Christopher Heil

Introduction to Harmonic Analysis

November 12, 2010

Springer

Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Contents

General Notation

We review some of the notational conventions that will be used throughout this volume.

We use the symbol $\square$ to denote the end of a proof, and the symbol $\diamondsuit$ to denote the end of a definition, remark, example, or exercise, or the end of the statement of a theorem whose proof will be omitted.

Unless otherwise specified, all vector spaces are taken over the complex field $\mathbb{C}$. In particular, functions whose domain is $\mathbb{R}^d$ (or a subset of $\mathbb{R}^d$) are generally allowed to take values in the complex plane $\mathbb{C}$.

If S is a subspace of a vector space V, then the finite linear span of S is denoted by $\mathrm{span}(S)$. If V is also a metric space, then the closed linear span $\overline{\mathrm{span}}(V)$ is defined to be the closure of $\mathrm{span}(S)$ in V. We say that S is *complete* if $\overline{\mathrm{span}}(S) = V$.

The extended real line is $\mathbb{R} \cup \{-\infty, \infty\} = [-\infty, \infty]$. We use the conventions that $1/0 = \infty$, $1/\infty = 0$, and $0 \cdot \infty = 0$.

Given a set X, the *characteristic function* of a subset $A \subseteq X$ is the function $\chi_A \colon X \to \mathbb{R}$ defined by

$$\chi_A(x) \;=\; \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

The *Kronecker delta* is

$$\delta_{ij} \;=\; \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

If $1 \le p \le \infty$ is given, then its *dual index* or *dual exponent* is the extended real number p' that satisfies

$$\frac{1}{p} + \frac{1}{p'} \;=\; 1.$$

Explicitly,

$$p' = \frac{p}{p-1}.$$

The dual index lies in the range $1 \le p' \le \infty$, and we have $1' = \infty$, $2' = 2$, and $\infty' = 1$.

Integrals with unspecified limits are taken over either the real line or $\mathbb{R}^d$, according to context. In particular, if $f\colon \mathbb{R} \to \mathbb{C}$, then we take

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

Unless otherwise specified, statements about measure and integration refer to Lebesgue measure and Lebesgue integration.

Given $f \in L^p(\mathbb{R})$ and $g \in L^{p'}(\mathbb{R})$ we use an inner-product-like notation to denote the following integral:

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

Note that the notation $\langle f, g \rangle$ is linear as a function of f, but is antilinear as a function of g. If $p = 2$ then $p' = 2$ and $\langle \cdot, \cdot \rangle$ is the ordinary inner product on $L^2(\mathbb{R})$.

The dual space of a Banach space X is denoted by X^*. Given a continuous linear functional $\mu \in X^*$, we denote its action on $f \in X$ by either

$$\mu(f) \qquad \text{or} \qquad \langle f, \mu \rangle.$$

The latter notation is especially convenient when dealing with $X = L^p(\mathbb{R})$ and with distributions. In particular, if $1 \le p < \infty$ then the dual space $L^p(\mathbb{R})^*$ is isomorphic to $L^{p'}(\mathbb{R})$ in the sense that given any $\mu \in L^p(\mathbb{R})^*$ there exists a unique function $g \in L^{p'}(\mathbb{R})$ such that

$$\langle f, \mu \rangle = \int f(x)\,\overline{g(x)}\,dx = \langle f, g \rangle, \qquad f \in L^p(\mathbb{R}).$$

The notation $\mu(f)$ is taken to be linear both as a function of f and as a function of μ, while the notation $\langle f, \mu \rangle$ is linear as a function of f but antilinear as a function of μ.

1

The Fourier Transform on $L^1(\mathbb{R})$

Harmonic analysis revolves around the Fourier transform. As we will discuss in Section 2.1, the Fourier transform can be defined on any locally compact abelian (LCA) group. However, for most of this volume we will concentrate on the setting of the real line $\mathbb{R}$. In many ways, the real line is among the most "complex" of the LCA groups, as it is neither compact nor discrete. Once we understand the Fourier transform on $\mathbb{R}$, it is easy to move to other settings, as we do in Chapter 2.

Even when we restrict our attention to a single underlying group such as the real line, there are many domains on which we can define the Fourier transform. In this chapter we focus on what is perhaps the most concrete domain, $L^1(\mathbb{R})$. In Chapter 3 we will see how the Fourier transform extends to $L^2(\mathbb{R})$. Chapter 4 takes us far beyond transformations of functions by defining the Fourier transform on the extraordinarily broad domain of the tempered distributions. We bring a little more concreteness back into the picture in Chapter 5 by considering the Fourier transform of bounded Radon measures.

Some basic notation was laid out in the General Notation section. Most of the other terminology that we will use in this chapter, such as the definitions of the spaces $L^1(\mathbb{R})$, $C_0(\mathbb{R})$, $C_b(\mathbb{R})$, $C_c^\infty(\mathbb{R})$ and so forth, are given in Appendix B, which briefly reviews definitions and major results from real analysis, operator theory, and functional analysis.

1.1 Definition and Basic Properties

1.1.1 The Fourier Transform on $L^1(\mathbb{R})$

We define the Fourier transform on $L^1(\mathbb{R})$ as follows.

Definition 1.1 (Fourier Transform on $L^1(\mathbb{R})$). The *Fourier transform* of $f \in L^1(\mathbb{R})$ is the function $\widehat{f}\colon \mathbb{R} \to \mathbb{C}$ defined by

$$\widehat{f}(\xi) \;=\; \int f(x)\, e^{-2\pi i\xi x}\, dx \;=\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi x}\, dx, \quad \xi \in \mathbb{R}. \qquad \diamondsuit \tag{1.1}$$

Note that even though f is only defined almost everywhere, $\widehat{f}(\xi)$ is a well-defined complex scalar for every $\xi \in \mathbb{R}$ since

$$\int |f(x)\, e^{-2\pi i\xi x}|\, dx \;=\; \int |f(x)|\, dx \;=\; \|f\|_1 \;<\; \infty.$$

In fact, we will soon see that $\widehat{f}$ is a continuous function on $\mathbb{R}$. We want to understand the properties of this function $\widehat{f}$, and the relation between f and $\widehat{f}$. When we wish to emphasize the role of the Fourier transform as an operator mapping f to $\widehat{f}$, we will write

$$\mathcal{F}f \;=\; \widehat{f}.$$

To avoid notionally ugly formulations, we sometimes write $f^\wedge$ or $(f)^\wedge$ instead of $\widehat{f}$.

Now we examine some of the properties of the Fourier transform, and explain why we are so interested in it.

Lemma 1.2. *The Fourier transform $\mathcal{F}$ is a bounded linear mapping of $L^1(\mathbb{R})$ into $L^\infty(\mathbb{R})$, and its operator norm satisfies*

$$\|\mathcal{F}\| \;=\; \sup_{\|f\|_1 = 1} \|\widehat{f}\|_\infty \;\le\; 1.$$

Proof. The fact that $\mathcal{F}$ is linear is the reader's first exercise. Given $f \in L^1(\mathbb{R})$, we have for any $\xi \in \mathbb{R}$ that

$$\begin{aligned}
|\widehat{f}(\xi)| \;&=\; \left| \int f(x)\, e^{-2\pi i\xi x}\, dx \right| \\
&\le\; \int |f(x)\, e^{-2\pi i\xi x}|\, dx \\
&=\; \int |f(x)|\, dx \;=\; \|f\|_1.
\end{aligned}$$

Hence

$$\|\widehat{f}\|_\infty \;=\; \operatorname*{ess\,sup}_{\xi \in \mathbb{R}} |\widehat{f}(\xi)| \;\le\; \|f\|_1.$$

Consequently $\mathcal{F}\colon L^1(\mathbb{R}) \to L^\infty(\mathbb{R})$ is bounded, and $\|\mathcal{F}\| \le 1$. $\square$

Thus, the Fourier transform maps the unit ball in $L^1(\mathbb{R})$ into the unit ball in $L^\infty(\mathbb{R})$. The following exercise shows that we actually have $\|\mathcal{F}\| = 1$, and that the supremum in the definition of the operator norm of $\mathcal{F}$ is achieved, i.e., there exists an $f \in L^1(\mathbb{R})$ with $\|f\|_1 = 1$ such that $\|\widehat{f}\|_\infty = 1$. Even so, the range of $\mathcal{F}$ is a proper subset of $L^\infty(\mathbb{R})$ (see Theorem 1.20).

Exercise 1.3. Show that if $f \in L^1(\mathbb{R})$ then we have the useful fact that

$$\widehat{f}(0) = \int f(x)\,dx.$$

Assume for now that $\widehat{f}$ is continuous (we will prove this in Exercise 1.8), and use this equality to show that if $f \geq 0$ a.e., then $\|\widehat{f}\,\|_\infty = \|f\|_1$. Conclude that $\|\mathcal{F}\| = 1$. $\diamondsuit$

Remark 1.4. There are many other standard normalizations for the Fourier transform. For example, each of

$$\int f(x)\,e^{2\pi i\xi x}\,dx, \qquad \int f(x)\,e^{-i\xi x}\,dx, \qquad \frac{1}{\sqrt{2\pi}}\int f(x)\,e^{-i\xi x}\,dx,$$

is a common choice. Each normalization makes certain formulas notationally nicer and other formulas more notationally complicated. $\diamondsuit$

We introduce a convenient notation for the dilation of a function.

Notation 1.5 (Dilation). Given a function $f\colon \mathbb{R} \to \mathbb{C}$ and given $r > 0$, we define the L^1-normalized *dilation* of f by r to be the function $f_r\colon \mathbb{R} \to \mathbb{C}$ given by

$$f_r(x) = rf(rx). \qquad \diamondsuit$$

The dilation f_r is L^1-normalized in the sense that it preserves the L^1 norm, i.e., $\|f_r\|_1 = \|f\|_1$. When $r < 1$, the function f_r is a "stretched out" version of f, while when $r > 1$, the function f_r is a "shrunken" version of f, both with a vertical scaling that ensure that the L^1-norm is preserved. When $r > 1$, it may be more descriptive to refer to f_r as a *compression* of f, but it is customary to use the term *dilation* to refer to f_r for all values of r.

Although there is the possibility of ambiguity between the notation for the dilation f_r of a function f and an element f_n of a sequence of functions $\{f_n\}_{n\in\mathbb{N}}$, the meaning will usually be clear from context.

It is about time that we see a specific example of a Fourier transform.

Definition 1.6 (Dirichlet and Sinc Functions). The *Dirichlet function* is

$$d(\xi) = \frac{\sin\xi}{\pi\xi}$$

(implicitly defined at the origin so as to be continuous on $\mathbb{R}$). The *sinc function* is

$$\operatorname{sinc}(\xi) = \frac{\sin\pi\xi}{\pi\xi} = \pi d(\pi\xi) = d_\pi(\xi). \qquad \diamondsuit$$

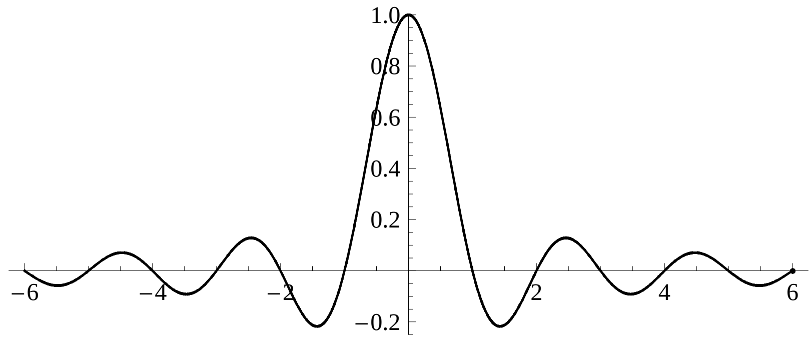

Fig. 1.1. Graph of the sinc function $d_\pi(\xi) = \frac{\sin \pi\xi}{\pi\xi}$.

The sinc function is also known as the *cardinal sine*, and indeed "sinc" is a contraction of *sinus cardinalis*. The "cardinal" nature of the sinc function is the fact that it is an interpolating function, because $\operatorname{sinc}(0) = 1$ while $\operatorname{sinc}(n) = 0$ for all integers $n \neq 0$. The Dirichlet function is named for Johann Dirichlet (1805–1859).

Note that the functions d and sinc do not belong to $L^1(\mathbb{R})$, because they only decay slowly at infinity (on the order of $1/|\xi|$). On the other hand, they do belong to $L^p(\mathbb{R})$ for every $1 < p \leq \infty$.

Exercise 1.7. Let $\chi_{[-T,T]}$ denote the characteristic function of the interval $[-T,T]$. Note that $\chi_{[-T,T]} \in L^1(\mathbb{R})$. Prove the following statements.

(a) $(\chi_{[-T,T]})^\wedge = \frac{\sin 2\pi T\xi}{\pi\xi} = d_{2\pi T}(\xi)$.

(b) $\|\chi_{[-T,T]}\|_1 = \|(\chi_{[-T,T]})^\wedge\|_\infty$.

(c) $d_{2\pi T}$ is uniformly continuous.

(d) $d_{2\pi T} \in C_0(\mathbb{R})$. $\diamondsuit$

The last two properties of $d_{2\pi T}$ are in fact shared by any function that is the Fourier transform of an L^1 function. The following exercise asks for a direct proof of the fact that $\widehat{f}$ is uniformly continuous if $f \in L^1(\mathbb{R})$. On the other hand, Theorem 1.20 below shows that if $f \in L^1(\mathbb{R})$ then $\widehat{f}$ must belong to $C_0(\mathbb{R})$, which also implies (see Exercise 1.17) that $\widehat{f}$ must be uniformly continuous.

Exercise 1.8. Show directly that if $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is uniformly continuous, i.e.,

$$\lim_{\eta\to 0} \|\widehat{f} - T_\eta \widehat{f}\|_\infty \;=\; \lim_{\eta\to 0} \left(\sup_{\xi\in\mathbb{R}} \bigl|\widehat{f}(\xi) - \widehat{f}(\xi-\eta)\bigr| \right) \;=\; 0. \qquad \diamondsuit$$

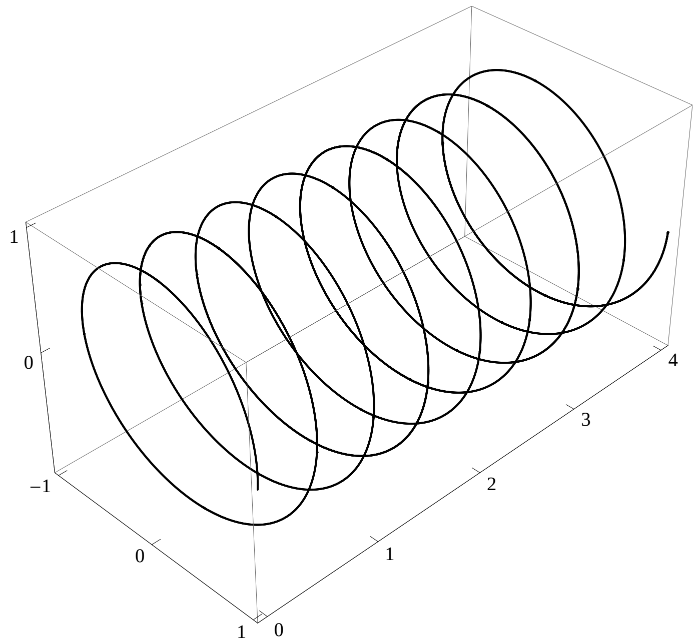

Fig. 1.2. Graph of $e_\xi(x) = e^{2\pi i \xi x}$ for $\xi = 2$ and $0 \le x \le 4$.

1.1.2 Motivation

We pause for a moment in our mathematical development of the Fourier transform to give some motivation for its definition and purpose. One of our main goals in this chapter is to prove the following *Inversion Formula* for the Fourier transform (see Theorem 1.89).

Theorem 1.9 (Inversion Formula). *If f, $\widehat{f} \in L^1(\mathbb{R})$ then f and $\widehat{f}$ are continuous and*

$$f(x) \;=\; \int \widehat{f}(\xi)\, e^{2\pi i \xi x}\, d\xi, \tag{1.2}$$

with equality holding pointwise everywhere. $\diamondsuit$

Thus, assuming the hypothesis that both f and $\widehat{f}$ are integrable, if we know all the values $\widehat{f}(\xi)$ of the Fourier transform of f, then we can recover f from these values via equation (1.2). To explain the significance of this, picture the complex exponential function

$$e_\xi(x) \;=\; e^{2\pi i \xi x} \;=\; \cos(2\pi\xi x) + i \sin(2\pi\xi x)$$

as a function of x. While x lies in $\mathbb{R}$, the function values $e_\xi(x)$ are complex numbers that lie on the unit circle S^1 in $\mathbb{C}$. As x ranges through the real line, the values $e_\xi(x) = e^{2\pi i \xi x}$ move around the unit circle S^1. If $\xi > 0$, then as x increases through an interval of length $1/\xi$, the values $e_\xi(x) = e^{2\pi i\xi x}$ move once around S^1 in the counter-clockwise direction. Thus e_ξ is periodic with period $1/\xi$. If ξ is negative, the same is true except that the values $e_\xi(x) = e^{2\pi i\xi x}$ circle around S^1 in the opposite direction. The graph of e_ξ is

$$\Gamma_\xi \;=\; \bigl\{(x, e^{2\pi i\xi x}) : x \in \mathbb{R}\bigr\} \;\subseteq\; \mathbb{R} \times \mathbb{C}.$$

Identifying $\mathbb{R} \times \mathbb{C}$ with $\mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3$, the graph Γ_ξ is a helix in $\mathbb{R}^3$ coiling around the x-axis, which runs down the center of the helix (see Figure 1.2). The function e_ξ is periodic with period $1/\xi$, and we therefore say that it has *frequency* ξ, i.e., the frequency is the reciprocal of the period. The function e_ξ is a "pure tone" or a "pure frequency" in some sense. Its real part is $\cos(2\pi\xi x)$, a cosine of period $1/\xi$ and frequency ξ. Its imaginary part is $\sin(2\pi\xi x)$, a sine of period $1/\xi$ and frequency ξ.

We turn to music for illustration. Restricting our attention for the moment to real-valued functions, imagine that the function $f(x)$ represents the displacement of the center of an ideal vibrating string, or the end of an ideal tuning fork, or the center of your stereo speaker, from its rest position. The vibrating string, tuning fork, or speaker creates a pressure wave in the air, which causes your eardrum to vibrate and your brain to perceive sound. For the function $\cos(2\pi\xi x)$, the sound that you would hear is a "pure tone" of frequency ξ. Real strings are of course much more complicated—a piano string or a violin string each vibrates in a very complicated way, resulting in their different sounds. But an ideal string or tuning fork whose displacement is exactly given by the function $\cos(2\pi\xi x)$ would be a pure tone, with no overtones or other complications (see the illustration in Figure 1.3). The function $e^{2\pi i\xi x}$ is a complex version of this pure tone.

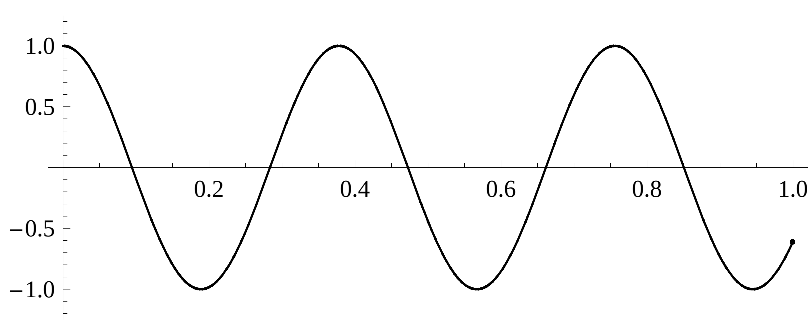

Fig. 1.3. Graph of $\varphi(x) = \cos(2\pi\sqrt{7}x)$.

For a given fixed ξ, the function $\widehat{f}(\xi)\, e^{2\pi i\xi x}$ is a pure tone whose *amplitude* is the scalar $\widehat{f}(\xi)$. The larger $\widehat{f}(\xi)$ is, the larger the vibrations of our string or tuning fork, and the louder the perceived sound.

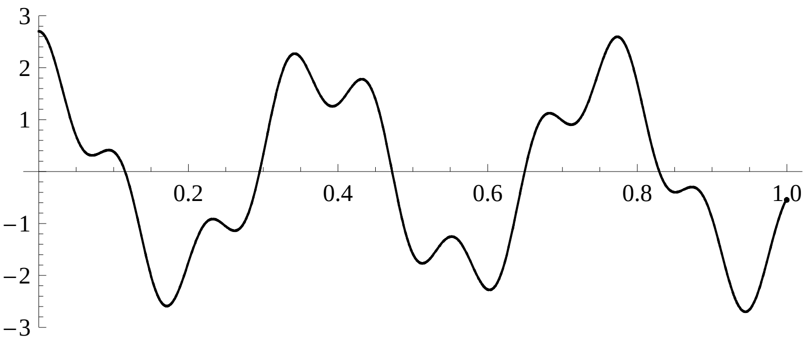

Fig. 1.4. Graph of $\varphi(x) = 2\cos(2\pi\sqrt{7}x) + 0.7\cos(2\pi 9x)$.

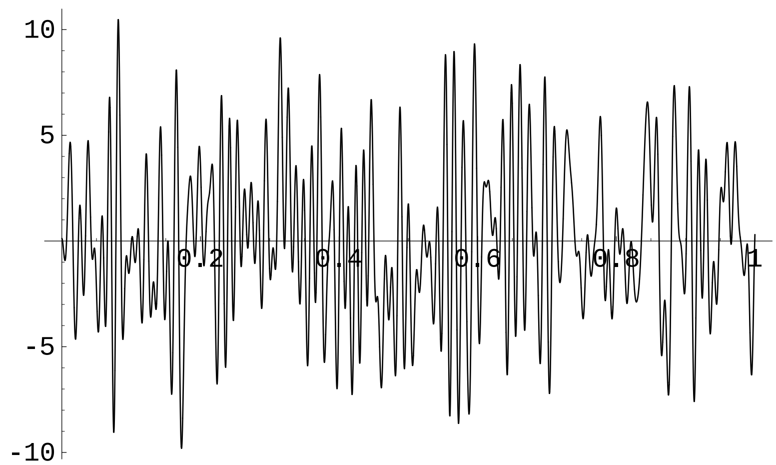

Fig. 1.5. Graph of 75 superimposed pure tones: $\varphi(x) = \sum_{k=1}^{75} \widehat{f}(\xi_k)\,\cos(2\pi\xi_k x)$.

Given two frequencies η, ξ and amplitudes $\widehat{f}(\eta)$, $\widehat{f}(\xi)$, a function φ of the form

$$\varphi(x) \;=\; \widehat{f}(\eta)\, e^{2\pi i\eta x} + \widehat{f}(\xi)\, e^{2\pi i\xi x}$$

is a superposition of two pure tones (see the illustration in Figure 1.4). With some caveats, if this function represents the displacement of our string, you would very likely be able to tell that the sound you hear is a superposition of two frequencies—the human ear is very well-adapted to sounds of this type. A superposition of 75 pure tones with randomly chosen frequencies and amplitudes is shown in Figure 1.5.

The Inversion Formula is an extreme version of such a superposition. It says that any function f (so long as f and $\widehat{f}$ are integrable) can be represented as an integral (in effect, a continuous sum) of pure tones $\widehat{f}(\xi)\, e^{2\pi i\xi x}$ over all possible frequencies $\xi \in \mathbb{R}$. By superimposing all the pure tones with the correct amplitudes, we create any sound that we like. The pure tones are our simple "building blocks," and by combining them we can create any sound (or signal, or function). Of course, the "superposition" is an integral, not a finite sum, but still we are combining our very simple special functions e_ξ to create very complicated functions f via the Inversion Formula.

Once we have a representation of f in terms of our pure tones, we can act on it. For example, assume that we measure time in seconds, in which case frequency is usually called hertz. If we don't like the annoying buzz in our signal f that is due to our 60 hertz overhead fluorescent lights, we might decide to modify f by creating a new function h whose Fourier transform is identical to $\widehat{f}$ except that $\widehat{h}(60) = 0$ (and most likely with some corresponding smooth modifications of the frequencies close to $\xi = 60$ as well). Once we know what we want $\widehat{h}$ to be, the function h that does this is given by the Inversion Formula as $h(x) = \int \widehat{h}(\xi)\, e^{2\pi i \xi x}\, d\xi$. In engineering jargon, we *filter* f to obtain h. We will see later that h can be obtained from f through the operation of convolution (see Section 1.3.4).

In light of the Inversion Formula, we make the following definition.

Definition 1.10 (Inverse Fourier Transform). The *inverse Fourier transform* of $f \in L^1(\mathbb{R})$ is

$$\check{f}(\xi) \;=\; \int f(x)\, e^{2\pi i \xi x}\, dx.$$

When we wish to denote the inverse Fourier transform as an operator acting on f, we write $\mathcal{F}^{-1}(f) = \check{f}$. $\diamondsuit$

Using this notation, the Inversion Formula says that if $f, \widehat{f} \in L^1(\mathbb{R})$ then $f = \big(\widehat{f}\big)^\vee$, and likewise $f = \big(\check{f}\big)^\wedge$ will hold under the same hypotheses.

Remark 1.11. If $f \in L^1(\mathbb{R})$ then $\hat{f}(\xi) = \check{f}(-\xi)$. Therefore, every result that we prove about the Fourier transform has an analogue for the inverse Fourier transform, simply by making a change of variables. We usually only state results for the Fourier transform, but it is a good idea for the reader to work out the corresponding formulas for the inverse Fourier transform (usually, a sign simply has to be moved from one place to another). Note, for example, that if $f, \widehat{f} \in L^1(\mathbb{R})$, then the Inversion Formula tells us that

$$f^{\wedge\wedge}(\xi) \;=\; \big(\widehat{f}\big)^\vee(-\xi) \;=\; f(-\xi). \qquad \diamondsuit$$

The musical discussion above may explain some terminology. In our example, the function $f(x)$ represented a displacement (of a string or speaker) that changed with time. Time is represented by the variable x, and thus we often speak of x as the *time variable*, and we say that values $f(x)$ describe the function f *in the time domain*. On the other hand, $\widehat{f}(\xi)$ represents the "amount" of frequency ξ present in f, and therefore we often refer to ξ as the *frequency variable*, and say that values $\widehat{f}(\xi)$ describe the function f *in the* *frequency domain.* We use this terminology even if x represents something other than time and ξ something other than frequency. For example, we often use this same terminology when we move to higher dimensions, i.e., even if f is defined on $\mathbb{R}^2$ instead of $\mathbb{R}$, we might refer to $\mathbb{R}^2$ as the "time domain." Or, if x is representing a spatial quantity, we may refer to $\mathbb{R}^2$ as the "spatial domain." There is no single perfect terminology since there are so many different contexts in which the Fourier transform can arise.

Additional Problems

1.1. Show that the Fourier transform of the *one-sided exponential* $f(x) = e^{-x}\,\chi_{[0,\infty)}(x)$ is

$$\widehat{f}(\xi) \;=\; \frac{1}{2\pi i\xi + 1}, \qquad \xi \in \mathbb{R},$$

and the Fourier transform of the *two-sided exponential* $g(x) = e^{-|x|}$ is

$$\widehat{g}(\xi) \;=\; \frac{2}{4\pi^2\xi^2 + 1}, \qquad \xi \in \mathbb{R}.$$

1.2. Prove that if $f \in L^1(\mathbb{R})$ is even, then $\widehat{f}$ is even, and if $f \in L^1(\mathbb{R})$ is odd, then $\widehat{f}$ is odd. The converse is also true, but is not as easy to prove, see Problem 1.46.

1.3. Prove that if $f \in L^1(\mathbb{R})$ is real-valued, then $\overline{\widehat{f}(\xi)} = \widehat{f}(-\xi)$. Conclude that if f is both real and even, then $\widehat{f}$ is both real and even as well.

1.4. Show that if $f \in L^1(\mathbb{R})$ is nonnegative almost everywhere (and is not the zero function), then $|\widehat{f}(\xi)| < \widehat{f}(0)$ for all $\xi \neq 0$.

1.5. (a) The *Gamma function* for complex numbers z satisfying $\mathrm{Re}(z) > 0$ is

$$\Gamma(z) \;=\; \int_0^\infty t^{z-1}\, e^{-t}\, dt.$$

Show that this is well-defined, i.e., $t^{z-1}\, e^{-t} \in L^1(0,\infty)$ whenever $\mathrm{Re}(z) > 0$.

Remark: The Gamma function is analytic on $\mathrm{Re}(z) > 0$, and has an analytic continuation to $\mathbb{C} \setminus \{0, -1, -2, \dots\}$. Also, $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$.

(b) Show that $f(x) = e^{-e^x} e^x \in L^1(\mathbb{R})$, and $\widehat{f}(\xi) = \Gamma(1 - 2\pi i\xi)$.

Remark: It can be shown that $\Gamma(z) \neq 0$ for every z where it is defined, so for this f we have $\widehat{f}(\xi) \neq 0$ for every $\xi \in \mathbb{R}$.

1.6. The *Riemann zeta function* for complex numbers s with $\mathrm{Re}(s) > 1$ is

$$\zeta(s) \;=\; \sum_{n=1}^\infty \frac{1}{n^s}.$$

The Riemann zeta function is analytic on $\mathrm{Re}(s) > 1$, and has an analytic continuation to $\mathbb{C} \setminus \{1\}$. The *Riemann hypothesis*, whose validity is one of the great open problems in mathematics, states that if $\zeta(s) = 0$ and $\mathrm{Re}(s) > 0$ then $\mathrm{Re}(s) = 1/2$.

(a) Show that

$$(1 - 2^{1-s})\,\zeta(s) \;=\; \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^s}, \qquad \mathrm{Re}(s) > 1. \tag{1.3}$$

(b) The right-hand side of equation (1.3) is an example of a *Dirichlet series*, and it converges and defines an analytic function on $\mathrm{Re}(s) > 0$ (see Theorem F.6). The left-hand side of equation (1.3) is analytic on $\mathrm{Re}(s) > 0$ except for $s = 1$, and can be defined so that it is analytic at $s = 1$ as well. Since both sides are analytic on $\mathrm{Re}(s) > 0$ and are equal on $\mathrm{Re}(s) > 1$, it follows from the properties of analytic functions that equation (1.3) holds for all s with $\mathrm{Re}(s) > 0$. For $s = 1$ we need to define the left-hand side to take the value $\sum \frac{(-1)^{n+1}}{n} = \ln 2$. Assuming these facts, show that

$$\Gamma(s)\,(1 - 2^{1-s})\,\zeta(s) \;=\; \sum_{n=1}^{\infty} (-1)^{n+1} \int_0^{\infty} x^{s-1}\, e^{-nx}\, dx, \quad \mathrm{Re}(s) > 0. \tag{1.4}$$

It is helpful to note that, by a change of variables, $\Gamma(s) = n^s \int_0^\infty x^{s-1}\, e^{-nx}\, dx$ for $\mathrm{Re}(s) > 0$.

