. Therefore

$$\begin{aligned} f(x) &= \lim_{t\to 0}\int_{\mathbb{R}} f(s)G_t(x-s)\,ds \\ &= \lim_{t\to 0}\int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}e^{-\pi|ts|^2}\,ds \\ &= \int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}\lim_{t\to 0} e^{-\pi|ts|^2}\,ds = \int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}\,ds =: (\widehat{f}\,)^{\vee}(x). \end{aligned}$$

Lo and behold, $f(x) = (\widehat{f}\,)^{\vee}(x)$, which is the inversion formula we were seeking. □

EXERCISE 7.34. Use the time–frequency dictionary to check the identity (7.26) used in the proof of the Fourier Inversion Formula. ♢

EXERCISE 7.35. Justify the following interchange of limit and integral for $f \in \mathcal{S}(\mathbb{R})$ using an argument other than the Dominated Convergence Theorem **??**:

$$\lim_{t\to 0}\int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}e^{-\pi|ts|^2}\,ds = \int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}\lim_{t\to 0} e^{-\pi|ts|^2}\,ds.$$

Hint: Break the integral in two pieces: where $|s| \le M$ and where $|s| > M$. For the bounded piece use uniform convergence of the integrands to interchange the limit and the integral. For the unbounded piece use the decay properties of $\widehat{f}$ to ensure that integral can be made small for M large enough. ♢

The inversion formula guarantees that the Fourier transform is in fact a *bijection* from the Schwartz space $\mathcal{S}(\mathbb{R})$ onto itself. More can be said: the Fourier transform is an *energy-preserving map*, also called a *unitary transformation*, on $\mathcal{S}(\mathbb{R})$. This is the content of *Plancherel's Identity*. First we must define our notion of the size of a function. We can equip $\mathcal{S}(\mathbb{R})$ with an inner product

$$\langle f, g\rangle := \int_{\mathbb{R}} f(x)\overline{g(x)}\,dx, \qquad \text{for all } f,\, g \in \mathcal{S}(\mathbb{R}),$$

with associated L^2-norm given by $\|f\|_2^2 = \langle f, f\rangle$. Thus the L^2-norm is defined by

$$\|f\|_2 := \left(\int_{\mathbb{R}} |f(x)|^2\,dx\right)^{1/2}.$$

For Schwartz functions f and g, the integral defining the inner product yields a well-defined finite complex number.

Plancherel's Identity says that each Schwartz function f and its Fourier transform $\widehat{f}$ have the same size; in other words that the Fourier transform preserves the L^2-norm.

THEOREM 7.36 (Plancherel's Identity). *If* $f \in \mathcal{S}(\mathbb{R})$, *then* $\|f\|_2 = \|\widehat{f}\|_2$.

PROOF. We use the time–frequency dictionary, the multiplication formula and the Fourier inversion formula to give a proof of Plancherel's Identity. Set $\widehat{g} = \overline{f}$ in the multiplication formula (7.24). Then, by the analogue $g = (\overline{f})^\vee = \overline{\widehat{f}}$ of property (f) for the inverse Fourier transform, we obtain

$$\int |f|^2 = \int f\overline{f} = \int f\widehat{g} = \int \widehat{f}g = \int \widehat{f}\overline{\widehat{f}} = \int |\widehat{f}|^2. \qquad \square$$

We end this section with two useful results. The first is the *polarization identity*, which says that the inner product of two Schwartz functions is equal to the inner product of their Fourier transforms. The second is an application to differential equations that relies crucially on Theorem 7.17.

EXERCISE 7.37. (*Polarization Identity*) Prove the polarization identity for real-valued functions f and g in $\mathcal{S}(\mathbb{R})$. Namely,

$$\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\rangle. \tag{7.27}$$

Hint: Use Plancherel's Identity for $f + g$ and $f - g$, then add. ◊

EXAMPLE 7.38. (*An Application to Solving Differential Equations*) Let us find a function f on $\mathbb{R}$ that satisfies the differential equation

$$f^{(3)}(x) + 3f''(x) - 2f'(x) - 6f(x) = e^{-\pi x^2}.$$

Taking the Fourier transform on both sides of the equation and using the time–frequency dictionary, we obtain

$$\widehat{f}(\xi)\big[(2\pi i\xi)^3 + 3(2\pi i\xi)^2 - 2(2\pi i\xi) - 6\big] = e^{-\pi\xi^2}. \tag{7.28}$$