(c) Justify interchanging the summation and integral in equation (1.4) to obtain

$$\Gamma(s)\,(1 - 2^{1-s})\,\zeta(s) \;=\; \int_0^{\infty} x^{s-1}\, \frac{1}{e^x + 1}\, dx, \qquad \mathrm{Re}(s) > 0.$$

(d) Given $t > 0$, define

$$f_t(x) \;=\; \frac{e^{tx}}{e^{e^x} + 1}.$$

Show that $f_t \in L^1(\mathbb{R})$. Given $\xi \in \mathbb{R}$, let $s = t - 2\pi i\xi$ and show that

$$\widehat{f_t}(\xi) \;=\; \Gamma(s)\,(1 - 2^{1-s})\,\zeta(s).$$

1.2 Translation, Modulation, Dilation, and Involution

A common theme of many of the issues that we will consider in this chapter concerns the dualities that occur between properties of f and those of $\widehat{f}$. In this section we examine the dualities that occur among several basic operators under the Fourier transform.

1.2.1 Four Fundamental Operators

Here are four operators that will pervade our study of the Fourier transform.

Definition 1.12. We define the following operators on functions $f : \mathbb{R} \to \mathbb{C}$.

$$\begin{array}{llcll} \text{Translation:} & (T_a f)(x) & = & f(x-a), & a \in \mathbb{R}. \\ \text{Modulation:} & (M_\theta f)(x) & = & e^{2\pi i\theta x} f(x), & \theta \in \mathbb{R}. \\ \text{Dilation:} & f_\lambda(x) & = & \lambda f(\lambda x), & \lambda > 0. \\ \text{Involution:} & \widetilde{f}(x) & = & \overline{f(-x)}. & \diamondsuit \end{array}$$

The scaling factor in the definition of dilation has been chosen so that dilation preserves the L^1-norm of a function. Translation, modulation, dilation, and involution are all isometries mapping $L^1(\mathbb{R})$ onto itself.

Exercise 1.13. Prove the following algebraic properties of the Fourier transform of $f \in L^1(\mathbb{R})$.

(a) $(T_a f)^\wedge(\xi) = (M_{-a}\widehat{f}\,)(\xi) = e^{-2\pi i a\xi}\widehat{f}(\xi)$, for $a \in \mathbb{R}$.

(b) $(M_\eta f)^\wedge(\xi) = (T_\eta \widehat{f}\,)(\xi) = \widehat{f}(\xi - \eta)$, for $\eta \in \mathbb{R}$.

(c) $(f_\lambda)^\wedge(\xi) = \lambda\,(\widehat{f}\,)_{1/\lambda}(\xi) = \widehat{f}(\xi/\lambda)$, for $\lambda > 0$.

(d) $(\bar{f}\,)^\wedge(\xi) = (\widehat{f}\,)^\sim(\xi) = \overline{\widehat{f}(-\xi)}$.

(e) $(\widetilde{f}\,)^\wedge(\xi) = \overline{\widehat{f}(\xi)}$.

Also derive analogous formulas relating the inverse Fourier transform to translation, modulation, dilation, and involution. $\diamondsuit$

In this sense, T_a is dual to M_{-a} under the Fourier transform, and likewise M_η is dual to T_η. Additionally, except for the normalizing scaling factor (which will be very important to us in Section 1.5!), dilation by λ is dual to dilation by $1/\lambda$ under the Fourier transform.

In order to derive some of the properties of the family of translation operators, it is helpful to know that $C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ when p is finite. We will prove this using standard real analysis techniques. By making use of convolutions, we will greatly refine this result in Section 1.5. For example, we will see that the seemingly "tiny" space $C_c^\infty(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for each $1 \le p < \infty$, and is dense in $C_0(\mathbb{R})$ with respect to the L^∞-norm.

The tool that we need for this proof is *Urysohn's Lemma*, a general topological result which states that if A and B are disjoint closed subsets of a normal topological space X then there exists a continuous function $f : X \to [0,1]$ that is identically 0 on A and identically 1 on B. We will prove Urysohn's Lemma for subsets of $\mathbb{R}^d$ (although the same simple proof can be used in any metric space). The key is the following lemma.

Lemma 1.14. *If $E \subseteq \mathbb{R}^d$ is nonempty, then*

$$f(x) \;=\; \operatorname{dist}(x, E) \;=\; \inf\bigl\{|x - z| : z \in E\bigr\}$$

is uniformly continuous on $\mathbb{R}^d$.

Proof. Fix $\varepsilon > 0$. Choose any $x, y \in \mathbb{R}^d$ with $|x-y| < \varepsilon/2$. By definition, there exist $a, b \in E$ such that $|x-a| < \operatorname{dist}(x,E)+\varepsilon/2$ and $|y-b| < \operatorname{dist}(y,E)+\varepsilon/2$. Hence

$$\begin{aligned} f(y) \;=\; \operatorname{dist}(y, A) \;&\le\; |y - a| \\ &\le\; |y - x| + |x - a| \\ &<\; \frac{\varepsilon}{2} + \operatorname{dist}(x, E) + \frac{\varepsilon}{2} \\ &=\; f(x) + \varepsilon. \end{aligned}$$

Similarly $f(x) < f(y) + \varepsilon$, so $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \varepsilon/2$. □

Theorem 1.15 (Urysohn's Lemma). *If E, F are disjoint closed subsets of $\mathbb{R}^d$, then there exists a continuous function $\theta\colon \mathbb{R}^d \to \mathbb{R}$ such that*

(a) $0 \le \theta \le 1$,

(b) $\theta = 0$ *on E, and*

(c) $\theta = 1$ *on F.*

Proof. Because E is closed, if $x \notin E$ then $\operatorname{dist}(x,E) > 0$. Also, by Lemma 1.14, $\operatorname{dist}(x,E)$ and $\operatorname{dist}(x,F)$ are each continuous functions of x. Therefore the function

$$\theta(x) \;=\; \frac{\operatorname{dist}(x, E)}{\operatorname{dist}(x, E) + \operatorname{dist}(x, F)}$$

has the required properties. □

Theorem 1.16. *$C_c(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d)$ for each $1 \le p < \infty$.*

Proof. First consider the function $f = \chi_E$ where $E \subseteq \mathbb{R}^d$ is bounded. If we fix $\varepsilon > 0$, then there exists a bounded open set $U \supseteq E$ such that $|U \backslash E| < \varepsilon$ and a compact set $K \subseteq E$ such that $|E \backslash K| < \varepsilon$. By Urysohn's Lemma (Theorem 1.15), we can find a continuous function $\theta\colon \mathbb{R}^d \to \mathbb{R}$ such that $0 \le \theta \le 1$, $\theta = 1$ on K, and $\theta = 0$ on $\mathbb{R}^d \backslash U$. Then $\theta \in C_c(\mathbb{R}^d)$, and we have

$$\|\chi_E - \theta\|_p^p \;=\; \int |\chi_E - \theta|^p \;=\; \int_{U\backslash K} |\chi_E - \theta|^p \;\le\; |U\backslash K| \;<\; 2\varepsilon.$$

Hence χ_E can be approximated arbitrarily closely in L^p-norm by elements of $C_c(\mathbb{R}^d)$. By forming finite linear combinations, it follows that every simple function in $L^p(\mathbb{R}^d)$ that has compact support can be approximated as well as we like by functions in $C_c(\mathbb{R}^d)$. Since the set of compactly supported simple functions is dense in $L^p(\mathbb{R}^d)$, we conclude that $C_c(\mathbb{R}^d)$ is dense as well. □

Now we establish some properties of the translation operator T_a. The "easy" way to prove part (c) of the next exercise is to first prove that it holds for functions in $C_c(\mathbb{R})$, and then use the fact that $C_c(\mathbb{R})$ is dense to extend to $L^p(\mathbb{R})$.

Exercise 1.17. (a) Prove that if $f \in C_0(\mathbb{R})$ then f is uniformly continuous, and show that this is equivalent to the statement

$$\lim_{a \to 0} \|T_a f - f\|_\infty \;=\; 0. \tag{1.5}$$

(b) Show that (1.5) can fail if we only assume that $f \in C_b(\mathbb{R})$.

(c) Show that if $1 \le p < \infty$ and $f \in L^p(\mathbb{R})$, then

$$\lim_{a \to 0} \|T_a f - f\|_p \;=\; 0. \qquad \diamondsuit$$

The function $\omega_p(a) = \|T_a f - f\|_p$ is often called the L^p*-modulus of continuity* of f.

Remark 1.18. If we consider the linear mapping $\tau\colon \mathbb{R} \to \mathcal{B}(L^p(\mathbb{R}))$ defined by $\tau(a) \;=\; T_a$, then Exercise 1.17 implies that for each $1 \le p < \infty$ we have

$$\forall\, f \in L^p(\mathbb{R}), \quad \lim_{t \to s} \|\tau(t)f - \tau(s)f\|_p \;=\; 0.$$

In another terminology, this says that $\tau(t) \to \tau(s)$ in the strong operator topology as $t \to s$. For $1 \le p < \infty$, we therefore say that $\{T_a\}_{a \in \mathbb{R}}$ is a *strongly continuous* one-parameter family of operators on $L^p(\mathbb{R})$. The same is true of $\{T_a\}_{a \in \mathbb{R}}$ on $C_0(\mathbb{R})$ when $p = \infty$. $\diamondsuit$

Exercise 1.19. Prove that the family $\{M_\eta\}_{\eta \in \mathbb{R}}$ of modulation operators is a strongly continuous family of operators on $L^p(\mathbb{R})$ for $1 \le p < \infty$. For $p = \infty$, show that $\{M_\eta\}_{\eta \in \mathbb{R}}$ is a strongly continuous family on $C_0(\mathbb{R})$, but not on $L^\infty(\mathbb{R})$. $\diamondsuit$

1.2.2 The Riemann–Lebesgue Lemma

We now use the strong continuity of the family of translation operators to prove that if $f \in L^1(\mathbb{R})$, then not only is $\widehat{f}$ continuous, but we also have decay of $\widehat{f}$ at infinity.

Theorem 1.20 (Riemann–Lebesgue Lemma). *If $f \in L^1(\mathbb{R})$, then $\widehat{f} \in C_0(\mathbb{R})$.*

Proof. Choose any $f \in L^1(\mathbb{R})$. We already know that $\widehat{f}$ is continuous, so we just have to show that it decays at infinity. Since $e^{-\pi i} = -1$, we have for $\xi \neq 0$ that

$$\begin{aligned}
\widehat{f}(\xi) &= \int f(x)\, e^{-2\pi i \xi x}\, dx && (1.6)\\
&= -\int f(x)\, e^{-2\pi i \xi x}\, e^{-2\pi i \xi (\frac{1}{2\xi})}\, dx\\
&= -\int f(x)\, e^{-2\pi i \xi (x+\frac{1}{2\xi})}\, dx\\
&= -\int f\Big(x-\frac{1}{2\xi}\Big)\, e^{-2\pi i \xi x}\, dx. && (1.7)
\end{aligned}$$

Averaging equalities (1.6) and (1.7), we obtain

$$\widehat{f}(\xi) = \frac{1}{2}\int \Big(f(x) - f\Big(x-\frac{1}{2\xi}\Big)\Big)\, e^{-2\pi i \xi x}\, dx.$$

The fact that translation is strongly continuous on $L^1(\mathbb{R})$ therefore implies that

$$|\widehat{f}(\xi)| \le \frac{1}{2}\int \Big|f(x) - f\Big(x-\frac{1}{2\xi}\Big)\Big|\, dx = \frac{1}{2}\, \|f - T_{\frac{1}{2\xi}} f\|_1 \to 0$$

as $|\xi| \to \infty$. □

While certainly elegant, and yielding the nice L^1-modulus of continuity estimate

$$|\widehat{f}(\xi)| \le \frac{1}{2}\, w_1\Big(\frac{1}{2\xi}\Big),$$

the preceding proof does have a less-than-satisfying "magical" feel to it. A different proof of the Riemann–Lebesgue Lemma is given in Problem 1.8, and a third proof appears in Section 1.4.

1.2.3 Position and Momentum

Two addition operators that play central roles in harmonic analysis are the mathematical versions of the position and momentum operators from quantum mechanics. These are defined as follows.

Definition 1.21 (Position and Momentum). The *position and momentum operators* P and M are

$$Pf(x) = xf(x), \qquad \text{and} \qquad Mf = \frac{1}{2\pi i}\, f'. \qquad \diamondsuit \tag{1.8}$$

The position operator is defined on all functions $f\colon \mathbb{R} \to \mathbb{C}$. The momentum operator is defined on all differentiable functions f, although often we only require that Mf be defined almost everywhere and hence only need to assume that f is differentiable at almost every point.

Unlike the translation, modulation, dilation, and involution operators, the position and momentum operators do not map $L^p(\mathbb{T})$ boundedly into itself, even if we restrict to domains where they are well-defined (see Problem 1.9). Even so, Problems 1.11 and 1.12 show that position and momentum are fundamentally tied to modulation and translation.

1.2.4 The HRT Conjecture

We close or discussion of translation and modulation by briefly discussing one of our favorite open mathematical problems.

Conjecture 1.22 (The HRT Conjecture). *If $g \in L^2(\mathbb{R})$ is not the zero function and $\Lambda = \{(p_k, q_k)\}_{k=1}^N$ is any set of finitely many distinct points in $\mathbb{R}^2$, then*

$$\mathcal{G}(g, \Lambda) \;=\; \bigl\{M_{q_k} T_{p_k} g\bigr\}_{k=1}^N$$

is a linearly independent set of functions in $L^2(\mathbb{R})$. $\diamondsuit$

Conjecture 1.22 was first made in [HRT96]. While many partial results related to this conjecture are known, as of the time of writing it is not known whether Conjecture 1.22 holds in the generality stated. For more details and background on this conjecture, we refer to the survey paper [Heil06] or Section 11.9 in the author's text [Heil11a].

Additional Problems

1.7. Let $f \colon \mathbb{R} \to \mathbb{C}$ be Lebesgue measurable. Prove that $T_a f \xrightarrow{\mathrm{m}} f$ (convergence in measure) as $a \to 0$ on any compact set, i.e.,

$$\forall \text{ compact } K \subseteq \mathbb{R}, \quad \forall\, \varepsilon > 0, \quad \lim_{a \to 0} \bigl|\{x \in K : |f(x) - T_a f(x)| > \varepsilon\}\bigr| \;=\; 0.$$

1.8. This problem provides an alternative proof to Theorem 1.20.

(a) Show that $\widehat{f} \in C_0(\mathbb{R})$ for every $f \in S = \mathrm{span}\{\chi_{[a,b]} : a < b \in \mathbb{R}\}$.

(b) Show that S is dense in $L^1(\mathbb{R})$ and use this to prove that $\widehat{f} \in C_0(\mathbb{R})$ for every $f \in L^1(\mathbb{R})$.

1.9. Let P, M be the position and momentum operators defined in equation (1.8), and fix $1 \le p < \infty$. These operators are not defined on all of $L^p(\mathbb{R})$. Instead, define domains

$$\begin{aligned} D_P \;&=\; \{f \in L^p(\mathbb{R}) : x f(x) \in L^p(\mathbb{R})\}, \\ D_M \;&=\; \{f \in L^p(\mathbb{R}) : f \text{ is differentiable and } f' \in L^p(\mathbb{R})\}, \end{aligned}$$

which are dense subspaces of $L^p(\mathbb{R})$. Restricted to these domains, P maps D_P into $L^p(\mathbb{R})$ and M maps D_M into $L^p(\mathbb{R})$. Show that P and M are unbounded even when restricted to these domains, i.e.,

$$\sup_{\substack{f \in D_P,\\ \|f\|_p = 1}} \|Pf\|_p \;=\; \infty \;=\; \sup_{\substack{f \in D_M,\\ \|f\|_p = 1}} \|Mf\|_p.$$

1.10. Let X be a Banach space, and fix $A \in \mathcal{B}(X)$. For $n > 0$ let A^n denote the usual nth power of A ($A^n = A \cdots A$, n times), and define $A^0 = I$ (the identity map on X).

(a) Given $x \in X$, show that the series $e^A(x) = \sum_{k=0}^{\infty} \frac{A^k x}{k!}$ converges absolutely in X, and show that e^A is a linear operator on X.

(b) Prove that the series $\sum_{k=0}^{\infty} \frac{A^k}{k!}$ converges absolutely in $\mathcal{B}(X)$, and equals the operator e^A defined in part (a). Conclude that $e^A \in \mathcal{B}(X)$ and $\|e^A\| \le e^{\|A\|}$.

(c) Prove that if $A, B \in \mathcal{B}(X)$ and $AB = BA$, then $e^A e^B = e^{A+B} = e^B e^A$.

(d) Let H be a Hilbert space. Show that if $A \in \mathcal{B}(H)$ is self-adjoint, then e^{iA} is unitary.

1.11. Show that for an appropriate dense subset of functions $f \in L^1(\mathbb{R})$ we have

$$M_\xi f \;=\; e^{2\pi i \xi P} f \;=\; \sum_{k=0}^{\infty} \frac{(2\pi i \xi P)^k f}{k!},$$

and similarly there is a set of $f \in L^1(\mathbb{R})$ such that

$$T_a f \;=\; e^{-2\pi i a M} f \;=\; \sum_{k=0}^{\infty} \frac{(-2\pi i a M)^k f}{k!},$$

where these series converge absolutely both in the pointwise sense and in L^1-norm. In contrast to Problem 1.10, note that the operators P and M are unbounded on $L^1(\mathbb{R})$.

1.12. Show that for an appropriate dense subset functions $f \in L^1(\mathbb{R})$ we have

$$2\pi i M f \;=\; \lim_{a \to 0} \frac{f - T_a f}{a} \quad \text{and} \quad -2\pi i P f \;=\; \lim_{\xi \to 0} \frac{f - M_\xi f}{\xi}, \tag{1.9}$$

where these limits converge both in the pointwise sense and in L^1-norm.

Remark: Equation (1.9) says that $2\pi i M$ is the *infinitesimal generator* of the strongly continuous family of operators $\{T_a\}_{a \in \mathbb{R}}$, and $-2\pi i P$ is the infinitesimal generator of the family $\{M_\xi\}_{\xi \in \mathbb{R}}$.

1.3 Convolution

Since $L^1(\mathbb{R})$ is a Banach space, we know that it has many useful properties. In particular the operations of addition and scalar multiplication are continuous. However, there are many other operations on $L^1(\mathbb{R})$ that we could consider. One natural operation is multiplication of functions, but unfortunately $L^1(\mathbb{R})$ is not closed under pointwise multiplication.

Exercise 1.23. Show that $f, g \in L^1(\mathbb{R})$ does not imply $fg \in L^1(\mathbb{R})$. $\diamondsuit$

In this section we will define a different "multiplication-like" operation under which $L^1(\mathbb{R})$ is closed. This operation, convolution of functions, will be one of the most important tools in our further development of harmonic analysis. Therefore, in this section we set aside the Fourier transform for the moment, and concentrate on developing the machinery of convolution.

1.3.1 Some Notational Conventions

Before proceeding, there are some technical issues related to the definition of elements of $L^p(\mathbb{R})$ that we need to clarify.

The basic source of difficulty is that an element f of $L^p(\mathbb{R})$ is not a function but rather denotes an equivalence class of functions that are equal almost everywhere. Therefore we cannot speak of the "value of $f \in L^p(\mathbb{R})$ at a point $x \in \mathbb{R}$," and consequently concepts such as continuity or support do not apply in a literal sense to elements of $L^p(\mathbb{R})$. For example, the zero function 0 and the function $\chi_{\mathbb{Q}}$ both belong to the zero element of $L^p(\mathbb{R})$, which is the equivalence class of functions that are zero a.e., yet 0 is continuous and compactly supported while $\chi_{\mathbb{Q}}$ is discontinuous and its support is $\mathbb{R}$. Even so, it is often essential to consider smoothness or support properties of functions, and we therefore adopt the following conventions when discussing the smoothness or support of elements of $L^p(\mathbb{R})$. More generally, these same issues and conventions apply to elements of

$$L^1_{\mathrm{loc}}(\mathbb{R}) \;=\; \bigl\{f\colon \mathbb{R} \to \mathbb{C} : f \cdot \chi_K \in L^1(\mathbb{R}) \text{ for every compact } K \subseteq \mathbb{R}\bigr\},$$

which is the *space of locally integrable functions* on $\mathbb{R}$. Note that $L^p(\mathbb{R}) \subseteq L^1_{\mathrm{loc}}(\mathbb{R})$ for every $1 \le p \le \infty$.

Notation 1.24 (Continuity for Elements of $L^1_{\mathrm{loc}}(\mathbb{R})$). We will say that $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ is continuous if there is a representative of f that is continuous, i.e., there exists some continuous function f_0 such that f is the equivalence class of all functions that equal f_0 almost everywhere.

Conversely, if g is a continuous function such that $\int_K |g(x)|\,dx < \infty$ for every compact $K \subseteq \mathbb{R}$, then we write $g \in L^1_{\mathrm{loc}}(\mathbb{R})$ with understanding that this means that the equivalence class of functions that equal g a.e. is an element of $L^1_{\mathrm{loc}}(\mathbb{R})$. In this sense we write statements such as $C_c(\mathbb{R}) \subseteq L^p(\mathbb{R})$ even though $C_c(\mathbb{R})$ is a set of functions while $L^p(\mathbb{R})$ is a set of equivalence classes of functions. $\diamondsuit$

Notation 1.25 (Support of Elements of $L^1_{\mathrm{loc}}(\mathbb{R})$). We will say that $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ has compact support if there is a representative of f that has compact support. Thus f has compact support if there exists an $N > 0$ such that $f(x) = 0$ for a.e. $|x| > N$.

In many situations, this definition of compact support is all that we need, but in some circumstances it is important to discuss the support of $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ explicitly. We define the support of $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ to be

$$\operatorname{supp}(f) \;=\; \bigcap \bigl\{F \subseteq \mathbb{R} : F \text{ is closed and } f(x) = 0 \text{ for a.e. } x \notin F\bigr\}.$$

In particular, if F is a closed subset of $\mathbb{R}$, then

$$\operatorname{supp}(f) \subseteq F \iff f(x) = 0 \text{ for a.e. } x \notin F.$$

However, if T is a generic subset of $\mathbb{R}$, then the statements $\operatorname{supp}(f) \subseteq \mathrm{T}$ and $f(x) = 0$ for a.e. $x \notin \mathrm{T}$ need not be equivalent.

In the language of Chapter 4, we are taking the support of $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ to be the support of the distribution in $\mathcal{D}'(\mathbb{R})$ that is determined by f, see Section 4.6. $\diamondsuit$

The reader should verify that if $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ is continuous (in the sense given in Notation 1.24), then the support of f in the sense of Notation 1.25 coincides with the usual definition of the support of f as the closure in $\mathbb{R}$ of the set $\{x \in \mathbb{R} : f(x) \neq 0\}$.

1.3.2 Definition and Basic Properties of Convolution

Now we can define convolution of functions.

Definition 1.26 (Convolution). Let $f\colon \mathbb{R} \to \mathbb{C}$ and $g\colon \mathbb{R} \to \mathbb{C}$ be Lebesgue measurable functions. Then the *convolution of f with g* is the function $f * g$ given by

$$(f * g)(x) \;=\; \int f(y)\, g(x-y)\, dy, \tag{1.10}$$

whenever this integral is well-defined. $\diamondsuit$

For example, suppose that $1 \le p \le \infty$ and let p' be the dual index to p. If $f \in L^p(\mathbb{R})$ and $g \in L^{p'}(\mathbb{R})$, then (as functions of y), $f(y)$ and $g(x-y)$ belong to dual spaces, and hence by Hölder's Inequality the integral defining $(f * g)(x)$ in equation (1.10) exists for every x, and furthermore is bounded as a function of x.

Exercise 1.27. Show that if $1 \le p \le \infty$, $f \in L^p(\mathbb{R})$, and $g \in L^{p'}(\mathbb{R})$, then $f * g \in L^\infty(\mathbb{R})$, and we have

$$\|f * g\|_\infty \le \|f\|_p\, \|g\|_{p'}. \tag{1.11}$$

We will improve on this exercise (in several ways) below. In particular, Exercise 1.27 does not give the only hypotheses on f and g which imply that $f * g$ exists—we will shortly see *Young's Inequality*, which is a powerful result that tells us that $f * g$ will belong to a particular Lebesgue space $L^r(\mathbb{R})$

whenever $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$, and we have the proper relationship among p, q, and r (specifically, $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$). Before turning to that general case, we prove the fundamental fact that $L^1(\mathbb{R})$ is closed under convolution, and that the Fourier transform interchanges convolution with multiplication.

Theorem 1.28. *If $f, g \in L^1(\mathbb{R})$ are given, then the following statements hold.*

(a) *$f(y)\, g(x-y)$ is Lebesgue measurable on $\mathbb{R}^2$.*

(b) *For almost every $x \in \mathbb{R}$, $f(y)\, g(x-y)$ is a measurable and integrable function of y, and hence $(f*g)(x)$ is defined for a.e. $x \in \mathbb{R}$.*

(c) *$f*g \in L^1(\mathbb{R})$, and*

$$\|f*g\|_1 \;\le\; \|f\|_1\, \|g\|_1.$$

(d) *The Fourier transform of $f*g$ is the product of the Fourier transforms of f and g:*

$$(f*g)^\wedge(\xi) \;=\; \widehat{f}(\xi)\, \widehat{g}(\xi), \qquad \xi \in \mathbb{R}.$$

Proof. (a) If we set $h(x,y) = f(x)$, then

$$h^{-1}(a,\infty) \;=\; \{(x,y) : h(x,y) > a\} \;=\; \{(x,y) : f(x) > a\} \;=\; f^{-1}(a,\infty) \times \mathbb{R},$$

which is a measurable subset of $\mathbb{R}^2$ since $f^{-1}(a,\infty)$ and $\mathbb{R}$ are measurable subsets of $\mathbb{R}$. Likewise $k(x,y) = g(y)$ is measurable. Since the product of measurable functions is measurable, we conclude that $F(x,y) = f(x)g(y)$ is measurable. Further, $T(x,y) = (y, x-y)$ is a linear transformation, so $H(x,y) = (F \circ T)(x,y) = F(y, x-y) = f(y)\, g(x-y)$ is measurable.

(b) Using the same notation as in part (a), we have

$$\begin{aligned} \iint |H(x,y)|\, dx\, dy \;&=\; \int \biggl(\int |g(x-y)|\, dx \biggr) |f(y)|\, dy \\ &=\; \int \|g\|_1\, |f(y)|\, dy \;=\; \|g\|_1\, \|f\|_1. \end{aligned}$$

Therefore $H(x,y) = f(y)\, g(x-y) \in L^1(\mathbb{R}^2)$, so Fubini's Theorem implies that the function $(f*g)(x) = \int f(y)\, g(x-y)\, dy$ exists for almost every x and is an integrable function of x.

(c) Using part (b),

$$\|f*g\|_1 \;=\; \int |(f*g)(x)|\, dx \;\le\; \iint |f(y)\, g(x-y)|\, dy\, dx \;=\; \|f\|_1\, \|g\|_1.$$

(d) Fubini's Theorem (exercise: justify its use) allows us to interchange integrals in the following calculation:

$$\begin{aligned} (f*g)^\wedge(\xi) \;&=\; \int (f*g)(x)\, e^{-2\pi i \xi x}\, dx \\ &=\; \iint f(y)\, g(x-y)\, e^{-2\pi i \xi x}\, dy\, dx \end{aligned}$$

$$= \int f(y)\, e^{-2\pi i\xi y} \left(\int g(x-y)\, e^{-2\pi i\xi(x-y)}\, dx \right) dy$$
$$= \int f(y)\, e^{-2\pi i\xi y} \left(\int g(x)\, e^{-2\pi i\xi x}\, dx \right) dy$$
$$= \int f(y)\, e^{-2\pi i\xi y}\, \widehat{g}(\xi)\, dy$$
$$= \widehat{f}(\xi)\, \widehat{g}(\xi). \qquad \square$$

In the proof of Theorem 1.28, we carefully addressed the measurability of $f * g$. We will usually take issues of measurability for granted from now on, but it is a good idea for the reader to consider wherever appropriate why the measurability of the functions we encounter is ensured.

Exercise 1.29. Establish the following basic properties of convolution. Given f, g, $h \in L^1(\mathbb{R})$, prove the following.

(a) Commutativity: $f * g = g * f$.

(b) Associativity: $(f * g) * h = f * (g * h)$.

(c) Distributive laws: $f * (g + h) = f * g + f * h$.

(d) Commutativity with translations: $f * (T_a g) = (T_a f) * g = T_a(f * g)$ for $a \in \mathbb{R}$.

(e) Behavior under involution: $(f * g)^\sim = \widetilde{f} * \widetilde{g}$. $\diamondsuit$

1.3.3 Young's Inequality

As we have seen, $L^1(\mathbb{R})$ is closed under convolution, which we write in short as

$$L^1(\mathbb{R}) * L^1(\mathbb{R}) \;\subseteq\; L^1(\mathbb{R}).$$

It is not true that $L^p(\mathbb{R})$ is closed under convolution for $p > 1$. Instead we have the following fundamental result, known as *Young's Inequality* (although, since Young proved many inequalities, it may be advisable to refer to this as *Young's Convolution Inequality*).

Exercise 1.30 (Young's Inequality). Prove the following statements.

(a) If $1 \le p \le \infty$ then $L^p(\mathbb{R}) * L^1(\mathbb{R}) \subseteq L^p(\mathbb{R})$, and we have

$$\forall\, f \in L^p(\mathbb{R}), \quad \forall\, g \in L^1(\mathbb{R}), \quad \|f * g\|_p \;\le\; \|f\|_p\, \|g\|_1. \tag{1.12}$$

(b) If $1 \le p, q, r \le \infty$ and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$ then $L^p(\mathbb{R}) * L^q(\mathbb{R}) \subseteq L^r(\mathbb{R})$, and we have

$$\forall\, f \in L^p(\mathbb{R}), \quad \forall\, g \in L^q(\mathbb{R}), \quad \|f * g\|_r \;\le\; \|f\|_p\, \|g\|_q. \qquad \diamondsuit \tag{1.13}$$

Of course, statement (a) in Young's Inequality is a special case of statement (b), but it is so useful that it is worth stating separately. It is also instructive on first try to attempt to prove statement (a) rather than the more general statement (b) in order to see the appropriate technique needed. There are many ways to prove Young's Inequality, e.g., via Hölder's Inequality or Minkowski's Integral Inequality (which is stated in Problem 1.25).

Remark 1.31. If $q = p'$ (in which case $r = \infty$), then Young's Inequality tells us that the convolution of $f \in L^p(\mathbb{R})$ with $g \in L^{p'}(\mathbb{R})$ belongs to $L^\infty(\mathbb{R})$. In fact, it follows from our later Exercise 1.39 that $f*g$ is continuous in this case, and therefore $(f*g)(x)$ is defined for *every* x. However, for general values of p, q, r satisfying the hypotheses of Young's Inequality, we are only able to conclude that $f*g \in L^r(\mathbb{R})$, and hence we usually only have that $f*g$ is defined pointwise almost everywhere. $\diamondsuit$

The inequalities given in Exercise 1.30 are in the form that we will most often need in practice, but it is very interesting to note that the implicit constant 1 on the right-hand side of equation (1.13) is not the optimal constant in general. Instead, if we define the *Babenko–Beckner constant* A_p by

$$A_p \;=\; \left(\frac{p^{1/p}}{p'^{1/p'}}\right)^{1/2}, \tag{1.14}$$

where we take $A_1 = A_\infty = 1$, then the optimal version of equation (1.13) is

$$\forall\, f \in L^p(\mathbb{R}), \quad \forall\, g \in L^q(\mathbb{R}), \quad \|f*g\|_r \;\le\; (A_p A_q A_{r'})\, \|f\|_p\, \|g\|_p, \tag{1.15}$$

and the constant $A_p A_q A_{r'}$ is typically not 1. The proof that $A_p A_q A_{r'}$ is the best constant in equation (1.15) is due to Beckner [Bec75] and Brascamp and Lieb [BrL76]. The Babenko–Beckner constant will make an appearance again when we consider the Hausdorff–Young Theorem in Chapter 3 (see Theorem 3.23).

1.3.4 Convolution as Filtering; Lack of an Identity

Theorem 1.28 gives us another way to view filtering (which was discussed in Section 1.1.2). Given $f \in L^1(\mathbb{R})$, we filter f by modifying its frequency content. That is, we create a new function h from f whose Fourier transform is

$$\widehat{h}(\xi) \;=\; \widehat{f}(\xi)\,\widehat{g}(\xi).$$

The Fourier transform of the function g tells us how to modify the frequency content of f. Assuming that the Inversion Formula applies, we can recover h by the formula

$$h(x) \;=\; \int \widehat{f}(\xi)\,\widehat{g}(\xi)\,e^{2\pi i\xi x}\,d\xi,$$

which is a superposition of the "pure tones" $e^{2\pi i\xi x}$ with the modified amplitudes $\widehat{f}(\xi)\,\widehat{g}(\xi)$. Assuming that $g \in L^1(\mathbb{R})$, Theorem 1.28 tells us that we can also obtain h by convolution: we have $h = f * g$. *Filtering is convolution.*

Obviously, there are many details that we are glossing over by assuming all of the formulas are applicable. Some of these we will address later, e.g., is it true that a function $h \in L^1(\mathbb{R})$ is uniquely determined by its Fourier transform $\widehat{h}$? (Yes, we will show that $f \mapsto \widehat{f}$ is an injective map of $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$, see Theorem 1.92.) Others we will leave for a course on digital signal processing. (For example, how do results for $L^1(\mathbb{R})$ relate to the processing of real-life digital signals whose domain is $\{1, \dots, n\}$ instead of $\mathbb{R}$?) In any case, keeping our attention on the real line, let us ask one interesting question. If our goal is to filter f, then one of the possible filterings should be the identity operation, i.e., do nothing to the frequency content of f. Is there a $g \in L^1(\mathbb{R})$ such that $f \mapsto f * g$ is the identity operation on $L^1(\mathbb{R})$?

Exercise 1.32. Suppose that there existed a function $\delta \in L^1(\mathbb{R})$ such that

$$\forall\, f \in L^1(\mathbb{R}), \quad f * \delta \;=\; f.$$

Show that δ would satisfy $\widehat{\delta}(\xi) = 1$ for all ξ, which contradicts the Riemann–Lebesgue Lemma. $\diamondsuit$

Consequently, *there is no identity element for convolution in* $L^1(\mathbb{R})$. This is problematic, and we will see several alternative ways of addressing this problem. In Section 1.5, we will construct functions which "approximate" an identity for convolution as closely as we like, according to a variety of meanings of approximation. In Chapter 4, we will create a distribution (or "generalized function") δ that is not itself a function but instead acts on functions and is an identity with respect to convolution. In Chapter 5, we will see that this distribution δ can also be regarded as a bounded measure on the real line. Thus, while there is no *function* δ that is an identity for convolution, an identity δ does exist in several generalized senses.

1.3.5 Convolution as Averaging; Introduction to Approximate Identities

Convolution can also be regarded as a kind of weighted averaging operator. For example, consider

$$\chi_T \;=\; \frac{1}{2T}\chi_{[-T,T]}, \qquad T > 0.$$

Given $f \in L^1(\mathbb{R})$, we have that

$$(f * \chi_T)(x) \;=\; \int f(y)\, \chi_T(x-y)\, dy \;=\; \frac{1}{2T}\int_{x-T}^{x+T} f(y)\, dy \;=\; \mathrm{Avg}_T f(x),$$

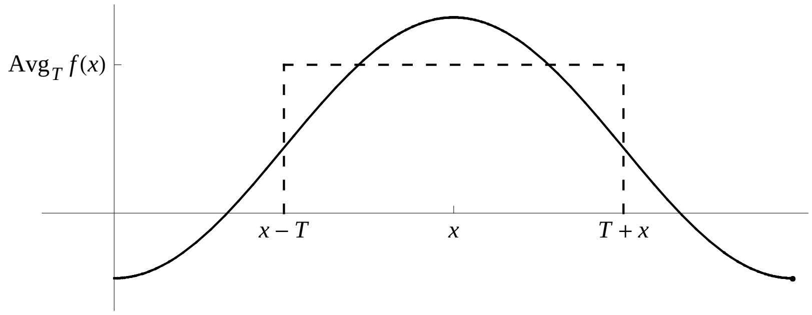

Fig. 1.6. The area of the dashed box equals $\int_{x-T}^{x+T} f(y)\,dy$, which is the area under the graph of f between $x - T$ and $x + T$.

where $\mathrm{Avg}_T f(x)$ is the average of f on the interval $[x - T, x + T]$ (see Figure 1.6).

For a general function g, the mapping $f \mapsto f * g$ can be regarded as a kind of weighted averaging of f, with g weighting some parts of the real line more than others. There is one technical point to observe in this viewpoint: while the function $\chi_T(x) = \frac{1}{2T}\chi_{[-T,T]}$ used in the preceding illustration is even, this will not be the case in general. Instead, in thinking of convolution as a weighted averaging, it is perhaps better to set $g^*(x) = g(-x)$ and write

$$(f * g)(x) \;=\; \int f(y)\, g^*(y - x)\, dy \;=\; \mathrm{Avg}_{g*} f(x),$$

the average of f around the point x corresponding to the weighting of the real line by $g^*(x) = g(-x)$. Alternatively, since convolution is commutative, we can equally view it as an averaging of g using the weighting corresponding to $f^*(x) = f(-x)$.

Looking ahead to Section 1.5, let us consider what happens to the convolution $f * \chi_T = \mathrm{Avg}_T f$ as $T \to 0$. The function $\chi_T = \frac{1}{2T}\chi_{[-T,T]}$ becomes a taller and taller "spike" centered at the origin, with the height of the spike being chosen so that the integral of χ_T is always 1. Intuitively, averaging over smaller and smaller intervals should give values $(f * \chi_T)(x)$ that are closer and closer to the original value $f(x)$. This intuition is made precise in Lebesgue's Differentiation Theorem (Theorem A.30), which implies that if $f \in L^1(\mathbb{R})$ then for almost every x (including all those in the Lebesgue set of f) we will have

$$f(x) \;=\; \lim_{T \to 0}\, (f * \chi_T)(x) \;=\; \lim_{T \to 0}\, \mathrm{Avg}_T f(x).$$

Thus $f \approx f * \chi_T$ when T is small. In this sense, while there is no identity element for convolution in $L^1(\mathbb{R})$, the function χ_T is *approximately* an identity for convolution, and the approximation becomes better and better the smaller T becomes.

Moreover, a similar phenomenon happens for the more general averaging operators $f \mapsto f * g = \mathrm{Avg}_{g*} f$. We can take any particular function $g \in L^1(\mathbb{R})$

and dilate it so that it becomes more and more compressed towards the origin, yet always keeping the total integral the same, by setting

$$g_\lambda(x) \;=\; \lambda g(\lambda x), \quad \lambda > 0,$$

as is done in Notation 1.5. Compressing g towards the origin corresponds to letting λ increase towards infinity (as opposed to $T \to 0$ in the discussion of χ_T above). Even if g is not compactly supported, it becomes more and more "spike-like" as λ increases (see the illustration in Figure 1.7).

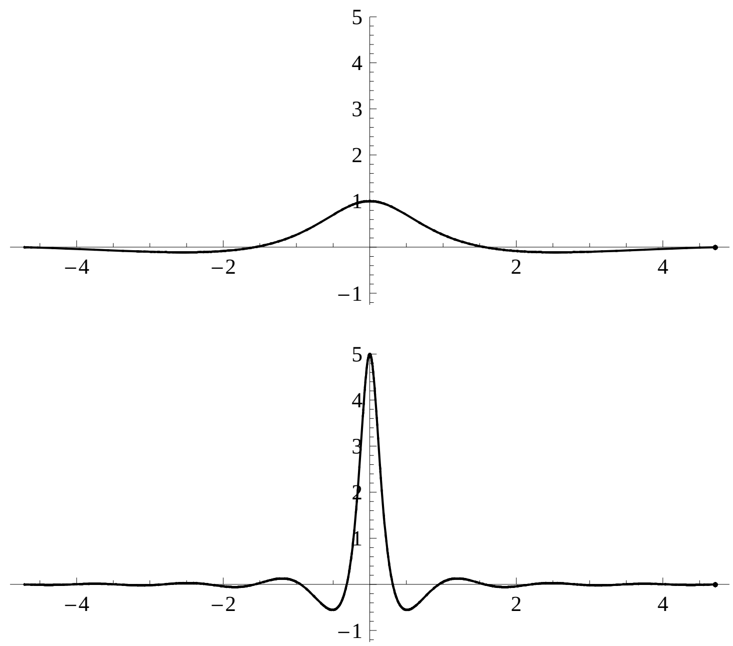

Fig. 1.7. Top: The function $g(x) = \cos(x)/(1 + x^2$. Bottom: The dilated function $g_5(x) = 5g(5x)$.

If it is the case that $\int g = 1$ (so $\int g_\lambda = 1$ also), then we will see in Section 1.5 that, for any $f \in L^1(\mathbb{R})$, the convolution $f * g_\lambda$ converges to f in L^1-norm (and possibly in other senses as well, depending on properties of f and g). The family $\{g_\lambda\}_{\lambda>0}$ is an example of what we will call an *approximate identity* in Section 1.5.

From this discussion we can see at least an intuitive reason why there is no identity function for convolution in $L^1(\mathbb{R})$. Consider the functions g_λ, each an integrable function with integral 1 that become more and more spike-like as λ increases. Suppose that we could let $\lambda \to \infty$ and obtain in the limit an integrable function δ that, like each function g_λ, has integral 1, but is indeed a spike supported entirely at the origin. Then we would hopefully have that $f * \delta = \lim_{\lambda\to\infty} f * g_\lambda = f$, and so δ would be an identity function for convolution. And indeed, it is not uncommon to see informal wording similar to the following.

> "Let δ be the function on $\mathbb{R}$ that has the property that $\delta(x) = 0$ for all $x \neq 0$ and $\int \delta(x)\,dx = 1$. Then
>
> $$\int f(y)\,\delta(x-y)\,dy \;=\; f(x).\text{"} \tag{1.16}$$

However, there is *no such function* δ. Any function that is zero for all $x \neq 0$ is zero almost everywhere, and hence is the zero element of $L^1(\mathbb{R})$. If $\delta(x) = 0$ for $x \neq 0$, then the Lebesgue integral of δ is $\int \delta(x)\,dx = 0$, not 1, *even if* we define $\delta(0) = \infty$. Thus $f * \delta = 0$, not f.

We cannot construct a *function* δ that has the property that $f*\delta = f$ for all $f \in L^1(\mathbb{R})$. However, we can construct families $\{g_\lambda\}_{\lambda>0}$ that have the property that $f * g_\lambda$ converges to f in various senses, and these are the approximate identities of Section 1.5. We can also construct objects that are not functions but which are identities for convolution—we will see the δ-*distribution* in Chapter 4 and the δ-*measure* in Chapter 5. In effect, the integral appearing in equation (1.16) is not a Lebesgue integral but rather is simply a shorthand for something else, namely, the action of the distribution or measure δ on the function f.

1.3.6 Convolution as an Inner Product

It is often useful to write a convolution in one of the following forms (the involution $\widetilde{g}(x) = \overline{g(-x)}$ was introduced in Definition 1.12):

$$\begin{aligned}
(f * g)(x) \;&=\; \int f(y)\,g(x-y)\,dy \\
&=\; \int f(y)\,\overline{\widetilde{g}(y-x)}\,dy \\
&=\; \int f(y)\,\overline{T_x\widetilde{g}(y)}\,dy \;=\; \langle f, T_x\widetilde{g}\rangle.
\end{aligned} \tag{1.17}$$

Thus we can view the convolution of f with g at the point x as the inner product of f with the function $\widetilde{g}$ translated by x.

Notation 1.33. In equation (1.17), we have used the notation $\langle\cdot,\cdot\rangle$, which in the context of functions usually denotes the inner product on $L^2(\mathbb{R})$. However, neither f nor $T_x\widetilde{g}$ need belong to $L^2(\mathbb{R})$, so we are certainly taking some poetic license in speaking of $\langle f, T_x\widetilde{g}\rangle$ as an inner product of f and $T_x\widetilde{g}$. We do this because in this volume we so often encounter integrals of the form $\int f(x)\,\overline{g(x)}\,dx$ and direct generalizations of these integrals that it is extremely convenient for us to retain the notation $\langle f, g\rangle$ for such an integral whenever it makes sense. Specifically, if f and g are any measurable functions on $\mathbb{R}$, then we will write

$$\langle f, g\rangle \;=\; \int f(x)\,\overline{g(x)}\,dx$$

whenever this integral exists. In another language, $\langle \cdot, \cdot\rangle$ is a *sesquilinear form* (linear in the first variable, antilinear in the second) that extends the inner product on $L^2(\mathbb{R})$. Although it is technically an abuse of terminology, we will often refer to $\langle f, g\rangle$ as the *inner product of f with g* even when f and g are not in $L^2(\mathbb{R})$ or another Hilbert space. $\diamondsuit$

In this volume we will encounter sesquilinear forms much more often than bilinear forms. A sesquilinear form is a function of two variables that is linear in the first variable but antilinear in the second, while a bilinear form is linear in both variables. The prefix "sesqui-" means "one and a half."

Note that in the calculation in equation (1.17), all that we know is that f and $T_x\widetilde{g}$ each belong to $L^1(\mathbb{R})$. Since the product of L^1 functions does not belong to L^1 in general, the integral appearing in equation (1.17) is not going to exist for every f, g, and x. Yet Theorem 1.28 implies the following unexpected fact.

Exercise 1.34. Show that

$$f, g \in L^1(\mathbb{R}) \implies f \cdot \overline{T_x\widetilde{g}} \in L^1(\mathbb{R}) \text{ for a.e. } x. \qquad \diamondsuit$$

Thus (thanks to Fubini and his theorem), even if we only assume that f and g are integrable, the "inner product" $(f * g)(x) = \langle f, T_x\widetilde{g}\rangle$ exists for almost every x.

1.3.7 Convolution and Smoothing

Since convolution is a type of averaging, it tends to be a smoothing operation. Generally speaking, a convolution $f*g$ inherits the "best" properties of both f and g. The following theorems and exercises will give several illustrations of this. We begin with an easy but very useful exercise.

Exercise 1.35. Show that

$$f \in C_c(\mathbb{R}),\; g \in C_c(\mathbb{R}) \implies f * g \in C_c(\mathbb{R}),$$

and in this case we have

$$\mathrm{supp}(f*g) \;\subseteq\; \mathrm{supp}(f)+\mathrm{supp}(g) \;=\; \bigl\{x+y : x \in \mathrm{supp}(f),\, y \in \mathrm{supp}(g)\bigr\}. \qquad \diamondsuit$$

Next we see an example of how a convolution $f * g$ can inherit smoothness from either f or g. This proof of this result uses a standard "extension by density" argument, which is a very useful technique for solving many of the exercises in this and other sections.

Theorem 1.36. *We have*

$$f \in L^1(\mathbb{R}),\ g \in C_0(\mathbb{R}) \implies f * g \in C_0(\mathbb{R}).$$

Proof. Note that if $f \in L^1(\mathbb{R})$ and $g \in C_0(\mathbb{R})$, then Exercise 1.27 implies that $f*g$ exists and is bounded. Also, since $g \in C_0(\mathbb{R})$, we know that g is uniformly continuous. Therefore, for $x,\, h \in \mathbb{R}$,

$$\begin{aligned}
&\bigl|(f*g)(x) - (f*g)(x-h)\bigr| \\
&\quad= \left|\int f(y)\, g(x-y)\, dy \;-\; \int f(y)\, g(x-h-y)\, dy\right| \\
&\quad\le \int |f(y)|\, |g(x-y) - g(x-h-y)|\, dy \\
&\quad\le \left(\sup_{u\in\mathbb{R}} |g(u) - g(u-h)|\right) \int |f(y)|\; dy \\
&\quad= \|g - T_h g\|_\infty\, \|f\|_1 \;\to\; 0 \quad \text{ as } h \to 0,
\end{aligned}$$

where the convergence follows from the fact that g is uniformly continuous. Thus $f*g \in C_b(\mathbb{R})$, and in fact $f*g$ is uniformly continuous. Actually, we can make this argument much more succinct by making use of the fact that convolution commutes with translation (Exercise 1.29). We need only write:

$$\begin{aligned}
\|f*g - T_h(f*g)\|_\infty &= \|f*g - f*(T_h g)\|_\infty \\
&= \|f*(g - T_h g)\|_\infty \\
&\le \|f\|_1\, \|g - T_h g\|_\infty \;\to\; 0 \quad \text{ as } h \to 0.
\end{aligned}$$

To show that $f*g \in C_0(\mathbb{R})$, consider first the case where $g \in C_c(\mathbb{R})$. Then $\operatorname{supp}(g) \subseteq [-N,N]$ for some $N > 0$. Hence

$$\begin{aligned}
|(f*g)(x)| &\le \int_{x-N}^{x+N} |f(y)|\, |g(x-y)|\, dy \\
&\le \|g\|_\infty \int_{x-N}^{x+N} |f(y)|\, dy \;\to\; 0 \quad \text{ as } |x| \to \infty.
\end{aligned}$$

This shows that $f*g \in C_0(\mathbb{R})$ whenever $g \in C_c(\mathbb{R})$.

Now we extend by density to all of $C_0(\mathbb{R})$. Choose an arbitrary $g \in C_0(\mathbb{R})$. Since $C_c(\mathbb{R})$ is dense in $C_0(\mathbb{R})$, we can find functions $g_n \in C_c(\mathbb{R})$ such that $g_n \to g$ in L^∞-norm. By our previous work we know that $f*g_n \in C_0(\mathbb{R})$ for every n, and, by equation (1.11),

$$\|f*g - f*g_n\|_\infty \;\le\; \|f\|_1\, \|g - g_n\|_\infty \;\to\; 0 \quad \text{ as } n \to \infty.$$

Thus $f*g_n \to f*g$ in L^∞-norm. Since $f*g_n \in C_0(\mathbb{R})$ for every n and since $C_0(\mathbb{R})$ is a closed subspace of $L^\infty(\mathbb{R})$, we conclude that $f*g \in C_0(\mathbb{R})$. □

Remark 1.37. It is important to observe that the function $f \in L^1(\mathbb{R})$ in the statement of Theorem 1.36 is really an equivalence class of functions that are equal almost everywhere. In this sense f is only defined a.e., yet the convolution $f * g$ is a continuous function that is defined everywhere. In particular, changing f on a set of measure zero has no effect on the values $(f * g)(x) = \int f(y)\, g(x-y)\, dy$ for $x \in \mathbb{R}$. $\diamondsuit$

There are other ways to go about proving Theorem 1.36. For example, the next exercise suggests a slightly different way of giving an extension by density argument.

Exercise 1.38. Let $f_n,\ g_n \in C_c(\mathbb{R})$ be such that $f_n \to f$ in L^1-norm while $g_n \to g$ in L^∞-norm, and show that $f_n * g_n \to f * g$ in L^∞-norm. Since $f_n * g_n \in C_c(\mathbb{R})$ by Exercise 1.35, and since $C_0(\mathbb{R})$ is closed in L^∞-norm, it follows that $f * g \in C_0(\mathbb{R})$. $\diamondsuit$

The following exercise extends the pairing (L^1, C_0) considered in Theorem 1.36 to pairings $(L^p, L^{p'})$ where $1 < p < \infty$ (and in so doing improves on Exercise 1.27). A weaker conclusion also holds for the pairing (L^1, L^∞). An extension by density argument similar to that of either Theorem 1.36 or Exercise 1.38 is useful in proving this next exercise.

Exercise 1.39. (a) Show that if $1 < p < \infty$, then

$$f \in L^p(\mathbb{R}),\ g \in L^{p'}(\mathbb{R}) \implies f * g \in C_0(\mathbb{R}).$$

(b) For $(p, p') = (1, \infty)$ or $(p, p') = (\infty, 1)$, show that

$$f \in L^1(\mathbb{R}),\ g \in L^\infty(\mathbb{R}) \implies f * g \in C_b(\mathbb{R}),$$

and $f * g$ is uniformly continuous. Show further that if g is compactly supported then $f * g \in C_0(\mathbb{R})$. However, give an example that shows that if g is not compactly supported then we need not have $f * g \in C_0(\mathbb{R})$, even if f is compactly supported. $\diamondsuit$

1.3.8 Convolution and Differentiation

Not only is convolution well-behaved with respect to continuity, but we can extend to higher derivatives.

Exercise 1.40. Given $1 \le p < \infty$ and $m \ge 0$, show that

$$f \in L^p(\mathbb{R}),\ g \in C_c^m(\mathbb{R}) \implies f * g \in C_0^m(\mathbb{R}),$$

and

$$f \in L^\infty(\mathbb{R}),\ g \in C_c^m(\mathbb{R}) \implies f * g \in C_b^m(\mathbb{R}).$$

Further, writing $D^j g = g^{(j)}$ for the jth derivative, show that differentiation commutes with convolution, i.e.,

$$D^j(f * g) \;=\; f * D^j g, \quad j = 0, \dots, m. \qquad \diamondsuit$$

In particular, for any $1 \le p \le \infty$, if $f \in L^p(\mathbb{R})$ is compactly supported and $g \in C_c^m(\mathbb{R})$, then $f * g \in C_c^m(\mathbb{R})$. This gives us an easy mechanism for generating new elements of $C_c^m(\mathbb{R})$ given any one particular element g. Moreover, if g is infinitely differentiable, then we can apply Exercise 1.40 for every m, and as a consequence obtain the following corollary.

Corollary 1.41. *If* $1 \le p < \infty$, *then*

$$f \in L^p(\mathbb{R}),\ g \in C_c^\infty(\mathbb{R}) \implies f * g \in C_0^\infty(\mathbb{R}),$$

and

$$f \in L^\infty(\mathbb{R}),\ g \in C_c^\infty(\mathbb{R}) \implies f * g \in C_b^\infty(\mathbb{R}).$$

Moreover, in either case, if f *is also compactly supported then we have* $f * g \in C_c^\infty(\mathbb{R})$. ◇

So from one function in $C_c^\infty(\mathbb{R})$ we can generate many others. But this begs the question: *Are there* any functions that are both compactly supported and infinitely differentiable? It is not at all obvious at first glance whether there exist *any* functions with such extreme properties, but the following exercise constructs some examples (see also Problem 1.20).

Exercise 1.42. Define $f(x) = e^{-1/x^2}\, \chi_{(0,\infty)}(x)$.

(a) Show that for every $n \in \mathbb{N}$, there exists a polynomial p_n of degree $3n$ such that

$$f^{(n)}(x) \;=\; p_n(x^{-1})\, e^{-x^{-2}}\, \chi_{(0,\infty)}(x).$$

Conclude that $f \in C_b^\infty(\mathbb{R})$, and, in particular, $f^{(n)}(0) = 0$ for every $n \ge 0$. Note that $\mathrm{supp}(f) = [0,\infty)$.

(b) Show that if $a < b$, then $g(x) = f(x-a)\, f(b-x)$ belongs to $C_c^\infty(\mathbb{R})$ and satisfies $\mathrm{supp}(g) = [a,b]$ and $g(x) > 0$ on (a,b). ◇

1.3.9 Convolution and Banach Algebras

A Banach algebra is a Banach space that is closed under an additional "multiplication-like" operation. Here is the precise abstract definition.

Definition 1.43 (Algebra). An *algebra* is a vector space $\mathcal{A}$ such that for each $f, g \in \mathcal{A}$ there exists a unique product $fg \in \mathcal{A}$ that satisfies the following conditions for all $f, g, h \in \mathcal{A}$ and $\alpha \in \mathbb{C}$:

(a) $(fg)h = f(gh)$,

(b) $f(g+h) = fg + fh$ and $(f+g)h = fh + gh$, and

(c) $\alpha(fg) = (\alpha f)g = f(\alpha g)$.

If $fg = gf$ for all $f, g \in \mathcal{A}$ then $\mathcal{A}$ is *commutative*. If there exists an element $e \in \mathcal{A}$ such that $ef = fe = f$ for every $f \in \mathcal{A}$ then $\mathcal{A}$ is an *algebra with identity*. ◇

Definition 1.44 (Banach Algebra). A *normed algebra* is a normed linear space $\mathcal{A}$ that is an algebra and also satisfies

$$\forall\, f, g \in \mathcal{A}, \quad \|fg\| \;\le\; \|f\|\,\|g\|.$$

A *Banach algebra* is a normed algebra that is a Banach space, i.e., it is a complete normed algebra. $\diamondsuit$

Since $\|f * g\|_1 \le \|f\|_1\,\|g\|_1$ and convolution is commutative, $L^1(\mathbb{R})$ is a commutative Banach algebra with respect to the operation of convolution. However, by Exercise 1.32 it has no identity element: There is no function $\delta \in L^1(\mathbb{R})$ that satisfies $f * \delta = f$ for all $f \in L^1(\mathbb{R})$. Here are some other examples of Banach algebras.

Example 1.45. (a) $C_b(\mathbb{R})$ is a commutative Banach algebra with identity under the operation of pointwise products of functions, $(fg)(x) = f(x)g(x)$.

(b) $C_0(\mathbb{R})$ is a commutative Banach algebra without identity with respect to pointwise products.

(c) Since the operator norm is submultiplicative, if X is a Banach space then the space $\mathcal{B}(X)$ of all bounded linear operators mapping X into itself is a noncommutative Banach algebra with identity with respect to composition of operators. $\diamondsuit$

We will explore some aspects of the Banach algebra structure of $L^1(\mathbb{R})$ and its relatives next. First we create another Banach algebra that isometrically inherits its structure from $L^1(\mathbb{R})$ via the Fourier transform.

Exercise 1.46. Define

$$A(\mathbb{R}) \;=\; \bigl\{\widehat{f} : f \in L^1(\mathbb{R})\bigr\},$$

and set $\|\widehat{f}\|_A = \|f\|_1$ for $f \in L^1(\mathbb{R})$. If we assume that the Fourier transform is injective on $L^1(\mathbb{R})$, which we prove later in Theorem 1.92, then $\|\cdot\|_A$ is well-defined. Given this, prove that $\|\cdot\|_A$ is a norm and that $A(\mathbb{R})$ is a Banach space with respect to this norm. Prove also that $A(\mathbb{R})$ is a commutative Banach algebra without identity with respect to the operation of pointwise products of functions. $\diamondsuit$

The space $A(\mathbb{R})$ is known by a variety of names, including both the *Fourier algebra* and the *Wiener algebra*, although it should be noted that both of these terms are sometimes used to refer to other spaces.

Our definition of $A(\mathbb{R})$ is *implicit*—it is the set of all Fourier transforms of L^1 functions, and in fact is isometrically isomorphic to $L^1(\mathbb{R})$. However, we can ask whether there is an *explicit* description of $A(\mathbb{R})$—is it possible to say that $F \in A(\mathbb{R})$ based directly on properties of F? For example, we know that $A(\mathbb{R}) \subseteq C_0(\mathbb{R})$ by the Riemann–Lebesgue Lemma (Theorem 1.20), so continuity and decay at infinity of F are necessary conditions for F to belong to $A(\mathbb{R})$, but are there further conditions that we can place on F that are both necessary and sufficient for membership in $A(\mathbb{R})$? No such explicit characterization is known. We will see later both implicit and explicit proofs that $A(\mathbb{R})$ is a *dense but proper* subset of $C_0(\mathbb{R})$, see Exercises 1.100–1.102.

As in abstract ring theory, the concept of an ideal plays an important role in the theory of Banach algebras. Ideals are the black holes of the algebra, sucking any product of an algebra element with an ideal element into the ideal. Since we are mostly interested in the algebra $L^1(\mathbb{R})$ we state the definition for commutative Banach algebras; in a noncommutative Banach algebra we would need to distinguish between left and right ideals.

Definition 1.47 (Ideals). A subspace I of a commutative Banach algebra $\mathcal{A}$ is an *ideal* in $\mathcal{A}$ if

$$x \in \mathcal{A},\, y \in I \implies xy \in I. \qquad \diamondsuit$$

In particular, a subspace I of $L^1(\mathbb{R})$ is an ideal in $L^1(\mathbb{R})$ if $f * g \in I$ whenever $f \in L^1(\mathbb{R})$ and $g \in I$.

Exercise 1.48. Suppose that $g \in L^1(\mathbb{R})$.

(a) Show that $g * L^1(\mathbb{R}) = \{g * f : f \in L^1(\mathbb{R})\}$ is an ideal in $L^1(\mathbb{R})$, called the *ideal generated by* g. Give an example that shows that g need not belong to $g * L^1(\mathbb{R})$.

(b) Show that

$$I(g) \;=\; \overline{g * L^1(\mathbb{R})}$$

is a closed ideal in $L^1(\mathbb{R})$, called the *closed ideal generated by* g. Problem 1.39 will later show that g always belongs to $I(g)$. Assuming this fact for now, prove that $I(g)$ is the *smallest* closed ideal that contains g. $\diamondsuit$

An ideal of the form $I(g)$ is also called a *principal ideal* in $L^1(\mathbb{R})$. Is every closed ideal in $L^1(\mathbb{R})$ a principal ideal? Atzmon [Atz72] answered this longstanding question in 1972, showing that there exist closed ideals in $L^1(\mathbb{R})$ that are not of the form $I(g)$.

Some Banach algebras also have an additional operation that has properties similar to that of conjugation of complex numbers.