Let $P(t) = t^3 + 3t^2 - 2t - 6 = (t + 3)(t^2 - 2)$. Then the polynomial inside the brackets in equation (7.28) is $Q(\xi) := P(2\pi i\xi)$. Solving for $\widehat{f}(\xi)$ (fortunately $Q(\xi)$ has no real zeros), and using the Fourier inversion formula in $\mathcal{S}(\mathbb{R})$, we obtain

$$\widehat{f}(\xi) = \frac{e^{-\pi\xi^2}}{Q(\xi)} \qquad \text{and so} \qquad f(x) = \left(\frac{e^{-\pi\xi^2}}{Q(\xi)}\right)^\vee (x).$$

We have expressed the solution $f(x)$ as the inverse Fourier transform of a known function. ◊

In general, let $P(x) = \sum_{k=0}^n a_k x^k$ be a polynomial of degree n with constant complex coefficients a_k. If $Q(\xi) = P(2\pi i\xi)$ has no real zeros, and $u \in \mathcal{S}(\mathbb{R})$, then the linear differential equation

$$P(D)f = \sum_{k=0}^n a_k D^k f = u$$

has a solution given by

$$f = \big(\widehat{u}(\xi)Q(\xi)^{-1}\big)^\vee.$$

Note that since $Q(\xi)$ has no real zeros, the function $\widehat{u}(\xi)Q(\xi)^{-1}$ is in the Schwartz class. Therefore we may compute its inverse Fourier transform and, by Theorem 7.17, the result is a function in $\mathcal{S}(\mathbb{R})$. In many cases $\big(\widehat{u}(\xi)Q(\xi)^{-1}\big)^{\vee}$ can be computed explicitly.

7.8. L^p-norms on $\mathcal{S}(\mathbb{R})$

In this subsection we define the L^p-norm $\|f\|_p$ of a Schwartz function f for $1 \leq p \leq \infty$, and state Minkowski's inequality. The heart of the subsection is the proof of Theorem 7.40, showing that for Schwartz functions f, the convolution of f with a suitable approximation to the identity converges to f in the L^p sense for each real number p such that $1 \leq p < \infty$.

We equip the Schwartz space $\mathcal{S}(\mathbb{R})$ with a family of norms, called the *L^p-norms*, indexed by the real numbers p such that $1 \leq p < \infty$ and by $p = \infty$. For $f \in \mathcal{S}(\mathbb{R})$ let

$$\|f\|_p := \left(\int_{\mathbb{R}} |f(x)|^p \, dx \right)^{1/p}, \qquad \text{for} \quad 1 \leq p < \infty,$$

and

$$\|f\|_\infty := \sup_{|x| \in \mathbb{R}} |f(x)|, \qquad \text{for } p = \infty.$$

EXERCISE 7.39. (*Schwartz Functions Have Finite L^p-norm*) Verify that if $f \in \mathcal{S}(\mathbb{R})$, then $\|f\|_p < \infty$ for each p with $1 \leq p \leq \infty$. ◊

The Schwartz space $\mathcal{S}(\mathbb{R})$ with the L^p-norm is a normed space. In particular, the *triangle inequality* holds:

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p, \qquad \text{for all } f,\, g \in \mathcal{S}(\mathbb{R}) \text{ and all } p \text{ with } 1 \leq p \leq \infty.$$

This triangle inequality for the L^p-norm is also known as *Minkowski's inequality*. We prove it in Chapter 11 (Lemma 11.60).

When $\|f\|_p < \infty$, we say that $f \in L^p(\mathbb{R})$. Thus for each p with $1 \leq p \leq \infty$, the function space $L^p(\mathbb{R})$ consists of those functions $f : \mathbb{R} \to \mathbb{R}$ whose L^p-norms are finite. To make this precise we have to appeal to Lebesgue integration on $\mathbb{R}$, as we did when defining $L^p(\mathbb{T})$ in Chapter 2. Exercise 7.39 shows that $\mathcal{S}(\mathbb{R}) \subset L^p(\mathbb{R})$. A short-cut, which we have already used when talking about $L^p(\mathbb{T})$, would be to define $L^p(\mathbb{R})$ as the *completion of $\mathcal{S}(\mathbb{R})$ with respect to the L^p-norm.* Both definitions coincide. The second has the advantage that it immediately implies that the Schwartz class is dense in $L^p(\mathbb{R})$ with respect to the corresponding L^p-norm.

We show now that given an approximation of the identity $\{K_t\}_{t \in \Lambda}$ in $\mathcal{S}(\mathbb{R})$ and a function $f \in \mathcal{S}(\mathbb{R})$, their convolutions $K_t * f$ converge to f not only uniformly but also in each L^p-norm.