Definition 1.49 (Involution). An *involution* on a Banach algebra $\mathcal{A}$ is a mapping $x \mapsto x^*$ of $\mathcal{A}$ into itself that satisfies the following for all $x,\, y \in \mathcal{A}$ and all scalars $\alpha \in \mathbb{C}$:

(a) $(x^*)^* = x$,

(b) $(xy)^* = y^*x^*$,

(c) $(x+y)^* = x^* + y^*$, and

(d) $(\alpha x)^* = \bar{\alpha} x^*$. $\diamondsuit$

For example, if H is a Hilbert space then the adjoint operation is an involution on $\mathcal{B}(H)$, since we $(AB)^* = B^*A^*$ for all $A, B \in \mathcal{B}(H)$. In the language of operator theory, because this involution satisfies $\|A^*A\| = \|A\|^2$ we say that $\mathcal{B}(H)$ is a *C^*-algebra*.

There is also an involution on $L^1(\mathbb{R})$.

Exercise 1.50. Given $f \in L^1(\mathbb{R})$, define $\widetilde{f}(x) = \overline{f(-x)}$. Show that $f \mapsto \widetilde{f}$ defines an involution on $L^1(\mathbb{R})$ with respect to convolution. $\diamondsuit$

1.3.10 Convolution on General Domains

Convolution can be defined for functions with domains other than the real line. We give two specific examples.

First, consider the discrete analogue of functions, i.e., sequences.

Definition 1.51 (Convolution of Sequences). Let $a = (a_k)_{k\in\mathbb{Z}}$ and $b = (b_k)_{k\in\mathbb{Z}}$ be sequences of complex scalars. Then the *convolution of a with b* is the sequence $a * b = \big((a*b)_k\big)_{k\in\mathbb{Z}}$ given by

$$(a*b)_k \;=\; \sum_{j\in\mathbb{Z}} a_j\, b_{k-j}, \tag{1.18}$$

whenever this series converges. $\diamondsuit$

Exercise 1.52. Wherever it makes sense, derive analogues of the theorems of this section for convolution of sequences. For example, results dealing with the continuity or differentiability of convolution of functions will have no analogue for sequences, but results dealing with integrability will have an analogue in terms of summability. In particular, prove analogues of Young's Inequality for sequences. Show that, unlike $L^1(\mathbb{R})$, there does exist an element $\delta \in \ell^1(\mathbb{Z})$ that is an identity for convolution, and we have

$$\forall\, 1 \le p \le \infty, \quad \forall\, f \in \ell^p(Z), \quad f * \delta = f. \qquad \diamondsuit$$

For a second example, consider periodic functions on the real line. To be precise, we say that a function $f\colon \mathbb{R} \to \mathbb{C}$ is *1-periodic* if

$$\forall\, x \in \mathbb{R}, \quad f(x+1) = f(x).$$

We define

$$L^p(\mathbb{T}) \;=\; \big\{f : f \text{ is 1-periodic and } f \in L^p[0,1)\big\}.$$

Definition 1.53 (Convolution of Periodic Functions). Let $f\colon \mathbb{R} \to \mathbb{C}$ and $g\colon \mathbb{R} \to \mathbb{C}$ be 1-periodic measurable functions. Then the *convolution of f with g* is the function $f * g$ given by

$$(f*g)(x) \;=\; \int_0^1 f(y)\, g(x-y)\, dy, \tag{1.19}$$

whenever this integral is well-defined. $\diamondsuit$

Note that $f * g$, if it exists, will be a 1-periodic function.

Exercise 1.54. Wherever it makes sense, derive analogues of the theorems of this section for convolution of periodic functions. In particular, prove analogues of Young's Inequality for periodic functions. Show that there is no function $\delta \in L^1(\mathbb{T})$ that is an identity for convolution. $\diamondsuit$

There is another way to view 1-periodic functions. If we define $\mathbb{T} = [0, 1)$, then $\mathbb{T}$ is an abelian group under the operation $\oplus$ of *addition modulo* 1,

$$x \oplus y \;=\; x + y \bmod 1,$$

where $a \bmod 1$ is the fractional part of a. Topologically, $\mathbb{T}$ is homeomorphic to the unit circle in $\mathbb{R}^2$ or $\mathbb{C}$ (thus, "$\mathbb{T}$" for "torus", though only a 1-dimensional torus in our setting). Hence we can identify 1-periodic functions on $\mathbb{R}$ with functions on the group $\mathbb{T}$. Writing $+$ for the operation on $\mathbb{T}$ instead of $\oplus$, convolution of functions on the domain $\mathbb{T}$ is again defined by equation (1.19).

Clearly, there must be a broader context here. The three domains for convolution that we have considered, namely, $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{T}$, are all *locally compact abelian groups*. That is, not only are they abelian groups, but they are also endowed with a topology that is locally compact (every point has a neighborhood that is compact). But convolution seems to involve more—we need the existence of Lebesgue measure on $\mathbb{R}$ and $\mathbb{T}$ in order to define the integral in the definition of convolution of functions, and we implicitly need counting measure on $\mathbb{Z}$ to define the series in the definition of the convolution of sequences. It is an amazing fact that every locally compact abelian group G has a positive, regular, translation-invariant Borel measure μ, and this measure is unique up to scaling by positive constants. This measure μ is called the *Haar measure* on G, see [Nac65] or [Fol99]. For $\mathbb{R}$ and $\mathbb{T}$ this is Lebesgue measure, and for $\mathbb{Z}$ it is counting measure. We are not going to define terms precisely here; we simply observe that convolution is part of a much broader universe.

Indeed, this is true of this entire volume: Much of what we do for the Fourier transform on $\mathbb{R}$ has analogues that hold for the setting of locally compact abelian groups. There is a general abstract theory of the Fourier transform on locally compact abelian groups, but not every result for the Fourier transform on $\mathbb{R}$ has an analogue in that general setting. We explore this briefly in Sections 2.1 and 2.2. In some ways, our chosen setting of the real line is among the most "complex" of the locally compact abelian groups, as $\mathbb{R}$ is neither compact (like $\mathbb{T}$) nor discrete (like $\mathbb{Z}$).

And the story does not end with *abelian* groups. If G is a locally compact group that is not abelian then we have to distinguish between left and right translations, but still G will have a unique left Haar measure μ_L (unique up to scale, left-translation invariant), and a unique right Haar measure μ_R (unique up to scale, right-translation invariant). The reader can consider how to define convolution in this setting, and what properties that convolution will possess—there will be a difference between "left" and "right" convolution. Still, convolution carries over without too many complications. On the other hand, the analogue of the Fourier transform becomes a very difficult (and interesting) object to define and study on nonabelian groups, and this is the topic of the representation theory of locally compact groups. For these generalizations we direct the reader to texts such as those by Folland [Fol95], Rudin [Rud62], or Hewitt and Ross [HR79].

Additional Problems

1.13. Let $f(x) = e^{-|x|}$, $g(x) = e^{-x^2}$, and $h(x) = xe^{-x^2}$. Show that

$$\begin{aligned}
(f * f)(x) &= (1 + |x|)\, e^{-|x|}, \\
(g * g)(x) &= \left(\frac{\pi}{2}\right)^{1/2} e^{-x^2/2}, \\
(h * h)(x) &= \frac{1}{4}\left(\frac{\pi}{2}\right)^{1/2} (x^2 - 1)\, e^{-x^2/2}.
\end{aligned}$$

1.14. Show that if $f,\, g \in L^1(\mathbb{R})$ and $f,\, g \geq 0$ a.e., then $\|f * g\|_1 = \|f\|_1\, \|g\|_1$. Find a function $h \in L^1(\mathbb{R})$ such that $\|h * h\|_1 < \|h\|_1^2$.

1.15. If $f,\, g \in L^1(\mathbb{R})$, then

$$\operatorname{supp}(f * g) \;\subseteq\; \overline{\operatorname{supp}(f) + \operatorname{supp}(g)}$$

(see Notation 1.25). Further, if f and g are both compactly supported then so is $f * g$ and, in this case, $\operatorname{supp}(f * g) \subseteq \operatorname{supp}(f) + \operatorname{supp}(g)$.

1.16. Let A_p be the Babenko–Beckner constant defined in equation (1.14), and prove the following facts.

(a) $A_p < 1$ for $1 < p < 2$.

(b) $A_p > 1$ for $2 < p < \infty$.

(c) $A_2 = \lim\limits_{p \to 1^+} A_p = \lim\limits_{p \to \infty} A_p = 1$.

1.17. Give an example of $f \in L^p(\mathbb{R})$ with $1 < p < \infty$ and $g \in C_0(\mathbb{R})$ such that $f * g$ is not defined. Compare this to Theorem 1.36 and Exercise 1.40, which show that $f * g \in C_0(\mathbb{R})$ if either $f \in L^1(\mathbb{R})$ and $g \in C_0(\mathbb{R})$, or if $f \in L^p(\mathbb{R})$ and $g \in C_c(\mathbb{R})$.

1.18. Prove the following variation on Exercise 1.40: If $f \in L^1(\mathbb{R})$ and $g \in C_b^m(\mathbb{R})$, then $f * g \in C_b^m(\mathbb{R})$.

1.19. Let $1 \leq p \leq \infty$ and $f \in L^p(\mathbb{R})$ be given. If there exists an $h \in L^p(\mathbb{R})$ such that

$$\lim_{a \to 0} \left\| h - \frac{f - T_a f}{a} \right\|_p \;=\; 0$$

then we call h a *strong* L^p *derivative of* f and denote it by $h = \partial_p f$. Show that, in this case, if $g \in L^{p'}(\mathbb{R})$ then $f * g$ is differentiable pointwise everywhere, and $(f * g)' = \partial_p f * g$.

1.20. Set $f(x) = e^{-1/x}\,\chi_{[0,\infty)}(x)$.

(a) Show that there exists a polynomial p_n of degree $2n$ such that $f^{(n)}(x) = p_n(x^{-1})\,e^{-1/x}\,\chi_{[0,\infty)}(x)$.

(b) Let $g(x) = f(1 - |x|^2)$, and show that $g \in C_c^\infty(\mathbb{R})$ with $\mathrm{supp}(g) = [-1, 1]$.

1.21. Create explicit examples of functions $f \in C_c^\infty(\mathbb{R})$ supported in the interval $[-1, 1]$ that have the following properties.

(a) $0 \le f(x) \le 1$ and $f(0) = 1$.

(b) $\int f(x)\,dx = 0$.

(c) $\int f(x)\,dx = 1$.

(d) $\int f(x)\,dx = 1$, $\int x f(x)\,dx = 0$.

1.22. Let $E \subseteq \mathbb{R}$ be measurable, and show that

$$I(E) \;=\; \{f \in L^1(\mathbb{R}) : \mathrm{supp}(\widehat{f}\,) \subseteq E\}$$

is a closed ideal in $L^1(\mathbb{R})$ (under the operation of convolution).

1.23. Define $A_c(\mathbb{R}) = \{F \in A(\mathbb{R}) : \mathrm{supp}(F) \text{ is compact}\}$. Show that $A_c(\mathbb{R})$ is an ideal in $A(\mathbb{R})$ (under the operation of pointwise multiplication). Compare Problem 1.56, which shows that $A_c(\mathbb{R})$ is dense in $A(\mathbb{R})$, and hence is not a closed ideal in $A(\mathbb{R})$.

1.24. (a) Let $E \subseteq \mathbb{R}$ be measurable with $0 < |E| < \infty$. By considering $\chi_E * \chi_{-E}$, show that $E - E = \{x - y : x, y \in E\}$ contains an open interval $(-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$.

(b) Consider the equivalence relation $x \sim y \iff x - y \in \mathbb{Q}$ on $\mathbb{R}$. By the Axiom of Choice, there exists a set E that contains exactly one element of each of the equivalence classes of $\sim$. Use part (a) to show that E is not measurable.

(c) Show that every subset of $\mathbb{R}$ that has positive exterior Lebesgue measure contains a nonmeasurable subset.

1.25. Let $f(x, y)$ be a measurable function on $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$, and fix $1 \le p < \infty$. Prove *Minkowski's Integral Inequality*:

$$\left(\int\left(\int |f(x,y)|\,dy\right)^{p} dx\right)^{1/p} \;\le\; \int\left(\int |f(x,y)|^p\,dx\right)^{1/p} dy. \tag{1.20}$$

Remark: This equation may be more revealing if we rewrite it as

$$\left\| \int |f(\cdot, y)|\, dy \right\|_p \leq \int \|f(\cdot, y)\|_p \, dy.$$

Thus, Minkowski's Integral Inequality is an integral version of the Triangle Inequality (also known as Minkowski's Inequality) on $L^p(\mathbb{R}^m)$.

1.4 The Duality Between Smoothness and Decay

A fundamental duality under the Fourier transform is the interchange between smoothness and decay. We explore this duality in this section.

1.4.1 Decay in Time Implies Smoothness in Frequency

To begin, we show that if f has sufficient decay, then $\widehat{f}$ must have a corresponding amount of smoothness. Although it is an abuse of notation, we will write $\big((-2\pi i x)^k f(x)\big)^\wedge$ to denote the Fourier transform of the function $g(x) = (-2\pi i x)^k f(x)$.

Theorem 1.55 (Smoothness and Decay I). *Let $f \in L^1(\mathbb{R})$ and $m \in \mathbb{N}$ be given. Then*

$$x^m f(x) \in L^1(\mathbb{R}) \implies \widehat{f} \in C_0^m(\mathbb{R}),$$

i.e., $\widehat{f}$ is m-times differentiable and $\widehat{f}, \widehat{f}\,', \dots, \widehat{f}^{\,(m)} \in C_0(\mathbb{R})$. Furthermore, in this case we have that $x^k f(x) \in L^1(\mathbb{R})$ for $k = 0, \dots, m$, and the kth derivative of $\widehat{f}$ is the Fourier transform of $(-2\pi i x)^k f(x)$:

$$\widehat{f}^{\,(k)} = \frac{d^k}{d\xi^k}\widehat{f} = \big((-2\pi i x)^k f(x)\big)^\wedge, \qquad k = 0, \dots, m. \qquad \diamondsuit \tag{1.21}$$

Before proving Theorem 1.55, we remark that the hypothesis $x^m f(x) \in L^1(\mathbb{R})$ is a kind of "average decay" condition. As x increases, the value of $|x^m f(x)|$ becomes increasing large compared to the value of $|f(x)|$ (except at points where $f(x) = 0$), yet still the area under the graph of $|x^m f(x)|$ must remain finite. However, this does not imply that f decays pointwise at infinity.

Exercise 1.56. Let $m \in \mathbb{N}$ be fixed. Give an example of a continuous function $f \in L^1(\mathbb{R})$ such that $x^m f(x) \in L^1(\mathbb{R})$ but $\lim_{x\to\pm\infty} f(x)$ does not exist. Show that we can even construct examples where f is continuous and unbounded. On the other hand, show that for any $h > 0$ we will have $\lim_{R\to\pm\infty} \int_R^{R+h} |x^m f(x)|\, dx = 0$. $\diamondsuit$

Equation (1.21) does not come out of the blue. We can guess the correct equation by formally exchanging a derivative and an integral:

$$\begin{aligned}
\frac{d}{d\xi}\widehat{f}(\xi) &= \frac{d}{d\xi}\int f(x)\, e^{-2\pi i\xi x}\, dx \\
&= \int f(x)\, \frac{d}{d\xi} e^{-2\pi i\xi x}\, dx \\
&= \int f(x)\, (-2\pi i x)\, e^{-2\pi i\xi x}\, dx \\
&= (-2\pi i x f(x))^{\wedge}(\xi).
\end{aligned}$$

Essentially, the proof of Theorem 1.55 is the justification of this interchange.

Proof (of Theorem 1.55). Consider the case $m = 1$. Assume that $f(x)$, $xf(x) \in L^1(\mathbb{R})$. We must show that the limit

$$\begin{aligned}
\widehat{f}\,'(\xi) &= \lim_{\eta\to 0} \frac{\widehat{f}(\xi+\eta) - \widehat{f}(\xi)}{\eta} \\
&= \lim_{\eta\to 0} \int f(x)\, \frac{e^{-2\pi i(\xi+\eta)x} - e^{-2\pi i\xi x}}{\eta}\, dx \qquad (1.22)
\end{aligned}$$

exists. We do this by applying the Lebesgue Dominated Convergence Theorem. Define $e_x(\xi) = e^{-2\pi i\xi x}$. Then the integrand in equation (1.22) converges pointwise for almost every x, because

$$\lim_{\eta\to 0} f(x)\, \frac{e_x(\xi+\eta) - e_x(\xi)}{\eta} = f(x)\, e_x'(\xi) = f(x)\, (-2\pi i x)\, e^{-2\pi i\xi x}.$$

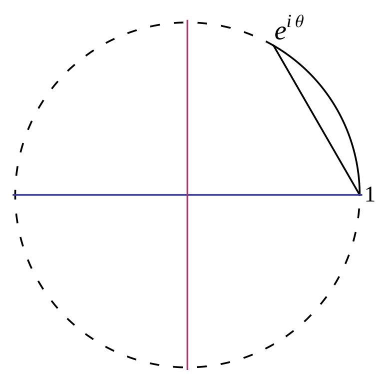

Fig. 1.8. The distance $|e^{i\theta} - 1|$ from $e^{i\theta}$ to 1 is smaller than the arclength θ along the unit circle from $e^{i\theta}$ to 1.

Further, since we always have $|e^{i\theta} - 1| \le |\theta|$ (see the "proof by picture" in Figure 1.8 for the case $0 \le \theta \le \frac{\pi}{2}$), we have

$$\begin{aligned}\left|f(x)\,\frac{e^{-2\pi i(\xi+\eta)x}-e^{-2\pi i\xi x}}{\eta}\right| &= |f(x)|\,|e^{-2\pi i\xi x}|\left|\frac{e^{-2\pi i\eta x}-1}{\eta}\right| \\ &\le 2\pi|xf(x)| \in L^1(\mathbb{R}).\end{aligned}$$

Hence the Lebesgue Dominated Convergence Theorem implies that the limit in equation (1.22) exists, and

$$\begin{aligned}\widehat{f}\,'(\xi) &= \int \lim_{\eta\to 0} f(x)\,\frac{e^{-2\pi i(\xi+\eta)x}-e^{-2\pi i\xi x}}{\eta}\,dx \\ &= \int f(x)\,(-2\pi i x)\,e^{-2\pi i\xi x}\,dx \\ &= \big((-2\pi i x)f(x)\big)^{\wedge}(\xi).\end{aligned}$$

Technically, the Dominated Convergence Theorem applies to sequences indexed by the natural numbers, whereas above we need to let the continuous variable η converge to zero. However, this can be justified by considering all possible sequences $\eta_k \to 0$ (see Problem 1.29). In any case, we conclude that $\widehat{f}$ is differentiable. Further, since $\widehat{f}\,'$ is the Fourier transform of the integrable function $(-2\pi i x)f(x)$, the Riemann–Lebesgue Lemma implies that $\widehat{f}\,' \in C_0(\mathbb{R})$. Since we also have $\widehat{f} \in C_0(\mathbb{R})$, we have shown that $\widehat{f} \in C_0^1(\mathbb{R})$.

Exercise: Apply induction to extend to all $m > 1$. $\square$

In operator notation, if we recall the *position and momentum operators* P and M defined by

$$Pg(x) = xg(x), \qquad \text{and} \qquad Mg = \frac{1}{2\pi i}\,g',$$

then Theorem 1.55 says that if f, $Pf \in L^1(\mathbb{R})$ then we have

$$M\widehat{f} = (-Pf)^{\wedge}.$$

In other words, $M\mathcal{F} = -\mathcal{F}P$, and we will shortly see that $P\mathcal{F} = \mathcal{F}M$ as well. Thus P and M are dual under the Fourier transform.

1.4.2 A Primer on Absolute Continuity

In order to prove our next main result, we will need to apply both the Fundamental Theorem of Calculus and the technique of integration by parts on an interval $[a, b]$. Of course, if f is differentiable and f' is continuous on $[a, b]$, then these results can certainly be applied. However, the Fundamental Theorem of Calculus and integration by parts hold more generally; in fact, they apply as long as f is an *absolutely continuous function*. Therefore we pause to briefly review the definition and basic properties of absolute continuity.

Definition 1.57 (Absolute Continuity). A function $f\colon [a,b] \to \mathbb{C}$ is *absolutely continuous on* $[a,b]$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any finite or countably infinite collection of nonoverlapping subintervals $\{[a_j,b_j]\}_j$ of $[a,b]$, we have

$$\sum_j (b_j - a_j) < \delta \quad \Longrightarrow \quad \sum_j |f(b_j) - f(a_j)| < \varepsilon.$$

We denote the class of absolutely continuous functions on $[a,b]$ by $\mathrm{AC}[a,b]$. The space of *locally absolutely continuous functions on* $\mathbb{R}$ is

$$\mathrm{AC}_{\mathrm{loc}}(\mathbb{R}) \;=\; \bigl\{f\colon \mathbb{R} \to \mathbb{C} : f \in \mathrm{AC}[a,b] \text{ for every } a<b\bigr\}. \qquad \diamondsuit$$

The next exercise shows that the antiderivative of an integrable function is absolutely continuous.

Exercise 1.58. Show that if $f \in L^1[a,b]$, then $g(x) = \int_a^x f(t)\,dt$ belongs to $\mathrm{AC}[a,b]$, and furthermore $g' = f$ a.e. $\diamondsuit$

In fact, much more holds. We refer to texts such as [BBT97], [Fol99], [WZ77] for proof of the next result.

Theorem 1.59 (Fundamental Theorem of Calculus). *If $g\colon [a,b] \to \mathbb{C}$, then the following statements are equivalent.*

(a) $g \in \mathrm{AC}[a,b]$.

(b) *There exists $f \in L^1[a,b]$ such that*

$$g(x) - g(a) \;=\; \int_a^x f(t)\,dt, \qquad x \in [a,b].$$

(c) *g is differentiable at almost every $x \in [a,b]$, $g' \in L^1[a,b]$, and*

$$g(x) - g(a) \;=\; \int_a^x g'(t)\,dt, \qquad x \in [a,b]. \qquad \diamondsuit$$

Some basic properties of absolutely continuous functions are given in Problem 1.32. In particular, every Lipschitz function on $[a,b]$ is absolutely continuous. A function f that is differentiable at every point and has a bounded derivative f' is Lipschitz and hence absolutely continuous, even if f' is not continuous.

It is a more subtle fact that if f is differentiable at every point on $[a,b]$ and f' is merely *integrable* on $[a,b]$ then f is absolutely continuous. The main steps needed to prove this are the following two results, which we state without proof.

First we need a lemma that relates the size of $f(E)$ to the integral of $|f'|$ on E. Since a measurable function need not map measurable sets to measurable sets, this result must be worded in terms of the *exterior* Lebesgue measure of $f(E)$. This subtlety of this result is that E is an arbitrary measurable set, not just an interval (for proof, see [BBT97, Lem. 7.10]).

Lemma 1.60. *Let $f\colon [a,b] \to \mathbb{R}$ be measurable. If $E \subseteq [a,b]$ is measurable and f is differentiable at every point of E, then the exterior Lebesgue measure of $f(E)$ satisfies the bound*

$$|f(E)|_e \;\le\; \int_E |f'|. \qquad \diamondsuit$$

The next fact that we need is the *Banach–Zarecki Theorem*, which gives characterization of absolutely continuous functions (see [BBT97, Thm. 7.11] for proof). Bounded variation is defined in Problem 1.30.

Theorem 1.61 (Banach–Zarecki Theorem). *Given $f\colon [a,b] \to \mathbb{C}$, the following statements are equivalent.*

(a) $f \in \mathrm{AC}[a,b]$.

(b) *f is continuous, $f \in \mathrm{BV}[a,b]$, and $|f(A)| = 0$ for every $A \subseteq [a,b]$ with $|A| = 0$.*

(c) *f is continuous and is differentiable a.e., $f' \in L^1[a,b]$, and $|f(A)| = 0$ for every $A \subseteq [a,b]$ with $|A| = 0$.* $\diamondsuit$

Combining these two results, we obtain a useful sufficient condition for absolute continuity.

Theorem 1.62. *If $f\colon [a,b] \to \mathbb{C}$ is everywhere differentiable and $f' \in L^1[a,b]$, then $f \in \mathrm{AC}[a,b]$.* $\diamondsuit$

Proof. By splitting into real and imaginary parts, we may assume that f is real-valued. Suppose that $A \subseteq [a,b]$ and $|A| = 0$. Then since f is differentiable at every point of A, we have by Lemma 1.60 that $|f(A)|_e \le \int_A |f'| = 0$. Theorem 1.61 therefore implies that f is absolutely continuous. $\square$

The hypotheses of Theorem 1.62 can be relaxed somewhat. For example, if f is differentiable at all but *countably* many points and $f' \in L^1[a,b]$, then f will be absolutely continuous (see [Ben76]). However, the assumptions that f is differentiable a.e. and $f' \in L^1[a,b]$ are by themselves not sufficient to ensure that f is absolutely continuous (the Cantor–Lebesgue function is a counterexample, see Problem 1.33).

As we can see from Theorem 1.61 (and it is easy to prove directly), every absolutely continuous function has bounded variation. Although we will not need to use it, we state a fundamental decomposition for functions of bounded variation.

Definition 1.63 (Singular Function). A function h is *singular* if h is differentiable at almost every point in its domain and $h' = 0$ a.e. $\diamondsuit$

Theorem 1.64. *If $f \in \mathrm{BV}[a,b]$, then $f = g + h$ where $g \in \mathrm{AC}[a,b]$ and h is singular on $[a,b]$. Moreover, g and h are unique up to additive constants, and we can take*

$$g(x) \;=\; \int_a^x f', \qquad x \in [a,b]. \qquad \diamondsuit \tag{1.23}$$

1.4.3 Smoothness in Time Implies Decay in Frequency

Now we show how the Fourier transform converts smoothness into decay.

Theorem 1.65 (Smoothness and Decay II). *Let $f \in L^1(\mathbb{R})$ and $m \in \mathbb{N}$ be given. If f is everywhere m-times differentiable and $f, f', \dots, f^{(m)} \in L^1(\mathbb{R})$, then*

$$(f^{(k)})^\wedge(\xi) \;=\; (2\pi i\xi)^k\, \widehat{f}(\xi), \qquad k = 0, \dots, m.$$

Consequently,

$$|\widehat{f}(\xi)| \;\le\; \frac{\|f^{(m)}\|_1}{|2\pi\xi|^m}, \qquad \xi \ne 0. \tag{1.24}$$

Proof. Consider the case $m = 1$. Assume that f is everywhere differentiable and that $f,\, f' \in L^1(\mathbb{R})$. Then $f \in \mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$ by Theorem 1.62, so f is absolutely continuous on every interval $[a, b]$ in $\mathbb{R}$. Consequently the Fundamental Theorem of Calculus (Theorem 1.59) holds for f, so

$$f(x) - f(0) \;=\; \int_0^x f'(t)\, dt, \qquad x \in \mathbb{R}.$$

Since f' is integrable, it follows that

$$\lim_{x\to\infty} f(x) \;=\; f(0) + \lim_{x\to\infty} \int_0^x f'(t)\, dt \;=\; f(0) + \int_0^\infty f'(t)\, dt$$

exists. Since f is continuous and integrable, this limit has to be zero: $\lim_{x\to\infty} f(x) = 0$. Similarly $\lim_{x\to-\infty} f(x) = 0$, so we conclude that $f \in C_0(\mathbb{R})$.

Integration by parts is valid for absolutely continuous functions (Problem 1.32), so for every $a < b$ we have

$$\int_a^b f'(x)\, e^{-2\pi i\xi x}\, dx \;=\; e^{-2\pi i\xi b} f(b) - e^{-2\pi i\xi a} f(a) + 2\pi i\xi \int_a^b f(x)\, e^{-2\pi i\xi x}\, dx.$$

Consequently, since f' is integrable,

$$\begin{aligned}
\widehat{f'}(\xi) \;&=\; \int_{-\infty}^{\infty} f'(x)\, e^{-2\pi i\xi x}\, dx \\
&=\; \lim_{\substack{a\to-\infty\\ b\to\infty}} \int_a^b f'(x)\, e^{-2\pi i\xi x}\, dx \\
&=\; \lim_{\substack{a\to-\infty\\ b\to\infty}} \left(e^{-2\pi i\xi b} f(b) - e^{-2\pi i\xi a} f(a) + 2\pi i\xi \int_a^b f(x)\, e^{-2\pi i\xi x}\, dx \right) \\
&=\; 2\pi i\xi \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi x}\, dx \\
&=\; 2\pi i\xi\, \widehat{f}(\xi).
\end{aligned}$$

Finally, for $\xi \neq 0$ we have

$$|\widehat{f}(\xi)| \;=\; \frac{|\widehat{f'}(\xi)|}{|2\pi i\xi|} \;\le\; \frac{\|\widehat{f'}\|_\infty}{|2\pi\xi|} \;\le\; \frac{\|f'\|_1}{|2\pi\xi|}.$$

Exercise: Use induction to extend to all $m > 1$. $\square$

We extract and reformulate one useful fact that played a role in the proof of Theorem 1.65.

Exercise 1.66. If $g \in L^1(\mathbb{R})$ and its antiderivative $f(x) = \int_{-\infty}^x g(t)\,dt$ also belongs to $L^1(\mathbb{R})$, then $\int g = 0$ and $f \in C_0(\mathbb{R})$. $\diamondsuit$

The essential property used in the proof of Theorem 1.65 is absolute continuity. For example, if we consider $m = 1$, the assumption in Theorem 1.65 that f is everywhere differentiable and f, f' are integrable is made solely so that we will know that f is absolutely continuous. Absolute continuity is certainly necessary, as is shown by the example of the reflected Cantor–Lebesgue function (see Problem 1.33). That particular function f is continuous and compactly supported, is differentiable a.e., and both f and f' belong to $L^1(\mathbb{R})$. However, f is singular so $f' = 0$ a.e., and hence $\widehat{f'}$ is identically zero while $(2\pi i\xi)\,\widehat{f}(\xi)$ is not.

On the other hand, as long as f is absolutely continuous and both f, f' are integrable, the proof of Theorem 1.65 can be carried over, and so we obtain the following result.

Theorem 1.67. *Suppose $f \in L^1(\mathbb{R}) \cap \mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$, which implies in particular that f is differentiable a.e. If $f' \in L^1(\mathbb{R})$, then $\widehat{f'}(\xi) = 2\pi i\xi\,\widehat{f}(\xi)$ for every $\xi \in \mathbb{R}$.* $\diamondsuit$

This implies the following variation on Theorem 1.65, dealing with the Fourier transform of an antiderivative of an integrable function.

Corollary 1.68. *If $g \in L^1(\mathbb{R})$ and $f(x) = \int_{-\infty}^x g(t)\,dt$ also belongs to $L^1(\mathbb{R})$, then $\widehat{g}(\xi) = 2\pi i\xi\,\widehat{f}(\xi)$ for $\xi \in \mathbb{R}$.* $\diamondsuit$

1.4.4 The Riemann–Lebesgue Lemma Revisited

By Theorem 1.65, if g is an everywhere differentiable function such that g and g' are both integrable, then

$$|\widehat{g}(\xi)| \;\le\; \frac{\|g'\|_1}{2\pi|\xi|} \;\to\; 0 \quad \text{as } |\xi| \to \infty.$$

Hence the Riemann–Lebesgue Lemma (Theorem 1.20) holds for these particular functions g. A consequence of Exercise 1.78, which comes later in Section 1.5, is that the set

$$\{g \in L^1(\mathbb{R}) : g \text{ is everywhere differentiable and } g' \in L^1(\mathbb{R})\}$$

is a dense subset of $L^1(\mathbb{R})$. Assuming this fact for the moment, we can use an extension by density argument to give another proof that the Riemann–Lebesgue Lemma is valid for all $f \in L^1(\mathbb{R})$.

Proof (of Theorem 1.20). Choose any $f \in L^1(\mathbb{R})$ and $\varepsilon > 0$. Then there exists a differentiable function g with both $g, g' \in L^1(\mathbb{R})$ such that $\|f-g\|_1 < \varepsilon$. By Theorem 1.65, and equation (1.24) in particular, we have $\widehat{g} \in C_0(\mathbb{R})$, so there exists an $R > 0$ such that $|\widehat{g}(\xi)| < \varepsilon$ for all $|\xi| > R$. Now, for every $\xi \in \mathbb{R}$ we have

$$|\widehat{f}(\xi) - \widehat{g}(\xi)| \;\le\; \|\widehat{f} - \widehat{g}\,\|_\infty \;\le\; \|f-g\|_1 \;<\; \varepsilon,$$

so $|\widehat{f}(\xi)| < 2\varepsilon$ for all $|\xi| > R$. Hence $\widehat{f} \in C_0(\mathbb{R})$ as well. □

Additional Problems

1.26. Illustrate the connection between smoothness and decay given in Theorem 1.65 by computing the following Fourier transforms explicitly. Then make the same comparisons for the functions given in Problem 1.1.

(a) Show that if $f(x) = (\cos 6\pi x)\, \chi_{[-1/2,1/2]}(x)$, then

$$\widehat{f}(\xi) \;=\; \frac{\xi \sin \pi\xi}{\pi\,(9-\xi^2)}.$$

Observe that f is discontinuous, which is reflected in the fact that $\widehat{f}$ decays slowly at infinity, on the order of $1/|\xi|$.

(b) Show that if $g(x) = (\sin 6\pi x)\, \chi_{[-1/2,1/2]}(x)$, then

$$\widehat{g}(\xi) \;=\; \frac{3i \sin \pi\xi}{\pi\,(\xi^2-9)}.$$

The function g is continuous, and $\widehat{g}$ decays on the order of $1/|\xi|^2$.

1.27. Let $f(x) = xe^{-x}\chi_{[0,\infty)}(x)$. Show both directly and by using Theorem 1.65 that $\widehat{f}(\xi) = (1+2\pi i\xi)^{-2}$.

1.28. Assuming the hypotheses of Theorem 1.65, improve the conclusion given in equation (1.24) by showing that $\lim_{|\xi|\to\infty} |\xi^m\, \widehat{f}(\xi)| = 0$.

1.29. Let X be a normed space, and let $\{f_t\}_{t\in\mathbb{R}}$ be a sequence in X indexed by a real parameter. Show that $f_t \to f$ as $t \to 0$ if and only if $f_{t_k} \to f$ for every sequence of real numbers $\{t_k\}_{k\in\mathbb{N}}$ such that $t_k \to 0$.

1.30. Fix $f\colon [a,b] \to \mathbb{C}$. Let $\Gamma = \{a = x_0 < \cdots < x_n = b\}$ denote a finite partition of $[a,b]$. Set

$$S_\Gamma \;=\; \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \qquad \text{and} \qquad V[f;a,b] \;=\; \sup_\Gamma S_\Gamma,$$

where the supremum is taken over all partitions Γ of $[a,b]$. We call $V[f;a,b]$ the *variation* of f over $[a,b]$ and say that f has *bounded variation* on $[a,b]$ if $V[f;a,b] < \infty$. The set of functions with bounded variation on $[a,b]$ is $\mathrm{BV}[a,b]$. Prove the following statements.

(a) If $f,\ g \in \mathrm{BV}[a,b]$, then $\alpha f + \beta g \in \mathrm{BV}[a,b]$ and $fg \in \mathrm{BV}[a,b]$. If $|g(x)| \geq \varepsilon > 0$ for all $x \in [a,b]$ then $f/g \in \mathrm{BV}[a,b]$.

(b) Set $f(x) = x^2 \sin(1/x)$ and $g(x) = x^2 \sin(1/x^2)$ for $x \neq 0$, and $f(0) = g(0) = 0$. Then f and g are differentiable everywhere, $f \in \mathrm{BV}[-1,1]$, and $g \notin \mathrm{BV}[-1,1]$.

(c) For $y \in \mathbb{R}$ set $y^+ = \max\{y,\,0\}$ and $y^- = \max\{-y,\,0\}$, so $y^+ - y^- = y$ and $y^+ + y^- = |y|$.

Given a real-valued $f \in \mathrm{BV}[a,b]$, define

$$S_\Gamma^+ \;=\; \sum_{i=1}^{n} \bigl[f(x_i) - f(x_{i-1})\bigr]^+ \qquad \text{and} \qquad V^+[f;a,b] \;=\; \sup_\Gamma S_\Gamma^+,$$

and similarly define S_Γ^- and $V^-[f;a,b]$. Prove that

$$V^+[f;a,x] + V^-[f;a,x] \;=\; V[f;a,x], \qquad x \in [a,b],$$

and

$$V^+[f;a,x] - V^-[f;a,x] \;=\; f(x) - f(a), \qquad x \in [a,b].$$

(d) Prove the *Jordan Decomposition*: If $f \in \mathrm{BV}[a,b]$ is real-valued then $f = g - h$ where g, h are monotone increasing on $[a,b]$. Conclude that the left- and right-hand limits $f(x+) = \lim_{y \searrow x} f(y)$ and $f(x-) = \lim_{y \nearrow x} f(y)$ exist for all $x \in (a,b)$, as do $f(a+)$ and $f(b-)$.

1.31. A function $f\colon [a,b] \to \mathbb{C}$ is *Lipschitz* on $[a,b]$ if there exists a constant $C > 0$ such that

$$|f(x) - f(y)| \;\leq\; C\,|x - y|, \qquad x, y \in [a,b].$$

The class of Lipschitz functions on $[a,b]$ is denoted by $\mathrm{Lip}[a,b]$.

(a) Show that if f is Lipschitz on $[a,b]$ then f is uniformly continuous, has bounded variation, and $V[f;a,b] \leq C\,(b-a)$.

(b) Exhibit a Lipschitz function that is not differentiable at every point of $[a,b]$.

(c) Show that if f is differentiable everywhere on $[a,b]$ and f' is bounded on $[a,b]$, then f is Lipschitz with constant $C = \|f'\|_\infty$. In particular, if f, f' are both continuous on $[a,b]$, then f is Lipschitz.

1.32. Prove the following properties of absolutely continuous functions.

(a) If $g \in \mathrm{AC}[a,b]$, then g is uniformly continuous on $[a,b]$.

(b) $\mathrm{Lip}[a,b] \subsetneq \mathrm{AC}[a,b] \subsetneq \mathrm{BV}[a,b]$.

(c) Integration by parts is valid for absolutely continuous functions, i.e., if if $f,\, g \in \mathrm{AC}[a,b]$, then

$$\int_a^b f(x)\, g'(x)\, dx \;=\; f(b)g(b) - f(a)g(a) - \int_a^b f'(x)\, g(x)\, dx.$$

(d) $\mathrm{AC}[a,b]$ is closed under pointwise products: If $f,\, g \in \mathrm{AC}[a,b]$ then $fg \in \mathrm{AC}[a,b]$.

(e) If $f \in \mathrm{AC}[a,b]$ and $f' = 0$ a.e. then f is constant.

(f) If $f \in \mathrm{AC}_{\mathrm{loc}}[a,b]$ and there is a continuous function g such that $f' = g$ a.e. then f is differentiable everywhere and $f'(x) = g(x)$ for all $x \in [a,b]$.

1.33. Consider the functions φ_1, φ_2 pictured in Figure 1.9. The function φ_1 takes the constant value $1/2$ on the interval $(1/3, 2/3)$ that is removed in the first stage of the construction of the Cantor middle-thirds set, and is linear on the remaining intervals. The function φ_2 also takes the same constant $1/2$ on the interval $(1/3, 2/3)$ but additionally is constant with values $1/4$ and $3/4$ on the two intervals that are removed in the second stage of the construction of the Cantor set. Continue this process, defining $\varphi_3, \varphi_4, \dots$, and prove the following facts.

(a) Each φ_k is monotone increasing on $[0,1]$.

(b) $|\varphi_{k+1}(x) - \varphi_k(x)| < 2^{-k}$ for every $x \in [0,1]$.

(c) $\varphi(x) = \lim_{k\to\infty} \varphi_k(x)$ converges uniformly on $[0,1]$.

The function φ constructed in this manner is called the *Cantor–Lebesgue function* or, more picturesquely, the *Devil's staircase*. If we extend φ to $\mathbb{R}$ by reflecting it about the point $x = 1$ and declaring it to be zero outside of $[0,2]$, we obtain the continuous function φ pictured in Figure 1.10. Prove the following facts.

(d) φ is continuous and monotone increasing on $[0,1]$, but φ is not Lipschitz.

(e) φ is singular on $[0,1]$, i.e., φ is differentiable a.e. and $\varphi'(x) = 0$ a.e.

(f) The Fundamental Theorem of Calculus does not apply to φ:

$$\varphi(1) - \varphi(0) \;\neq\; \int_0^1 \varphi'(x)\, dx.$$

Conclude that φ is not absolutely continuous on $[0,1]$.

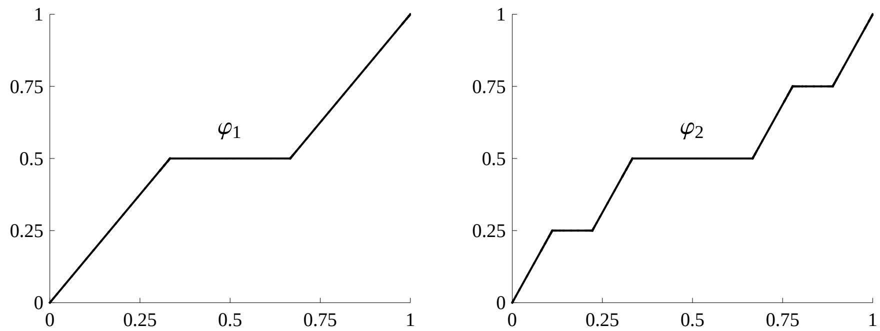

Fig. 1.9. First stages in the construction of the Cantor–Lebesgue function. Left: The function φ_1. Right: The function φ_2.

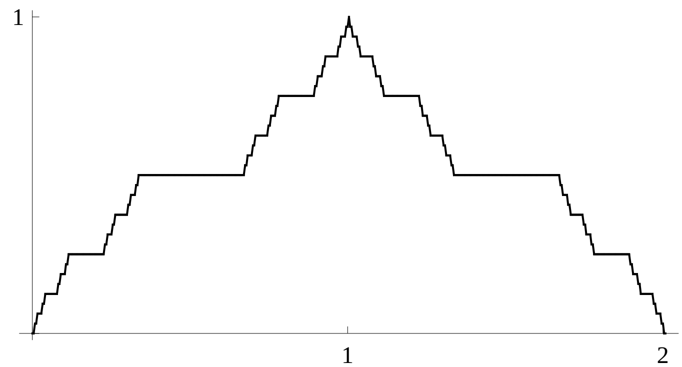

Fig. 1.10. The reflected Devil's staircase (Cantor–Lebesgue function).

1.5 Approximate Identities

Although $L^1(\mathbb{R})$ is closed under convolution, we have seen that it has no identity element. In this section we will show that there are functions in $L^1(\mathbb{R})$ that are "almost" identity elements for convolution. We will construct families of functions $\{k_\lambda\}_{\lambda>0}$ which have the property that $f * k_\lambda$ converges to f in $L^1(\mathbb{R})$ (and in fact in many other senses, depending on what space f belongs to). If k_λ is a "nice" function, then $f * k_\lambda$ will inherit that niceness, and so we will have a nice function that is arbitrarily close to f. Using this procedure we will be able to show that many spaces of nice functions are dense in $L^1(\mathbb{R})$, including $C_c(\mathbb{R})$, $C_c^m(\mathbb{R})$ for each $m \in \mathbb{N}$, and even $C_c^\infty(\mathbb{R})$.

1.5.1 Definition and Existence of Approximate Identities

The properties that a family $\{k_\lambda\}_{\lambda>0}$ will need to possess in order to be an approximate identity for convolution are listed in the next definition.

Definition 1.69. An *approximate identity* or a *summability kernel* is a family $\{k_\lambda\}_{\lambda>0}$ of functions in $L^1(\mathbb{R})$ satisfying

(a) *Normalization*: $\int k_\lambda(x)\,dx = 1$ for every $\lambda > 0$,

(b) L^1-*boundedness*: $\sup_\lambda \|k_\lambda\|_1 < \infty$, and

(c) L^1-*concentration*: For every $\delta > 0$,

$$\lim_{\lambda\to\infty} \int_{|x|\geq\delta} |k_\lambda(x)|\,dx \;=\; 0. \qquad \diamondsuit$$

By definition, an approximate identity is a family of integrable functions. If it is the case that $k_\lambda \geq 0$ for each λ, then requirement (b) in Definition 1.69 follows from requirement (a). However, in general the elements of an approximate identity need not be nonnegative functions.

The "easy" way to create an approximate identity is through dilation of a single function.

Exercise 1.70. Let $k \in L^1(\mathbb{R})$ be any function that satisfies

$$\int k(x)\,dx \;=\; 1.$$

Define k_λ by an L^1-normalized dilation:

$$k_\lambda(x) \;=\; \lambda\,k(\lambda x), \qquad \lambda > 0,$$

and show that the family $\{k_\lambda\}_{\lambda>0}$ forms an approximate identity. $\diamondsuit$

Note that there is an inherent ambiguity in our notation: We may use $\{k_\lambda\}_{\lambda>0}$ to indicate a generic family of functions indexed by λ, or, as introduced in Notation 1.5, we may use k_λ to denote the L^1-normalized dilation of a function k. The intended meaning is usually clear from context.

In any case, if we define k_λ by dilation, then, as λ increases, k_λ becomes more and more similar to our intuition of what a "δ-function" (a function that is an identity for convolution) would look like (see the illustration in Figure 1.7 and the related discussion in Section 1.3.5). While there is no such identity for convolution in $L^1(\mathbb{R})$, the collection of functions $\{k_\lambda\}_{\lambda>0}$ in some sense forms an approximation to this nonexistent δ-function, for requirement (c) implies that k_λ becomes more and more concentrated near the origin as λ increases, with $\int k_\lambda = 1$ for every λ.

Consider also the appearance of $k_\lambda(x) = \lambda k(\lambda x)$ in the frequency domain. By Exercise 1.13, we have

$$\widehat{k_\lambda}(\xi) \;=\; \widehat{k}\Big(\frac{\xi}{\lambda}\Big).$$

Since k is integrable with $\int k = 1$, we know that $\widehat{k}$ is continuous, and

$$\widehat{k}(0) \;=\; \int k \;=\; 1.$$

The continuity of $\widehat{k}$ therefore implies that for each $\xi \in \mathbb{R}$ we have

$$\lim_{\lambda\to\infty} \widehat{k_\lambda}(\xi) \;=\; \lim_{\lambda\to\infty} \widehat{k}\left(\frac{\xi}{\lambda}\right) \;=\; \widehat{k}(0) \;=\; 1 \tag{1.25}$$

(see the illustration in Figure 1.11 using the Gaussian function $g(x) = e^{-\pi x^2}$, which by Exercise 1.109 has the interesting property that $\widehat{g}(\xi) = e^{-\pi\xi^2}$). Thus $\widehat{k_\lambda}$ converges pointwise everywhere to the constant function 1. This again matches our intuition for what the Fourier transform of a "δ-function" would be if there was one, for if there was a function δ that satisfied both $\delta(x) = 0$ for $x \neq 0$ *and* $\int \delta = 1$, then $\widehat{\delta}$ would be identically constant:

$$\widehat{\delta}(\xi) \;=\; \int \delta(x)\, e^{-2\pi i\xi x}\, dx \;=\; \int \delta(x)\, dx \;=\; 1.$$

This contradicts the Riemann–Lebesgue Lemma, so no such function δ can exist. However, when we define δ not as a function but rather as a distribution in Chapter 4 or as a measure in Chapter 5, we will see that $\widehat{\delta}$ does exist and is precisely the constant function.

In any case, given any $f \in L^1(\mathbb{R})$ and an approximate identity $\{k_\lambda\}_{\lambda>0}$ of the form $k_\lambda(x) = \lambda k(\lambda x)$, we have that

$$(f * k_\lambda)^\wedge(\xi) \;=\; \widehat{f}(\xi)\,\widehat{k_\lambda}(\xi) \;\to\; \widehat{f}(\xi), \qquad \xi \in \mathbb{R}.$$

So, we at least have that $(f * k_\lambda)^\wedge(\xi)$ converges pointwise to $\widehat{f}(\xi)$, and this gives us hope that $f * k_\lambda$ should converge to f in other senses as well. Our goal in this section is understand in what sense this hope holds true.

1.5.2 Approximation in $L^p(\mathbb{R})$ by an Approximate Identity

We begin by quantifying the notion that an approximate identity is approximately an identity for convolution in $L^1(\mathbb{R})$. The proof of this theorem illustrates the "standard trick" of introducing k_λ into an equation by virtue of the fact that $\int k_\lambda = 1$, and also the division of the integral into small and large parts in order to make use of the defining properties of an approximate identity.

Theorem 1.71. *Let $\{k_\lambda\}_{\lambda>0}$ be an approximate identity. Then*

$$\forall\, f \in L^1(\mathbb{R}), \qquad \lim_{\lambda\to\infty} \|f - f * k_\lambda\|_1 \;=\; 0.$$

That is, $f * k_\lambda \to f$ in L^1-norm as $\lambda \to \infty$.

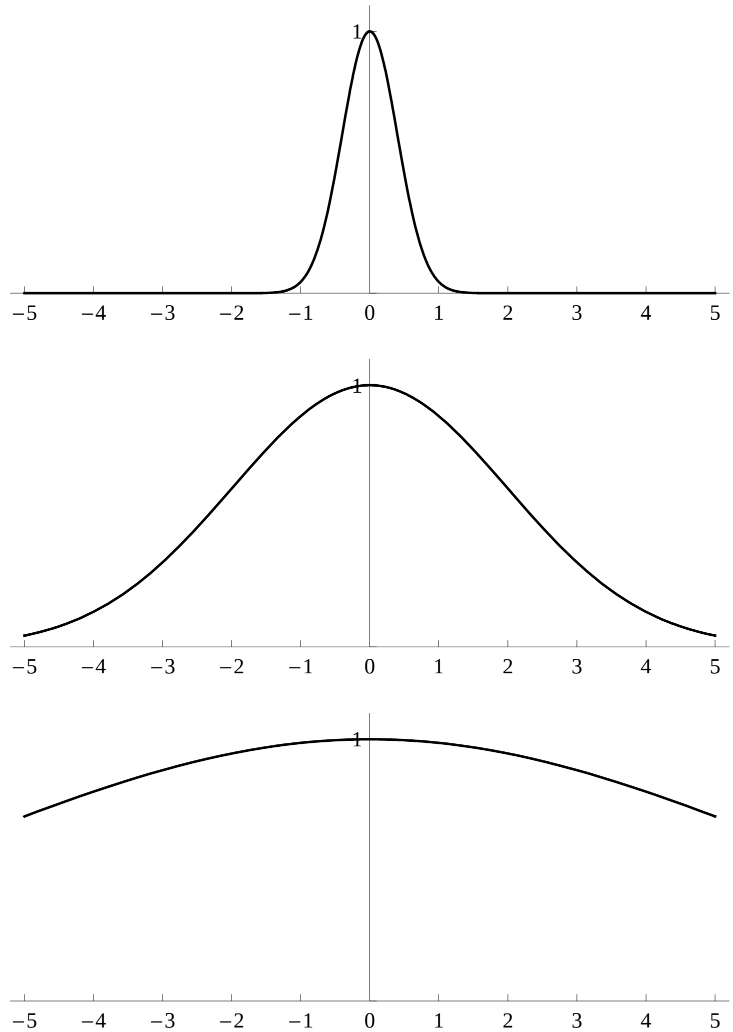

Fig. 1.11. Top: The Fourier transform $\widehat{g}(\xi) = e^{-\pi\xi^2}$ of the function $g(x) = e^{-\pi x^2}$. Middle: The Fourier transform $\widehat{g_5}(\xi) = \widehat{g}(\xi/5)$ of the dilated function $g_5(x) = 5g(5x)$. Bottom: The Fourier transform $\widehat{g_{15}}(\xi) = \widehat{g}(\xi/15)$ of the dilated function $g_{15}(x) = 15g(15x)$.

Proof. Fix any $f \in L^1(\mathbb{R})$. Since $k_\lambda \in L^1(\mathbb{R})$, we know that $f * k_\lambda \in L^1(\mathbb{R})$, and we wish to show that it approximates f well in L^1-norm. Using the fact that $\int k_\lambda = 1$, we compute that

$$\begin{aligned}
\|f - f * k_\lambda\|_1 &= \int |f(x) - (f * k_\lambda)(x)|\, dx \\
&= \int \left| f(x) \int k_\lambda(t)\, dt \;-\; \int f(x-t)\, k_\lambda(t)\, dt \right| dx \\
&\le \iint |f(x) - f(x-t)|\, |k_\lambda(t)|\, dt\, dx \\
&= \iint |f(x) - f(x-t)|\, |k_\lambda(t)|\, dx\, dt \\
&= \int |k_\lambda(t)| \int |f(x) - T_t f(x)|\, dx\, dt \\
&= \int |k_\lambda(t)|\, \|f - T_t f\|_1\, dt, \qquad (1.26)
\end{aligned}$$

where the interchange in the order of integration is permitted by Tonelli's Theorem since the integrands are nonnegative. We want to show that the quantity in equation (1.26) is small when λ is large.

Choose any $\varepsilon > 0$. Since translation is strongly continuous on $L^1(\mathbb{R})$, there exists a $\delta > 0$ such that

$$|t| < \delta \implies \|f - T_t f\|_1 < \varepsilon.$$

Also, by definition of approximate identity, we know that

$$K \;=\; \sup_\lambda \|k_\lambda\|_1 \;<\; \infty,$$

and that there exists some λ_0 such that

$$\lambda > \lambda_0 \implies \int_{|t| \ge \delta} |k_\lambda(t)|\, dt < \varepsilon.$$

Therefore, for $\lambda > \lambda_0$ we can continue equation (1.26) as follows:

$$\begin{aligned}
(1.26) &= \int_{|t|<\delta} |k_\lambda(t)|\, \|f - T_t f\|_1\, dt \;+\; \int_{|t|\ge\delta} |k_\lambda(t)|\, \|f - T_t f\|_1\, dt \\
&\le \int_{|t|<\delta} |k_\lambda(t)|\, \varepsilon\, dt \;+\; \int_{|t|\ge\delta} |k_\lambda(t)|\, \big(\|f\|_1 + \|T_t f\|_1\big)\, dt \\
&\le \varepsilon \int |k_\lambda(t)| \;+\; 2\|f\|_1 \int_{|t|\ge\delta} |k_\lambda(t)|\, dt \\
&\le \varepsilon K \;+\; 2\|f\|_1\, \varepsilon.
\end{aligned}$$

Thus $\|f - f * k_\lambda\|_1 \to 0$ as $\lambda \to \infty$. $\square$

To illustrate the convergence proved in the preceding theorem, consider the particular function $\chi = \chi_{[0,1]}$ and a particular approximate identity that will be of considerable use to us later. This is the *Fejér kernel* $\{w_\lambda\}_{\lambda>0}$, which is produced by dilating the *Fejér function* $w(x) = \left(\frac{\sin \pi x}{\pi x}\right)^2$ depicted in Figure 1.14. In Figure 1.12, we see the convolutions $\chi * w$, $\chi * w_5$, and $\chi * w_{25}$. In addition to the convergence apparent in these figures, note that the convolved functions appear to be continuous, while χ is discontinuous. This is due to the smoothing effect of convolution, which was discussed in Section 1.3.7.

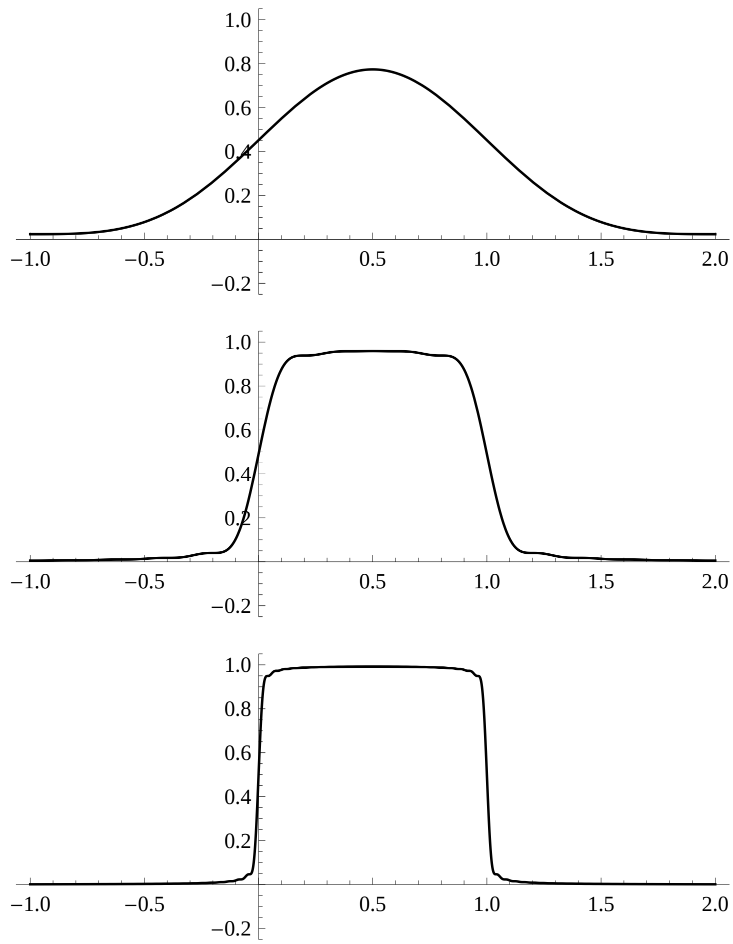

Fig. 1.12. Convolution with an approximate identity. Top: $\chi * w$. Middle: $\chi * w_5$. Bottom: $\chi * w_{25}$.

There are many variations on the theme of Theorem 1.71. To begin, since $L^p * L^1 \subseteq L^p$, we expect that we may be able to extend to other values of p, and indeed for p finite we have the following result.

Exercise 1.72. Let $\{k_\lambda\}_{\lambda>0}$ be an approximate identity. Prove that if $1 \le p < \infty$, then

$$\forall\, f \in L^p(\mathbb{R}), \qquad \lim_{\lambda\to\infty} \|f - f * k_\lambda\|_p \;=\; 0.$$

That is, $f * k_\lambda \to f$ in L^p-norm as $\lambda \to \infty$. $\diamondsuit$

As usual, we do need to be careful when $p = \infty$, and we consider this issue next.

1.5.3 Uniform Convergence

Now we turn to approximation by approximate identities in spaces such as $L^\infty(\mathbb{R})$, $C_0(\mathbb{R})$, and $C_b(\mathbb{R})$. Considering that the elements k_λ of an approximate identity are integrable functions, if f belongs to $L^\infty(\mathbb{R})$ then $f * k_\lambda$ belongs to $C_b(\mathbb{R})$ by Exercise 1.39. Hence it is simply too much to expect that $f * k_\lambda$ will converge to f in L^∞-norm for all $f \in L^\infty(\mathbb{R})$. However, if we impose some smoothness on f, then we can obtain convergence in senses that are appropriate to f.

If we assume that f belongs to $C_0(\mathbb{R})$, then $f * k_\lambda$ also belongs to $C_0(\mathbb{R})$ by Theorem 1.36. By replacing $L^\infty(\mathbb{R})$ with $C_0(\mathbb{R})$, we therefore obtain for $p = \infty$ an exact analogue of the convergence result of Exercise 1.72.

Exercise 1.73. Let $\{k_\lambda\}_{\lambda>0}$ be an approximate identity. Prove that

$$\forall\, f \in C_0(\mathbb{R}), \qquad \lim_{\lambda\to\infty} \|f - f * k_\lambda\|_\infty \;=\; 0.$$

That is, $f * k_\lambda$ converges uniformly to f as $\lambda \to \infty$. $\diamondsuit$

Suppose that we relax the decay hypothesis on f in Exercise 1.73, and assume only that $f \in C_b(\mathbb{R})$ instead of in $C_0(\mathbb{R})$. In this case we do have that $f * k_\lambda \in C_b(\mathbb{R})$ by Exercise 1.39, so we can hope that $f * k_\lambda$ will still converge to f. This is true, but in general the convergence will be uniform only when restricted to compact sets.

Exercise 1.74. Let $\{k_\lambda\}_{\lambda>0}$ be an approximate identity. Prove that if $f \in C_b(\mathbb{R})$, then $f * k_\lambda$ converges uniformly to f on compact sets, i.e.,

$$\forall \,\text{compact } K \subseteq \mathbb{R}, \qquad \lim_{\lambda\to\infty} \big\|(f - f * k_\lambda)\,\chi_K\big\|_\infty \;=\; 0.$$

Give an example of $f \in C_b(\mathbb{R})$ such that $f * k_\lambda$ does not converge uniformly to f on $\mathbb{R}$. $\diamondsuit$

On the other hand, if we impose even a slight amount of "extra smoothness" on $f \in C_b(\mathbb{R})$, then we can restore the uniform convergence of $f * k_\lambda$ to f on the entire real line. We can quantify this extra smoothness in terms of Hölder continuity, which is a generalization of Lipschitz continuity.

Definition 1.75. (a) A function $f : \mathbb{R} \to \mathbb{C}$ is *Hölder continuous with exponent* $\alpha > 0$ *at a point* x if there exists a constant $K > 0$ such that

$$\forall\, y \in \mathbb{R}, \quad |f(x) - f(y)| \;\le\; K\,|x-y|^\alpha.$$

(b) f is *Hölder continuous with exponent* $\alpha > 0$ if there exists a constant $K > 0$ such that

$$\forall\, x, y \in \mathbb{R}, \quad |f(x) - f(y)| \;\le\; K\,|x-y|^\alpha.$$

(c) Given $0 < \alpha < 1$, we set

$$C^\alpha(\mathbb{R}) \;=\; \bigl\{f \in C(\mathbb{R}) : f \text{ is Hölder continuous with exponent } \alpha\bigr\}. \qquad \diamondsuit$$

The reader should verify that the only functions that are Hölder continuous with exponent $\alpha > 1$ are the constant functions. Lipschitz continuity is Hölder continuity with $\alpha = 1$. Every differentiable function f such that f' is bounded is Lipschitz, but a Lipschitz function need not be differentiable everywhere (consider $f(x) = |x|$). Hence $C^1(\mathbb{R})$, the space of function that are differentiable with a continuous derivative, is a proper subspace of the space of Lipschitz functions. A Lipschitz function f has bounded variation on any finite interval, and consequently f is differentiable at almost every point and f' is integrable on any finite interval.