THEOREM 7.40. *Let the family $\{K_t\}_{t \in \Lambda}$ be an approximation of the identity in $\mathcal{S}(\mathbb{R})$ as $t \to t_0$. If $f \in \mathcal{S}(\mathbb{R})$, then*

$$\lim_{t \to 0} \|K_t * f - f\|_p = 0 \qquad \text{for each } p \text{ with } 1 \leq p \leq \infty.$$

PROOF. We prove the result for $1 \leq p < \infty$. The proof below can be modified for the case $p = \infty$, which is the content of Theorem 7.23, whose proof was relegated to Exercise 7.24.

Define the translation operator τ_y by $\tau_y f(x) := f(x-y)$, for $y \in \mathbb{R}$. Notice that *translations preserve L^p-norms*, in other words $\|\tau_y f\|_p = \|f\|_p$ for all $f \in \mathcal{S}(\mathbb{R})$ and for all $h \in \mathbb{R}$. Thus for all $h \in \mathbb{R}$ we have the upper bound $\|\tau_y f - f\|_p \leq 2\|f\|_p$, by the triangle inequality.

The following estimate holds:

$$\|K_t * f - f\|_p = \left\|\int_{\mathbb{R}} [\tau_y f(x) - f(x)] K_t(y)\, dy\right\|_p \leq \int_{\mathbb{R}} \|\tau_y f - f\|_p |K_t(y)|\, dy. \tag{7.29}$$

We will justify (7.29) for $p = 1$ now, and we defer the sketch of the proof of inequality (7.29) for $p > 1$ until the end of our proof of Theorem 7.40, so as not to distract from the main line of the argument.

The following inequalities follow from property (i) of the approximation of the identity and the triangle inequality for integrals:

$$\begin{aligned} |K_t * f(x) - f(x)| &= \left|\int_{\mathbb{R}} f(x-y)\, K_t(y)\, dy - f(x)\right| \\ &= \left|\int_{\mathbb{R}} [\tau_y f(x) - f(x)]\, K_t(y)\, dy\right| \\ &\leq \int_{\mathbb{R}} |\tau_y f(x) - f(x)|\, |K_t(y)|\, dy. \end{aligned} \tag{7.30}$$

In the second equality we used property (i) of the kernels K_t that says that they have integral equal to one, therefore $f(x) = \int f(x) K_t(y)\, dy$. The inequality holds because the absolute value of an integral is smaller than the integral of the absolute value. This inequality is what we are calling the triangle inequality for integrals, and can be justified by noticing that the integrals can be approximated by Riemann sums (finite sums), for which the ordinary triangle inequality for absolute values can be applied. If instead of absolute values we were calculating with some other norm, in our case the L^p-norm, for which the triangle inequality holds, we would expect the analogous estimate (7.29) to hold. For $p = 1$, all that remains to be done is to integrate the pointwise inequality just derived and apply Fubini Theorem **??** to interchange the order of integration:

$$\|K_t * f - f\|_1 \leq \int_{\mathbb{R}}\int_{\mathbb{R}} |\tau_y f(x) - f(x)|\, |K_t(y)|\, dy\, dx = \int_{\mathbb{R}} \|\tau_y f - f\|_1 |K_t(y)|\, dy.$$

We estimate separately the contributions to the right-hand side of inequality (7.29) for large and small y.

First, the integral of the absolute values of K_t away from zero can be made arbitrarily small for t close to t_0; this is property (iii) of the approximation of the identity. More precisely, given $f \in \mathcal{S}(\mathbb{R})$, $\varepsilon > 0$, and $\delta > 0$, we can choose h small enough so that for all $|t - t_0| < h$,

$$\int_{|y|>\delta} |K_t(y)|\, dy \leq \frac{\varepsilon}{4\|f\|_p}.$$

Second, we can make $\|\tau_y f - f\|_p$ small provided y is small. More precisely, since $f \in \mathcal{S}(\mathbb{R})$, we know that f is continuous and $f \in L^p(\mathbb{R})$. For continuous functions in $L^p(\mathbb{R})$,

$$\lim_{y\to 0} \|\tau_y f - f\|_p = 0. \tag{7.31}$$

In other words, translation is a continuous operation at zero in L^p, at least for continuous functions f. See Exercise 7.42. A density argument shows that the statement is actually true for all functions in $L^p(\mathbb{R})$, not just the continuous ones.