The graph of a function that is Hölder continuous with exponent $\alpha < 1$ typically has a “fractal” appearance. The smaller that we must take α, the more “jagged” the graph of the function appears. For example, the Cantor–Lebesgue function is Hölder continuous but not Lipschitz (Problem 1.42).

Exercise 1.76. Let $\{k_\lambda\}_{\lambda>0}$ be an approximate identity. Prove that if $f \in C_b(\mathbb{R})$ is Hölder continuous for some exponent $0 < \alpha \le 1$, then $f * k_\lambda$ converges uniformly to f on $\mathbb{R}$, i.e.,

$$\lim_{\lambda\to\infty} \|f - f * k_\lambda\|_\infty \;=\; 0. \qquad \diamondsuit$$

1.5.4 Pointwise Convergence

If f is a locally integrable function, then $x \in \mathbb{R}$ is a *Lebesgue point* of f if

$$\lim_{h\to 0} \frac{1}{2h} \int_{x-h}^{x+h} |f(y) - f(x)|\, dy \;=\; 0,$$

and the set of Lebesgue points is called the *Lebesgue set* of f. Every point of continuity is a Lebesgue point of f. More interestingly, the Lebesgue Differentiation Theorem implies that *almost every* $x \in \mathbb{R}$ is a Lebesgue point (Theorem A.30).

Now, we have seen that if f belongs to $L^p(\mathbb{R})$ with p finite then $f * k_\lambda \to f$ in L^p-norm. If we impose some restrictions on k, then we can also show that we have pointwise convergence of $(f * k_\lambda)(x)$ to $f(x)$ at each Lebesgue point x of f.

Theorem 1.77. *Let k be a bounded, compactly supported function such that $\int k = 1$, and define $k_\lambda(x) = \lambda k(\lambda x)$. If $f \in L^1(\mathbb{R})$, then $(f * k_\lambda)(x) \to f(x)$ as $\lambda \to \infty$ for each point x in the Lebesgue set of f. In particular, $f * k_\lambda$ converges to f pointwise a.e.*

Proof. By hypothesis, $\mathrm{supp}(k) \subseteq [-R, R]$ for some R. If x is a Lebesgue point of f, then

$$\begin{aligned}
&\lim_{\lambda\to\infty} |f(x) - (f * k_\lambda)(x)| \\
&\quad = \lim_{\lambda\to\infty} \left| f(x) \int k_\lambda(x-t)\,dt \;-\; \int f(t)\,k_\lambda(x-t)\,dt \right| \\
&\quad \le \lim_{\lambda\to\infty} \lambda \int |f(x) - f(t)|\,|k(\lambda(x-t))|\,dt \\
&\quad = \lim_{\lambda\to\infty} \frac{2R\lambda}{2R} \int_{x-(R/\lambda)}^{x+(R/\lambda)} |f(x) - f(t)|\,|k(\lambda(x-t))|\,dt \\
&\quad \le 2R\,\|k\|_\infty \lim_{h\to 0} \frac{1}{2h} \int_{x-h}^{x+h} |f(x) - f(t)|\,dt \\
&\quad = 0,
\end{aligned}$$

where the limit is zero by definition of Lebesgue point. Finally, since almost every x is a Lebesgue point of f, we conclude that $f * k_\lambda$ converges to f pointwise a.e. □

The hypotheses on k in Theorem 1.77 can be relaxed. In particular, compact support of k is not required. For example, Stein and Weiss [SW71, p. 13] give a more intricate argument that shows that if $k \in L^1(\mathbb{R})$ satisfies $\int k = 1$ and there exists an even function $\phi \in L^1(\mathbb{R})$ that is decreasing and differentiable on $(0, \infty)$ that dominates k in the sense that $|k(x)| \le \phi(x)$ for all x, then $(f * k_\lambda)(x) \to f(x)$ for every x in the Lebesgue set of f.

1.5.5 Dense Sets of Nice Functions

Theorem 1.16 gave us a proof, based on Urysohn's Lemma, that the space $C_c(\mathbb{R})$ is dense in $L^1(\mathbb{R})$. It almost seems that we should be able to give a "simple" proof of this fact by using approximate identities and arguing as follows. Choose any $f \in L^1(\mathbb{R})$. Then we can find a compactly supported $g \in L^1(\mathbb{R})$ that is close to f, e.g., if we take R large enough and set $g = f\,\chi_{[-R,R]}$ then we will have $\|f - g\|_1 < \varepsilon$. If we convolve g with an element k_λ of an approximate identity, then $g * k_\lambda$ will be close to g if λ is large enough, say $\|g - g * k_\lambda\|_1 < \varepsilon$. Further, if we choose our approximate identity so that k_λ is a "nice" function then $g * k_\lambda$ will inherit this "niceness" as well. For example, if $k_\lambda \in C_c(\mathbb{R})$ then $g * k_\lambda \in C_c(\mathbb{R})$, and so we have found an element of $C_c(\mathbb{R})$ that lies within 2ε of f in L^1-norm.

The flaw in this reasoning is that our proof in Theorem 1.71 that $g*k_\lambda \to g$ in L^1-norm relies on the fact that translation is a strongly continuous family of operators in $L^1(\mathbb{R})$. The proof of this strong continuity of translation is the content of Exercise 1.17. *However*, the proof of that exercise (at least the proof we suggest in the hints) requires us to *already* know that $C_c(\mathbb{R})$ is dense in $L^1(\mathbb{R})$. Hence the reasoning of the preceding paragraph is circular.

We could take a different approach, e.g., by first arguing that simple functions are dense in $L^1(\mathbb{R})$ and trying to proceed from there to show that $C_c(\mathbb{R})$ is dense. But it doesn't really matter, one way or the other we essentially have to "get our hands dirty" and show that some particular special subset is dense. The power of approximate identities comes at the next step—once we know that one particular set is dense, we can use the spirit of the argument above (convolution with a "nice" approximate identity) to easily show that "much nicer" spaces are also dense. We even obtain results that almost seem too good to be true. For example, we will see that the space $C_c^\infty(\mathbb{R})$ consisting of *infinitely differentiable, compactly supported functions* is dense in $L^p(\mathbb{R})$ for every $1 \le p < \infty$! This fact is not just an abstract surprise, but will be of great utility to us throughout the remainder of this volume, especially when we turn to distribution theory in Chapter 4.

Exercise 1.78. (a) Show that $C_c^m(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for each $m \ge 0$ and $1 \le p < \infty$, and also is dense in $C_0(\mathbb{R})$ in L^∞-norm.

(b) Show that $C_c^\infty(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for each $1 \le p < \infty$, and also is dense in $C_0(\mathbb{R})$ in L^∞-norm. $\diamondsuit$

After proving the Inversion Formula in Section 1.6, we will also be able to show that many spaces of functions with nice Fourier transforms are also dense. For example, in Problem 1.66 we will see that $\bigl\{f \in L^1(\mathbb{R}) : \widehat{f} \in C_c^\infty(\mathbb{R})\bigr\}$ is dense in $L^p(\mathbb{R})$ for $1 \le p < \infty$.

1.5.6 The C^∞ Urysohn Lemma

We proved Urysohn's Lemma for the setting of the real line in Theorem 1.15. By using approximate identities, we now prove a much more refined C^∞-version of Urysohn's Lemma for subsets of $\mathbb{R}$.

Theorem 1.79 (C^∞ Urysohn Lemma). *Let $K \subseteq \mathbb{R}$ be compact, and let $U \supseteq K$ be an open set. Then there exists $f \in C_c^\infty(\mathbb{R})$ such that $0 \le f \le 1$, $f = 1$ on K, and $\operatorname{supp}(f) \subseteq U$.*

Proof. Since K is compact and $\mathbb{R}\backslash U$ is closed, the distance between these sets is positive, i.e.,

$$d \;=\; \operatorname{dist}(K, \mathbb{R}\backslash U) \;=\; \inf\bigl\{|x-y| : x \in K, y \notin U\bigr\} \;>\; 0.$$

Let

$$V = \Big\{y \in \mathbb{R} : \operatorname{dist}(y, K) < \frac{d}{3}\Big\},$$

and let k be any function in $C_c^\infty(\mathbb{R})$ such that $k \ge 0$, $\int k = 1$, and $\operatorname{supp}(k) \subseteq [-\frac{d}{3}, \frac{d}{3}]$ (for example, dilate the function constructed in Exercise 1.42). Set $f = k * \chi_V$. Since k and χ_V are both compactly supported, their convolution is also compactly supported, and hence it follows from Corollary 1.41 that $f \in C_c^\infty(\mathbb{R})$. Since

$$f(x) = \int_V k(x-y)\, dy \le \int k = 1,$$

we have $0 \le f \le 1$ everywhere. Also, if $x \in K$ and $y \notin V$ then $|x - y| \ge \frac{d}{3}$ and so $k(x - y) = 0$. Therefore for $x \in K$ we have

$$f(x) = \int_V k(x-y)\, dy = \int k(x-y)\, dy = 1.$$

Similarly if $x \notin U$ then it follows that $f(x) = 0$. □

1.5.7 Gibbs's Phenomenon

Gibbs's phenomenon refers to the behavior of the partial sums of the Fourier series of a periodic function near a jump discontinuity. Although this chapter focuses on the Fourier transform on $\mathbb{R}$ rather than Fourier series on $\mathbb{T}$, there is an analogous phenomenon for the Fourier transform that we will discuss. For the formulation and proof of Gibbs's phenomenon on the torus, we refer to [DM72].

To illustrate this, let $H = \chi_{[0,\infty)}$ be the *Heaviside function.* Although H does not belong to $L^1(\mathbb{R})$, the important fact for this example is that H has a jump discontinuity at $x = 0$. Pointwise convergence of Fourier series corresponds to convolution with the Dirichlet kernel on the torus (see Section 2.2.1, and equation (2.9) in particular). The Dirichlet kernel on the real line is $\{d_\lambda\}_{\lambda>0}$, which is obtained by dilating the Dirichlet function

$$d(\xi) = \frac{\sin \xi}{\pi \xi}.$$

Since $d \notin L^1(\mathbb{R})$, the Dirichlet kernel does not form an approximate identity, but even so let us consider the pointwise behavior of $H * d_\lambda$ as $\lambda \to \infty$. Using the fact that, as an improper Riemann integral, $\int_0^\infty \frac{\sin x}{x}\, dx = \frac{\pi}{2}$ (see Problem 1.43), we have

$$\begin{aligned}(H * d_\lambda)(x) &= \int_0^\infty \frac{\sin \lambda(x-y)}{\pi(x-y)}\, dy \\ &= \int_{-\infty}^{\lambda x} \frac{\sin y}{\pi y}\, dy\end{aligned}$$

$$= \frac{1}{2} + \int_0^{\lambda x} \frac{\sin y}{\pi y}\, dy.$$

Since $\frac{\sin y}{\pi y} > 0$ for $0 < y < \pi$, $H * d_\lambda$ is increasing on $(0, \pi/\lambda)$. Then it decreases on $(\pi/\lambda, 2\pi/\lambda)$, then increases on $(2\pi/\lambda, 3\pi/\lambda)$—but never back to the value it had at π/λ. Continuing in this way we see that $H * d_\lambda$ achieves its maximum at $x = \pi/\lambda$, and this maximum is

$$(H * d_\lambda)\Big(\frac{\pi}{\lambda}\Big) = \frac{1}{2} + \int_0^{\pi} \frac{\sin y}{\pi y}\, dy \approx 1.089\ldots.$$

Note that this maximum is *independent of* λ. Although $(H * d_\lambda)(x)$ converges pointwise to $H(x)$ as $\lambda \to \infty$ for all $x \neq 0$, this convergence is *not uniform.* In particular, the maximum distance between $(H * d_\lambda)(x)$ and $H(x)$ for $x > 0$ is a constant (approximately 0.089...) that is independent of λ, although its *location* at $x = \pi/\lambda$ decreases with λ. Figure 1.13 displays a plot of $H * d_\lambda$ for $\lambda = 16\pi$.

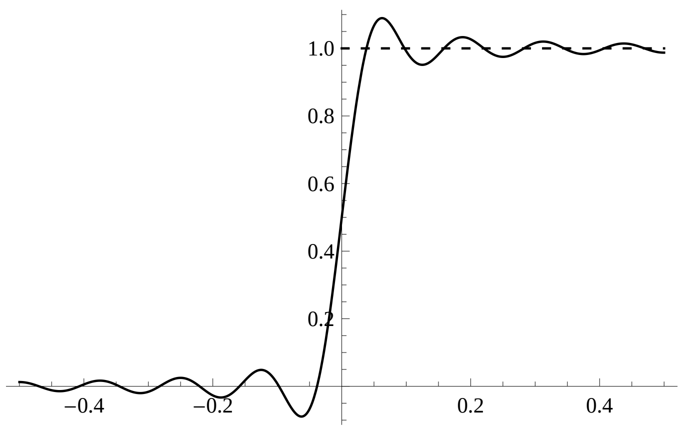

Fig. 1.13. Graph of $H * d_\lambda$ for $\lambda = 16\pi$.

1.5.8 Translation-Invariant Subspaces of $L^1(\mathbb{R})$

The following characterization of the closed, translation-invariant subspaces of $L^1(\mathbb{R})$ is due to Norbert Wiener (1894–1964); compare this to the characterization of the closed, translation-invariant subspaces of $L^2(\mathbb{R})$ given in Problem 3.10.

Definition 1.80 (Translation-Invariant Subspaces). We say that a subset J of $L^1(\mathbb{R})$ is *translation-invariant* if J is closed under all translations, i.e., if

$$f \in J,\ a \in \mathbb{R} \implies T_a f \in J. \qquad \diamondsuit$$

Remark 1.81. A *shift-invariant* space V is one that is closed under *integer* translations, i.e., if $f \in V$ then $T_k f \in V$ for all integer k. These spaces play important roles in sampling theory and wavelet theory, see [Dau92]. $\diamondsuit$

Exercise 1.82. Let J be a closed subspace of $L^1(\mathbb{R})$, and prove that the following statements are equivalent.

(a) J is translation-invariant.

(b) J is an ideal in $L^1(\mathbb{R})$ under convolution. $\diamondsuit$

Note that closedness is important here. For example, $C_c(\mathbb{R})$ is a translation-invariant subspace of $L^1(\mathbb{R})$, but it is not an ideal with respect to convolution.

By using Exercise 1.82, we can give another characterization of the principal ideal $I(g) = \overline{g * L^1(\mathbb{R})}$ generated by a function $g \in L^1(\mathbb{R})$. By Exercise 1.48 and Problem 1.39, $I(g)$ is the smallest closed ideal in $L^1(\mathbb{R})$ that contains g.

Exercise 1.83. Given $g \in L^1(\mathbb{R})$, show that $I(g)$ is the closure of the finite linear span of the translations of g:

$$I(g) \;=\; \overline{\operatorname{span}}\big(\{T_a g\}_{a\in\mathbb{R}}\big). \qquad \diamondsuit$$

Thus, the smallest closed ideal that contains g is precisely the smallest closed subspace that contains all translates of g.

Recall that a set $S \subseteq L^1(\mathbb{R})$ is *complete* in $L^1(\mathbb{R})$ if $\operatorname{span}(S)$ is dense in $L^1(\mathbb{R})$. It therefore follows from Exercise 1.83 that $\{T_a g\}_{a\in\mathbb{R}}$ is complete in $L^1(\mathbb{R})$ if and only if $I(g) = L^1(\mathbb{R})$. But when does this happen? The next exercise gives a necessary condition.

Exercise 1.84. Given $g \in L^1(\mathbb{R})$, show that

$$\{T_a g\}_{a\in\mathbb{R}} \text{ is complete in } L^1(\mathbb{R}) \quad\Longrightarrow\quad \widehat{g}(\xi) \neq 0 \text{ for all } \xi \in \mathbb{R}. \qquad \diamondsuit$$

The converse of Exercise 1.84 is also true, but is a much deeper fact that is part of *Wiener's Tauberian Theorem*, which we discuss in Section 2.10. In contrast, the analogous question in $L^2(\mathbb{R})$ is much simpler, see Problem 3.9.

Additional Problems

1.34. This problem constructs an example of an approximate identity $\{k_\lambda\}_{\lambda>0}$ where k_λ need not be a dilation of a single function k. For each $\lambda > 0$, let $k_\lambda \in L^1(\mathbb{R})$ be any function satisfying $k_\lambda \geq 0$, $\|k_\lambda\|_1 = \int k_\lambda = 1$, and $\operatorname{supp}(k_\lambda) \subseteq [-\frac{1}{\lambda}, \frac{1}{\lambda}]$. Show that $\{k_\lambda\}_{\lambda>0}$ is an approximate identity.

1.35. Show that if $\{k_\lambda\}_{\lambda>0}$ is an approximate identity then $\widehat{k_\lambda}(\xi) \to 1$ pointwise as $\lambda \to \infty$.

1.36. Let $\{k_\lambda\}_{\lambda>0}$ be an approximate identity and fix $f \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. Show that if f is continuous at some point $x \in \mathbb{R}$, then $(f * k_\lambda)(x) \to f(x)$ as $\lambda \to \infty$.

1.37. Fix $g \in L^1(\mathbb{R})$ and define $L_g\colon L^1(\mathbb{R}) \to L^1(\mathbb{R})$ by $L_g(f) = f * g$. Show that the operator norm of L_g is $\|L_g\| = \|g\|_1$. This gives the best constant in equation (1.15) for the case $p = q = r = 1$.

1.38. Assume $k \in L^1(\mathbb{R})$ is given and we set $r = \int k$ and $k_\lambda(x) = \lambda k(\lambda x)$. Prove that if $1 \le p < \infty$, then for each $f \in L^p(\mathbb{R})$ we have that $f * k_\lambda \to rf$ in L^p-norm as $\lambda \to \infty$. Note that this includes the possibility that r may be complex or zero.

1.39. Fix $g \in L^1(\mathbb{R})$, and consider the ideals $g * L^1(\mathbb{R})$ and $I(g)$ introduced in Exercise 1.48. Show that we need not have $g \in g * L^1(\mathbb{R})$, but we always have $g \in I(g)$.

1.40. Let $k \in L^1(\mathbb{R})$ be any function such that $\int k = 1$ and $xk(x) \in L^1(\mathbb{R})$, and define an approximate identity by setting $k_\lambda(x) = \lambda k(\lambda x)$. Fix $1 \le p < \infty$.

(a) Show that $\|Pk_\lambda\|_1 \to 0$ as $\lambda \to \infty$, where P is the position operator.

(b) Show that if $f \in L^p(\mathbb{R})$ then $f * Pk_\lambda \to 0$ in L^p-norm. Further, if we also have $xf(x) \in L^p(\mathbb{R})$, then $P(f * k_\lambda) \to Pf$ in L^p-norm.

1.41. Given $0 < \alpha < 1$, let $C^\alpha(\mathbb{R})$ be the space of Hölder continuous functions given in Definition 1.75. Show that

$$\|f\|_{C^\alpha} \;=\; |f(0)| \;+\; \sup_{x \neq y} \frac{f(x) - f(y)}{|x - y|^\alpha}$$

is a norm on $C^\alpha(\mathbb{R})$, and $C^\alpha(\mathbb{R})$ is complete with respect to this norm.

1.42. Show that the Cantor–Lebesgue function is Hölder continuous precisely for exponents α in the range $0 < \alpha \le \log_3 2 \approx 0.6309\dots$.

1.43. (a) Show that $\int_0^\infty \left|\frac{\sin x}{x}\right| dx = \infty$.

(b) If $x > 0$, then $\int_0^\infty e^{-tx}\, dt = \frac{1}{x}$. Combine this with Fubini's Theorem to evaluate $\int_0^a \frac{\sin x}{x}\, dx$. Then apply the Lebesgue Dominated Convergence Theorem to show that $\lim_{a\to\infty} \int_0^a \frac{\sin x}{x}\, dx = \frac{\pi}{2}$.

Thus, even though $\frac{\sin x}{x}$ is not integrable, the improper Riemann integral $\int_0^\infty \frac{\sin x}{x}\, dx$ exists and equals $\frac{\pi}{2}$ (this integral can also be evaluated by using contour integration).

(c) Show that $\sup_{a<b} \left|\int_a^b \frac{\sin x}{x}\, dx\right| < \infty$.

1.44. This problem will illustrate one of the many different possible generalizations of the results of this section by considering particular *weighted L^p spaces*. Given $s \in \mathbb{R}$ let $v_s(x) = (1+|x|)^s$; we refer to v_s as a *polynomial weight* (since it has polynomial-like growth if $s \ge 0$, or decays like the reciprocal of a polynomial if $s \le 0$). For this problem fix $1 \le p < \infty$ and $s \in \mathbb{R}$. Then we define $L^p_s(\mathbb{R})$ to be the space of all measurable functions $f\colon \mathbb{R} \to \mathbb{C}$ such that

$$\|f\|_{p,s} \;=\; \|f\,v_s\|_p \;=\; \left(\int |f(x)|^p\,(1+|x|)^{ps}\,dx\right)^{1/p} \;<\; \infty.$$

If we identify functions that are equal a.e. then $L^p_s(\mathbb{R})$ is a Banach space. Prove the following statements.

(a) If $s \ge 0$ then v_s is *submultiplicative*, i.e., $v_s(x+y) \le v_s(x)\,v_s(y)$ for $x, y \in \mathbb{R}$. If $s \le 0$ then v_s is *v_{-s}-moderate*, i.e., $v_s(x+y) \le v_s(x)\,v_{-s}(y)$ for $x, y \in \mathbb{R}$.

(b) For each $a \in \mathbb{R}$, the translation operator T_a is a continuous map of $L^p_s(\mathbb{R})$ into itself, with operator norm $\|T_a\|_{L^p_s \to L^p_s} \le v_{|s|}(a) = (1+|a|)^{|s|}$.

(c) Translation is strongly continuous on $L^p_s(\mathbb{R})$, i.e., for each $f \in L^p_s(\mathbb{R})$ we have $\lim_{a\to 0} \|f - T_a f\|_{p,s} = 0$.

(d) If $\{k_\lambda\}_{\lambda>0}$ is an approximate identity, then for each $f \in L^p_s(\mathbb{R})$ we have $\lim_{\lambda\to\infty} \|f - f * k_\lambda\|_{p,s} = 0$.

(e) $C_c^\infty(\mathbb{R})$ is dense in $L^p_s(\mathbb{R})$.

(f) Do parts (b)–(d) still hold if we replace v_s by an arbitrary weight $w\colon \mathbb{R} \to (0,\infty)$? What properties do we need w to possess?

1.6 The Inversion Formula

In this section we will prove Theorem 1.9, the Inversion Formula for the Fourier transform on $L^1(\mathbb{R})$, which states that if f and $\widehat{f}$ both belong to $L^1(\mathbb{R})$, then $f = \big(\widehat{f}\,\big)^{\vee} = \big(\check{f}\,\big)^{\wedge}$.

Unfortunately, the Inversion Formula does not apply to every function in $L^1(\mathbb{R})$. For example, by Exercise 1.7 we have that $\chi = \chi_{[-1/2,1/2]}$ belongs to $L^1(\mathbb{R})$ but $\widehat{\chi}(\xi) = d_\pi(\xi) = \frac{\sin \pi\xi}{\pi\xi}$ is not integrable. Another example is given in Problem 1.1: The one-sided exponential function $f(x) = e^{-x}\,\chi_{[0,\infty)}(x)$ is integrable, but its Fourier transform $\widehat{f}(\xi) = \frac{1}{2\pi i\xi+1}$ is not.

1.6.1 The Fejér Kernel

To prove the Inversion Formula, we will use the machinery of approximate identities that we developed in Section 1.5. Specifically, we will use a particular approximate identity, named after Lipót Fejér (1880–1959), that has some useful special properties.

Definition 1.85 (Fejér Kernel). The *Fejér function* is the square of the sinc function d_π:

$$w(x) \;=\; \left(\frac{\sin \pi x}{\pi x}\right)^2 \;=\; d_\pi(x)^2.$$

The *Fejér kernel* is $\{w_\lambda\}_{\lambda>0}$ where $w_\lambda(x) = \lambda w(\lambda x)$. $\diamondsuit$

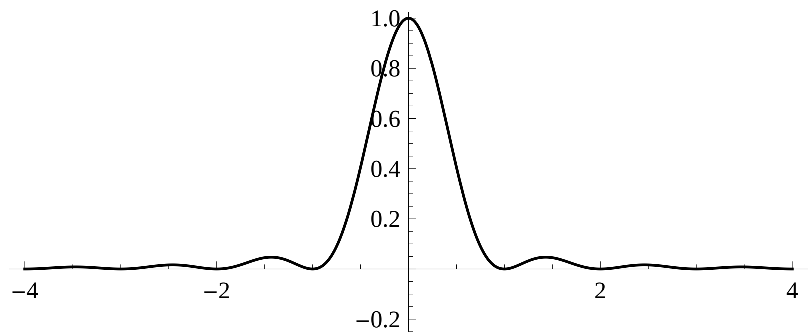

Fig. 1.14. Graph of the Fejér function w.

The letter "w" is for "Weiss," which was Fejér's surname at birth.

In order to conclude that the Fejér kernel is an approximate identity, we need to know that the integral of w is 1.

Exercise 1.86. Show that $w \in L^1(\mathbb{R})$ and $\int w = 1$. $\diamond$

Let $\chi = \chi_{[-1/2,1/2]}$. Then we know from Exercise 1.7 that $\widehat{\chi}(\xi) = \frac{\sin \pi\xi}{\pi\xi} = d_\pi(\xi)$. Hence if we let

$$W \;=\; \chi * \chi,$$

then we have

$$\widehat{W} \;=\; \widehat{\chi}^{\,2} \;=\; d_\pi^2 \;=\; w. \tag{1.27}$$

Also $\overset{\vee}{W} = w$ since W is even. Since both W and $\widehat{W} = w$ belong to $L^1(\mathbb{R})$, if we had already proved that the Inversion Formula holds, then we could apply it to W to conclude that w's Fourier transform is W, i.e.,

$$\widehat{w} \;=\; (\overset{\vee}{W})^\wedge \;=\; W \qquad \text{and} \qquad \overset{\vee}{w} \;=\; (\widehat{W})^\vee \;=\; W. \tag{1.28}$$

We will see eventually that this is true, but first we must prove the Inversion Formula. And we will prove the Inversion Formula by making use of the Fejér kernel, although we will not need equation (1.28) to do this. Obviously, this is a good thing, since otherwise the argument would be circular.

It is not so much the Fejér kernel itself that will be important to us, but rather some of the properties that it happens to have, including:

(a) $w,\, W \geq 0$,

(b) $w(0) = 1$,

(c) $W,\, \widehat{W} \in L^1(\mathbb{R})$,

(d) $\int W = \widehat{W}(0) = 1$.

Even these are not all essential, and the reader can consider what other kernels we might use instead to prove the Inversion Formula.

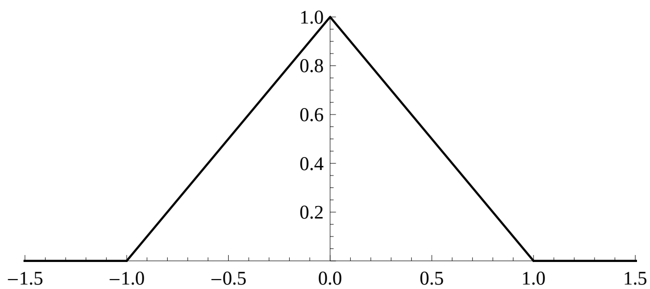

Fig. 1.15. Graph of the hat function W.

Exercise 1.87. Prove that W is the *hat function* or *tent function* on the interval $[-1,1]$ defined by

$$W(\xi) \;=\; \max\{1-|\xi|,\, 0\}. \qquad \diamondsuit$$

Consequently,

$$W(\xi/\lambda) \;=\; \max\Bigl\{1-\frac{|\xi|}{\lambda},\, 0\Bigr\} \tag{1.29}$$

is the dilated hat function with height 1 supported on $[-\lambda, \lambda]$. In particular, $W(\xi/\lambda) \to 1$ pointwise as $\lambda \to \infty$.

1.6.2 Proof of the Inversion Formula

To motivate our first step towards the proof of the Inversion Formula, choose any $f \in L^1(\mathbb{R})$. Then $f * w_\lambda \in L^1(\mathbb{R})$ and also $(f * w_\lambda)^\wedge = \widehat{f}\,\widehat{w_\lambda}$. We don't know this yet, but we will show that $\widehat{w_\lambda}(\xi) = W(\xi/\lambda) \in L^1(\mathbb{R})$. Once we establish this, it follows (since $\widehat{f}$ is bounded) that

$$(f * w_\lambda)^\wedge \;=\; \widehat{f}\,\widehat{w_\lambda} \;\in\; L^1(\mathbb{R}).$$

Hence, if only we already knew that the Inversion Formula was valid, we could compute that

$$(f * w_\lambda)(x) \;=\; \bigl((f * w_\lambda)^\wedge\bigr)^\vee(x) \;=\; (\widehat{f}\,\widehat{w_\lambda})^\vee(x) \;=\; \int \widehat{f}(\xi)\, W(\xi/\lambda)\, e^{2\pi i \xi x}\, d\xi.$$

Unfortunately, these calculations are not yet justified since, among other things, they rely on the Inversion Formula, which has not yet been proved. However, instead of trying to justify the full Inversion Formula right now, we begin with a much smaller step: We show *directly* that the specific equality $(f * w_\lambda)(x) = \int \widehat{f}(\xi)\, W(\xi/\lambda)\, e^{2\pi i \xi x}\, d\xi$ holds when $f \in L^1(\mathbb{R})$.

Theorem 1.88. *If $f \in L^1(\mathbb{R})$, then for each $\lambda > 0$ we have*

$$
\begin{aligned}
(f * w_\lambda)(x) &= \int \widehat{f}(\xi)\, W(\xi/\lambda)\, e^{2\pi i \xi x}\, d\xi \\
&= \int_{-\lambda}^{\lambda} \widehat{f}(\xi) \left(1 - \frac{|\xi|}{\lambda}\right) e^{2\pi i \xi x}\, d\xi. \qquad (1.30)
\end{aligned}
$$

Proof. We already know that

$$w(x) = \overset{\vee}{W}(x) = \int W(\xi)\, e^{2\pi i \xi x}\, d\xi,$$

and therefore by making a change of variables it follows that

$$w_\lambda(x) = \lambda w(\lambda x) = \int W(\xi/\lambda)\, e^{2\pi i \xi x}\, d\xi.$$

Suppose that $f \in L^1(\mathbb{R})$ is given. Assuming for the moment that the use of Fubini's Theorem in the following calculation is justified, we have that

$$
\begin{aligned}
(f * w_\lambda)(x) &= \int f(y)\, w_\lambda(x-y)\, dy \\
&= \int f(y) \int W(\xi/\lambda)\, e^{2\pi i \xi (x-y)}\, d\xi\, dy \\
&= \int \int_{-\lambda}^{\lambda} f(y) \left(1 - \frac{|\xi|}{\lambda}\right) e^{2\pi i \xi (x-y)}\, d\xi\, dy \\
&= \int_{-\lambda}^{\lambda} \left(\int f(y)\, e^{2\pi i \xi (x-y)}\, dy \right) \left(1 - \frac{|\xi|}{\lambda}\right) d\xi \\
&= \int_{-\lambda}^{\lambda} \widehat{f}(\xi) \left(1 - \frac{|\xi|}{\lambda}\right) e^{2\pi i \xi x}\, d\xi.
\end{aligned}
$$

Of course, the applicability of Fubini's Theorem does have to be justified, and we assign this task to the reader as an exercise. $\square$

Now we obtain the Inversion Formula by taking the limit in equation (1.30).

Theorem 1.89 (Inversion Formula). *If $f,\ \widehat{f} \in L^1(\mathbb{R})$, then f and $\widehat{f}$ are continuous, and*

$$f(x) = \bigl(\widehat{f}\,\bigr)^{\vee}(x) = \int \widehat{f}(\xi)\, e^{2\pi i \xi x}\, d\xi, \qquad (1.31)$$

with equality holding pointwise everywhere, and similarly

$$f(x) = \bigl(\overset{\vee}{f}\,\bigr)^{\wedge}(x) = \int \overset{\vee}{f}(\xi)\, e^{-2\pi i \xi x}\, d\xi$$

for every x.

Proof. Since $f \in L^1(\mathbb{R})$, we know that $\widehat{f}$ is a continuous function on $\mathbb{R}$. Therefore, for any fixed x, we have the pointwise limit

$$\forall\, \xi \in \mathbb{R}, \quad \lim_{\lambda\to\infty} \widehat{f}(\xi)\, W(\xi/\lambda)\, e^{2\pi i \xi x} \;=\; \widehat{f}(\xi)\, e^{2\pi i \xi x}.$$

Further, since $0 \le W \le 1$,

$$|\widehat{f}(\xi)\, W(\xi/\lambda)\, e^{2\pi i \xi x}| \;\le\; |\widehat{f}(\xi)| \;\in\; L^1(\mathbb{R}).$$

Since $f * w_\lambda$ is continuous by Theorem 1.36, it is defined pointwise everywhere. By Theorem 1.88 and the Lebesgue Dominated Convergence Theorem, we therefore have for each x that

$$\begin{aligned} \lim_{\lambda\to\infty} (f * w_\lambda)(x) \;&=\; \lim_{\lambda\to\infty} \int \widehat{f}(\xi)\, W(\xi/\lambda)\, e^{2\pi i \xi x}\, d\xi \\ &=\; \int \widehat{f}(\xi)\, e^{2\pi i \xi x}\, d\xi \;=\; \big(\widehat{f}\,\big)^{\vee}(x). \end{aligned}$$

On the other hand, $f * w_\lambda \to f$ in L^1-norm, so there is a subsequence such that $(f * w_{\lambda_k})(x) \to f(x)$ for almost every x. Therefore $\big(\widehat{f}\,\big)^{\vee}(x) = f(x)$ a.e. Since $\big(\widehat{f}\,\big)^{\vee}$ is continuous, by redefining f on a set of measure zero we can assume that equality holds pointwise everywhere. $\square$

Observe that the hypothesis $f,\, \widehat{f} \in L^1(\mathbb{R})$ can be equivalently restated as

$$f,\, \widehat{f} \in L^1(\mathbb{R}) \qquad \Longleftrightarrow \qquad f \in L^1(\mathbb{R}) \cap A(\mathbb{R}).$$

Corollary 1.90. *If $f,\, \widehat{f} \in L^1(\mathbb{R})$, then $f,\, \widehat{f} \in C_0(\mathbb{R})$.* $\diamondsuit$

One consequence of the Inversion Formula is that we can now justify our hope, presented in equation (1.28), that $\widehat{w} = W$. Note that since w and W are even, their Fourier and inverse Fourier transforms coincide.

Corollary 1.91. $\widehat{w} = (\check{W})^{\wedge} = W = (\widehat{W})^{\vee} = \check{w}$.

Proof. This follows from the Inversion Formula, the fact that w and W both belong to $L^1(\mathbb{R})$, and our proof in equation (1.27) that $\widehat{W} = w$. $\square$

An additional consequence is that the Fourier transform is an injective map on $L^1(\mathbb{R})$.

Theorem 1.92 (Uniqueness Theorem). *If $f,\, g \in L^1(\mathbb{R})$, then*

$$f = g \;\; a.e. \qquad \Longleftrightarrow \qquad \widehat{f} = \widehat{g} \;\; a.e.$$

In particular,

$$f = 0 \;\; a.e. \qquad \Longleftrightarrow \qquad \widehat{f} = 0 \;\; a.e.$$

Proof. Since the Fourier transform is linear, the first equivalence is a consequence of the second. If $f = 0$ a.e., then $\widehat{f} = 0$ everywhere by definition of the Fourier transform. On the other hand, if $\widehat{f} = 0$ a.e., then we have both $f, \widehat{f} \in L^1(\mathbb{R})$, so the Inversion Formula applies, and we obtain $f = \big(\widehat{f}\,\big)^{\vee} = \check{0} = 0$. □

Note that we could also appeal to Theorem 1.88 to give another proof of the Uniqueness Theorem.

1.6.3 Summability

If $\widehat{f} \notin L^1(\mathbb{R})$ then the Inversion Formula need not hold. However, as an immediate consequence of Theorem 1.88, we have the following *approximation formula* for f in terms of $\widehat{f}$ that is valid for all $f \in L^1(\mathbb{R})$, even if $\widehat{f}$ is not integrable.

Theorem 1.93. *If $f \in L^1(\mathbb{R})$, then*

$$\int_{-\lambda}^{\lambda} \widehat{f}(\xi)\left(1 - \frac{|\xi|}{\lambda}\right) e^{2\pi i \xi x}\, d\xi \;\to\; f \qquad \text{in } L^1\text{-norm as } \lambda \to \infty. \tag{1.32}$$

Proof. Simply note that, by Theorem 1.88, the left-hand side of equation (1.32) is $f * w_\lambda$. □

Thus, even if $\widehat{f} \notin L^1(\mathbb{R})$, we have an "approximate" interpretation of the formula $f = \big(\widehat{f}\,\big)^{\vee}$ in the sense that equation (1.32) will hold. This is analogous to using summability conditions for evaluating divergent series. For example, consider a formal[1] bi-infinite series of the form $\sum_{k=-\infty}^{\infty} a_k$. Let us say that this series converges and equals L if the symmetric partial sums

$$s_N = \sum_{k=-N}^{N} a_k$$

converge to L. The *Cesàro means* or *arithmetic means* of these partial sums are

$$\sigma_N \;=\; \frac{s_0 + s_1 + \cdots + s_N}{N+1}.$$

If the series $\sum_{k=-\infty}^{\infty} a_k$ converges (for example, if $a = (a_k)_{k\in\mathbb{Z}}$ is a summable sequence), then the Cesàro means will converge to the same limit. However, the Cesàro means may converge even when the partial sums do not; if the Cesàro means converge then we say that $\sum_{k=-\infty}^{\infty} a_k$ is *Cesàro summable.* Further, the form of the Cesàro means given in the next exercise suggests an interesting analogy with Theorem 1.93.

[1] It is an amusing linguistic footnote that a *formal statement* in mathematics means that the statement is entirely *informal.* In particular, we use the term "formal" at this point because the symbols $\sum_{k=-\infty}^{\infty} a_k$ denote a completely arbitrary series, without any requirement that it converge in any sense whatsoever.

Exercise 1.94. Given a sequence of scalars $a = (a_k)_{k\in\mathbb{Z}}$, let the partial sums s_N and Cesàro means σ_N be as above.

(a) Show that

$$\sigma_N \;=\; \sum_{k=-N}^{N} \Bigl(1 - \frac{|k|}{N+1}\Bigr)\, a_k.$$

(b) Show that if the partial sums s_N converge, then the Cesàro means σ_N converge to the same limit, i.e.,

$$\lim_{N\to\infty} \sum_{k=-N}^{N} \Bigl(1 - \frac{|k|}{N+1}\Bigr)\, a_k \;=\; \lim_{N\to\infty} s_N \;=\; \sum_{k=-\infty}^{\infty} a_k. \tag{1.33}$$

(c) Set $a_k = (-1)^k$ for $k \ge 0$ and $a_k = 0$ for $k < 0$. Show that the series $\sum a_k$ is Cesàro summable even though the partial sums do not converge, and find the limit of the Cesàro means.

(d) Show that if $a_k \ge 0$ for every k then σ_N converges if and only if s_N converges. $\diamondsuit$

Comparing Theorem 1.93 to equation (1.33), we see that Theorem 1.93 is a continuous version of Cesàro summability. Essentially, Theorem 1.93 says that if $f \in L^1(\mathbb{R})$, then the formal integral

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

is Cesàro summable to f even if $\widehat{f}$ is not integrable.

It is possible to make this analogy even more precise. Just as the integral in equation (1.32) is related to the convolution of f with the Fejér kernel, specifically,

$$(f * w_\lambda)(x) \;=\; (\widehat{f}\,\widehat{w_\lambda})^{\vee}(x) \;=\; \int_{-\lambda}^{\lambda} \widehat{f}(\xi)\, \Bigl(1 - \frac{|\xi|}{\lambda}\Bigr)\, e^{2\pi i \xi x}\, d\xi,$$

it is also true that Cesàro means of an infinite series are related to a discrete version of the Fejér kernel. Indeed, when comparing the Fourier transform to Fourier series, the two Fejér kernels play entirely analogous roles. We expand on this in Section 2.2.

As an application, we use Theorem 1.93 to obtain useful results regarding functions with compactly supported Fourier transforms. Such functions appear in a variety of contexts. For example, we will see them again in Chapter 3 when we prove the Paley–Wiener Theorem (Section 3.5) and the energy concentration theorems (Section 3.8).

Definition 1.95 (Bandlimited Function). We say that a function f is *bandlimited* if $\operatorname{supp}(\widehat{f}\,)$ is compact. $\diamondsuit$

Exercise 1.96. (a) Show that the space of bandlimited functions in $L^1(\mathbb{R})$,

$$\{f \in L^1(\mathbb{R}) : \operatorname{supp}(\widehat{f}\,) \text{ is compact}\},$$

is dense in $L^1(\mathbb{R})$.

(b) Show that if Ω is fixed, then the space of functions in $L^1(\mathbb{R})$ bandlimited to $[-\Omega, \Omega]$,

$$\{f \in L^1(\mathbb{R}) : \operatorname{supp}(\widehat{f}\,) \subseteq [-\Omega, \Omega]\},$$

is a closed proper subspace of $L^1(\mathbb{R})$. $\diamondsuit$

1.6.4 Pointwise Inversion

As we have seen, the Cesàro means σ_N of a formal infinite series may converge even when the partial sums s_N do not. For the Fourier transform, the analogue of the Cesáro means are convolutions with the Fejér kernel:

$$(f * w_\lambda)(x) \;=\; \int_{-\lambda}^{\lambda} \widehat{f}(\xi) \Big(1 - \frac{|\xi|}{\lambda}\Big)\, e^{2\pi i \xi x}\, d\xi.$$

We always have $f * w_\lambda \to f$ in L^1-norm when $f \in L^1(\mathbb{R})$.

Similarly, the analogue of the partial sums for the Fourier transform are convolutions with the Dirichlet kernel. To see this, write $\chi_\lambda = \chi_{[-\lambda,\lambda]}$. Then $\check{\chi}_\lambda = d_{2\pi\lambda}$, so

$$\begin{aligned}
(f * d_{2\pi\lambda})(x) \;&=\; \int f(y)\, \check{\chi}_\lambda(x-y)\, dy \\
&=\; \int f(y) \int_{-\lambda}^{\lambda} e^{2\pi i \xi (x-y)}\, d\xi\, dy \\
&=\; \int_{-\lambda}^{\lambda} \Big(\int f(y)\, e^{2\pi i \xi y}\, dy\Big)\, e^{2\pi i \xi x}\, d\xi \\
&=\; \int_{-\lambda}^{\lambda} \widehat{f}(\xi)\, e^{2\pi i \xi y}\, dy.
\end{aligned}$$

The interchange in the order of integration in this computation is justified by Fubini's Theorem, since $f(y)\, e^{2\pi i \xi y}\, \chi_\lambda(\xi) \in L^1(\mathbb{R} \times [-\lambda, \lambda])$.

Unfortunately, the Dirichlet kernel is not an approximate identity, and the question of convergence of $f * d_{2\pi\lambda}$ is much more delicate than the convergence of $f * w_\lambda$. In general, if we know only that $f \in L^1(\mathbb{R})$ then $f * d_{2\pi\lambda}$ need not converge to f in L^1-norm. However, we will show that if we impose a modest amount of local regularity on f near a given point x then $(f * d_{2\pi\lambda})(x)$ will converge to $f(x)$ for that x.

The "local regularity" that we need is bounded variation on an interval around x, and the main part of the argument is contained in the following lemma. The proof of this lemma uses the Second Mean Value Theorem for integrals. Problem 1.55 gives a short proof of a special case of the Second Mean Value Theorem, and for the proof of the general case we refer to [Fol99, Lem. 8.4.1].

Lemma 1.97. *If* $g \in \mathrm{BV}[0,\delta]$, *then*

$$\lim_{\lambda\to\infty} \int_0^\delta g(x)\, d_{2\pi\lambda}(x)\, dx \;=\; \frac{g(0+)}{2}.$$

Proof. The Jordan Decomposition Theorem (Problem 1.30) tells us that any function $g \in \mathrm{BV}[0,\delta]$ can be written as $g = g_1 - g_2$ where g_1, g_2 are increasing on $[0,\delta]$. Therefore it suffices to assume that g is increasing. Further, by replacing $g(x)$ with $g(x) - g(0+)$, we may also assume that $g(0+) = 0$.

By Problem 1.43,

$$C \;=\; \sup_{a<b} \left| \int_a^b \frac{\sin x}{x}\, dx \right| \;<\; \infty.$$

Fix $\varepsilon > 0$. Then there exists an $\eta > 0$ such that $|g(x)| < \varepsilon$ for all $0 < x < \eta$. Since g is continuous and $d_{2\pi\lambda}$ is continuous, the Second Mean-Value Theorem for Integrals tells us that there exists some point $c \in [0,\eta]$ such that

$$\int_0^\eta g(x)\, d_{2\pi\lambda}(x)\, dx \;=\; g(0+) \int_0^c d_{2\pi\lambda}(x)\, dx \;+\; g(c-) \int_c^\eta d_{2\pi\lambda}(x)\, dx.$$

Since $g(0+) = 0$, this implies that

$$\begin{aligned} \left| \int_0^\eta g(x)\, d_{2\pi\lambda}(x)\, dx \right| \;&=\; \left| g(c-) \int_c^\eta \frac{\sin 2\pi\lambda x}{\pi x}\, dx \right| \\ &=\; |g(c-)| \left| \int_{2\pi\lambda c}^{2\pi\lambda\eta} \frac{\sin x}{\pi x}\, dx \right| \\ &\le\; \varepsilon\, \frac{C}{\pi}. \end{aligned} \tag{1.34}$$

Since $\eta > 0$, the function $f(x) = \frac{g(x)}{2\pi i x}\, \chi_{[\eta,\delta]}(x)$ is integrable on $\mathbb{R}$. Applying the Riemann–Lebesgue Lemma, we therefore have

$$\begin{aligned} \lim_{\lambda\to\infty} \int_\eta^\delta g(x)\, d_{2\pi\lambda}(x)\, dx \;&=\; \lim_{\lambda\to\infty} \int_\eta^\delta \frac{g(x)}{2\pi i x}\, (e^{2\pi i\lambda x} - e^{-2\pi i\lambda x})\, dx \\ &=\; \lim_{\lambda\to\infty} \big(\widehat{f}(\lambda) - \widehat{f}(-\lambda)\big) \;=\; 0. \end{aligned} \tag{1.35}$$

Combining equations (1.34) and (1.35), we see that

$$\limsup_{\lambda\to\infty} \left| \int_0^\delta g(x)\, d_{2\pi\lambda}(x)\, dx \right| \;\le\; \frac{C\varepsilon}{\pi}.$$

Since ε is arbitrary, the result follows. □

Now we can prove Jordan's Theorem on the pointwise convergence of the "partial sums" $f * d_{2\pi\lambda}$. In the statement of this result we are perform a common abuse of notation by combining an almost everywhere requirement ($f \in L^1(\mathbb{R})$) with a pointwise everywhere requirement (bounded variation on $[x-\delta, x+\delta]$). As usual, this means that there is function g defined everywhere such that $f = g$ a.e. and $g \in \mathrm{BV}[x-\delta, x+\delta]$.

Theorem 1.98 (Jordan's Theorem). *If $f \in L^1(\mathbb{R})$ and $f \in \mathrm{BV}[x-\delta, x+\delta]$, then*

$$\lim_{\lambda\to\infty} (f * d_{2\pi\lambda})(x) \;=\; \frac{f(x+) - f(x-)}{2}.$$

Proof. Noting that x is fixed, define

$$g(t) \;=\; f(x-t) + f(x+t) \qquad \text{and} \qquad h(t) \;=\; \frac{g(t)}{2\pi i t}\, \chi_{[\delta,\infty)}(t).$$

The hypotheses imply that $g \in \mathrm{BV}[0,\delta]$ and $h \in L^1(\mathbb{R})$. Applying Lemma 1.97 and the Riemann–Lebesgue Lemma, we obtain

$$\begin{aligned}
&\lim_{\lambda\to\infty} (f * d_{2\pi\lambda})(x) \\
&\qquad = \lim_{\lambda\to\infty} \int f(x-t)\, d_{2\pi\lambda}(t)\, dt \\
&\qquad = \lim_{\lambda\to\infty} \int_0^\infty \big(f(x-t) + f(x-t)\big)\, d_{2\pi\lambda}(t)\, dt \\
&\qquad = \lim_{\lambda\to\infty} \int_0^\delta g(t)\, d_{2\pi\lambda}(t)\, dt \;+\; \lim_{\lambda\to\infty} \int_\delta^\infty \frac{g(t)}{2\pi i t}\, (e^{2\pi i\lambda t} - e^{-2\pi i \lambda t})\, dt \\
&\qquad = \frac{g(0+)}{2} \;+\; \lim_{\lambda\to\infty} \big(\widehat{h}(\lambda) - \widehat{h}(-\lambda)\big) \\
&\qquad = \frac{f(x+) - f(x-)}{2}. \quad \square
\end{aligned}$$

1.6.5 Decay and Smoothness Revisited

There are many variations on the theme of duality between smoothness and decay under the Fourier transform. Some of these were presented in Section 1.4. As an application of the Inversion Formula, we now prove another version, relating decay in frequency to smoothness in time.

Recall that $A(\mathbb{R}) = \{\widehat{f} : f \in L^1(\mathbb{R})\}$. Hence if $f \in L^1(\mathbb{R})$ then $\widehat{f} \in A(\mathbb{R})$ by definition, and to say that $2\pi i\xi\, \widehat{f}(\xi)$ belongs to $A(\mathbb{R})$ means that there exists a function $g \in L^1(\mathbb{R})$ whose Fourier transform is $\widehat{g}(\xi) = 2\pi i \xi\, \widehat{f}(\xi)$. Note that this is more than just a decay requirement on $\widehat{f}$, for not only must we have $2\pi i\xi\, \widehat{f}(\xi) \in C_0(\mathbb{R})$, but we must have that $2\pi i \xi\, \widehat{f}(\xi)$ belongs to the smaller space $A(\mathbb{R})$.

Theorem 1.99. *If $f \in L^1(\mathbb{R})$ and $2\pi i\xi\widehat{f}(\xi) \in A(\mathbb{R})$, then f is differentiable a.e. and $f' = g$ a.e. where $g \in L^1(\mathbb{R})$ is the function that satisfies $\widehat{g}(\xi) = 2\pi i\xi\,\widehat{f}(\xi)$.*

Proof. First consider the case where we have the additional hypothesis that $\widehat{g} \in L^1(\mathbb{R})$. Since $2\pi i\xi\,\widehat{f}(\xi) = \widehat{g}(\xi) \in L^1(\mathbb{R})$ and $\widehat{f}$ is continuous, it follows that $\widehat{f} \in L^1(\mathbb{R})$. Therefore we can apply the inverse Fourier transform analogue of Theorem 1.55 to the function $\widehat{f}$ and conclude that $\big(\widehat{f}\,\big)^\vee \in C_0^1(\mathbb{R})$. Since the Inversion Formula applies to f, we know that $f = \big(\widehat{f}\,\big)^\vee$ is differentiable everywhere and $f' \in C_0(\mathbb{R})$. Furthermore,

$$f' \;=\; (f^{\wedge\vee})' \;=\; \big(2\pi i\xi\,\widehat{f}(\xi)\big)^\vee \;=\; \big(\widehat{g}\,\big)^\vee \;=\; g,$$

with equality holding everywhere.

Now consider the general case, i.e., $f,\, g \in L^1(\mathbb{R})$ and $\widehat{g}(\xi) = 2\pi i\xi\,\widehat{f}(\xi)$. Let k be any function in $C_c^2(\mathbb{R})$ that satisfies $\int k = 1$. By Theorem 1.65 (and equation (1.24) in particular), we have $\widehat{k} \in L^1(\mathbb{R})$. Define $k_\lambda(x) = \lambda k(\lambda x)$, so $\{k_\lambda\}_{\lambda>0}$ is an approximate identity. We have $f * k_\lambda,\, g * k_\lambda \in L^1(\mathbb{R}),$ and also, since $\widehat{k}$ is integrable and $\widehat{g}$ is bounded, $(g * k_\lambda)^\wedge \in L^1(\mathbb{R})$. Further,

$$(g * k_\lambda)^\wedge(\xi) \;=\; \widehat{g}(\xi)\,\widehat{k_\lambda}(\xi) \;=\; 2\pi i\xi\,\widehat{f}(\xi)\,\widehat{k_\lambda}(\xi) \;=\; 2\pi i\xi\,(f * k_\lambda)^\wedge(\xi).$$

The previous case therefore implies that $f * k_\lambda$ is differentiable everywhere and $(f * k_\lambda)' = g * k_\lambda$. Since $g * k_\lambda \in L^1(\mathbb{R})$, we conclude that $f * k_\lambda$ is absolutely continuous on every interval $[a, b]$ (see Theorem 1.62). Consequently, the Fundamental Theorem of Calculus applies, so we have for any $a < b$ that

$$(f * k_\lambda)(b) - (f * k_\lambda)(a) \;=\; \int_a^b (g * k_\lambda)(t)\,dt \;\to\; \int_a^b g(t)\,dt,$$

the convergence following from the fact that $g * k_\lambda \to g$ in $L^1(\mathbb{R})$ as $\lambda \to \infty$. On the other hand, by Theorem 1.77,

$$(f * k_\lambda)(x) \;\to\; f(x) \quad \text{for a.e. } x.$$

Hence we have for almost every $a < b$ that

$$f(b) - f(a) \;=\; \int_a^b g(t)\,dt. \tag{1.36}$$

By redefining f on a set of measure zero, we can assume that equation (1.36) holds for all $a < b$. Therefore f is absolutely continuous on every interval $[a, b]$, f is differentiable a.e., and $f' = g$ a.e. (see Theorem 1.59). □

Additional Problems

1.45. Show that if f, $\widehat{f} \in L^1(\mathbb{R})$, then $f \in L^p(\mathbb{R})$ for every $1 \le p \le \infty$.

1.46. Suppose that $f \in L^1(\mathbb{R})$. Show that if $\widehat{f}$ is even then f is even, and if $\widehat{f}$ is odd then f is odd. Compare Problem 1.2.

1.47. Show that if $g \in L^1(\mathbb{R})$ and f, $\widehat{f} \in L^1(\mathbb{R})$, then $(fg)^\wedge = \widehat{f}\,\widehat{g}$.

1.48. Show that

$$\int \frac{\sin \pi x}{x}\, e^{-2\pi|x|+\pi i x}\, dx \;=\; \frac{\pi}{4}.$$

1.49. Show that the only function in $L^1(\mathbb{R})$ that satisfies $f = f * f$ is $f = 0$ a.e. (compare Problem 3.5).

1.50. Show that if $g \in L^1(\mathbb{R})$, $g \neq 0$, then $\{T_a g\}_{a \in \mathbb{R}}$ is finitely linearly independent, i.e., every finite subset is linearly independent.

1.51. Prove the following variation on the theme "decay in frequency implies smoothness in time": If $f \in L^1(\mathbb{R})$ and there exists $C > 0$ and $0 < \alpha < 1$ such that

$$\forall\, \xi \in \mathbb{R}, \quad |\widehat{f}(\xi)| \;\le\; \frac{C}{|\xi|^{1+\alpha}},$$

then f is Hölder continuous with exponent α (see Definition 1.75).

1.52. Show that the Fourier transform maps $L^1(\mathbb{R}) \cap A(\mathbb{R})$ bijectively onto itself, and that $L^1(\mathbb{R}) \cap A(\mathbb{R}) \subseteq C_0(\mathbb{R})$.

1.53. For x, $\xi \in \mathbb{R}$, define

$$m_x \;=\; \inf\bigl\{\|f\|_1 + \|\widehat{f}\|_1 : f \in L^1(\mathbb{R}) \cap A(\mathbb{R}) \text{ and } f(x) = 1\bigr\},$$

$$m_\xi \;=\; \inf\bigl\{\|f\|_1 + \|\widehat{f}\|_1 : f \in L^1(\mathbb{R}) \cap A(\mathbb{R}) \text{ and } \widehat{f}(\xi) = 1\bigr\}.$$

Show that $m_x = m_\xi = 1$ for every x and ξ.

1.54. Let $B_0 = \chi_{[0,1]}$, and recursively define the nth *B-spline* B_n by

$$B_n \;=\; B_{n-1} * \chi_{[0,1]}.$$

B-splines and more general splines have applications in numerical analysis, computer graphics, and many other areas.

(a) Find explicit formulas for B_1, B_2, and B_2'. Find an explicit formula for $\widehat{B_n}$, and show that $\widehat{B_n} \in L^1(\mathbb{R})$ for all $n > 0$.

(b) Prove that $B_n' = T_1 B_{n-1} - B_{n-1}$ for $n > 1$ (T_1 is the translation operator).

(c) Show that $B_n \in C_c^{n-1}(\mathbb{R})$ for $n > 0$, and that $B_n^{(n-1)}$ is piecewise linear.

(d) Prove that there exist scalars c_{kn} such that B_n satisfies the *refinement equation*

$$B_n(x) \;=\; \sum_{k=0}^{n+1} c_{kn}\, B_n(2x-k).$$

(e) Prove that there exists a 1-periodic function m_0 (in fact, a trigonometric polynomial) such that $\widehat{B_n}(\xi) = m_0(\xi/2)\,\widehat{B_n}(\xi/2)$.

1.55. This problem will prove a special case of the Second Mean Value Theorem for Integrals. Assume that h is both continuous and nonnegative on $[a,b]$ and g is monotone increasing on $[a,b]$ with $g(a+) \geq 0$. Define

$$G(x) \;=\; g(a+)\int_a^x h(t)\,dt \;+\; g(b-)\int_x^b h(t)\,dt,$$

and show that $G(b) \leq \int_a^b g(t)\,h(t)\,dt \leq G(a)$. Apply the Intermediate Value Theorem to show there exists a point $c \in [a,b]$ such that

$$\int_a^b g(t)\,h(t)\,dt \;=\; G(c) \;=\; g(a+)\int_a^c h(t)\,dt + g(b-)\int_c^b h(t)\,dt.$$

1.7 The Range of the Fourier Transform

We know that $A(\mathbb{R})$, the range of the Fourier transform acting on $L^1(\mathbb{R})$, is a subspace of $C_0(\mathbb{R})$. We will show that it is a dense, but proper, subspace of $C_0(\mathbb{R})$.

To begin, the fact that $A(\mathbb{R})$ is dense is a consequence of the smoothness versus decay dualities of the Fourier transform combined with the Inversion Formula—any sufficiently smooth function $f \in L^1(\mathbb{R})$ will have a Fourier transform $\widehat{f}$ that decays quickly enough that we must have $\widehat{f} \in L^1(\mathbb{R})$.

Exercise 1.100. Show that if $f \in C_c^2(\mathbb{R})$ then $\widehat{f} \in L^1(\mathbb{R})$. Conclude that $C_c^2(\mathbb{R}) \subseteq A(\mathbb{R})$, and that $A(\mathbb{R})$ is dense in $C_0(\mathbb{R})$. $\diamondsuit$

The next exercise (taken from [Fol99]) shows that $A(\mathbb{R})$ is a proper subset of $C_0(\mathbb{R})$, although the argument is implicit in the sense that it does not construct a specific example of a function in $C_0(\mathbb{R})\backslash A(\mathbb{R})$. The main point is that if the Fourier transform $\mathcal{F}$ was a bounded map of $L^1(\mathbb{R})$ *onto* $C_0(\mathbb{R})$, then, since both of these are Banach spaces, the Inverse Mapping Theorem (Theorem C.14) would imply that $\mathcal{F}$ had a bounded inverse. However, the exercise shows that $\mathcal{F}^{-1}$ is not a bounded map of $A(\mathbb{R})$ (under the L^∞-norm) back to $L^1(\mathbb{R})$.

Exercise 1.101. Define $f_k = \chi_{[-1,1]} * \chi_{[-k,k]}$ for $k \in \mathbb{N}$.

(a) Find an explicit formula for f_k and show that $\|f_k\|_\infty = 2$.

(b) Show that $\lim_{k\to\infty} \|\widehat{f_k}\|_1 = \infty$.

(c) Show that $\mathcal{F}^{-1}\colon (A(\mathbb{R}), \|\cdot\|_\infty) \to L^1(\mathbb{R})$ is unbounded, and conclude that $A(\mathbb{R}) \subsetneq C_0(\mathbb{R})$. $\diamondsuit$

While we now know that $A(\mathbb{R})$ is a proper subset of $C_0(\mathbb{R})$, we do not yet have any explicit examples of functions in $C_0(\mathbb{R})\backslash A(\mathbb{R})$. The next exercise will construct such an example (this construction is based on Goldberg's text [Gol61, p. 8]).

Exercise 1.102. (a) Show that if $f \in L^1(\mathbb{R})$ and f is odd, then

$$\sup_{b\ge 1} \left| \int_1^b \frac{\widehat{f}(\xi)}{\xi}\, d\xi \right| \;<\; \infty.$$

(b) Show that if $f \in L^1(\mathbb{R})$ is odd, $\widehat{f}$ is differentiable at $\xi = 0$, and $\widehat{f} \ge 0$ on $(0,\infty)$, then $\widehat{f}(\xi)/\xi \in L^1(\mathbb{R})$.