Hence, given $f \in \mathcal{S}(\mathbb{R})$ and $\varepsilon > 0$, there exists $\delta > 0$ (depending on f) such that

$$\|\tau_y f - f\|_p \leq \frac{\varepsilon}{2C}, \qquad \text{for all } |y| < \delta. \tag{7.32}$$

For $|y| < \delta$ we can apply estimate (7.32), with C the constant that controls uniformly the L^1-norms of K_t, that is $\int |K_t(y)|dy \leq C$ for all $t \in \Lambda$. Altogether, we have

$$\begin{aligned}
\|K_t * f - f\|_p &\leq \int_{|y|<\delta} |K_t(y)| \|\tau_y f - f\|_p \, dy + \int_{|y|\geq\delta} |K_t(y)| \|\tau_y f - f\|_p \, dy \\
&\leq \frac{\varepsilon}{2C} \int_{|y|<\delta} |K_t(y)| \, dy + 2\|f\|_p \int_{|y|\geq\delta} |K_t(y)| \, dy \\
&\leq \frac{\varepsilon}{2} + 2\|f\|_p \frac{\varepsilon}{4\|f\|_p} = \varepsilon.
\end{aligned}$$

In the last inequality we have used properties (ii) and (iii) of the approximation of the identity.

This completes the proof of Theorem 7.40, modulo inequality (7.29) for $p > 1$. □

It remains to establish inequality (7.29). The inequality (7.29) is a consequence of an integral version of the *triangle inequality* that goes by the name *Minkowski's Integral Inequality*, which essentially says that the *norm of the integral is the integral of the norms*. At the end of the book you can find a precise statement and some applications (page 239). Since f is a Schwartz function, both integrals in inequality (7.29) are limits of Riemann integrals on larger and larger intervals $[-M, M]$. The Riemann integral $\int_{-M}^{M} |\tau_y f(x) - f(x)| \, |K_t(y)| \, dy$ is a limit of finite sums of the type

$$\sum_{n=1}^{N} |\tau_{y_n} f(x) - f(x)| \, |K_t(y_n)| \, \Delta y.$$

We can now apply the triangle inequality[3] for the L^p-norm to this sum:

$$\left\| \sum_{n=1}^{N} |\, \tau_{y_n} f(\cdot) - f(\cdot)| \, |K_t(y_n)| \, \Delta y \right\|_p \leq \sum_{n=1}^{N} \|\tau_{y_n} f - f\|_p \, |K_t(y_n)| \, \Delta y.$$

The right-hand side is a Riemann sum for the integral $\int_{-M}^{M} \|\tau_y f - f\|_p \, |K_t(y)| \, dy$.

Taking appropriate limits on the partitions defining the Riemann sums, we conclude that for all $M > 0$,

$$\left\| \int_{-M}^{M} (\tau_y f(x) - f(x)) \, K_t(y) \, dy \right\|_p \leq \int_{-M}^{M} \|\tau_y f - f\|_p \, K_t(y) \, dy.$$

[3]More precisely, the triangle inequality for N summands $f_n \in L^p(\mathbb{R})$, $a_n \in \mathbb{C}$,

$$\left\| \sum_{n=1}^{N} a_n f_n \right\|_p \leq \sum_{n=1}^{N} |a_n| \|f_n\|_p.$$

In our case, choose $f_n = \tau_{y_n} f - f$, and $a_n = k(y_n)\Delta y$.

Finally, let $M \to \infty$ to obtain inequality (7.29).

EXERCISE 7.41. Show that if $f \in \mathcal{S}(\mathbb{R})$, then $\tau_y f \in \mathcal{S}(\mathbb{R})$, and for all $1 \leq p < \infty$,

$$\|f\|_p = \|\tau_y f\|_p.$$

$\Diamond$

EXERCISE 7.42. Show that for $f \in \mathcal{S}(\mathbb{R})$, we can interchange the following limit and integral:

$$\lim_{y\to 0} \int_{\mathbb{R}} |f(x-y) - f(x)|^p \, dx = \int_{\mathbb{R}} \lim_{y\to 0} |f(x-y) - f(x)|^p \, dx = 0. \tag{7.33}$$

Thus equation (7.31) is valid. $\Diamond$

We have not used the fact that $K \in \mathcal{S}(\mathbb{R})$, except to feel comfortable when writing down the integrals. We did use the defining properties of an approximation of an identity. As for f, all that is really required is for f to be L^p-integrable. There are stronger versions of the theorem; one can weaken the hypotheses and still get the same conclusion.

We encountered this theorem earlier in the context of Fourier series. It implied that the Cesàro and Abel sums converge to f in $L^p(\mathbb{T})$, at least for continuous functions f. See Section 4.7.