(c) Define

$$F(\xi) \;=\; \begin{cases} 1/\ln \xi, & \xi > e, \\ \xi/e, & -e \le \xi \le e, \\ -1/\ln \xi, & \xi < -e, \end{cases}$$

and show that $F \in C_0(\mathbb{R})\backslash A(\mathbb{R})$. $\diamondsuit$

In fact, there exist compactly supported functions in $C_0(\mathbb{R})$ that do not belong to $A(\mathbb{R})$. An example is constructed in [Her85], where it is shown that

$$B(\xi) \;=\; \begin{cases} \frac{1}{n}\sin(2\pi 4^n \xi), & \frac{1}{2^{n+1}} \le |\xi| \le \frac{1}{2^n}, \\ 0, & \xi \le 0 \text{ or } |\xi| > \frac{1}{2}, \end{cases}$$

(B for "butterfly") belongs to $C_c(\mathbb{R})\backslash A(\mathbb{R})$.

Although it is dense in $C_0(\mathbb{R})$, in a topological sense $A(\mathbb{R})$ is only a "small" part of $C_0(\mathbb{R})$. One consequence of the Open Mapping Theorem is that if $T\colon X \to Y$ is a bounded linear map on a Banach space X and its range $T(X)$ is a dense but proper subspace of a Banach space Y, then $T(X)$ is a *meager* or *first category* subset of Y (Problem 1.57). Applying this to the mapping $\mathcal{F}\colon L^1(\mathbb{R}) \to C_0(\mathbb{R})$, we conclude that $A(\mathbb{R})$ is only a meager subset of $C_0(\mathbb{R})$.

Additional Problems

1.56. Define $A_c(\mathbb{R}) = \{F \in A(\mathbb{R}) : \operatorname{supp}(F) \text{ is compact}\}$. Show that $A_c(\mathbb{R})$ is a dense subspace of $A(\mathbb{R})$ in the norm of $A(\mathbb{R})$. Compare Problem 1.23, which shows that $A_c(\mathbb{R})$ is an ideal in $A(\mathbb{R})$.

1.57. Let X and Y be Banach spaces. Show that $T \in \mathcal{B}(X,Y)$ is surjective if and only if range(T) is not meager in Y.

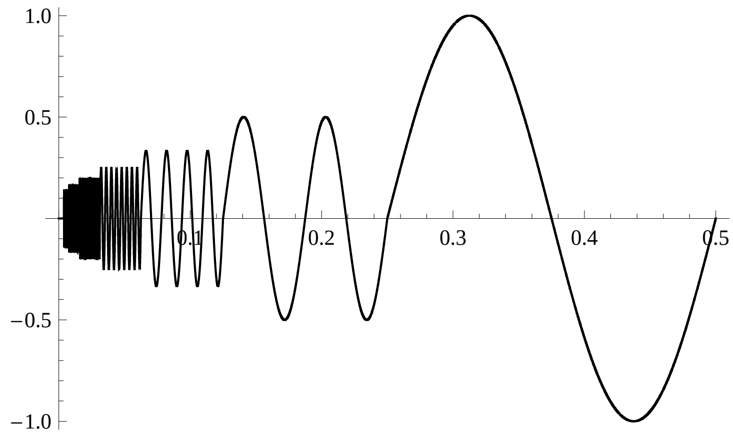

Fig. 1.16. Graph of the butterfly function.

1.8 Some Special Kernels

In addition to the Fejér function, there are several very useful special functions that can be used to generate approximate identities.

First, we modify the Fejér kernel to obtain an approximate identity $\{v_\lambda\}_{\lambda>0}$ that has the appealing property that $\widehat{v_\lambda} = 1$ on $[-\lambda, \lambda]$. As we have seen, any approximate identity $\{k_\lambda\}_{\lambda>0}$ has the property that $(f * k_\lambda)^\wedge(\xi) = \widehat{f}(\xi)\,\widehat{k_\lambda}(\xi)$ converges pointwise to $\widehat{f}$. Using the approximate identity $\{v_\lambda\}_{\lambda>0}$ we will have the extra property that $\widehat{f}(\xi)$ is actually equal to $\widehat{f}(\xi)\,\widehat{v_\lambda}(\xi)$ on the compact set $[-\lambda, \lambda]$.

Definition 1.103 (de la Vallée–Poussin Kernel). Let w be the Fejér function. Then the *de la Vallée–Poussin function* is

$$v(x) \;=\; 2w_2(x) - w(x) \;=\; 4w(2x) - w(x),$$

and the *de la Vallée–Poussin kernel* is $\{v_\lambda\}_{\lambda>0}$ where $v_\lambda(x) = \lambda v(\lambda x)$. $\diamondsuit$

Charles Jean de la Vallée–Poussin (1866–1962) is perhaps best known for the fact that he and Jacques Hadamard (1865–1963) independently gave, in 1896, the first proofs of the Prime Number Theorem.

Exercise 1.104. Show that $\int v = 1$ (so $\{v_\lambda\}_{\lambda>0}$ is an approximate identity), and show that

$$\widehat{v_\lambda}(\xi) \;=\; \begin{cases} 1, & |\xi| \le \lambda, \\ \text{linear}, & \text{on } [-2\lambda, -\lambda] \text{ and } [\lambda, 2\lambda], \\ 0, & |\xi| \ge 2\lambda. \qquad \diamondsuit \end{cases}$$

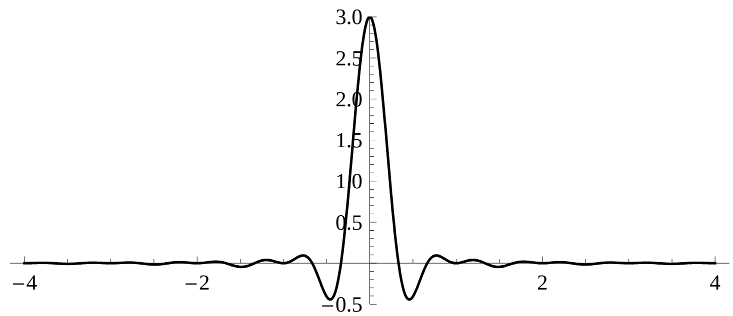

Fig. 1.17. Graph of the de la Vallée–Poussin function v.

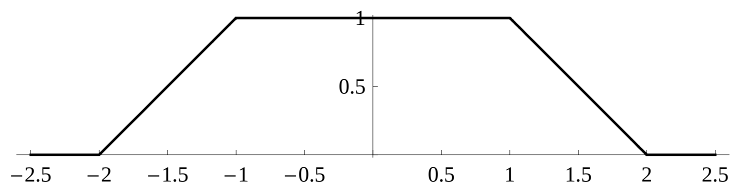

Fig. 1.18. Graph of the Fourier transform $\widehat{v}$.

To illustrate its use, we hint that an "easy" solution to the following exercise can be obtained by using the de la Vallée–Poussin kernel.

Exercise 1.105. Prove that if $f \in L^1(\mathbb{R})$, then

$$\operatorname{supp}(\widehat{f}\,) \text{ is compact} \quad \Longleftrightarrow \quad \exists\, g \in L^1(\mathbb{R}) \text{ such that } f = f * g. \qquad \diamondsuit$$

Note that the preceding exercise characterizes those functions $f \in L^1(\mathbb{R})$ such that f belongs to $f * L^1(\mathbb{R})$; compare Exercise 1.48 and Problem 1.39.

Poisson's kernel, named after Siméon-Denis Poisson (1781–1840) is related to the two-sided exponential function (see Problem 1.1).

Definition 1.106 (Poisson Kernel). The *Poisson function* is

$$p(x) \;=\; \frac{1}{\pi(x^2+1)},$$

and the *Poisson kernel* is $\{p_\lambda\}_{\lambda>0}$ where $p_\lambda(x) = \lambda p(\lambda x)$. $\diamondsuit$

Exercise 1.107. Show that $\int p = 1$ and $\widehat{p}(\xi) = e^{-2\pi|\xi|}$. $\diamondsuit$

A *Gaussian function* is a function of the form $ae^{-(x-b)^2/(2c^2)}$. In Chapter 3 we will see that Gaussian functions are the functions that are "best concentrated" with respect to the Uncertainty Principle for the Fourier transform. For this reason, Gaussian functions are ubiquitous in Fourier analysis. In particular, we can use an appropriately normalized Gaussian to generate an approximate identity. Of course, Gaussian functions are named after Carl Friedrich Gauss (1777–1855).

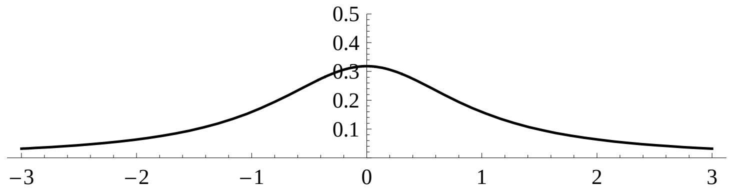

Fig. 1.19. Graph of the Poisson function p.

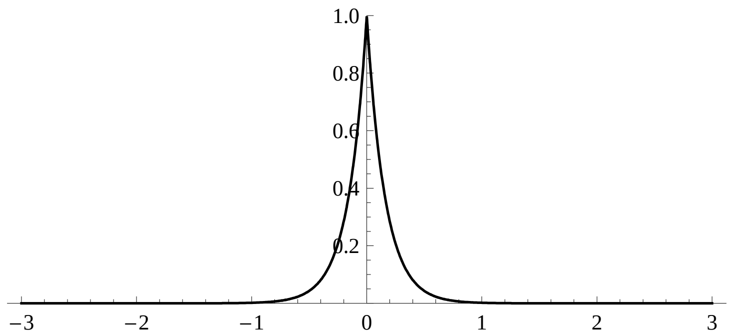

Fig. 1.20. Graph of the Fourier transform $\widehat{p}$.

Definition 1.108 (Gauss Kernel). Let ϕ denote the Gaussian function

$$\phi(x) \;=\; e^{-\pi x^2}.$$

Then the *Gauss kernel* is $\{\phi_\lambda\}_{\lambda>0}$ where $\phi_\lambda(x) = \lambda\phi(\lambda x)$. $\diamondsuit$

In order to know that $\{\phi_\lambda\}_{\lambda>0}$ is an approximate identity, we must show that $\int \phi = 1$. Fortunately, this is not hard to prove. Unfortunately, it is more difficult to compute an explicit formula for the Fourier transform $\widehat{\phi}$. Aficionados of complex analysis will recognize that one way to do this is by using contour integration, and we encourage those readers familiar with this technique to use it to compute $\widehat{\phi}$. On the other hand, a real-variable approach to computing $\widehat{\phi}$ is given in the next exercise.

Exercise 1.109. Let $\Phi = \widehat{\phi}$, where $\phi(x) = e^{-\pi x^2}$.

(a) Use Theorems 1.55 and 1.65 to show that

$$\Phi'(\xi) \;=\; -2\pi\xi\Phi(\xi).$$

(b) Solve the differential equation in part (a), and show that

$$\Phi(\xi) \;=\; \Phi(0)\,\phi(\xi).$$

(c) Show that $\Phi(0) = 1$. Conclude that $\Phi = \phi$, and therefore the Gauss kernel is an approximate identity. $\diamondsuit$

Thus the Gaussian function ϕ has the interesting property that $\widehat{\phi} = \phi$, and hence is a 1-eigenvector for the Fourier transform! Are there any other such functions? Yes, the *Hermite functions* are eigenfunctions of the Fourier transform, and they form an orthogonal basis for $L^2(\mathbb{R})$. We will investigate the Hermite functions in Section 3.2.

As an application, we will use the Gauss kernel to prove the Weierstrass Approximation Theorem (the proof we give is taken from Stein and Shakarchi [SS03a]).

Theorem 1.110 (Weierstrass Approximation Theorem). *If $f \in C[a,b]$ and $\varepsilon > 0$, then there exists a polynomial p such that*

$$\|f - p\|_\infty = \sup_{x \in [a,b]} |f(x) - p(x)| < \varepsilon.$$

Proof. Choose $f \in C[a,b]$, and fix R large enough that $[a,b] \subseteq (-R,R)$. Let g be any continuous function on $\mathbb{R}$ supported in $[-R,R]$ that equals f on $[a,b]$. Let $\{\phi_\lambda\}_{\lambda>0}$ be the Gauss kernel. By Exercise 1.73, $g*\phi_\lambda$ converges uniformly to g, so we can choose a λ such that

$$\|g - g*\phi_\lambda\|_\infty < \frac{\varepsilon}{2}. \tag{1.37}$$

Since the Taylor series for e^x converges uniformly on compact sets, the series

$$\phi_\lambda(x) = \lambda e^{-\pi\lambda^2x^2} = \sum_{n=0}^{\infty} \lambda \frac{(-\pi\lambda^2x^2)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n\,\pi^n\,\lambda^{2n+1}}{n!}\,x^{2n}$$

converges uniformly on $[-2R,2R]$. Therefore, there exists an N such that if we set

$$q(x) = \sum_{n=0}^{N} \frac{(-1)^n\,\pi^n\,\lambda^{2n+1}}{n!}\,x^{2n}$$

then

$$\sup_{x\in[-2R,2R]} |\phi_\lambda(x) - q(x)| < \frac{\varepsilon}{2\|g\|_1}.$$

Therefore, for $x \in [-R,R]$ we have

$$\begin{aligned} \big|(g*\phi_\lambda)(x) - (g*q)(x)\big| &\le \int_{-R}^{R} |g(y)|\,|\phi_\lambda(x-y) - q(x-y)|\,dy \\ &\le \int_{-R}^{R} |g(y)|\,\frac{\varepsilon}{2\|g\|_1}\,dy = \frac{\varepsilon}{2}. \end{aligned} \tag{1.38}$$

Since g and f are equal on $[a,b]$, by combining equations (1.37) and (1.38), we see that

$$\sup_{x\in[a,b]} |f(x) - (g*q)(x)| < \varepsilon.$$

Finally, if we write out $g * q$ and expand using the binomial theorem, we see that $g * q$ is actually a polynomial of degree at most $2N$:

$$\begin{aligned}(g*q)(x) &= \sum_{n=0}^{N} \frac{(-1)^n\,\pi^n\,\lambda^{2n+1}}{n!} \int_{-R}^{R} g(y)\,(x-y)^{2n}\,dy \\ &= \sum_{n=0}^{N} \frac{(-1)^n\,\pi^n\,\lambda^{2n+1}}{n!} \sum_{j=0}^{2n} \binom{2n}{j} \left(\int_{-R}^{R} g(y)\,(-y)^{2n-j}\,dy\right) x^j.\end{aligned}$$

The integrals appearing above are finite since $g \in C_c(\mathbb{R})$. Thus $p = g * q$ is the polynomial that we sought. $\square$

Additional Problems

1.58. There are many functions that equal their own Fourier transforms. Show that if $f, \widehat{f} \in L^1(\mathbb{R})$, then $f^{\wedge\wedge\wedge\wedge} = f$, and consequently $g = f + f^{\wedge} + f^{\wedge\wedge} + f^{\wedge\wedge\wedge}$ satisfies $\widehat{g} = g$.

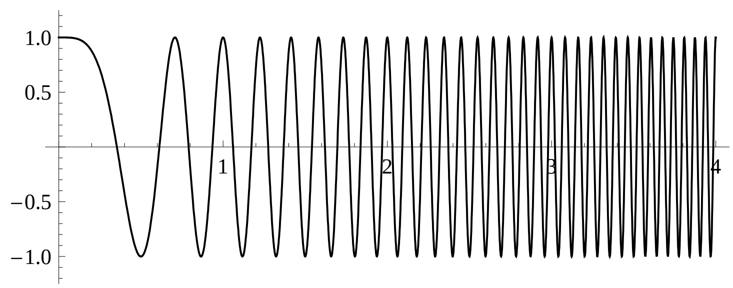

Fig. 1.21. Graph of the real part of the chirp $e^{2\pi i 2x^2}$.

1.59. Let $\phi(x) = e^{-\pi x^2}$ be the Gaussian function.

(a) For $r > 0$, set $\varphi_r(x) = e^{-\pi r x^2}$. Show that $\widehat{\varphi_r} = r^{-1/2}\varphi_{1/r}$.

(b) Now we extend the definition of φ_r to complex parameters. Let $c = a + ib \in \mathbb{C}$ be complex, and set

$$\varphi_c(x) = e^{-\pi c x^2} = e^{-\pi i b x^2}\, e^{-\pi a x^2}.$$

In engineering jargon, φ_c is a Gaussian multiplied by a *chirp* $e^{-\pi i b x^2}$ (and the resulting function is often also called a Gaussian function). Show that part (a) extends to complex parameters with positive real part, i.e., if $c = a + ib$ and $a > 0$, then $\widehat{\varphi_c} = c^{-1/2}\varphi_{1/c}$. This fact will be useful to us in Section 3.3.

Note: We take a complex square root that extends the square root of the positive real numbers. For example, since $c = a + ib$ with $a > 0$, we can write $c = re^{i\theta}$ with $-\pi/2 < \theta < \pi/2$. We take $c^{1/2} = r^{1/2}e^{i\theta/2}$ and $c^{-1/2} = r^{-1/2}e^{-i\theta/2}$.

Remark: If we regard $e^{2\pi i\xi x}$ as a "pure tone" of constant frequency, then we can think of $e^{2\pi i\xi x^2}$ as a function whose frequency increases with time, see the illustration in Figure 1.21. If played through a speaker, such a function sounds something like a bird's chirp.

1.60. Suppose that $f \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ has a jump discontinuity at the origin, i.e.,

$$f(0+) = \lim_{t\to 0^+} f(t), \qquad f(0-) = \lim_{t\to 0^-} f(t)$$

both exist, but $f(0-) \neq f(0+)$. Let $\{\phi_\lambda\}_{\lambda>0}$ be the Gauss kernel. Prove that

$$\lim_{\lambda\to\infty} (f * \phi_\lambda)(0) = \frac{f(0+) + f(0-)}{2}.$$

Can other kernels be used to obtain the same result?

1.61. Let $f \in L^1(\mathbb{R})$ be given. Show that if there exist δ, $R > 0$ such that f is bounded on $[-\delta, \delta]$ and $\widehat{f}(\xi) \geq 0$ for $|\xi| > R$, then $\widehat{f} \in L^1(\mathbb{R})$.

1.62. Suppose that $f \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ is real and even (so $\widehat{f}$ is real and even as well). Show that if f is not continuous, then $\widehat{f}$ must change sign infinitely often. Thus, a "local" (jump) discontinuity of f forces a global "reaction" in $\widehat{f}$.

On the other hand, show that $g(x) = ie^{-|x|}\left(\chi_{[0,\infty)} - \chi_{(-\infty,0)}\right)$ has a single jump discontinuity and that $\widehat{g}$ is real and changes sign only once. Still, the fact that g has a discontinuity implies that $\widehat{g}$ decays slowly.

1.63. Prove the following version of *Bernstein's Inequality.* Suppose that $f \in L^1(\mathbb{R})$ is differentiable and $\mathrm{supp}(\widehat{f}\,) \subseteq [-R, R]$. Show that $\|f'\|_p \leq CR\,\|f\|_p$ for each $1 \leq p \leq \infty$, where C is a fixed constant independent of f, R, or p.

1.9 The Schwartz Space

The Schwartz space will play an important role throughout the rest of this volume. In this section, after giving its definition and basic properties, we will discuss its special place in harmonic analysis.

The Schwartz space is a space of infinitely differentiable functions that, together with all their derivatives, "decay rapidly" at infinity. It is named in honor of Laurent Schwartz (1915–2002), because of its role in the theory of distributions, which he developed. Distributions will be the focus of our attention in Chapter 4.

1.9.1 Definition and Basic Properties

The precise definition of the Schwartz space is as follows.

Definition 1.111 (The Schwartz Space). The *Schwartz space* is

$$\mathcal{S}(\mathbb{R}) \;=\; \bigl\{f \in C^\infty(\mathbb{R}) : x^m f^{(n)}(x) \in L^\infty(\mathbb{R}) \text{ for all } m, n \geq 0\bigr\}.$$

An element of $\mathcal{S}(\mathbb{R})$ is called a *Schwartz-class function.* $\diamondsuit$

Thus, "rapid decay" means decay faster than the reciprocal of any polynomial: For each m, $n \geq 0$ there exists a constant C_{mn} such that

$$|f^{(n)}(x)| \;\leq\; \frac{C_{mn}}{|x^m|}, \qquad x \in \mathbb{R}.$$

The constants C_{mn} need not be uniformly bounded in m and n.

A consequence of rapid decay is that f and every derivative of f is integrable.

Exercise 1.112. Show that if $f \in \mathcal{S}(\mathbb{R})$ then for every m, $n \geq 0$ we have $x^m f^{(n)}(x) \in L^1(\mathbb{R}) \cap C_0(\mathbb{R})$, with

$$\|x^m f^{(n)}(x)\|_1 \;\leq\; 2\,\|x^m f^{(n)}(x)\|_\infty \;+\; 2\,\|x^{m+2} f^{(n)}(x)\|_\infty. \qquad \diamondsuit$$

Since $C_c^\infty(\mathbb{R}) \subseteq \mathcal{S}(\mathbb{R})$, we know that $\mathcal{S}(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for every $1 \leq p < \infty$, and also is dense in $C_0(\mathbb{R})$ in L^∞-norm. The Gaussian function e^{-x^2} is an example of a function in $\mathcal{S}(\mathbb{R})$ that is not compactly supported. On the other hand, while the two-sided exponential $e^{-|x|}$ has rapid decay, it is not a Schwartz-class function since it is not differentiable at the origin.

1.9.2 Topology and Convergence in the Schwartz Space

The Schwartz space is a *topological vector space* but not a normed vector space. Instead of being determined by a single norm, the topology on $\mathcal{S}(\mathbb{R})$ is determined by the infinite collection of seminorms

$$\rho_{mn}(f) \;=\; \|x^m f^{(n)}(x)\|_\infty, \qquad m, n \geq 0.$$

The Schwartz space and related spaces determined by families of seminorms will be discussed in more detail in Section 4.2. For us, the meaning of *convergence* in $\mathcal{S}(\mathbb{R})$ is more important than the associated topology, and convergence simply means simultaneous convergence with respect to each of the seminorms ρ_{mn}.

Definition 1.113 (Convergence in $\mathcal{S}(\mathbb{R})$). Given f_k, $f \in \mathcal{S}(\mathbb{R})$, we say that f_k *converges to* f *in* $\mathcal{S}(\mathbb{R})$ if

$$\forall\, m, n \geq 0, \quad \lim_{k\to\infty} \rho_{mn}(f - f_k) \;=\; 0,$$

Writing out the seminorms explicitly, convergence in $\mathcal{S}(\mathbb{R})$ means that

$$\forall\, m, n \geq 0, \quad \lim_{k\to\infty} \|x^m f^{(n)}(x) - x^m f_k^{(n)}(x)\|_\infty \;=\; 0.$$

We write $f_k \to f$ *in* $\mathcal{S}(\mathbb{R})$ to denote that f_k converges to f in $\mathcal{S}(\mathbb{R})$. $\diamondsuit$

Because $\mathcal{S}(\mathbb{R})$ is determined by only countably many seminorms, it is possible to create a metric on $\mathcal{S}(\mathbb{R})$ that determines the convergence criterion. In particular, if we define

$$\mathrm{d}(f,g) \;=\; \sum_{m,n\geq 0} \frac{1}{2^{m+n}} \, \frac{\rho_{mn}(f-g)}{1+\rho_{mn}(f-g)},$$

then $f_k \to f$ in $\mathcal{S}(\mathbb{R})$ if and only if $\mathrm{d}(f, f_k) \to 0$ (see Exercise 4.12). Unfortunately, there is no norm that induces this metric on $\mathcal{S}(\mathbb{R})$, but the fact that there is an underlying metric means that we can define convergence in terms of ordinary sequences indexed by the natural numbers instead of having to use abstract nets, as is required in generic topological spaces (see [Heil11b, Chap. 1] for details).

Since we have a metric, we can ask whether $\mathcal{S}(\mathbb{R})$ is complete with respect to this metric. The answer is yes, every Cauchy sequence in $\mathcal{S}(\mathbb{R})$ converges in $\mathcal{S}(\mathbb{R})$ (see Exercise 4.14). Thus $\mathcal{S}(\mathbb{R})$ is a vector space whose topology is generated by a metric, and it is complete with respect to that metric. Such a space is called a *Fréchet space*.

Since we have defined convergence in $\mathcal{S}(\mathbb{R})$, we also have a notion of continuity. For example, a functional $\mu\colon \mathcal{S}(\mathbb{R}) \to \mathbb{C}$ is continuous if $f_k \to f$ in $\mathcal{S}(\mathbb{R})$ implies that $\langle f_k, \mu\rangle \to \langle f, \mu\rangle$, where $\langle f, \mu\rangle$ denotes the action of the functional μ on the vector f. The space of all continuous linear functionals on $\mathcal{S}(\mathbb{R})$, i.e., the dual space of $\mathcal{S}(\mathbb{R})$, is the *space of tempered distributions*, denoted $\mathcal{S}'(\mathbb{R})$. It will be a major center of attention in Chapter 4.

1.9.3 Invariance of the Schwartz Space

Now that we have introduced the Schwartz space, let us explain why it is so interesting to us. The definition of the Schwartz space incorporates both smoothness and decay—an element f of $\mathcal{S}(\mathbb{R})$ is both infinitely smooth and decays rapidly at infinity (as do all of its derivatives). The Fourier transform interchanges smoothness and decay. Thus, since $f \in \mathcal{S}(\mathbb{R})$ has both smoothness and decay, $\widehat{f}$ has both decay and smoothness, suggesting that we may have $\widehat{f} \in \mathcal{S}(\mathbb{R})$. And in fact this is true: *The Schwartz space is invariant under the Fourier transform.* In order to prove this, we first need the reader to verify that an application of Theorems 1.55 and 1.65 implies the following relation between derivatives and products. To simplify notation, we let D denote the differentiation operator:

$$Df(x) \;=\; f'(x). \tag{1.39}$$

Exercise 1.114. Show that if $x^j f^{(k)}(x) \in L^1(\mathbb{R})$ for $j = 0, \dots, m$ and $k = 0, \dots, n$, then

$$\Big(D^n\big((-2\pi i x)^m f(x)\big)\Big)^{\wedge}(\xi) \;=\; (2\pi i\xi)^n\, D^m \widehat{f}(\xi). \qquad \diamondsuit$$

Theorem 1.115 (Invariance of the Schwartz Space). *The Fourier transform is a bijection of $\mathcal{S}(\mathbb{R})$ onto itself.*

Proof. Suppose that $f \in \mathcal{S}(\mathbb{R})$. Then, by the product rule, we have for any $m, n \geq 0$ that

$$D^n\big((-2\pi i x)^m f(x)\big) \;=\; \sum_{j=0}^{n} \binom{n}{j} D^j (-2\pi i x)^m\, f^{(n-j)}(x) \;\in\; L^1(\mathbb{R}).$$

Hence,

$$(2\pi i\xi)^n\, D^m \widehat{f}(\xi) \;=\; \Big(D^n\big((-2\pi i x)^m f(x)\big)\Big)^{\wedge}(\xi) \;\in\; L^\infty(\mathbb{R}).$$

Since this is true for every m and n, we conclude that $\widehat{f} \in \mathcal{S}(\mathbb{R})$. Thus the Fourier transform maps $\mathcal{S}(\mathbb{R})$ into itself, and we know that it is injective by the Uniqueness Theorem.

On the other hand, we also have $\check{f} \in \mathcal{S}(\mathbb{R})$, and therefore $(\check{f})^\wedge \in \mathcal{S}(\mathbb{R})$. Hence $f = (\check{f})^\wedge$ by the Inversion Formula, so the Fourier transform is surjective. $\square$

Not only is the Fourier transform a bijection, but it is a *topological isomorphism* (or *homeomorphism*) of $\mathcal{S}(\mathbb{R})$ onto itself, i.e., both $\mathcal{F}$ and $\mathcal{F}^{-1}$ are continuous with respect to the topology of the Schwartz space.

Exercise 1.116 (Continuity of the Fourier Transform on $\mathcal{S}(\mathbb{R})$). Given any $m, n \geq 0$, show that there exist constants $C_{j\ell} > 0$ such that

$$\|\xi^n \widehat{f}^{(m)}(\xi)\|_\infty \;\leq\; \sum_{j=0}^{m} \sum_{\ell=0}^{n} C_{j\ell}\, \|x^j f^{(\ell)}(x)\|_1.$$

Use this to show that $\mathcal{F}\colon \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is continuous, i.e.,

$$f_k \to f \text{ in } \mathcal{S}(\mathbb{R}) \implies \widehat{f_k} \to \widehat{f} \text{ in } \mathcal{S}(\mathbb{R}),$$

and similarly $\mathcal{F}^{-1}\colon \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is continuous. $\diamondsuit$

As a final remark, we note that since $C_c^\infty(\mathbb{R}) \subseteq \mathcal{S}(\mathbb{R})$, the Fourier transform maps $C_c^\infty(\mathbb{R})$ into $\mathcal{S}(\mathbb{R})$:

$$f \in C_c^\infty(\mathbb{R}) \implies \widehat{f} \in \mathcal{S}(\mathbb{R}).$$

However, we will later see the Paley–Wiener Theorem (Theorem 3.49), which implies that f and $\widehat{f}$ cannot both be compactly supported (unless $f = 0$). Hence the Fourier transform does not map $C_c^\infty(\mathbb{R})$ into itself:

$$\mathcal{F}\big(C_c^\infty(\mathbb{R})\big) \subseteq \mathcal{S}(\mathbb{R}) \qquad \text{but} \qquad \mathcal{F}\big(C_c^\infty(\mathbb{R})\big) \cap C_c^\infty(\mathbb{R}) = \{0\}.$$

Similarly, if $f \in L^1(\mathbb{R})$ and $\widehat{f} \in C_c^\infty(\mathbb{R})$, then $\widehat{f} \in \mathcal{S}(\mathbb{R})$, so we know that $f \in \mathcal{S}(\mathbb{R})$, but f cannot be compactly supported.

Additional Problems

1.64. Show that if $f \in \mathcal{S}(\mathbb{R})$ and $g \in C_b^\infty(\mathbb{R})$, then $fg \in \mathcal{S}(\mathbb{R})$. In particular, $\mathcal{S}(\mathbb{R})$ is closed under pointwise products.

1.65. Show that if $f \in \mathcal{S}(\mathbb{R})$ and $x^m g(x) \in L^1(\mathbb{R})$ for every $m \geq 0$, then $f * g \in \mathcal{S}(\mathbb{R})$. In particular, $\mathcal{S}(\mathbb{R})$ is closed under convolution.

1.66. Show that $\big\{f \in L^1(\mathbb{R}) : \widehat{f} \in C_c^\infty(\mathbb{R})\big\}$ is dense in $L^p(\mathbb{R})$ for each $1 \leq p < \infty$.

1.67. Construct a function $f \in \mathcal{S}(\mathbb{R})$ that satisfies $\lim_{n\to\infty} |f^{(n)}(n)| = \infty$.